Ottawa, Ontario, Canada
April 25-29, 2014

Transforming Healthcare through Excellence in Assessment and Evaluation

2014

16th OTTAWA CONFERENCE

12th CANADIAN CONFERENCE ON MEDICAL EDUCATION

OTTAWA CONFERENCE ABSTRACTS

Ottawa Tourism

www.ottawaconference.org   www.mededconference.ca

Saturday April 26
4:00PM – 5:30PM

PLENARY: AFMC-AMS J. WENDELL MACLEOD MEMORIAL LECTURE

Assessment Transformation of Health Care

Speaker: *Dr. John Norcini, President and Chief Executive Officer, Foundation of Advancement for Medical Education and Research*

Changes to health professions education across the continuum hold the promise of improving population health outcomes and the quality of patient care while reducing costs. Assessment can play a central role in achieving this promise. It offers the ability to ensure, direct, motivate and create learning. This session will explore how assessment can support the transformation of health care in terms of protecting patients and improving the health of communities, while at the same time creating and supporting the learning of health care providers throughout their careers.

Learning Objectives
After attending this session, the participants will:
1. Understand the transformative possibilities for assessment in terms of improving population health outcomes and the quality of patient care while reducing costs
2. Be exposed to some of the assessment methods available to support reform
3. Understand the resources required to support learning and create change
4. Understand the limitations of assessment in supporting learning and creating change

Sunday April 27
8:30AM – 10:00AM

PLENARY: NEWER PRINCIPLES OF TEACHING, LEARNING AND ASSESSMENT

Planning the disposition and training of healthcare workers for an uncertain future

Speaker: *Professor Des Gorman MD PhD, Professor of Medicine and Associate Dean, Faculty of Medical and Health Sciences, the University of Auckland (Head of the School of Medicine 2005-2010)*

The future health care milieu is uncertain. Predictably, conventional approaches to planning future health care systems, necessary health care workforces and their underpinning training is frequently "wrong". The participation of some elements of the health care workforce in work is more dependent on general economic conditions than it is on either health need or plans.
As such, the health care workforce often transitions quickly from "feast to famine" and back. A different approach to planning the necessary future health care workforce has been successfully implemented in New Zealand. This is based on addressing as many as is possible of scenarios of how health care might be delivered. The scenarios are generated by groups of subject matter experts and professional opinion leaders, are aggregated into services and solutions are integrated capital, IT and workforce projections.

Scylla and Charybdis: Caught between examination and reflection in medical education

Speaker: Brian D. Hodges, MD, PhD, FRCPC
Vice-Presi*dent Education, University Health Network Professor, Department of Psychiatry, University of Toronto*

A peculiar disjunction is apparent in the assessment of the health professionals. On one hand there has been an explosion of testing technologies such that health professionals undergo an almost endless series of written and performance examinations during training and practice. They live in what Michel Foucault called the 'examined society' in which constant surveillance and testing locates the responsibility for competence externally to individuals. Simultaneously a different discourse about assessment is gaining adherents, one that is tethered to a 'trinity' of reflective technologies: self-assessment, self-direction and self-regulation. This approach places the locus for control of competence internally, requiring portfolios and reflective diaries. Both conceptions of assessment have significant advantages but also limitations and even adverse effects. How does the health professional educator navigate between the Scylla of excessive external examination and the Charybdis of naïve reliance on self-assessment

SUNDAY APRIL 27
10:30AM – 12:00 PM

SYMPOSIA

OTT-SA-1
NBME Stemmler Grants: Demonstrating Excellence in Assessment and Evaluation

Presenters: Kevin Eva (*University of British Columbia*), Larry Gruppen (*University of Michigan*), Maxine Papadakis (*UCSF School of Medicine*), David B Swanson (*National Board of Medical Examiners (NBME) Moderator*)
(*Symposium organised by National Board of Medical Examiners*)

Summary: 2015 marks the 100th anniversary of the NBME. The mission of the NBME is improving healthcare around the world through assessment and that mission supports directly the theme of the 2014 Ottawa Conference, "Transforming Healthcare through Excellence in Assessment and Evaluation".

One mechanism the NBME uses to transform healthcare through excellence in assessment is the awarding of grant funds to support research in assessment through the Stemmler Fund. The goal of the Stemmler Fund is to provide support for research or development of innovative assessment approaches that will enhance the evaluation of those preparing to, or continuing to, practice medicine. The symposium will highlight the work of three Stemmler recipients and consider the impact their work has had on assessment and evaluation of medical professionals and suggestions for continued research in assessment and evaluation.

OTT-SA-2
Work Place Based Assessment in UK general practice - How do we build on our 5 year experience?

Presenters: Jane Mamelok (*MRCGP core group for Workplace Based Assessment*), Jill Edwards (*MRCGP core group for Workplace Based Assessment*), Simon Street (*MRCGP core group for Workplace Based Assessment*)

Summary: There are changes happening in the NHS and in medical education in the UK. Inevitably this has repercussions in general practice. There are new clinical and administrative responsibilities for GPs. Training for general practice must now address an expanded curriculum and that expansion is reflected in new demands on the programme of assessment. The symposium will discuss issues including: (1) The quality management of the programme of workplace based learning and assessment in the UK since 2007; (2) The developing GP curriculum and the use of "blueprinting" in developing the assessments; (3) The changes in the assessment of psychomotor skills in general practice 4.

The regulatory constraints on developments in the assessment programme.

WORKSHOPS

OTT-WA-1
Critical Perspectives on Assessment through Socially Accountable Lenses

Presenter(s): Jeff Bachiu (*Northern Ontario School of Medicine, Undergraduate Medical Education, Sudbury, ON, Canada*), Rachel Ellaway (*Northern Ontario School of Medicine, Undergraduate Medical Education, Sudbury, ON, Canada*), Stacey Ritz (*Northern Ontario School of Medicine, Undergraduate Medical Education, Sudbury, ON, Canada*), Kathleen Beatty (*Northern Ontario School of Medicine, Undergraduate Medical Education, Sudbury, ON, Canada*)

Background: Assessment in medical education is multidimensional and involves varying levels of precision, validity and reliability depending on the constructs and contexts involved (Epstein, 2007). Assessment has political and sociocultural dimensions and it intersects with ideological and cultural constructs, such as social accountability.
This workshop engages in a critical evaluation of student assessment practices in medical education focusing on how social accountability can impact assessment in medical education. This workshop will use critical perspectives to deconstruct the concept of assessment in the context of socially accountable medical education, and contrast this with the realities of student learning and its complex interactions with contemporary assessment.
Intended Outcomes: At the end of this workshop participants will be able to:
• critically evaluate the role of assessment in medical education with a focus on interactions with social accountability.
• identify the semantic and political dimensions of student assessment and how they create and inflect the hidden curriculum.
• apply a range of critical analytical techniques to student assessment tools and protocols that are linked to social accountability.
Structure: Following small group discussion on assessment within medical education and an introduction to critical theory and its application to assessment medical education, participants will engage in a series of practical exercises critiquing different models of assessing student learning using a social accountability lens. The workshop will end in a discussion and consensus building exercise looking at how the concepts and themes of the workshop can be translated to participants' own practices.
Reference: Epstein RM. (2007). Assessment in Medical Education. NEJM 356(4); 387-96.

Who Should Attend: Medical educators and administrators involved in assessment
Level of Workshop: Intermediate

OTT-WA-2
Principles of Survey Design in Health Professions Education

Presenter(s): Anthony R. Artino (*Uniformed Services University of the Health Sciences, Preventive Medicine & Biometrics, Bethesda, MD, United States*), Ingrid Philibert (*Journal of Graduate Medical Education, Editorial Office, Chicago, United States*)

Background: Surveys are commonly used in graduate medical education. Unfortunately, few educators are familiar with the best practices of survey design. The purpose of this workshop is to provide medical educators with an introduction to a systematic process for creating valid and reliable surveys that can be used as assessment or research tools.
Intended Outcomes: Upon completion of the workshop, participants will be able to
1) Recognize how to use a systematic, 7-step process as the framework for survey design;
2) Demonstrate how to develop an appropriate set of items (a survey "scale") to characterize the educational construct being measured;
3) Identify common item-writing pitfalls in survey design; and
4) Define the purpose of expert validation, cognitive interviews, and pilot testing.
Structure: Lecture and interactive workshop
Who Should Attend: Faculty, evaluators, medical education researchers
Level of Workshop: Intermediate

OTT-WA-3
Learn to love the validity argument: How to plan, collect and organize validity evidence

Presenter(s): Rose Hatala(*University of British Columbia, Vancouver, Canada*), Ryan Brydges (*University of Toronto, Toronto, Canada*), David A. Cook (*Mayo Clinic, Rochester, United States)*

Background: A key consideration in the adoption and use of any quantitative assessment is the validity of (i) its scores and (ii) the decisions that result. Such validity judgments are facilitated by a carefully crafted validity argument. This workshop will help participants construct a persuasive validity argument for rigorously studying an assessment tool. We use Kane's unifying view of validity, wherein a validity argument is built in four domains: scoring, generalization, extrapolation and decision.(Kane,2006) This framework considers all validity as construct validity, and assimilates evidence from at least five sources: content, response process, internal structure, relations with other variables, and consequences.(Downing,2003) This workshop focuses on these evolving concepts, with emphasis on how to build validity arguments for clinical performance assessments.
Intended Outcomes: Participants will (i) gain an understanding of how to construct a validity argument by examining the evidence supporting the Mini-CEX and (ii) begin to develop a validity argument for an assessment tool used in their own setting.
Structure: 30 minutes: Interactive discussion: Kane's validity framework and the sources of validity evidence; 30 minutes: Small group activity: examine original research studies for evidence supporting the mini-CEX, and use that evidence to create a validity argument; 30 minutes: Small group activity: for an assessment tool commonly used by participants, use a structured approach to plan the collection, organization, and interpretation of validity evidence.
Who Should Attend: Faculty involved in assessment with an interest in validity concepts, such as program directors, assessment chairs, clinical educators.
Level of Workshop: Intermediate

OTT-WA-4
Introducing a practical guide to sequential testing: Educational & economic benefits

Presenter(s): Richard Fuller(*Leeds Institute of Medical Education, Leeds, United Kingdom*), Godfrey Pell (*Leeds Institute of Medical Education, Leeds, United Kingdom*)

Background: Institutions have conflicting pressures of quality, feasibility and cost in the assessment of students. Students desire fair assessment processes ideally completed by all in a single academic year. This usually takes the format of assessment of the whole student cohort, followed by a period of remediation and retesting for those who underperform. However, underperforming candidates are not a homogenous group and considerable difficulty persists in accurately identifying and profiling these 'at risk' students. Emergent work also reveals that traditional models of test-remediate-rest may not lead to sustained long term improvement in performance for these students, raising major implications in assessment policy for educational institutions (Pell et al 2009; Hauer et al 2009).
Intended Outcomes: Participants will explore theoretical and practical applications of sequential testing methodology, and explore approaches to managing cost-benefit analyses and quality issues in assessment policy. The workshop will conclude with a 'toolkit' to help participants in modelling & implementing sequential assessment in their own institutions
Structure: This interactive workshop will use practical exercises and round-table discussion in an overview of existing literature in this area, and the impact on assessment strategies and methodologies for underperforming students vs. competent students. The development of sequential test methodologies will explore a way of meeting this challenge. Modelling of assessment data will illustrate practical benefits on assessment quality, cost and student/faculty viewpoints.

Who Should Attend: This workshop has particular significance for those responsible for the design and delivery performance based assessment
Level of Workshop: Intermediate

OTT-WA-6
Applying an evidence-based rubric to assess the performance of educators: You be the judge!

Presenter(s): Maryellen Gusic(*Indiana University School of Medicine, Dean's Office, Indianapolis, United States*), Brian Mavis (*Michigan State University College of Human Medicine, Office of Medical Education Research and Development, East Lansing, United States*), Thomas Viggiano (*Mayo Medical School, Gastroenterology and Hepatology, Rochester, United States*)

Background: Significant progress has been made in defining the value of educational scholarship. Educators at many academic health centers continue to struggle with advancement and promotion because evaluators lack a standardized process to assess educator contributions. In response to this challenge, the AAMC Task Force on Educator Evaluation created resources to aid decision-makers in using clear, consistent and efficient evaluation processes for faculty whose career focus is in education. The product of a 3 year iterative process, the Toolbox for Evaluating Educators, provides indicators for assessment in each of the five domains of educator activity, teaching, curriculum development, assessment, advising/mentoring and educational leadership, using accepted frameworks and a common format.
Intended Outcomes: Attendees will:
1. Apply tools from the AAMC Task Force on Educator Evaluation to assess sections of a faculty member's portfolio
2. Justify decision-making based on the rigorous application of criteria for evaluation
3. Create a plan to disseminate the Toolbox at one's home institution to promote the use of performance metrics in processes to recognize and reward educators.
Structure: In this workshop, attendees will participate in interactive, hands-on activities to apply the criteria in the Toolbox to evaluate sample dossiers. To demonstrate the flexibility of the Toolbox, in facilitated large group discussions, the audience will explore, how the evidence-based rubrics can be employed to make fair and rigorous assessments consistent with their own institutional contexts and guidelines. Each participant will leave the session with ideas about how one can implement use of the Toolbox at his/her home institution.
Who Should Attend: Those who evaluate educators, those who mentor educators, and educators themselves.
Level of Workshop: Advanced

OTT-WA-7

How best to respond where doctors in training face problems with professionalism

Presenter(s): Richard B Hays(*Bond University, Faculty of Health Sciences and Medicine, Gold Coast, Australia*), Roger P Worthington (*Yale University, General internal Medicine, New Haven, United States*)

Background: Professionalism is not just an established part of medical education - it is something that clinicians need to reflect on throughout their training and subsequent careers. Education directors need to respond effectively to serious and/or significant unprofessional conduct displayed by some doctors, and interventions need to be in place before formal training is complete. In workplace settings, the stakes are always high and available options may not be the same as in student settings (e.g., where career outcomes could be a determining factor).
The authors have been working collaboratively on these issues for some years and seek to interact with other educators to share and build practical experience. The quality of postgraduate programs could be judged partly on how such matters are addressed.
Intended Outcomes: To allow delegates to debate practical approaches and share experiences of dealing with unprofessional conduct in the workplace through small-group discussion and case studies.
To consider how to factor in differences in culture and jurisdiction when assessing professionalism issues that appear generic.
To help the transition from theory to practice and enable educators to respond effectively to professionalism problems that arise.
Structure: 10 minute introduction, putting the issues in context; 40 minute small-group discussion (with cases brought by workshop leads and delegates); 40 minute plenary discussion and wrap-up.
Who Should Attend: Postgraduate clinical educators and program leads
Level of Workshop: Advanced

ORALS

OTT-OA-1: PLANNING ASSESSMENT

OTT-OA-1-1
Towards nationwide, collaborative assessment of medical student learning outcomes in Australia

David Wilkinson (*Macquarie University, Sydney, Australia)*

Background: Historically, medical schools in Australia have assessed student learning outcomes largely in isolation, through individual school's exams. There has been limited inter-school collaboration, and no benchmarking. A National Senior Teaching Fellowship

program aimed to developed a nationwide and collaborative model for Australia.

Summary of Work: Global models, ranging from national licensing exams, through collaborative models, were studied and evaluated. Wide ranging national consultation occurred, together with substantial lobbying.

Summary of Results: A nationwide model of shared assessment, open to all schools was developed, tested and refined. A business cases and governance structure was presented at a national consensus meeting. Pilot studies were developed with increasing numbers of schools involved and engaged.

Conclusions: A nationwide model of shared assessment has been developed, and is being implemented. It is owned by schools, and open to all. The model does not replace schools' ownership of and responsibility for assessment, but does improve collaboration, and enables quantitative benchmarking.

Take-home Messages: Substantial change in the approach to and delivery of assessment, at a national level is possible, but takes substantial effort and resources. Models need to be locally contextual.

OTT-OA-1-2
Programmatic longitudinal integrated assessment in a Longitudinal Integrated Clerkship

Denese Playford (*The University of Western Australia, The Rural Clinical School of Western Australia, Crawley, Western Australia*)

Background: Programmatic assessment allows students to be assessed across multiple clinical disciplines using multiple instruments over an extended time period. The main aim of such assessment is to give ongoing, timely feedback to students, so as to allow them to chose how to further develop their skills in relation to end of year performance expectations. The Rural Clinical School of WA Longitudinal Integrated Clerkship (LIC) introduced programmatic assessment in 2007.

Summary of Work: All students' assessment records from 2008 to 2012 were imported into SPSS. Correlations and logistic regression were performed on these raw scores.

Summary of Results: There were 365 students in the period 2008 – 2012. Their longitudinal integrated assessment positively predicted end of year performance in standard-set OSCE (R 0.538, p=0.000) and written papers (R 0.702, p=0.000). Individual pieces of assessment were also predictive, albeit less strongly. Clinical assessment correlated positively with the end of year OSCE: early formative OSCE (rxy 0.43), videotaped consultation (rxy 0.25), classic long case presentation (rxy 0.28). Subjectively marked narrative papers correlated well with the written exam: obstetrics (rxy 0.46), adult medicine (rxy 0.31).

Conclusions: Programmatic assessment is well used by students in a LIC to develop towards agreed end of year performance goals. The way they perform during the year positively predicts their end of year performance.

Take-home Messages: In a longitudinal assessment programme even non-standardised assessments, such as written work that allows students to explore issues in patient care, contributes meaningfully to predicting end of year results.

OTT-OA-1-4
Blueprinting assessment to enhance constructive alignment

Josephine Boland (*National University of Ireland Galway, School of Medicine, Galway, Ireland),* Yvonne Finn (*National University of Ireland Galway, School of Medicine, Galway*), Rosemary Geoghegan (*National University of Ireland Galway, School of Medicine, Galway*)

Background: Blueprinting of assessment is advocated as an essential step for enhancing constructive alignment and validity of assessment. It is of particular value in medical education, given the high stakes of the examinations and the need to ensure that we graduate 'fit-for-purpose' doctors. Blueprinting represents one of the elements of a comprehensive curriculum mapping strategy. It can be addressed at a range of levels and is promoted as a precursor to effective assessment/test construction.

The process of academic simplification of the undergraduate curriculum in the School of Medicine, National University of Ireland Galway has had particular consequence for the clinical years. The associated challenges represented an opportunity to implement blueprinting of assessment.

Summary of Work: Medical educators and clinicians from across the clinical years collaborated on a blueprinting process – at macro level (i.e against mandated Medical Council outcomes), at meso/module level (i.e. mapping assessment techniques against learning outcomes) and at a micro level (i. e. mapping assessments against domain, content areas or cases of importance). Initial work has focussed on OSCEs and MCQ/EMQs.

Summary of Results: Our experience suggests that blueprinting can serve as an effective quality assurance tool and provides evidence that informs assessment design. Moreover, the collaborative process has fostered greater integration within a spiral curriculum. It has helped identify duplication and omissions of key content areas or under-represented domains.

Conclusions: Blueprinting is not just an academic or theoretical exercise. It supports deliberative and collaborative assessment planning, at programmatic (rather than module) level, ultimately, enhancing student learning.

Take-home Messages: Blueprinting is a key component of any curriculum mapping strategy.

OTT-OA-1-5
Putting the pieces together to create a defensible blueprint: A subject matter expert-based process

Claire Touchie (*Medical Council of Canada, Ottawa, Canada),* Andrea Gotzmann (*Medical Council of Canada, Research and Development, Ottawa*), Cindy Streefkerk

(*Medical Council of Canada, Ottawa*), Andre De Champlain (Medical Council of Canada*, Research and Development, Ottawa*)

Background: In order to provide valid interpretations of test scores, validity evidence of appropriate content must be obtained. In the case of a high-stakes licensing examination, the process should ensure that the necessary knowledge, skills and behaviors required for entry into practice is included in the overall blueprint.

Summary of Work: A panel of 12 physician subject matter experts (SMEs) was asked to develop a blueprint for assessments leading up to two decision points: (1) entry into supervised practice (residency), and (2) entry into unsupervised (independent) practice.

In order to better inform this process, the following sources were presented to our SMEs for review: (1) a study on resident entrustable professional activities; (2) the results of a national survey of physicians and of the public; (3) a study on the incidence and prevalence of diseases; (4) and a paper on the future of assessment.

Summary of Results: The outcome was a two-dimensional blueprint defined by the following axes: (1) Dimensions of Care (Health Promotion and Illness Prevention, Acute Care, Chronic Care, and Psychosocial Aspects of Care) and (2) Physician Activity (Assessment/Diagnosis, Management, Communication, and Professional Behaviors). Clear definitions, substantiated with observable behaviors, were provided by our SMEs for these eight broad categories. The SMEs also suggested test specifications for each of the blueprint dimensional categories as a function of both decision points.

Conclusions: The present study provides a clear, defensible process by which a blueprint can be developed and defended to candidates, practicing physicians as well as the public.

OTT-OA-2: THE EXAMINER

OTT-OA-2-1
Recognising the responsibility of senior clinicians in assessment

David CM Taylor (*University of Liverpool, School of Medicine, Liverpool, UK*)

Background: Clinicians are under increasing pressure to meet clinical targets, and academics are under increasing pressure to improve the student experience and meet curriculum outcomes. Clinical academics face both challenges, but have the ultimate responsibility for ensuring that prospective graduates are likely to be safe and effective practitioners.

Summary of Work: A series of semi-structured interviews were held with senior clinical academics, with a view to determining the factors that motivate or demotivate them when it comes to educating medical students. Study outcomes include restructuring the recognition and support given to clinical academics. Three of the question strands related to a clinical academic's responsibility to the students, the medical school and their specialty.

Summary of Results: The discussions revealed the tensions between supporting students, and potentially acting as the final arbiter of progression. This was overlaid with the desire to give the students the best experience possible, help identify future members of the clinician's specialty, and of course to ensure the development of competence and the appropriate standards of professionalism

Conclusions: Assessing knowledge and technical skills presented the participants with few problems. In contrast assessing the professional ability of weak students in the workplace is demanding in terms of both time and emotional energy, and is an essential role for senior clinicians.

Take-home Messages: Assessing the clinical and professional ability of students takes time, and is important enough to be a real demand on the time of senior clinical academics. Mutual support and focussed training is needed but it is essential that time taken in assessment is protected, recognized and rewarded.

OTT-OA-2-2
Is there consistency of judgement between examiners in performance-based assessment? A case study in medical education

Amy (Wai Yee) Wong (*The University of Queensland, School of Medicine, Centre for Medical Education Research and Scholarship, Brisbane, Australia*)

Background: Clinical competency assessment evaluates medical students' ability in undertaking clinical tasks. There are a number of reliability and consistency issues when a grading approach is applied to assess student performance in clinical competency examinations. These issues are exacerbated by the fact that the majority of examiners are clinical practitioners rather than educators and they may have differing standards regarding student performance. This study aimed to explore the factors which may affect examiners' judgement of medical students' performance, and to recommend possible effective strategies in examiner training.

Summary of Work: A mixed-methods approach was adopted. Descriptive statistics were used to compare the marks given by the examiners in the Year 2 (pre-clinical) and 4 (final year) Objective Structured Clinical Examination (OSCE). The size of each cohort was approximately 500 students. Qualitative data were collected by semi-structured interviews with examiners. Open and analytical coding was used to explore the factors that influenced examiners' judgement of medical students' performance in clinical examinations.

Summary of Results: Data revealed there were marking discrepancies among examiners in the OSCE. Characteristics of examiners play an important role in determining the marks given to the students. Provision of effective and adequate training to examiners on performance assessment is necessary, particularly in assessing medical students' clinical competency in high-stakes assessment.

Take-home Messages: Evidence-based recommendations for, and effective strategies for use in, the faculty development of teachers in assessment, and examiner training are needed to improve the consistency and reliability of examiner judgement on performance-based assessment.

OTT-OA-2-3
Trainers' Understanding and Application of Assessment in Training Programmes

Helen M Goodyear (*Health Education West Midlands, Medical Education, Birmingham, UK),* Taruna Bindal (*Alexandra Hospital, Paediatrics, Redditch*), David Wall (*University of Dundee, Centre for Medical Education, Dundee*)

Background: Doctors have a professional obligation to contribute to education and training of health care professionals and students. It is essential that trainers acquire the appropriate knowledge, behaviours and continue to develop to carry out this role. Since 2009, all UK doctors involved in clinical training have been encouraged to attend a Teaching the Teachers course (TtT) as well as gaining medical education qualifications so as to be more effective in trainee assessment. The General Medical Council requires trainers to attend an update course every 3 years.

Summary of Work: A 20-item questionnaire was completed by attendees at Paediatric regional assessment update meetings for educational supervisors.

Summary of Results: Of 41 respondents, 37 (88%) had completed a generic TtT course, 15 (36%) an assessment and appraisal course and only 4 (10%) had a medical education qualification. 8 (20%) trainers had helped with the regional annual review of competence progression. No-one confused the concepts of appraisal and assessment. Trainers felt confident in using workplace based assessments (WBAs) and handling educational supervision meetings but were less confident in ePortfolio use and dealing with doctors in difficulty. There was no difference in any results by age, gender, years as a consultant or place of basic qualification.

Conclusions: Trainers felt confident in their understanding of assessment and using WBAs which they utilise regularly. However trainers needed the refresher courses for updating themselves on ePortfolio use and dealing with doctors in difficulty which occurs less frequently.

Take-home Messages: The key factors affecting trainers confidence in assessment is frequency of use of tools and experience.

OTT-OA-2-4
Theatrical dove or hawkish wassock? The consistency of examiner behaviour in a national postgraduate clinical examination

Richard Wakeford (*University of Cambridge, Hughes Hall, Cambridge, UK)*

Background: Since Hartog and Rhodes' work in the 1930s, examiners have been known to vary in their behaviour on at least three continua: their stringency ('the hawk-dove effect'); the breadth of their marking distribution, whether restrained or 'theatrical'; and the extent to which their candidates' rank order agrees with that under other examiners. (We have coined the term 'wassock' for the examiner whose marks correlate negatively or trivially with others'.) The last is particularly important, as its level will impact on reliability and have implications for test length. Whilst examiner stringency has been studied, we have not found any relevant studies of the consistency of examiner performance, year on year.

Summary of Work: In the MRCGP Clinical Skills Assessment, taken by all UK GP trainees towards the end of their postgraduate training, candidates' 13 consultations are each marked by a single trained examiner. We studied three years' of examiner (n=187) marks of individual consultations, rank ordering examiners on all three continua in each year. We calculated the mean between-year correlation for each continuum.

Summary of Results: The mean correlation on 'level' was 0.51, on 'spread' (SD) 0.49, and on 'agreement with colleagues' marks' 0.22. (All individual correlations p<.01)

Conclusions: Whilst examiners vary quite consistently regarding the level and spread of the marks they award, the extent of their agreement with other examiners is less consistent. This may be due to examiner 'case specificity of performance'.

Take-home Messages: The psychometric implications of the level and consistency of inter-examiner agreement are considerable and need further study.

OTT-OA-2-5
Rating Demands and the Impact on Rater Performance and Behavior: A Mixed Methods Study

Walter Tavares (*University of Toronto, Wilson Centre, Toronto)*, Shiphra Ginsburg (*University of Toronto, Toronto*), Kevin Eva (*University of British Columbia, Vancouver*)

Background: Research suggests that optimization of assessments of competence requires a reliance on a rater's inherent cognitive abilities and behaviors. In this study we asked: In assessing clinical competence, what effect do rating demands have on indicators of rating quality? Secondly, what cognitive strategies do raters engage under high demand conditions?

Summary of Work: We conducted a sequential mixed-methods study by randomly assigning participants to one of four conditions (in a 2x2 design) and having them rate 3 pre-recorded clinical performances. One factor involved the number of dimensions to be rated (7vs2). The second was the requirement (or lack thereof) to conduct additional extraneous, ecologically valid tasks. Outcome measures included observational completeness and inter-rater reliability. "High-load"

participants were interviewed post-task to explore their rating strategies.

Summary of Results: Examining the 2 dimensions common to both groups, ANOVA revealed a significant main effect of the number of dimensions to be rated on both observational completeness and inter-rater reliability. Relative to those in the 7-dimension group, participants in the 2-dimension group identified more features (Dimension 1=F(1,80)=31.6, p=<.001; Dimension 2=F(1,80)=94.3, p=<.001) and achieved higher reliability (G=0.75 vs. 0.63). "High-load" participants' observational strategies suggested they minimized extraneous load and managed intrinsic load in idiosyncratic ways.

Conclusions: As rating demands increase, observational completeness declines leading to decreased inter-rater reliability. Rating strategies in high-load conditions are varied and appear to include attempts to reduce the amount of mental workload.

Take-home Messages: To optimize rater-based assessments of competence, rating demands may need to be aligned with inherent cognitive abilities.

OTT-OA-2-6
Complexities in multidisciplinary assessment: Whose variance matters?

Rachel Fisher (*McGill University, Anesthesia, Montreal, Canada),* Lily HP Nguyen (*McGill University, Otolaryngology Head and Neck Surgery, Montreal*), Ilana Bank (*McGill University, Pediatric Emergency Medicine, Montreal*), Meredith Young (McGill University*, Centre for Medical Education, Montreal*)

Background: There are increasing numbers of educational programs and innovations focusing on interprofessional and interdisciplinary teams. As these programs fill a strong educational need, they are more likely to be integrated into formalized assessment programs. In addition to the complexities of assessing teams rather than individuals, the move towards including multiple health professions and specialties begs the question – who is the 'right' assessor?

Summary of Work: An interdisciplinary team simulation-based course on managing crises was offered to pediatric emergency, anesthesia and otolaryngology senior residents. Residents' performance was video recorded and assessed using both the Ottawa GRS and Mayo Team scales. Raters included two from each participating specialty and one from a non-participating specialty but with non-technical skills content expertise. Rater training was done on two standard videos, and raters discussed any significant differences.

Summary of Results: Analysis was focused on categorizing sources of rater variance. Foci included amount of variance attributed to: different raters (hawks vs. doves), specialty of rater, and specialty of individual being assessed. Preliminary results suggest that raters from different specialties differ in their assessment of residents, particularly when specialty of the rater and the individual being assessed align.

Conclusions: An important contributor to rater variance in a multidisciplinary environment may be the specialty of the rater and of the individual being assessed.

Take-home Messages: With increased use of interdisciplinary training programs, careful attention must be paid to rater characteristics. This remains a particular concern for assessment equivalence as these programs become included in formalized assessment.

OTT-OA-3: PROGRESS TESTING

OTT-OA-3-2
Moving mountains in medical curriculum: Two-years' experience with progress test

Viktor Riklefs (*Karaganda State Medical University, Karaganda),* Raushan Dosmagambetova (*Karaganda State Medical University, Karaganda, Kazakhstan*)

Background: Designing and improving medical curriculum is a very complex task which is too often oversimplified. The judgment is made using fragmentary opinions of students and professors without considering the objective evidence. The progress test might become an invaluable tool in monitoring curricular efficiency, since it provides unjaundiced data on achieving final learning outcomes by trainees.

Summary of Work: For two years we have been administering formative progress test to more than 4000 medical students of our University. We also surveyed them on learning styles and analyzed factors such as instruction language (Russian, Kazakh), rural or urban home region, gender, ethnicity, etc. Focus-groups among higher and lower achieving students helped to explore the reasons behind students' academic achievements.

Summary of Results: Progress test and following research allowed collecting enough evidence of students achieving learning outcomes differently depending on their social, lingual background, and style of learning. In particular, the lack of convergent thinking, i.e. the ability to find the best possible solution in given situation, significantly deteriorated achievement of learning outcomes.

Conclusions: Based on the evidence gathered from progress test, we were able to come with recommendations on improving the medical curriculum. The evidence persuaded administration and professors on the necessity of curricular reform, and the shift to more open academic environment where the students were not limited in their critical thinking.

Take-home Messages: The mountains of curricular innovations could be moved, if the direction of change is well defined and based on evidence. The progress test could serve as a tool to direct and guide curricular reforms.

OTT-OA-3-3
Growth models and progress tests: A happy marriage?

Chris Ricketts (*Plymouth University, Peninsula Schools of Medicine & Dentistry, Plymouth, UK)*, Lee Coombes (*Plymouth University, Peninsula Schools of Medicine & Dentistry, Plymouth*), Adrian Freeman (*University of Exeter, Medical School, Exeter*)

Background: Most assessment systems base their decisions on a person's current level of achievement. However, educators have long considered that growth of ability should also be taken into account. Thus, a person who has not achieved the current standard but who shows a strong growth rate towards a later standard could be rewarded as well as one who has reached the current standard. Accordingly, a number of simple growth models have been proposed and some are used in the American school system (Castellano & Ho, 2013).

Summary of Work: We have applied the growth models most commonly used in school assessment systems to progress tests of medical knowledge to investigate their utility.

Summary of Results: Test-to-test correlations are about 0.6. Projection methods based upon multiple tests provide good predictions but with an R2 value ranging from 45-65%. However, they are better than simple time-based predictive methods. A Bayesian approach also provides good predictions. The quality of the prediction depends upon the scale used.

Conclusions: Growth models can be used with progress tests, but there is no single method for estimating growth and hence predicting future performance which is clearly better than any other.

Take-home Messages: Prediction of student performance continues to be difficult, even with good longitudinal data from progress tests.

OTT-OA-3-4
Progress Testing: An innovative approach to assessment of competency-based milestones in clinical skills

Lauren Block (*Hofstra North Shore-LIJ School of Medicine, General Internal Medicine, Hempstead, NY, USA)*, R. Ellen Pearlman (*Hofstra North Shore-LIJ School of Medicine)*, Judith Brenner (*Hofstra North Shore-LIJ School of Medicine*), Joseph Weiner (Hofstra North Shore-LIJ School of Medicine), Alice Fornari (*Hofstra North Shore-LIJ School of Medicine*), Marie Barilla-Labarca (*Hofstra North Shore-LIJ School of Medicine*)

Background: Growing evidence suggests that medical student assessment should be developmentally sequenced, competency-based, individualized, and integrated across disciplines. To our knowledge, use of a single standardized case as a progress test to compare developmentally-based clinical skills competencies has not been described.

Summary of Work: As the culmination of a longitudinal clinical skills curriculum, we designed a single standardized case administered in a clinical skills center to assess competency in communication, patient care, and professionalism among first and second year students. A committee of standardized patients and

faculty reviewed student performance based on proposed milestones expected for the student's year.

Summary of Results: 58 first year and 33 second year students at Hofstra North Shore-LIJ School of Medicine completed the examination. 81% (47) of first-year and 85% (28) of second-year students met developmental expectations in communication, 90% (52) and 94% (31) in patient care, and 84% (49) and 76% (25) in professionalism. The 23 students who did not meet expectations were offered remediation through individual feedback, videotape review, and a repeat exam. The implications of these data will be discussed.

Conclusions: A single standardized case served as a developmental progress test to identify students attaining milestones in clinical skills and highlight areas for feedback and remediation.

Take-home Messages: As developmental milestones are proposed, it is critical to evaluate whether students meet the recommended competencies. We designed a single case to serve as an integrated clinical skills progress test. Future work will include prospective development of a competency-based instrument for longitudinal assessment in clinical skills.

OTT-OA-3-5
Can an OSCE be used as a progress test?

Debra Pugh (*University of Ottawa, Medicine, Ottawa, Canada)*, Claire Touchie (*University of Ottawa, Medicine, Ottawa*), Susan Humphrey-Murto (*University of Ottawa, Medicine, Ottawa*), Timothy J. Wood (University of Ottawa*, Medicine, Ottawa*)

Background: Written progress tests are frequently used to assess learner performance over time because they allow for a standardized benchmark of achievement. However, there are no published studies on the use of an OSCE as a progress test.

At the University of Ottawa, first to fourth-year Internal Medicine residents participate in an annual OSCE (IM-OSCE). Similar to written progress tests, content and difficulty are set at the highest level of ability (i.e., at the PGY-4 level). The purpose of this study was to explore the use of the IM-OSCE as a progress test, by demonstrating an increase in individual scores over time.

Summary of Work: Data from eight administrations of the IM-OSCE were analyzed. OSCE scores were scaled to standard scores with a mean of 500 and SD of 100. Progression of resident performance over their three years of core training was compared to see if individual scores increased with time. For each administration of the exam, differences in scores by PGY level were also compared.

Summary of Results: Of the residents who participated in the IM-OSCE in each of their three core years of training (n = 42), individual performance improved from a mean of 438.2 (83.6) for PGY-1s, to 531.1 (71.8) for PGY-3s (p<0.001). A total of 243 residents participated in at least one IM-OSCE from 2006-2013. Scores increased significantly with level of training and ranged from 431.5 (83.1) for PGY-1s to 572.0 (83.7) for PGY-4s (p<0.001).

Conclusions: Individual IM-OSCE scores increase with training.
Take-home Messages: This study provides support for the use of the IM-OSCE as a progress test for trainees from different levels of training.

OTT-OA-3-6
Progress Testing and Rasch Modelling

Vernon Mogol (*University of Auckland, Medical Programme Directorate, Auckland, New Zealand).* Steven Lillis (*University of Auckland, Medical Programme Directorate, Auckland*), Lee Coombes (*University of Plymouth, Peninsula Schools of Medicine and Dentistry, Plymouth*), Adrian Freeman (University of Exeter*, Medical School, Exeter*), Warwick Bagg (*University of Auckland, Medical Programme Directorate, Auckland*)

Background: Progress tests (PTs) are written assessments using single best answer format and are taken several times a year by all students in a medical programme regardless of seniority. There is considerable educational value in such testing but two contentious issues remain: use of norm referencing and use of formula scoring (negative marking). Computerised adaptive testing (CAT) using Rasch modelling may represent a solution to both of these. However, there is little published research on how PT data fit the Rasch model.
Summary of Work: A test with linking design was simulated using data from a PT administered to Years 2 and 4 students of University of Auckland. The data were subjected to Rasch analysis using the RUMM2030 software.
Summary of Results: The data fit the model better when the Don't Know response is treated as a missing and not an incorrect response (fewer analysis runs with fewer item misfits). Differential item functioning was also examined and when addressed further improved fit. The simulated PT was well-targeted (pitched correctly) and had acceptable internal consistency.
Conclusions: This research demonstrates that data from PT with linking design conform to the Rasch model. The method produced useful psychometric data that indicated adequate fit and good internal consistency.
Take-home Messages: Rasch modelling may represent a viable option for analysis of PT data. The technique may allow for CAT as the mode of delivery with benefits in removing formula scoring and norm referencing.

OTT-OA-4: TECHNOLOGY AND ASSESSMENT

OTT-OA-4-1
Point-of-view High Definition Video Assessment: The Future of Technical Skills Training

D Leslie (*Morriston Hospital, Anaesthetics, Swansea, UK),* M Oliver (*University Hospital of Wales, Anaesthetics, Cardiff*), M R Stacey (*University Hospital of Wales, Anaesthetics, Cardiff*)

Background: Junior doctor training in the UK has undergone significant changes, with reduced working hours being a driving force. This necessitates novel assessment tools, to allow out-of-hours work to provide more training opportunities. Workplace-based assessments, in their current form, are time consuming to arrange, and not treated uniformly by assessors. Video assessment is well established in general practice, and we believe it could complement current anaesthetic training.
Summary of Work: We have trialled high definition video sport glasses to record anaesthetists performing central neuraxial blockade. The glasses are modified for recording close up procedures, rotating the camera sensor to portrait wide-angle, and removing the lenses to make them less intrusive. Procedures are also filmed by an observer wearing a second pair of glasses with simultaneous playback enabling detailed feedback.
Summary of Results: All trainees believe it is fairer, more accurate and more useful, than current assessment tools. It provides a non-intrusive way of recording procedures, enabling trainees to perform as they usually would. Regular use allows development of a portfolio of procedures, demonstrating consistent levels of performance, and therefore competence.
Conclusions: Our experience suggests it is reliable, valid, more educationally beneficial than current assessment tools. The video produced represents excellent value, with negligible running costs, more than justifying the initial, relatively low cost of purchasing the glasses.
Take-home Messages: Changes to medical training are inevitable and necessary. We believe video analysis of practical procedures using recordable video sports glasses would be a valid, reliable, and educationally beneficial way of improving assessment, and therefore training.

OTT-OA-4-2
The Use of Student Response System in High Stakes Examinations - A Feasibility Study

Kalyani Premkumar (*University of Saskatchewan, Department of Community Health & Epidemiology, College of Medicine, Saskatoon, Canada)*, Cyril Coupal (*University of Saskatchewan, Saskatoon*)

Background: Student Response Systems (SRS) are primarily used by instructors for student engagement and/or formative assessment and feedback. With the advent of SRS with a display screen and built in data storage capacity,
there is scope for its usage in summative examinations.
Summary of Work: We examined the feasibility of its use in highstakes summative assessments in three different courses across two colleges at the University of Saskatchewan. In our study we identified student and instructor acceptance of the
technology, the benefits and barriers in delivering examinations using the SRS. End-of-course online surveys to students with closed and open-ended questions and instructor interviews were used to gather data.

Summary of Results: In general students were accepting of the use of this technology in high stakes examinations and found it engaging and satisfying, primarily because of instant feedback. The Instructor found the process less time-consuming, efficient, and more secure as compared to the use of scan sheets.

Conclusions: SRS is best used for examinations with multiple choice
type questions or questions with minimal text/mathematical symbols entry and examinations with up to 100 questions.

OTT-OA-4-3
Using Tablets for Written and OSCE Exams: Principles and Experiences

Jörn Heid (*Heidelberg University, International Consortium for Assessment Networks, Heidelberg, Germany)*, Konstantin Brass (*Heidelberg University, International Consortium for Assessment Networks, Heidelberg*), Achim Hochlehnert (*Heidelberg University, International Consortium for Assessment Networks, Heidelberg*), Andreas Möltner (Heidelberg University, *International Consortium for Assessment Networks, Heidelberg*), Jobst-Hendrik Schulz (*Heidelberg University, International Consortium for Assessment Networks, Heidelberg*), Jana Jünger (*Heidelberg University, International Consortium for Assessment Networks, Heidelberg*)

Background: Due to the shortcomings of normal PCs educational institutions still use pen and paper for written exams and OSCE documentation, although manually transferring the results is often erroneous. Among the major problems of PC-based exams are the need for computer rooms and the distraction when grading students with laptops at OSCE stations. To solve these issues and to address the lack of modern question types and multimedia support, tablets offer a great advantage because of their small size and usability.

Summary of Work: The International Consortium for Assessment Networks (ICAN) developed an app for written exams used by students and an app for OSCE documentation used by the examiners. Both apps are built upon HTML5 to support different platforms like iOS, Android and Windows 8. Additionally we focused on security, flexibility and legal certainty.
To be entirely location-independent, the apps can run offline without a WIFI connection during the exams. The tablets are handed out and brought back after the exams so no technical staff is needed on-site.
Because the apps automatically save screenshots of the selections, a neutral documentation is generated for legal certainty.

Summary of Results: The apps are successfully in use since 2012 with up to 200 iPads in written exams and in OSCEs with up to 160 students in 20 stations.

Conclusions: The usage of tablets has proven to be more flexible and it has reduced the amount of errors and needed staff compared to paper- or desktop based assessment.

Take-home Messages: Tablets offer the perspective of location-independence without compromising on security or neglecting financial budgets.

OTT-OA-4-4
Contrasting three automated scoring approaches for Bilingual Medical Licensing Examination

Syed Muhammad Fahad Latifi (*University of Alberta, Educational Psychology, Edmonton, Canada)*, Mark J. Gierl (*University of Alberta, Educational Psychology, Edmonton*), André-Philippe Boulais (*Medical Council of Canada, Research and Development, Ottawa*), André De-Champlain (Medical Council of Canada*, Research & Development, Ottawa*)

Background: Most medical licensing examinations target some form of clinical decision-making (CDM) ability, often via constructed-response (CR) or open-ended questions. Although these types of questions are essential for measuring knowledge in specific clinical situations, the scoring process can be time-consuming and resource intensive. With the adoption of computer-based testing in medical licensing, the demand for consistent and time-efficient scoring has increased. We propose a framework for technology enhanced scoring in medical licensing examination.

Summary of Work: Open-source natural language processing technology was employed for developing an automated scoring process. Pre-scored English and French responses from eight CDM-CR questions were extracted. Three machine learning algorithms were then employed for training and evaluating the automated score prediction models, across two linguistics groups. To assess the viability of this automated scoring methodology, exact agreement percentages and Kappa coefficients were computed between human (physician) and computer scores.

Summary of Results: Across eight CDM-CRs, the prediction accuracy for English responses ranged from 90.6% to 98.7%, with corresponding Kappa values of 0.79 and 0.97. Similarly, the accuracy for French responses ranged from 88.2% to 98.8%, with corresponding Kappa values of 0.62 and 0.90. Overall, average human-computer agreement was established at 96% for English, and 95% for French, CDM cases.

Conclusions: Automated scoring of CDM-CR questions highly concurred with physician-based scoring.The suggested framework is adaptable and has provided strong evidence to support the use of automated scoring in high-stakes multilingual medical licensing context.

Take-home Messages: Inexpensive technology is available for testing organizations willing to make the transition towards automated scoring systems.

OTT-OA-4-5
Going paperless? Issues in converting a surgical assessment tool to an electronic version

Nancy Dudek (*University of Ottawa, Medicine, Ottawa, Canada)*, Steven Papp (*University of Ottawa, Surgery,*

Ottawa), Wade Gofton (*University of Ottawa, Surgery, Ottawa*)

Background: The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) is a paper-based tool that assesses a trainee's competence to perform an entire surgical procedure. Paper systems have concerns that may be addressed by converting to an electronic system. This study describes issues related to the implementation of the electronic version, the O-SCORE(e).

Summary of Work: Partnering with Resilience Software, we developed the O-SCORE(e). The tool itself was not changed. Trainees logged cases electronically. Automatically an evaluation was posted to the staff account. Orthopedic Surgery, General Surgery and Urology participated. Six cases in Orthopedic Surgery, five cases in General Surgery and four cases in Urology were identified as "O-SCORE cases". We identified an extremely low participation rate. Therefore, we surveyed the participants to better understand why the O-SCORE(e) was not being completed.

Summary of Results: Over the first three months 440 "O-SCORE cases" were logged (Orthopedic Surgery-173, General Surgery-179, Urology-88). There were 27 O-SCORE(e)s completed in Orthopedic Sugery, 4 in Urology, and 0 in General Surgery. Both staff and residents indicated that the O-SCORE is a useful tool that improves their day-to-day feedback. However, they identified several barriers in using the O-SCORE(e) such as increased time compared to the paper version and technological issues related to the log-in procedure.

Conclusions: Despite the perceived usefulness of the O-SCORE, the electronic version was not well adopted by users as it was more time consuming and took more effort.

Take-home Messages: Electronic assessment tools need to be as convenient as paper in order for them to be accepted by users in residency training programs.

OTT-OA-4-6
Incorporating an external assessment program into the educational continuum: The ARDMS/WINFOCUS cardiac ultrasound pilot

Elizabeth Langston (*ARDMS, , Chicago, IL, USA)*, R. Eugene Zierler (*University of Washington, Seattle, WA*), Gabriele Via (*WINFOCUS, Milan*), Enrico Storti (*WINFOCUS, Milan*), Ellen Julian (*ARDMS, Rockville, MD*)

Background: The American Registry for Diagnostic Medical Sonography (ARDMS) developed an assessment program for integration into the WINFOCUS Focused Cardiac Ultrasound (point-of-care ultrasound) course.

Summary of Work: In contrast to the traditional high-stakes ARDMS sonography credentialing examinations, this program was created for remote delivery on personal computers and tablets. Item types included hot-spots, multiple response, drag-and-drop, drop-down, and fill-in, and many incorporated still images or videos.

Twenty-one participants, mostly physicians, completed both pre-course and post-course assessments; 13 completed the six-month follow-up assessment.

Summary of Results: The participants who completed all three assessments averaged 70% correct pre-course, 84% post-course, and 81% after six months. The pre- and post-course scores indicate increased knowledge, which appears to have been sustained six-months out. The domain related to ultrasound clinical implementation was consistently more difficult than the other domains, suggesting to the faculty a need to strengthen education in this area and adding credence to the importance of clinical application.

Conclusions: The integration of external assessments into the point-of-care ultrasound curriculum documented participant learning and identified content areas needing emphasis.

Take-home Messages: The test-building and psychometric skills required to create high-stakes credentialing exams can be repurposed to providing assessments for integration into a training course, documenting acquisition and retention of the ability to use ultrasound in non-traditional settings and applications, and providing feedback to faculty on course quality.

OTT-OA-5: SIMULATED/STANDARDIZED PATIENTS

OTT-OA-5-1
Capturing Communication Skills: Creating an online self-guided module for teaching communication skills to Standardized Patients

Carrie K Bernat (*University of Michigan Medical School, Ann Arbor, USA)*, Stacie L Buckler (*University of Michigan Medical School, Ann Arbor*)

Background: The process of training Standardized Patients in communication skills lent itself to a lack of standardization and redundancy in training. We developed an online communication skills training module with the primary objective of training SPs to assess students on their ability to communicate effectively.

Summary of Work: Using the tenets of adult and online learning theories, we hypothesized that the module would improve standardization of training and increase the accuracy of SP scoring, increase the portability of training and allow for tailoring the module to accommodate interprofessional audiences.

Summary of Results: SPs (N=85/86) completing the module were surveyed for their feedback. We found that 66% felt that an online module is an effective method for learning communication skills, 87% found the module easy/very easy to use and 89% felt prepared to score communication skills after completing the module.

We measured overall accuracy scores of SPs at three points: SPs trained prior to the module's existence, veteran SPs after completing the module and new SPs

not trained previously. We found that SPs in all three groups scored with approximately the same degree of accuracy.

Conclusions: We found SPs scored as accurately after implementing the module as they did before. This result, combined with the cost and time savings realized by our Program indicates a positive intermediate outcome. We were able to identify areas for ongoing development for Standardized Patient training as well as use for faculty development and student learning.

Take-home Messages: An online communication skills module is an effective and efficient method for training SPs to assess student performance.

OTT-OA-5-2
How to Evaluate the Role-play of Simulated Patients: Development and Validation of a New Questionnaire

S Kujumdshiev (*University of Zurich, Dean`s Office of the Faculty of Medicine, Zurich, Switzerland)*, Ch Conrad (*University of Frankfurt, Frankfurt*), M Zupanic (*University Witten-Herdecke, Witten*), TOF Wagner (University of Frankfurt*, Frankfurt*), M Fischer (*University of Munich, Munich*)

Background: Reliability and validity of an OSCE depend amongst others on the highly standardized role-play of the simulated patient.

What are the criteria for an authentic role-play of simulated patients in an OSCE? How is it possible to operationalize these criteria in a short questionnaire to be able to assess the role-play? We developed a questionnaire and tested its reliability and validity.

Summary of Work: In collaboration with experts, teachers, simulated patients and students we collected and prioritized criteria of good role-play and operationalized them.

In pre-test A 18 raters and in pre-test B 18 other raters used the questionnaire for 5 different video role-plays. After the first pre-test items were refined and subscales enhanced by improved item alignment.

The revised questionnaire was validated through an OSCE with parallel use of the validated MaSP questionnaire. All examiners were trained to use the MaSP, our questionnaire and the OSCE rating.

Summary of Results: In pre-test A the Cronbach's alpha of the subscales was 0.79 for communication, 0.58 for information and 0.90 for realism. Pre-test B showed an increased internal consistency of 0.82 for information. Cronbach's alpha of the OSCE was 0.86. MaSP2 reliability was 0.63.

Analysis of subscales revealed a Cronbach's alpha of 0.77 for communication, 0.85 for information and 0.69 for realism.

Conclusions: We defined appropriate criteria of an authentic role-play of simulated patients, operationalized them and created a valid tool with reproducible ratings. After extensive refinements in subscale information ratings improved further.

Take-home Messages: Our questionnaire is a highly reliable tool for quality management of simulated patients' role-play.

OTT-OA-5-3
Use of an "exam readiness tool" to ensure quality of standardized/simulated patient performance in high stakes objective structured clinical exams (OSCEs)

Carol OByrne (*Pharmacy Examining Board of Canada, Toronto, Canada)*, Cathy Smith (*Pharmacy Examining Board of Canada, Toronto*), John Pugsley (*Pharmacy Examining Board of Canada, Toronto*)

Background: To meet quality standards for high-stakes OSCEs, it is necessary to ensure high quality, standardized performance of the Standardized Patients (SPs) One way this can be assured is through providing explicit training goals and performance criteria, and assessing the quality of SPs' performance in and after training and during the assessment. There is some literature concerning validated instruments that have been used to assess SP performance in formative contexts but very little related to high stakes contexts.

Summary of Work: The Pharmacy Examining Board of Canada developed SP performance criteria and an "exam readiness assessment tool", to improve SP performance quality and consistency within and across multiple exam sites for a national high stakes OSCE. Selected video recordings of SP performance were reviewed and assessed over a 3-year period, following implementation. This assessment tool is also being adapted for use in a medical education context.

Summary of Results: Using the "exam readiness tool", SP Trainers/Educators (SPTs)(Trainers) assessed SP performance, and provided constructive feedback to improve and standardized SPs' performance. SPs and SPTs were clearer about the training goals; SP accuracy and consistency improved within and across sites.

Conclusions: The exam readiness assessment tool is useful and effective to improve the quality and consistency of SP performance and can be adapted for different contexts.

Take-home Messages: High quality, standardized SP performance can be achieved by applying quality assurance processes in and after training and during the assessment.

OTT-OA-5-4
Is it possible to train standardized patients for high stake OSCE in an hour?

Rachmadya Nur Hidayah (*Faculty of Medicine, Universitas Gadjah Mada, Department of Medical Education, Yogyakarta, Indonesia)*, Gandes Retno Rahayu (*Faculty of Medicine, Universitas Gadjah Mada, Department of Medical Education, Yogyakarta*), Yoyo Suhoyo (*Faculty of Medicine, Universitas Gadjah Mada, Department of Medical Education, Yogyakarta*), Mohammad Adrian Hasdianda (Faculty of Medicine, Universitas Gadjah Mada*, Department of Medical Education, Yogyakarta*)

Background: Medical licensure examination in Indonesia consists of multiple choice questions and OSCE. Six OSCE tryouts were carried out before the national OSCE was established as licensure examination. With 9 SP encounters from 12 stations, SP performance is essential for the OSCE.

Summary of Work: We compared SP performance during tryouts and the licensure examination. In tryouts, the SP were trained for 4 hours in the previous day before the exam and briefed an hour before the tryouts. In the licensure examination, they were trained for an hour prior to the examination. The training used specific scenarios prepared by the committee. Sets of questionnaires were used to assess SP performance, completed by examiners, examinee, and SP trainers. SP were assessed for their global performance, their act, interaction with examinee, attire, and intervention during the encounter.

Summary of Results: There were no significant differences for SP global performance (p=0.449) between the try out and the examination. In both, the range of examinees who agreed: the SP act was in line with the case are 70-90%, had a good interaction with examinees 60-90%, wore suitable attire/ make up 55-75%, and performed only necessary intervention during the encounter 55-80%. As for examiners, the range was 75-90%, 75-95%, 60-80%, and 55-80%, respectively. SP trainers agreed that scenarios and previous general preparation were helpful to train within the limited time of training.

Conclusions: It is possible to train standardized patients for high stake OSCE in a limited time prior to examination.

Take-home Messages: Training standardized patients for high stake OSCE does not need a long duration of training. Clear and comprehensive scenarios, well-trained SP trainers, and previous general preparation are the key points for success.

OTT-OA-5-5
An RCT comparing simulated patients to usual learning in the acquisition by medical students of musculoskeletal examination skills

James Parle (*University of Birmingham, Health and Population Sciences, Birmingham, UK)*, Edward Davis (*Royal Orthopaedic Hospital, , Birmingham*), Jacqueline Beavan (*University of Birmingham, Health and Population Sciences, Birmingham*), Celia Taylor (University of Birmingham*, Medical Education Unit, Birmingham*)

Background: Acquisition of the skills to perform musculoskeletal (MSK) examinations is core to doctors' training but health service pressures and the (morally correct) desire not to inflict more pain on patients appear to be acting to reduce the competence of newly qualified doctors. We hypothesised that structured, scenario based training (SSBT) using simulated patients (SPs) would result in better acquisition of MSK examination skills than 'usual learning' i.e. clinical experience and teaching on wards and in outpatients.

Summary of Work: We performed an RCT in two year cohorts in a single medical school, randomising students to SSBT with SPs or to usual learning; the MSK module is in year 4 of the 5 year course. All students were assessed in a single OSCE station on their MSK examination skills; for cohort 1 (n = 208) the OSCE was in their 5th year; for cohort 2 (n = 379) the OSCE was in 4th year. Delay between MSK module and OSCE was ~8 months (IQR 4-12) for intervention group and ~7 (4-10) for control group. OSCE examiners and statistician were blind to allocations.

Summary of Results: There was a difference in mean MSK OSCE station scores of 3 percentage points (t=2.50, p=0.013) comparing intervention vs. control arm. There was no attenuation of difference in OSCE scores with time between module and OSCE.

Conclusions: Scenario-based structured MSK skills training with SPs produces significant improvement in medical students' MSK OSCE scores which does not appear to reduce over time.

Take-home Messages: SPs are effective in training medical students in MSK examination skills.

OTT-OA-6: OSCE 1

OTT-OA-6-1
"Give us the OSCE" – attitudes and perceptions of year 1 and 2 medical students towards a new integrated formative OSCE

DS Furmedge (*University College London Medical School, London, UK)*, LJ Smith (*UCL Medical School, London*), V Balakrishnan (*UCL Medical School, London*), A Sturrock (UCL Medical School*, London*)

Background: OSCE is a core component of undergraduate medical student assessment. With increased emphasis on integrated programmes, more courses include OSCEs in the traditionally 'preclinical' years. In line with this and to ensure constructive alignment, University College London Medical School (UCLMS) identified a need for such an OSCE in their course.

Summary of Work: In 2013 UCLMS piloted a formative, integrated OSCE in years 1 and 2 of the MBBS programme. 654 students completed an OSCE testing professional and clinical skills, integrated with core knowledge. Students completed questionnaires (Likert scales and free-text responses) exploring views of the OSCE (response rate >95%). Four focus groups were subsequently held to further explore views and emerging themes. Data was independently transcribed and coded by two researchers using NVivo 10.

Summary of Results: Six overarching themes were identified: application of knowledge and skills; OSCE as an experience; OSCE as a process; a learning curve; becoming a doctor; and creating an effective OSCE. Students were overwhelmingly supportive of the OSCE as an addition to existing assessments. They enjoyed the opportunity to put skills into practice that were not

tested elsewhere, felt the process motivated learning and importantly, made them feel like trainee doctors.

Conclusions: Results strongly support using OSCEs early in the medical course with many benefits reported by students. An OSCE at this stage aligns with the vision of integrated medical education which includes early patient contact and early introduction of clinical and professional skills.

Take-home Messages: Medical schools should strongly consider the inclusion of an OSCE in the early years of medical school.

OTT-OA-6-2
A Palliative Care OSCE for Medical Student Assessment

Andrew Putnam (*Yale University School of Medicine, Internal Medicine, New Haven, USA)*, Michael Green (*Yale University School of Medicine, Internal Medicine, New Haven*), Carol Pfeiffer (*University of Connecticut School of Medicine, Internal Medicine, Farmington*), Margaret Bia (Yale University School of Medicine*, Internal Medicine, New Haven*), Matthew Ellman (*Yale University School of Medicine, Internal Medicine, New Haven*)

Background: The observed structured clinical examination (OSCE) is an important tool to assess clinical skills; however there is no reported palliative care OSCE for medical students.

Summary of Work: With input from experts in palliative care and medical education, we created a representative case and a checklist of 16 history items in five palliative care domains. Rising 4th year medical students completed this OSCE and 7 others. We performed standard item analyses on the history items, determined the internal consistency, and compared performance on the palliative care history with overall performance on the history and communication items from the other 7 stations.

Summary of Results: There was 94% agreement in ratings on the history items between the standardized patient and a remote observer. The ninety-four students scored 64% (SD 12) on the 16 history items. The difficulty index ranged from 0.20 to 0.98. Only one item had a point biserial correlation < 0. The remaining ranged from 0.16 to 0.52. The internal consistency reliability of the scale was α = 0.34. The palliative care history scores did not correlate with history scores (R = 0.12, p = 0.24) but correlated modestly with communication scores (R = 0.29, p = 0.005) on the overall skills examination.

Conclusions: A palliative care OSCE is feasible to implement with high inter-rater reliability. The majority of items showed favorable discrimination. Palliative care history scores correlated with overall communication scores.

Take-home Messages: We created and implemented a new OSCE to assess student palliative care skills. Analysis of its performance demonstrates successful aspects and the need for further refinement.

OTT-OA-6-3

Feasibility and validity evidence using a simulation-based OSCE to assess readiness for independent practice in Pediatric Critical Care trainees

Briseida Mema (*Hospital for Sick Children, Critical Care Medicine, Toronto, Canada)*, Yoon Soo Park (*University of Illinois at Chicago, Department of Medical Education, Chicago*), Anna Theresa Lobos (*University of Ottawa, Critical Care Medicine, Ottawa*), Hilary Writer (University of Ottawa*, Critical Care Medicine, Ottawa*), Afrothite Kotsakis (*Hospital for Sick Children, Critical Care Medicine, Toronto*)

Background: A simulation-based evaluation tool to assess competence has not been developed nor has the best method of evaluating readiness for independent practice been examined in Pediatric Critical Care. The purpose of this study was to provide validity and feasibility evidence for the use of a Simulation-based OSCE in the assessment of Pediatric Critical Care Medicine (PCCM) trainees.

Summary of Work: Validity evidence was based on Messick's five-point, unified construct validity framework. Simulation scenarios were developed for content validity by PCCM and education experts using CanMEDS competencies. Scenarios were trialed before the study to ensure fidelity and feasibility. Each scenario was evaluated by an inter-professional team of 2.

Summary of Results: Seventeen trainees were assessed. Inter-rater agreement, measured using intra-class correlations, was 0.907 (SE=0.086) across stations. Generalizability theory was used to evaluate internal structure and reliability. Reliability was moderate (G coefficient=0.668, phi coefficient=0.517). The greatest source of variability was from participant by station variance (40.6%).

Pearson's correlation coefficients were used to evaluate the relationship of OSCE with each traditional assessment instrument (360, ITER, SAQ, MCQ) and length of training. Performance on the OSCE correlated with performance on the MCQ (r=0.521, p=0.03) and 360 (r=0.589, p=0.017) but not with performance on the ITER (r=0.371, p=0.143) or SAQ (r=0.081, p=0.767). Strongest correlation was with Manager aspect of 360 evaluation (r= 0.74, p<0.01). Standard setting was done with Hofstee method.

Conclusions: Validity and feasibility evidence in this study indicate that the use of the OSCE can be a useful way to assess CanMEDS competencies required for independent practice in PCCM.

Take-home Messages: The use of a simulation-based OSCE can be used to assess competencies integral to PCCM.

OTT-OA-6-4
Continued Validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): Use in the simulated environment

Matt MacEwan (*University of Ottawa, Ottawa, Canada)*, Wade Gofton (*University of Ottawa, Ottawa*), Dudek

Nancy (*University of Ottawa, Ottawa*), Tim Wood (University of Ottawa*, Ottawa*)

Background: The O-SCORE assesses a surgical trainee's competence to perform an entire surgical procedure. It accurately identifies trainee level in the operating room. However, the evaluating surgeon is not blinded to trainee level in that environment leading to a potential bias. This study assesses if the O-SCORE remains valid when the rater is blinded to trainee level.

Summary of Work: Nineteen orthopaedic surgery residents and 2 staff surgeons participated in an objective structured assessment of technical skills (OSATS) station (open reduction and internal fixation of a radius) which was expanded to include a questionnaire addressing the pre and post procedure plan. The station was videotaped with the participants' faces removed to ensure anonymity. Two independent orthopaedic surgeons rated the performance using the O-SCORE, an OSATS global rating scale (GRS) and checklist. The order of the assessments was randomized.

Summary of Results: Inter-rater reliability for the O-SCORE was 0.63.The g-coefficient using the two raters was 0.90 indicating that the O-SCORE reliably differentiates between trainee level. The O-SCORE was highly correlated with the GRS (0.96).

Conclusions: The O-SCORE when used in a blinded fashion was able to differentiate trainee level further validating this tool. We also found good inter-rater reliability between observers with the O-SCORE suggesting that the tool is appropriate for formative assessment. Additionally, the O-SCORE demonstrated a high correlation with the current laboratory gold standard technical evaluation tool, the OSATS.

Take-home Messages: The O-SCORE demonstrates accurate and reproducible results when used in a blinded fashion further strengthening the validity evidence for this tool as a measure of surgical competence.

OTT-OA-6-5
An indigneous OSCE: Do we need it?

Suzanne Pitama (*University of Otago, Maori/Indigenous Health Institute (MIHI), Christchurch, New Zealand)*, Tania Huria (*University of Otago, Maori/Indigenous Health Instiute (MIHI), Christchurch*), Cameron Lacey (*University of Otago, Maori/Indigenous Health Instiute (MIHI), Christchurch*)

Background: Only one article to date (Ewen et al, 2011) discusses in depth how simulated patients/OSCE are used as a component of the indigenous health curriculum within different countries. What is included in an indigenous health OSCE scenario? How is this content determined? What learning outcomes should be expected?

Summary of Work: I undertook, as a component of my PhD, interviews with multiple sites (6 medical schools) and multiple stakeholders (based at the University of Otago, Campus), which involved 75 interviews in total. Qualitative data was analysed using inductive analysis (including structural analysis and descriptive analysis).

From this data emerged specific findings about the content included in measuring indigenous health competencies.

Summary of Results: The results identified how the use of an indigenous health framework provided the specific content for what was included within the OSCE scenario, and also assisted in the framing of the marking schedule. The OSCE provided two specific opportunities, the first to measure the students' ability to demonstrate indigenous health competencies, the second was for the indigenous health community workers (who 'acted' during the OSCEs) to use the OSCE experience to provide peer critique on indigenous health content (as identified by student behaviours in the OSCE) to the indigenous health teaching team.

Conclusions: The OSCE is a way to measure indigenous health competencies, but also is being used as a way to link social accountability between the indigenous health community and the medical school.

Take-home Messages: That an indigenous health OSCE enhances a medical curriuclum and provides a model for how a component of social accountability may be included within the curriuclum.

OTT-OA-7: STUDENT CHARACTERISTICS AND SELECTION

OTT-OA-7-1
Hot or Not: The Effects of Physical Attractiveness on Rater Decisions in Admissions

Kulamakan Kulasegaram (*University of Toronto, Mississauga, Ontario, Canada)*, Lindsay Jackowetz (*University of Toronto, Toronto, Ontario*), Lisa DeBruine (*University of Glasgow, Glasgow*), Daniel Krupp (Queen's University, *Kingston, Ontario*), Mark Hanson (*University of Toronto, Toronto, Ontario*), Nicole Woods (*University of Toronto, Toronto, Ontario*)

Background: Evidence from evolutionary psychology suggests physical attractiveness influences evaluation of competence independently of ability. We examined the effect of the physical attractiveness of mock medical school applicants on evaluations by admissions raters. Physical attractiveness was manipulated using traits identified as evolutionarily significant: masculinity/femininity and healthiness.

Summary of Work: 8 mock applicant profiles and portrait images were created (4 female, 4 male). The profiles were part of an online training module for medical school interviewers. Profiles provided short bios, responses to 2 questions on motivation and ethics. Interviewers rated motivation, ethical reasoning, and suitability for medicine on a 7-point scales. Responses were consistent for all raters, but accompanying portraits were manipulated with face morphing software to be masculinised or feminized (2 female, 2 male) and healthy or unhealthy (2 female, 2 male). Raters were randomized to conditions with masculine and unhealthy (MU) portraits or feminine and healthy

(FH). We analyzed the relationship profile gender, condition (MU vs. FH), and rater gender on ratings.
Summary of Results: 85 raters completed training (MU=46, FH=39). A significant interaction was seen between profile masculinity/femininity and rater gender (F(1,81)=8.0, p<0.006). Post-hoc testing showed male raters gave significantly higher mean(SD) ratings to hyper-feminized female profiles (4.3(0.8)) compared to masculinised females (3.7(0.8)). Female raters gave higher ratings to masculinised females (4.1(0.6) to 3.8(1.1)). No difference was seen for the healthiness manipulation.
Conclusions: This study replicates effects on the role of physical attractiveness in judgement. However, effects were small and limited to one trait. Future research will investigate the effects during interviewing.

OTT-OA-7-2
Ethnicity and social background as predictors of successful application for medical school

Karen M Stegers-Jager (*Erasmus MC, Institute of Medical Education Research Rotterdam, Rotterdam, Netherlands*), Susanna M Lucieer (*Erasmus MC, Institute of Medical Education Research Rotterdam, Rotterdam*), Axel PN Themmen (*Erasmus MC, Institute of Medical Education Research Rotterdam / Department of Internal Medicine, Rotterdam*)

Background: Medical schools worldwide seek measures to increase the diversity of their students. Currently, little is known about the effects of different selection procedures on student diversity. The aim of this prospective study was to identify socio-demographic factors that predict successful application for a selection procedure that combines non-academic (=quality and extent of extracurricular activities) and academic (=cognitive tests) selection criteria.
Summary of Work: Selection procedure participants in 2013 (n=692;92.3%) completed a questionnaire on ethnicity (Dutch, Surinamese/Antillean, Turkish/Moroccan/African, Asian, Western), first-generation immigrant, first language, parental education (first-generation university student) and parental profession (parent doctor). Outcome measures were: "selected", "failed on non-academic criteria", "failed on academic criteria". Age and gender were included as moderator variables. Data were analysed using logistic regression.
Summary of Results: Participants more often failed on academic criteria when they came from Surinam/Antilles (OR=3.59; 95%CI 1.76-7.31), were first-generation immigrants (OR=2.53; 95%CI 1.22-5.21) or first-generation university students (OR=1.48; 95%CI 1.01-2.15) and therefore non-Dutch participants (34%) were selected less often than Dutch participants, especially when they came from Surinam/Antilles (OR=0.32; 95%CI 0.14-0.73). The outcome measure "failed on non-academic criteria" showed no significant differences between the ethnic or social subgroups.
Conclusions: The results on the non-academic criteria were promising regarding increasing social and ethnic diversity; however, it cannot be ruled out that self-selection instigated by the selection procedure is stronger for students from non-traditional backgrounds. Further research should also focus on why cognitive tests might favour traditional medical students.
Take-home Messages: There is a need for studies that explore the impact of different weightings of criteria within selection procedures on student diversity.

OTT-OA-7-3
Characters of medical school applicants: What do they report and does it influence selection?

Steven Churchill (*University of Sheffield, Sheffield, UK),* Ben Holden (*University of Sheffield, Sheffield*), Sam Meadows (*University of Sheffield, Sheffield*), Kathryn Graves (University of Sheffield*, Sheffield*), Isabelle Pierry (*University of Sheffield, Sheffield*), Philip Chan (*University of Sheffield, Sheffield*)

Background: This medical school is moving towards character-based selection of applicants. The criteria studied were a list of 12 characters, based on Peterson & Seligman's 'Character Strengths and Virtues' with 4 additional skills (time management, public performance, organisational skills, self-directed study skills).
Summary of Work: Applicants coming to admissions interview completed a self-evaluation of their character from the list of 16 items, selecting four strengths and one weakness. The interview panel – a faculty member, a clinician and a student – conducted the interview along standard lines, without specific training in character assessment. The panel assessed each applicant's strengths and weakness the same way, blinded to the self-evaluation. Offers and rejections were recommended based on global impression.
Summary of Results: Applicants reporting self-directed study skills, time management and initiative as strengths had the highest chance of receiving offers (Odds Ratio 1.19, 1.16 and 1.15 respectively), while those who reported leadership and initiative as weaknesses had the lowest chance of receiving offers (OR 0.773 and 0.859 respectively). Agreement between interviewers on applicants' strengths was more common in those given offers compared to those rejected (95% CI 0.54 to 1.93, p=0.0005).
Conclusions: Initiative is the most important character; as a strength, it increases the chance of an offer while as a weakness this chance decreases. Skills are more predictive than characters in determining the outcome of our current format of admissions interviews. Inter-interviewer agreement on strengths is a positive indicator for an offer whereas agreement on weaknesses is indicative of rejection.
Take-home Messages: An approach based on assessment of character strengths can be developed for medical selection.

OTT-OA-7-4
The Development of the Cambridge Personal Styles Questionnaire (CPSQ)

Lyn Dale (*Cambridge Assessment, Admissions Testing Service, Cambridge, UK)*

Background: In occupational testing, personality assessment is a well-established method for providing standardised information on candidates' non-cognitive attributes.

Using the approach of occupational testing, The Admissions Testing Service set out to develop a personality or 'personal styles' assessment. The goal is to help admissions staff in charge of vocational subjects (e.g. healthcare) to 'match the right candidate to the right course'. The paper covers the rationale for the model of personality used, trialling stages and techniques used to develop the Cambridge-Personal Styles Questionnaire (CPSQ).

The CPSQ is a self-report assessment that uses thirteen scales or facets to measure five major style domains: Thinking, Learning, Self-Management, Emotional Resources and Interpersonal

Summary of Work: An extensive four year programme of research trialling has been conducted to ensure that the CPSQ is accessible to diverse student groups, whether studying at school (16 years+) or participating in higher education. It is a computer-based assessment which uses a dynamic rating and ranking system to reduce and, in some cases, prevent response biases (e.g. acquiescence, socially desirable responding), which can potentially affect the validity of personality scores.

Summary of Results: Key measures of reliability show that respondents answer the questionnaire in a consistent manner even after a period of time suggesting that CPSQ scores are likely to reflect real underlying tendencies in behavioural style. Furthermore analysis of scores shows that the assessment format (using both rating and ranking response options) helps to control for social desirability responding ('faking good')

Conclusions: The outcomes of the latest trialling and validity research show that the assessment has been developed to the highest standards of reliability and internal validity. Further external trialling will continue to ensure CPSQ is an accurate predictor of future behaviour.

Take-home Messages: CPSQ can be used to gain a more holistic, evidence-based insight into applicants for healthcare education.

OTT-OA-7-5
Academic performance of cognitive versus non-cognitive selected students

Susanna M. Lucieer (*Erasmus MC, Institute of Medical Education Research Rotterdam, Rotterdam, Netherlands)*, Karen M. Stegers-Jager (*Erasmus MC, Institute of Medical Education Research Rotterdam, Rotterdam*), Remy M.J.P. Rikers (*Erasmus University Rotterdam, Department of Psychology, Rotterdam*), Axel P.N. Themmen (Erasmus MC*, Institute of Medical Education Research Rotterdam / Department of Internal Medicine, Rotterdam*)

Background: Previous research showed that students selected by a combination of a cognitive and a non-cognitive procedure were academically more successful than those admitted by lottery. We aimed to determine the contribution of both procedures to this performance-difference in a controlled experiment. It was predicted that cognitive-selected students' pre-clinical performance would be better than non-cognitive-selected students' performance, and that non-cognitive selected students' clinical-performance would be better.

Summary of Work: Students selected exclusively by the non-cognitive procedure (2007-2008;N=277), students selected exclusively by the cognitive procedure (2009-2010;N=289) and lottery-admitted students (2007-2010;N=788) were compared in GPA, dropout, clinical-course grades and number of credits at 1, 2 and 3 years after enrolment.

Summary of Results: ANCOVA (cohort as covariate) showed that cognitive-selection had a positive significant effect over lottery on first-year GPA $(t(1341)=-3.909,p<.001,$ partial $\eta2=.013)$ and number of credits $(t(1354)=-4.113,p<.001,$ partial $\eta2=.013)$. First-year GPA of non-cognitive-selected students (M=6.29) was higher than of lottery-admitted (M=6.07) and cognitive-selected students (M=6.2) and the number of credits of non-cognitive-selected students (M=48.32) was higher than those of lottery-admitted (M=46.24) but lower than those of cognitive-selected students (M=51.37), albeit not significant after correction.

Conclusions: Pre-clinical performance of cognitive-selected students was better than performance of other students. Since our bachelor is mainly cognitive-based, it is not striking that the cognitive-selected earned more credits than the non-cognitive-selected. Our next analysis within this study will therefore focus on the relation between clinical-skills courses' scores and the non-cognitive selection procedure.

Take-home Messages: The mode of selection determines academic success in a specific context and should therefore be carefully chosen.

OTT-OA-7-6
Assessing selection for healthcare training: Finding the best starting point

Elana Curtis (*University of Auckland, Te Kupenga Hauora Maori, Auckland, New Zealand)*, Erena Wikaire (*University of Auckland, Te Kupenga Hauora Maori, Auckland*), Papaarangi Reid (*University of Auckland, Te Kupenga Hauora Maori, Auckland*), Robert Loto (University of Auckland*, Te Kupenga Hauora Maori, Auckland*), Yannan Jiang (*University of Auckland, Department of Statistics, Auckland*), Louise McMillan (*University of Auckland, Department of Statistics, Auckland*)

Background: Equity-targeted admission programmes utilize multiple assessment tools to select students into health professional programmes of study. Recent changes to assessment within MAPAS (Māori and Pacific Admission Scheme) at the University of Auckland (UoA) have included the introduction of a Multiple Mini

Interview and mathematics and English testing. Understanding whether the assessment tools and admission processes used by equity programmes predict tertiary academic success requires investigation.

Summary of Work: A Kaupapa Māori Research (KMR) methodological approach was used to analyse the predictive effect of MAPAS assessment variables (Multiple Mini Interview, Maths/English testing, school results, MAPAS recommendations) on tertiary academic outcomes (passing all courses and Grade Point Average) within First Year Tertiary Study (includes bridging/foundation study) and First Year Bachelor study at UoA between the years of 2008-2012. Univariate and multivariate logistic and linear regression analysis was completed.

Summary of Results: 918 Māori and Pacific applicants were assessed by MAPAS between 2008-2012. MAPAS Advice Followed; MAPAS Mathematics Test Result; NCEA Rank Score and Exposure to Any 2 Science subjects showed significant predictive effects after controlling for year of admission, age, gender, ancestry and secondary school decile. The odds of passing all first year courses was 5.4 times higher for those students who followed MAPAS advice versus those students who did not follow MAPAS advice (p<0.0001; CI:2.355-12.393).

Conclusions: The combination of assessment tools utilised by the MAPAS admissions process contribute to entry recommendations that predict academic success.

Take-home Messages: Comprehensive assessment of indigenous and ethnic minority applicants for healthcare training via equity admission programmes can improve academic outcomes in the first year of tertiary study.

OTT-OA-8: FEEDBACK UNDERGRADUATE 1

OTT-OA-8-2
Why do students not make more use of feedback after summative assessment?

Christopher Harrison (*Keele University, School of Medicine, Staffordshire, UK)*, Karen Könings (*Maastricht University, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht*), Lambert Schuwirth (*Flinders University, School of Medicine, Adelaide*), Val Wass (*Keele University, School of Medicine, Staffordshire*), Cees van der Vleuten (*Maastricht University, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht*)

Background: Despite calls for feedback to be incorporated in all assessments we know very little about its effective delivery in the summative setting. Previous work suggests it is not always used effectively by learners in this context. In this study we explored the reasons for this.

Summary of Work: We conducted individual interviews with 17 students in a UK medical school who had recently received web based feedback following a summative Year 3 OSCE. Constant comparative analysis of the interview transcripts was conducted with iterative discussion to identify recurring themes.

Summary of Results: The summative assessment culture, with a focus on avoiding failure, was a dominant and negative influence on the use of feedback. Strong emotions were prevalent throughout the period of assessment and feedback, which reinforced the focus on the need to pass, rather than excel. These affective factors were heightened by interactions with others. The influence of prior learning experiences affected expectations about achievement and the need to use feedback. The summative assessment and subsequent feedback appeared disconnected from future clinical workplace learning.

Conclusions: Our study has shed light on factors that need to be understood and manipulated in assessment in order to successfully integrate assessment of learning with assessment for learning. Simply providing the feedback and expecting students to engage is insufficient when the summative assessment culture is so dominant.

Take-home Messages: A move away from the potentially punitive summative assessment culture may be needed in order to maximise the learning potential of assessments.

OTT-OA-8-3
Student responsibility for seeking feedback

JR Rudland (*University of Otago, Faculty of Medicine, Dunedin, New Zealand)*, TJ Wilkinson (*University of Otago*), M Meeks (*University of Otago*), P Blyth (University of Otago), J Swan (*University of Otago*)

Background: It has been proposed that students need to take more responsibility in the feedback process (Rudland et al 2012). This presentation reports on the current state of student understanding of responsibility for feedback and how this responsibility is practised by the student body.

Summary of Work: Data were collected from medical students, using an electronic survey, asking about who was responsible for ensuring they receive feedback and what hindered gaining this feedback. Ranking data were analysed using SPSS and team members coded qualitative data manually.

Summary of Results: Half the respondents ranked themselves in the top three groups of people as having responsibility for ensuring that they receive feedback about their progress. A tenth of respondents ranked themselves in the bottom three groups of people with this responsibility.

More than half the respondents reported having asked for feedback, and most reported that this was a positive experience. Regardless of having asked for feedback or not, respondents also reported lacking the skills to ask for feedback; not seeing a need to ask for feedback; concerns at the reaction to a request for feedback; and perceiving that clinicians are too busy to deal with student requests for feedback.

Conclusions: Medical students value feedback and many recognise their responsibility to ask for feedback. Reasons for students not being more proactive in asking

for feedback need to be considered by faculty, as well as assisting those less proactive students to take greater responsibility.

Take-home Messages: There are barriers to students being proactive in asking for feedback and these barriers need to be addressed.

OTT-OA-8-4
Feedback in the OSCE, do examinees remember?

Marika Mihok (*University of Ottawa, Ottawa, Canada)*, Debra Pugh (*University of Ottawa, Ottawa*), Claire Touchie (*University of Ottawa, Ottawa*), Timothy J Wood (University of Ottawa*, Ottawa*), Samantha Halman (*University of Ottawa, Ottawa*), Susan Humphrey-Murto (*University of Ottawa, Ottawa*)

Background: Feedback (FB) is an essential component of student or resident training. Immediate FB after objective structured clinical examination (OSCE) stations is thought to improve student performance but many questions surrounding FB in this context remain unanswered. We conducted a study to 1) determine how well trainees recall FB and 2) determine the accuracy of the FB recalled.

Summary of Work: Internal Medicine residents (n= 18) participated in a 10-station OSCE testing clinical skills. Residents received 2 minutes of verbal FB from a physician examiner after each station. Examiner FB was recorded and transcribed. Residents were asked to describe the elements of the FB immediately after the OSCE and again one month later. The number and accuracy of FB points recalled was compared to the FB points made by the examiner.

Summary of Results: Six stations had sufficient data for analysis. For each station, the mean number of feedback points provided by the examiner was higher than FB immediately recalled by the resident (9.7 vs 2.7, 21 vs 2.7, 12.6 vs 2.7, 10.1 vs 2.7, 16.8 vs 2.8, 26.2 vs 2.2; $p<0.005$ for each station). At one month, the mean number of points recalled ranged from 1.7 to 2.7. Results for accuracy of the FB points will be presented.

Conclusions: Residents recall very few points of feedback provided by physician examiners in an OSCE.

Take-home Messages: Providing immediate feedback in an OSCE may not serve the purpose intended.

OTT-OA-9: COMPETENCY BASED ASSESSMENT 1

OTT-OA-9-1
Evaluation of Competence: Interface between psychiatrists and general practice physicians

Izabel Cristina Meister Coelho (*Faculdades Pequeno Principe, Research, Curitiba, Brazil),* Giseli Cipriano Rodacoski (*Faculdades Pequeno Principe, Curitiba*), Anna Beatriz Naumes (*Faculdades Pequeno Principe*)

Background: The Health System, established in Brazil in 1990, has decentralization as a constitutional principle.

In primary care doctors attend in Family Health, which must establish a line of medical care with specialists in secondary care. In a city of 264.000 inhabitants, located in the southern region of the country, the line of mental health care was not established for several reasons, especially the lack of psychiatrists and difficulties presented by primary care physicians to diagnose and treat mental disorders.

Summary of Work: This qualitative research was conducted through recorded interviews with eight psychiatrists and five doctors of primary care. The question was: What competences related to the psychiatrist does the primary care physician need to develop in working in mental healthcare?

Summary of Results: The result showed that the network of mental health services will be the most effective if the primary care physicians identify the most common psychiatric syndromes in the community where they operate, follow a treatment protocol, known therapeutic modalities mainly psychoeducation, develop the ability to listen and to conduct an interview in order to discriminate the complaint to avoid the diagnoses made by check-list with the DSM and know the mechanism of action of psychiatric medication to encourage adherence of users.

Conclusions: Among the conflicts are the difficulty of interaction and negotiation to integrate work-processes among specialists and generalists. New protocols have been developed and there was a systematization of work-processes between primary care and specialized units.

Take-home Messages: In order to ensure integrality in the care of mental disorders a patient-centered focus is necessary from all professionals that perform the approach and can share expertise, transcending their specialties.

OTT-OA-9-2
The Competency-Based Achievement System (CBAS): Trustworthy assessment of resident progress to competency

Michel Donoff (*University of Alberta, Family Medicine, Edmonton, Canada)*, Shelley Ross (*University of Alberta, Family Medicine, Edmonton*), Paul Humphries (*University of Alberta, Family Medicine, Edmonton*), Shirley Schipper (University of Alberta*, Family Medicine, Edmonton*)

Background: The Competency-Based Achievement System (CBAS) is an innovative competency-based assessment framework. CBAS is guided by the key principles of assessment FOR learning, the value of formative feedback and regular reviews of progress, guided self-assessment, direct observation, and that the focus of assessment should be on the habits being developed and demonstrated by learners.

Summary of Work: We evaluated the uptake and educational value of CBAS using mixed methods. Residents (n=244), program directors (n=4), and preceptors (n=56) participated in focus groups and/or completed surveys. Quantitative analysis of CBAS included examining FieldNotes (documented formative feedback) and summative progress reports. We also

explored the efficacy of CBAS in identifying and serving as a remediation tool for residents in difficulty.

Summary of Results: Over 16000 FieldNotes are in the CBAS electronic system. Most FieldNotes are about medical knowledge and clinical reasoning. While many participants express frustration with the interface, most respondents find CBAS more transparent and effective than traditional assessment. CBAS has been very useful in remediation with residents who are not progressing well. Feedback from users indicates both general faculty development and specifically training in the use of CBAS for summative assessments is still needed.

Conclusions: Implementation is proceeding well, but more training/faculty development is needed.

Take-home Messages: CBAS is an effective and powerful alternative to traditional resident assessment practices.

OTT-OA-9-3
Pan-Canadian Practice Ready Assessment for International Medical Graduate Physicians: A competency-based assessment for provisional licensure in family medicine

Cindy Streefkerk (*Medical Council of Canada, Ottawa, Canada)*, Dan Faulkner (*College of Physicians and Surgeons of Ontario, Toronto*), Lauren Copp (*Medical Council of Canada, Ottawa*), Sydney Smee (Medical Council of Canada*, Ottawa*), André De Champlain (*Medical Council of Canada, Ottawa*), Timothy Allen (*College of Family Physicians of Canada, Mississauga*)

Background: The National Assessment Collaboration (NAC) Practice Ready Assessment (PRA) initiative was established to design a pan-Canadian PRA for the provisional licensure of international medical graduate (IMG) physicians requiring a competency-based assessment. The initiative seeks to design common screening and comparable PRA processes across jurisdictions to assure regulators of a physician's ability and enable mobility under the Agreement on Internal Trade (AIT).

Summary of Work: The development of standards was based on the results of a previously-completed review of the current program state, refined and formalized through a collaborative and consultative approach engaging subject matter experts representing wide-ranging stakeholders. Results of the consultations were validated by participants during workshops.

Summary of Results: The outcome of the collaborative and consultative process resulted in the creation of pan-Canadian standards for family medicine PRAs covering initial screening, selection assessments (as PRA is a capacity limited process), over-time (workplace-based) assessment, decision making and orientation.

Conclusions: While research and exploration identify needs to evolve thinking on competency-based assessment, the model was based on a hybrid of existing materials and incremental, evolutionary thinking. Bringing together appropriate subject matter experts to discuss pan-Canadian standards facilitated the process

and resulted in the creation of achievable, common family medicine standards.

Take-home Messages: The development of standards is a critical first step in establishing a national process in an environment of multiple stakeholders with various subject matter expertise, different regional jurisdictions and of jurisdictions at different stages of implementation.

OTT-OA-9-4
Delivering the leadership curriculum using a mandatory assessment for learning

Lindsay Hadley (*Health Education Kent Surrey and Sussex, School of Leadership, London, UK)*, David Black (*Health Education Kent Surrey and Sussex, Postgraduate Dean*), Clare Penlington (*Health Education Kent Surrey and Sussex, Education, London*), Patrick Marshall (Health Education Kent Surrey and Sussex*, Education, London*)

Background: Clinical leadership is considered essential for maintaining and improving patient care. In the UK leadership is incorporated into the curriculum for all doctors in training and is part of the GMC's duties of a doctor. As an organisation with responsibility for ensuring the delivery of the curriculum, Health Education Kent Surrey and Sussex has used a multifaceted approach to teaching leadership with a progression of different initiatives supported by an infrastructure based in the local education providers. Trainees and their supervisors have had a target of one formative assessment of their leadership skills each year and trainees need to submit written evidence for their annual review of competency progression (ARCP). To support the target, workshops, training days and on-line resources have been developed and a formative assessment tool LEADER is used widely.

Summary of Work: Following ARCP, the trainee e-portfolios were searched for the leadership assessments. The names of the supervisors performing the assessments were logged and the form was examined for evidence of feedback.

Summary of Results: The assessments provide evidence of more than 80% coverage in most specialties. Only 5% of supervisors completed more than one form. 98% of forms included a description of the learning situation. 80% of the forms had some evidence of personalised feedback.

Conclusions: All trainees can receive some feedback on their leadership skills with a mandatory assessment and large numbers of educational supervisors are learning to assess leadership skills.

Take-home Messages: A mandatory formative assessment of leadership skills can raise awareness of leadership and how to assess it across an organisation.

OTT-OA-9-5
Health Advocate: The impact of our underlying assumptions on assessment of this role

Maria M Hubinette (*University of British Columbia, Family Practice, Vancouver, Canada)*, Sarah Dobson

(*University of British Columbia, Vancouver*), Angela Towle (*University of British Columbia, Vancouver*), Cynthia Whitehead (University of Toronto*, Toronto*)

Background: Health Advocate has been accepted as an important physician role in competency-based medical education. By examination of the development of this role, we have uncovered assumptions about health advocacy which can help us begin to understand the impact on the assessment of Health Advocate competence in trainees.

Summary of Work: Three pivotal documents provide a successive account of health advocacy in medical education in Canada: Educating Future Physicians for Ontario (1993), CanMEDS 2000 and CanMEDS 2005. We performed a textual analysis using a series of questions to explore differences in emphasis of these documents. We used Carlisle's (2000) model as a conceptual lens to identify different approaches to advocacy.

Summary of Results: Several assumptions are revealed by our analysis of the evolution of the Health Advocate role. It is currently framed as the responsibility of individual physicians rather than of the profession as a whole and focuses on prescription and protection (rather than empowerment and facilitation). This significantly shapes our assessment approaches and may limit our ability to explore competence in health advocacy from the perspective of collective action.

Conclusions: Analysis of key documents representing the development of the health advocate role over time elucidate underlying our assumptions about health advocacy in physicians which have an impact on the way we assess the health advocate role in our trainees. As a medical education community, we need to consider whether current assumptions sufficiently incorporate our core advocacy values.

Take-home Messages: Assumptions about health advocacy have an impact on the way we assess the the associated competencies in our trainees.

OTT-OA-9-6
What Knowledge, Skills, and Attitudes Do Physicians Judge to Be Important for Entry Into Supervised and Unsupervised Practice?

Andrea Gotzmann (*Medical Council of Canada, Research and Development, Ottawa, Canada)*, André F. De Champlain (*Medical Council of Canada, Research and Development, Ottawa*), Claire Touchie (*Medical Council of Canada, Ottawa*), John R. Boulet (Educational Commission for Foreign Medical Graduates*, Philadelphia*)

Background: The purpose of this survey was to collect physician judged importance of knowledge, skills and attitudes (KSAs) required at entry into supervised practice (residency) as well as unsupervised practice (initial licensure).

Summary of Work: Our survey included: (1) medical expert (e.g., "anxiety"), and; (2) non-medical expert (e.g., "appropriately develop and maintain ethical relationships") KSA questions, as defined by the CanMEDS framework (Frank, 2005). The 487 KSA questions were repeated for the unsupervised and supervised decision points. Most of the medical expert questions were presented for determining cause and initiating management. A four-point Likert scale was used to collect importance judgments. A total of 1,836 physician respondents completed the survey resulting in an overall response rate of about 10%.

Summary of Results: The four categories ("Not Important", "Moderately Important", "Very Important", "Extremely Important") were collapsed to "Not/Moderately Important" and "Very/Extremely Important". There was complete overlap in questions that were deemed "Very/Extremely Important" across the supervised and unsupervised levels. Non-medical expert questions were proportionally as important as the medical expert questions at the supervised level. For the medical expert questions, slightly more emphasis was placed on questions for determining cause at the supervised level whereas more prominence was given to questions for initiating management at the unsupervised level.

Conclusions: The survey results suggested that a common blueprint could be adopted for physicians at the supervised and unsupervised level.

Take-home Messages: The results of this survey were critical in aiding subject matter experts to develop a defensible blueprint and accompanying test specifications for a national medical licensing examination program that reflect expectations of practicing physicians across the country.

POSTERS

OTT-PB3: POT POURRI

OTT-PB3-01
Validation of a novel Problem Based Learning assessment tool

Eoghan McCarthy (*University College Dublin, School of Medicine and Medical Sciences, Dublin, Ireland)*, Paula Heaphy, Annette Liston, Pat Murray, Suzanne Donnelly (*University College Dublin, School of Medicine and Medical Sciences, Dublin, Ireland*)

Background: Previously we have described a novel feedback tool for students participating in PBL:BRIE/OPHS with each letter describing the role students undertook during the reporting(BRIE) and discussion(OPHS) phase of PBL. We sought to develop and validate a numeric tool based on BRIE/OPHS that can be used in the summative assessment of PBL.

Summary of Work: Each letter on the BRIE/OPHS was assigned a numeric value. Differences between the numeric scores generated between our currently used validated tool(CT) and the novel BRIE/OPHS tool(NT) were assessed by Mann Whitney whilst correlation was assessed using Spearmans test. The relationship

between the component parts of both tools was also analysed.

Summary of Results: 372 PBL episodes were analysed. A strong correlation was observed between the novel tool(NT) and currently used(CT) assessment format(r=0.85, p<0.0001). The report and discussion sections when analysed alone revealed similar correlations(Report:r =0.87, p<0.001; Discussion:r = 0.81, p<0.001). When total scores were assessed a significant difference was observed between tools such that students scored lower when assessed using the new instrument (Median CT 0.6 v NT 0.5, p<0.001). Analysis of the components revealed that this was due to a lower median mark in the discussion phase of the PBL session(NT 0.4 vs CT 0.6,p<0.001) with no differences observed in scores between tools in the report phase(NT 0.6 vs CT 0.6,p=0.22).

Conclusions: The BRIE/OPHS, a novel feedback tool, can be successfully adapted to provide a numeric score that demonstrates strong correlation with currently available PBL assessment tools.

OTT-PB3-02
Assessment of Problem Based Learning using a 'criterion referenced system' for year-1 medical students at RAK Medical & Health Sciences, a multicultural university in U.A.E.

Anand Srinivasan (*RAK Medical & Health Sciences University, Anatomy, Ras Al Khaimah, UAE)*, Saidunnisa Begum (*RAK Medical & Health Sciences University, Biochemistry, Ras Al Khaimah*)

Background: RAK Medical & Health Sciences University (RAKMHSU), a leading medical university in U.A.E. follows a system based horizontally integrated curriculum. PBL (problem based learning) is one of the instructional methodologies and assessment tool encompass both course objectives and PBL process.

Summary of Work: In 2006, PBL assessment based on specific criteria and a numeric scale was created. Overall 25% of the internal assessment was from PBL which included self and peer assessment (each given weightage of 25%) and facilitator assessment (50%). Based on the observations for 2 years, self and peer assessments were mostly subjective and biased, thus requiring modification in the method.

In the year 2008, self & peer together were given a weightage of 25%, facilitator 25% and a separate written paper based on learning objectives of the PBL formed 50%. Since the knowledge component got more weightage, students were concentrating more on the written PBL paper rather than PBL process.

As we moved from semester to yearly assessment system, PBL was given a total weightage of 15% in which only the PBL process, were assessed. The knowledge gained was tested in continuous and end year exams.

Summary of Results: Criterion referenced assessment (CRA) helped in identifying and assisting students having trouble developing generic skills, professional attitudes, and set objective standards for student performance.

Conclusions: CRA provides students with enhanced opportunities to accept increasing levels of responsibility for their learning.

Take-home Messages: PBL assessment validity is enhanced if consistent not only with curricular goals and learning objectives but also with its learning process.

OTT-PB3-04
Using QI Tools to Demonstrate QI Design, Selection and Implementation in a Community

Joanne Laine-Gossin (*North York General Hospital and University of Toronto, Family and Community Medicine, Toronto, Canada)*, Kimberly Lazare (*North York General Hospital and University of Toronto, Family and Community Medicine, Toronto*), Alan Monavvari (*North York General Hospital and University of Toronto, Family and Community Medicine, Toronto*), Harvey Blankenstein (North York General Hospital and University of Toronto*, Family and Community Medicine, Toronto*)

Background: Quality Improvement (QI) is now recognized as an integral part of primary care. Our poster outlines the processes of creating, implementing and promoting a QI culture using QI tools. It identifies enablers, challenges and lessons learned in order to effectively incorporate QI into the daily work of residents and physicians practicing in community settings.

Summary of Work: In the two years since its implementation 31 residents, 54 faculty and 40 allied health/reception staff within the North York General Hospital area have been trained. The lessons learned in the selection of projects, getting buy-in from preceptors and residents, and aligning interests is discussed in this poster.

Summary of Results: This poster is of interest to faculty, residents and community-based physicians. We suggest the development of a QI framework as a guiding tool to choose QI projects and align these to the goals of the department at large, the faculty and individual offices . We have learned that it is challenging to get buy-in from faculty and allied health personnel, but that this is essential for the learning experience.

Conclusions: Through the use of the QI model community-based faculty, allied health and residents became familiar with Quality Improvement as an overarching goal. Various enablers and challenges were identified and improved upon. Using the QI axiom of doing small tests of change, the residents' projects, comfort in doing QI and potential for incorporating QI into their future practice were all markedly improved. This was true for preceptors too.

Take-home Messages: Using the QI framework and tools in residents' teaching and every day office practice helps to imbed the QI culture. We think this then results in improved patient care and physician satisfaction. Our poster shows that this is possible, and effective, in a community-based teaching hospital.

OTT-PB3-05

An Assessment Model to Ensure Doctorateness in a Ph.D. Health Professions Education Qualification

M.M. Nel (*Faculty of Health Sciences, University of the Free State, Health Sciences Education, Bloemfontein, South Africa*)

Background: The requirements that theses in the Health Professions Education programme should meet, are that they should demonstrate the candidate's familiarity with the relevant literature, their research skills, as well as their ability to write a proper report that bear evidence of analytical skills, critical stance and substantive insight. The candidates should make an original and significant contribution to the subject field and prove that they have mastered and are able to use the research methods and techniques of the particular subject field. The aim of the study was to analyse the feedback of examiners with the view to compile an assessment model to ensure doctorateness.

Summary of Work: Quantitative and qualitative approaches were used. An empirical, non-experimental research design was followed in this descriptive study.

Summary of Results: The findings are used to make recommendations on the quality of Ph.D. qualifications and are presented in an assessment model that includes an emphasis on philosophical underpinnings, seeking evidence of scholarship, projecting postdoctoral capabilities as independent researchers and looking for maturity in research understanding by candidates.

Conclusions: The role that defined criteria and clear recommendations can play is of utmost importance. With this study an attempt was made to emphasise feedback of examiners as learning opportunities for supervisors and postgraduate students as well as to ensure quality in the achieving of a doctorate.

Take-home Messages: The role that an assessment model, consisting of defined criteria and clear recommendations can play, is of utmost importance.

OTT-PB3-06
Time frame as substitute for the perceived quality of oral assessments

Jos M T Draaisma (*Radboud University Medical Centre, Pediatrics, Nijmegen, Netherlands)*, Edith Janssen (*Radboud University Medical centre, Pediatrics, Nijmegen*), Harry van Goor (*Radboud University Medical Centre, Surgery, Nijmegen*)

Background: In our clerkships, three kinds of oral assessments are performed: an interim formative assessment of performance and two summative assessments: one of performance and one of knowledge. The latter two define the summative results of the clerkship. For quality evaluation, the time frame of the assessment is used as a quality guide. An exam of 30 minutes or more is considered as best practice. However, many teachers raised the question of the general principle behind this initiative. In their opinion, quality is not at all dependent on the length of time of the oral assessment. The aim of this study was to test this hypothesis.

Summary of Work: A retrospective cohort study was conducted using data obtained by monthly surveys that interns complete at the end of each clerkship. Questions about the time frame and perceived quality were included in the study. The perceived quality was considered to be the gold standard.

Summary of Results: There was a correlation of r=0.353 for the interim assessment, r=0.333 for the final performance assessment and r=0.240 for the knowledge assessment. The positive predictive values for all the assessments were high (0.94, 0.89 and 0.84 respectively), but the negative predictive values were generally low (0.14, 0.19 and 0.32 respectively).

Conclusions: The time frame of an oral assessment cannot be used as a predictor of the perceived quality given the low negative predictive values. Further research is required to identify new tools for quality evaluation of oral assessments.

Take-home Messages: The time frame of an oral assessment cannot be used as a substitute for the perceived quality.

OTT-PB3-07
Assessment in Paediatrics: the status quo and options for change

Hannah Jacob (*Institute of Child Health, University College London, London, UK),* Caroline Fertleman (*Institute of Child Health, University College London*)

Background: Pressure is mounting to standardise assessment for medical undergraduates in the United Kingdom (UK). A national common syllabus for paediatrics is being developed which could form the basis for such assessment in child health yet little is known about how paediatrics is currently being assessed.

This study aimed to establish how child health is assessed in UK medical schools.

Summary of Work: The lead for undergraduate paediatrics at every UK medical school was contacted. Semi-structured interviews were conducted with all those agreeing to participate. Participants were asked about assessment tools used within their course and who performed this assessment.

Summary of Results: 18/31 (58%) undergraduate paediatric leads were interviewed between June and August 2013. 15 (83%) medical schools used OSCEs and all used multiple choice questions of varying formats. Workplace-based assessments and skills logs were both used by 10 (56%) and 5 (28%) used reflective logs. Multi-source feedback and discharge summaries were used in 1 (6%) medical school each. Assessment always involved Consultant paediatricians, with most including clinical academics (16, 89%) and some involving trainee paediatricians (5, 28%).

Conclusions: A wide range of tools is used to assess undergraduate paediatrics and it is not clear that such assessment is comparable. The itinerate nature of UK medical graduates makes parity of paediatric assessment essential to maintain standards nationally.

A common syllabus for undergraduate child health training could support standardised assessment.
Take-home Messages: A national syllabus for undergraduate paediatrics could be used to develop common assessment tools.
Standardised assessment would help promote high quality child health training nationwide.

OTT-PB3-08
Breastfeeding knowledge, confidence, beliefs and attitudes of Canadian Physicians

Catherine M Pound (*Children's Hospital of Eastern Ontario, Pediatrics, Ottawa, Canada*), Kathryn Williams (*Children's Hospital of Eastern Ontario, Clinical Research Unit, Ottawa*), Renee Grenon (*Children's Hospital of Eastern Ontario, Clinical Research Unit, Ottawa*), Mary Aplipay (Children's Hospital of Eastern Ontario*, Clinical Research Unit, Ottawa*), Amy C Plint (*Children's Hospital of Eastern Ontario, Pediatrics Emergency Medicine and Clinical Research Unit, Ottawa*)

Background: Physicians' attitudes and recommendations directly impact breastfeeding duration. Yet, previous studies have shown that physicians lack the skills to offer proper guidance to breastfeeding mothers. This study aims to identify gaps in breastfeeding knowledge, confidence, beliefs and attitudes of Canadian physicians, and factors associated with breastfeeding knowledge.
Summary of Work: A breastfeeding questionnaire was developed and piloted for validity. These questionnaires were sent to 1429 pediatricians (PED) and 1329 family physicians (FP), for a target sample size of 385 per group. Multiple linear regression was used to identify factors independently associated with knowledge score.
Summary of Results: The analysis included 397 PED, and 322 FP who completed the questionnaire. Mean overall correct knowledge score was 67.8% for PED, and 64.3% for FP. 285 PED (74.2%), and 228 FP (73.1%) felt confident with their breastfeeding counseling skills, but few PED and FP (5.1% and 11.3%) routinely observed breastfeeding in mother-infant pairs. Most participants (72.1%) felt that residency had prepared them poorly or somewhat poorly to support breastfeeding mothers. Independent predictors of breastfeeding knowledge included female gender, > 50% of one's practice under 1 year of age, breastfeeding one's own child, self-directed learning and confidence teaching mothers about breastfeeding.
Conclusions: Breastfeeding knowledge, confidence, beliefs and attitudes are suboptimal among Canadian physicians. Confidence and knowledge are closely intertwined, and currently develop through learning methods other than residency training.
Take-home Messages: This study identified specific areas of concerns in physicians' breastfeeding knowledge, confidence, beliefs and attitudes, allowing for the development of targeted educational tools that will help physicians better support breastfeeding.

OTT-PB3-09

The awareness of Baby Friendly Hospital Initiative by semester-10 students, prior to internship training

Kyin Win (*International Medical University, Malaysia, Paediatrics, Batu Pahat, Johor, Malaysia*), Ankur Barua (*International Medical University, Malaysia, Community Medicine, Kuala Lumpur*), Davendralingam Sinniah (*International Medical University, Malaysia, Paediatrics, Seremban, Negri Sembilan*)

Background: Breastfeeding delivers substantial benefits for child health with improvement in survival, and cognitive performance. The Baby Friendly Health Initiative (BFHI) supports and encourages breastfeeding as an optimal way of feeding in infants. During residency training, house officers must attain essential concepts of BFHI. Hence, acquiring that concept during undergraduate training is crucial. In the International Medical University semester-10 students accomplish clinical competencies by outcome-based curriculum through senior clerkship at teaching hospitals affiliated with the university.
Summary of Work: Objectives: To evaluate the readiness of the semester-10 students to integrate breastfeeding practice in clinical setting and adequate awareness of BFHI, prior to internship training. Methods: Cross-sectional descriptive study which was conducted for the 179 students, 2 cohorts of semester-10 of International Medical University, at the completion of undergraduate training. The questions are designed to fundamental of breastfeeding and key BFHI principles using 5 points Likert scale.
Summary of Results: The response rate was 95 %. The majority of the respondents have adequate awareness of the BFHI principles except only few areas which can be learned from the lactation management course.
Conclusions: The overall awareness of BFHI by final year medical students is adequate through clinical posting with assigned and supervised senior clerkship curriculum. However, only few areas of competencies must be experienced further.
Take-home Messages: The expected competency of BFHI principles can be achieved through senior clerkship curriculum at accredited baby friendly hospitals.

OTT-PB3-10
Knowledge and Attitude of Medical Students towards Inter-professional Collaboration

S M Tajdit Rahman (*Sir Salimullah Medical College, 5th year, Dhaka, Bangladesh*), Md Sazid Rezwan (*Sir Salimullah Medical College, 5th year, Dhaka*)

Background: Interdisciplinary healthcare teams are central to improving patient outcomes. Strong inter-professional education (IPE) is fundamental for effective team performance. Although academic and policy perspectives on IPE are often the sole lenses through which IPE is viewed, equally important is soliciting student perspectives on IPE. But, in Bangladesh, there is lack of knowledge and process to develop a positive

attitude towards Inter-professional education and collaboration.

Summary of Work: A cross sectional type of descriptive study was conducted to assess the knowledge and attitude of students of Sir Salimullah Medical College, Dhaka towards inter-professional collaboration. 700 students were purposively selected for study. Data were collected by face to face interview from the respondents through semi-structured questionnaire.

Summary of Results: Majority of them showed lack of knowledge (82.86%) about inter-profesional collaboration. Most of the Students did not know the importance (89.29%) of inter-professional collaboration and education. Very few students enjoy team work (12.86%). Majority of them do not know how to implement (68.57 %) inter-professional collaboration.

Conclusions: That can be concluded that knowledge and attitude of medical students of third world country like Bangladesh is very poor. So a large community cannot take part in the era of inter-professional collaboration which makes this worldwide collaboration incomplete.

Take-home Messages: Measures like campaigning, review of curriculum, collaboration with other countries, student exchange etc. should be taken. This can contribute a lot in making the term 'Inter-professional Collaboration' complete.

OTT-PB3-11
A comprehensive evaluation framework for an interprofessional program: systematically getting to student assessment

Sharla King (*University of Alberta, Health Sciences Education and Research Commons, Edmonton, Canada)*, Mary Roduta Roberts (*University of Alberta, Department of Occupational Therapy, Edmonton*), Ken Cor (*University of Alberta, Faculty of Pharmacy and Pharmaceutical Sciences, Edmonton*), Hollis Lai (University of Alberta*, Undergraduate Medical Education, Edmonton*)

Background: Interprofessional (IP) education is increasingly integrated within health science curricula. Assessing student IP competence after these learning experiences is often done in isolation and is removed from a broader, systematic approach that evaluates an IP program, including faculty development/experiences. Therefore, an institutional level, comprehensive evaluation framework is required to evaluate IP program outcomes.

Summary of Work: Representatives from 10 health science programs at the University of Alberta collaborated to develop an evaluation framework . Our framework was defined by three key assessment areas: curriculum and student learning, student experiences, and faculty development/experiences. Evaluation components of each IP activity were identified and linked to each area of assessment. Outcomes from the logic models served as a common basis for defining evaluable questions to address programs' accreditation needs.

Summary of Results: This process resulted in the identification of data required to evaluate criteria associated with the institution level IP activities. This information will be used to determine the existing data from individual programs that is useful for evaluation, guide revisions of existing assessment tools, and create new assessment tools to generate evaluation data that does not already exist.

Conclusions: The steps used to develop the evaluation plan for the IP experiences across the health sciences programs at the University of Alberta has resulted in a systematic framework for evaluating meaningful program outcomes.

Take-home Messages: A single evaluation framework is the most efficient means for assessing students' preparedness for collaborative practice. Our common evaluation framework outlines both short/long-term evaluation strategies within the IP program.

OTT-PB3-12
Measuring Measures: Assessing outcomes used in program evaluations for patient education on mental health problems and addiction

David Wiljer (*Centre for Addiction and Mental Health, Toronto, Canada)*, Andrew Johnson (*Centre for Addiction and Mental Health, Toronto*), Michael-Jane Levitan (*Centre for Addiction and Mental Health, Toronto*), Sandra Cunning (Centre for Addiction and Mental Health*, Toronto*), Ivan Silver (*Centre for Addiction and Mental Health, Toronto*), Karen MacCon (*Centre for Addiction and Mental Health, Toronto*)

Background: There is high demand for evidence-based, educational programs that promote recovery from mental health and addiction (MH&A). However, there is no universally accepted way to operationalize recovery as available outcome measures often fail to accurately capture the distinctly personal process. There is a pressing need to counteract potentially confounding variables with an evidence-based framework for classifying and validating outcomes in this field. To inform this framework, this review will map patterns of outcomes used in recent program evaluations for mental health education.

Summary of Work: A critical review was conducted of recovery-oriented MH&A educational programs to identify effective strategies for measuring impact. The review was limited to journal articles published in English in PsycINFO, Medline, and CINAHL between 2002 and 2013. Outcome domains, indicators and measurement tools were extracted from each included article.

Summary of Results: Of the 4,639 records generated from the search, 273 were included in the review. This sample produced a total of 121 outcome indicators that were distilled into 8 main domains. The most frequent of these were related to symptoms (n=107), recovery (n=89), knowledge (n=55), and functioning (n=48). There were 376 distinct tools used to measure these outcomes with a significant majority (70.5%) cited only once. All of the twenty most frequent tools have been scientifically validated.

Conclusions: This data demonstrates a wide range of outcomes that hamper cross study outcomes comparisons. A common framework is required to map outcome evaluation among educational programs.
Take-home Messages: Understanding the relationship between outcomes, indicators and tools will indicate trends in quantifying recovery-oriented concepts and inform a comprehensive framework to guide future evaluations.

OTT-PB3-13
Measurement of the clinical learning environment in anaesthesia

NA Smith (*Wollongong Hospital, Anaesthesia, Wollongong, NSW, Australia)* Damian Castanelli (*Monash Medical Centre, Anaesthesia and perioperative medicine, Melbourne*)

Background: Anaesthesia training occurs within individual department and hospitals settings, guided a by a formal College curriculum. Factors within the work environment are critically important for the quality of learning. The Australia and New Zealand College of Anaesthetists (ANZCA) implemented a new curriculum from 2013, with a much stronger evidence on workplace based learning and assessment than previously. We wished to gain a measurement of the learning environment throughout the ANZCA regions at the very beginning of this change process.
Summary of Work: We administered an on-line survey to a large random sample of anaesthesia trainees throughout Australia and New Zealand. The survey was developed from existing literature, and piloted with both trainees and trainers. Validity and reliability were explored using various psychometric tools. Data was analysed as overall scores, with comparisons between the themed question groups, between countries, and between levels of training.
Summary of Results: The revised questionnaire tool was found to have a high reliability and validity. Exploratory factor analysis revealed four themed groups that were slightly different from the original literature. Data showed that the trainees perception of their learning environment is generally high, with small differences between some groups.
Conclusions: We developed a new tool that demonstrates that it is feasible to gain a measurement of the learning environment for anaesthesia trainees on a widespread level. This information could be used to gauge the effects of curricular change, or to measure changes at departmental levels.
Take-home Messages: The clinical learning environment is important for the quality of learning in anaesthesia training. This can be measured in a reliable and valid manner. Currently, ANZCA trainees perceive the quality of their learning environments to be generally high.

OTT-PB7: POSTGRADUATE AND CONTINUING EDUCATION/REVALIDATION

OTT-PB7-01
Quality management systems in Postgraduate Medical Education (PGME) – an Associate Dean's role

Kevin Kelleher (*Health Education Kent Surrey & Sussex, Deputy Postgraduate Dean for Secondary Care, London, UK*)

Background: Since the advent of the Postgraduate Medical Education and Training Board (PMETB), now merged with the General Medical Council (GMC), deaneries are responsible for the quality management of delivery of postgraduate medical education.
Summary of Work: The Postgraduate Deanery for Kent, Surrey and Sussex (KSS) has recruited an Associate Dean (AD) specifically for the role of ensuring that quality management systems are standardised across the Deanery region. We describe the post holders main duties and responsibilities in the standardisation and integration of day-to-day quality management systems within the guidance of the Deanery's 'Graduate Educational and Assessment Regulations' (GEAR), which underpin quality management systems in KSS.
Summary of Results: We describe the detail of the ADs role as focussing on Local Education Provider (LEP) faculty function and visits with standardisation of report writing, dissemination of 'good practice' from LEP delivery of postgraduate programmes, and the delivery of annual Deanery and annual School reports which inform the regulator GMC.
Conclusions: To ensure a standardised approach to quality management systems, KSS has recruited an AD, and we describe the domains of their work since inception in 2009.
Take-home Messages: To ensure a quality management system standardisation assists the processes.

OTT-PB7-02
Development of a Tool to Guide Consistency and Rigor in Resident Scholarship

Rebecca Blanchard (*Baystate Health/Tufts University School of Medicine, Academic Affairs, Springfield, USA)*, Kevin Hinchey (*Baystate Health/Tufts University School of Medicine, Academic Affairs, Springfield*), Jeffrey LaRochelle (*WRNMMC/Uniformed Services University of the Health Sciences, Internal Medicine, Bethesda*)

Background: High quality patient care requires evidence-based practice. To develop and demonstrate skills in reviewing, synthesizing and applying literature, the ACGME requires residents to complete a scholarly project. Expectations and guidelines for scholarly activity vary greatly across institutions, departments, and mentors. This presentation describes the development of a checklist to improve the quality and consistency in resident scholarly activity.
Summary of Work: Program leaders from ten residency programs were brought together to identify the skills associated with a successful resident scholarly project.

Skills were then organized and mapped onto two frameworks; principles for applying evidence-based medicine (for research skills) and the criteria for a scholarly approach to teaching (for non-research skills). Feedback on mapped skills was obtained from research mentors in the institution and external experts.

Summary of Results: The final tool contains a checklist for both research and clinical scholarship. Residents identify a mentor, project type (research or clinical), and timeline on the tool. Mentors indicate which skills the resident should demonstrate with the project; and subsequently indicate when the skill is completed.

Conclusions: The development of a checklist for scholarship expectations explicitly defines skills that demonstrate a rigorous approach to both clinical and research scholarship. Next steps include a multi-institutional pilot of the checklist to establish completeness, utility, and feasibility.

Take-home Messages: This checklist for resident scholarship has strong potential to assist mentors in clarifying expectations, informing feedback, and maintaining a consistent level of rigor to sustain a scholarly approach to practice.

OTT-PB7-03
Indicators of those trainee doctors who require extended periods of training following annual assessments

Charlotte Rothwell (*Durham University, Centre for Medical Education Research, Durham, UK*), John McLachlan (*Durham University, Centre for Medical Education Research, Durham*), Jan Illing (*Durham University, Centre for Medical Education Research, Durham*), Simon Forrest (Durham University*, Centre for Medical Education Research, Durham*)

Background: In the UK trainee doctors have to pass an annual assessment; 'Annual Review Competency Progression Panel' (ARCP) to enable them to progress to the next level of training. An adverse ARCP outcome requires them to extend their training(i).
General Medical Council evidence shows that postgraduate trainees who qualified outside the UK are proportionally more likely to experience difficulties whilst progressing through training .
However it is not just International Graduates who have difficulty during training. Causes such as health, misconduct, behavioural, personal circumstances and clinical issues may impact on any doctor's performance(ii) .

Summary of Work: To investigate who requires extended periods of training and for what reasons. Data was interrogated in one UK Local Education and Training Board to look at patterns emerging about these trainees.

Summary of Results: Early results indicate that qualification outside the UK and age are factors which can delay progression through training.

Conclusions: Indicators have been identified associated with trainee doctors who have had extended periods of training. It is important to know which doctors are more likely to require extended periods of training which will

enable those in potential difficult to receive early interventions.

Take-home Messages: It is important to understand who these doctors are and what they are having difficulty with to help them to be identified and supported earlier in the process.

References:
i GMC. The State of Medical Education and Practice in the UK, GMC report 2012. http://www.gmc-uk.org/publications/10586.asp.
ii NPSA (2009) NCAS Casework – The First Eight Years Summary September 2009. National Patient Safety Agency

OTT-PB7-04
Large-scale assessment of visual diagnostic expertise with volumetric datasets in postgraduate radiology training

A. van der Gijp (*University Medical Center Utrecht, Department of Radiology, Utrecht, Netherlands)*, C.J. Ravesloot (*University Medical Center Utrecht, Department of Radiology)*, M. Maas (*Academic Medical Center Amsterdam, Department of Radiology*), Th. J. ten Cate (University Medical Center Utrecht*, Center for Research and Development of Education*), M.F. van der Schaaf (*Utrecht University, Faculty of social and behavioural sciences*), J.P.J. van Schaik (*University Medical Center Utrecht, Department of Radiology*)

Background: In current radiology practice radiologists interpret digital images, with a substantial amount of volumetric datasets. Postgraduate assessment of radiology trainees concerning expertise in visual diagnosing is still mainly based on 2D images. Digital assessments with volumetric datasets could improve authenticity. We aimed to implement and evaluate large-scale assessment with volumetric datasets in postgraduate radiology.

Summary of Work: In April 2013, 383 radiology residents simultaneously took the Dutch Radiology Progress Test (DRPT) in a digital assessment program VQuest. The test contained 200 true/false/don't-know questions. Thirty-four questions were image-based, including eleven volumetric datasets. Participants could navigate through datasets in different planes and contrast settings. Scores were calculated with formula scoring. Participants received a questionnaire concerning assessment authenticity and quality.

Summary of Results: The DRPT was completed by all participants. Eleven questions (containing two 2D and one volumetric image question) were eliminated afterwards based on item analysis and participant comments. The reliability of the test was high, Cronbach's α = .92. Reliabilities of 2D image questions and volumetric dataset questions (after Spearman-Brown corrections for k = 40) were .74 and .77 respectively. Questionnaire response rate was 58%. Participants graded authenticity and image quality of the digital test higher compared to the former (paper-based) test, t(205) = 6.03 (p < .001, r = .39) and t(202) = 9.98 (p < .001, r = .57) respectively.

Conclusions: Large-scale digital assessment of visual diagnostic expertise with volumetric datasets was conducted with a high reliability and improved perceived authenticity and image quality.

Take-home Messages: Digital volumetric datasets can improve assessment of visual expertise in postgraduate radiology.

OTT-PB7-05
Does a review course improve performance on structured oral examinations?

Yvonne Ying (*University of Ottawa, Surgery, Ottawa, Canada),* Timothy Wood (*University of Ottawa, AIME, Ottawa*)

Background: To prepare for high-stakes examinations, many residents participate in review courses. These are well attended, feedback is generally positive and the impression is that they improve examination performance. However, there is little evidence that such courses actually improve examination outcomes. There is a need to establish if these short courses can influence examination performance.

Summary of Work: A review course in Plastic Surgery consisting of a series of structured oral examinations (SOE) similar to those used on the certification examination was held. Each resident received one SOE question set (QS) consisting of 8 questions on the first day followed by feedback. They then received another SOE on the second day. QS-A was administered on Day one, while QS-B was administered on Day two. To explore if there was a confound with question set and difficulty, the same question sets were administered three years later in reverse order to a different cohort.

Summary of Results: Residents performed better on QS-B (M= 70.62) than QS-A (M=70.14). When QS-B was administered after QS-A, there was a significant improvement in scores of 0.65 (p=0.024) but when QS-B was administered before QS-A, there was no difference in scores.

Conclusions: It is possible that QS-B was easier than QS-A, and that was the only reason for improvement in score. If true, however, one would have expected QS-B scores to be higher in both conditions. It would appear that practicing an SOE may help improve performance on a subsequent examination.

Take-home Messages: A review course with mock SOEs can improve performance on subsequent SOEs.

OTT-PB7-06
Foundation doctors working at night: risks and benefits?

Ross Coomber (*Ipswich Hospital, Trauma and Orthopaedics, Ipswich, UK*), Daniel Smith (*Research and Analytics Manager, London Deanery, London*), Daniel McGuiness (*Royal Free Hospital, Medicine, London*), Emily Shao (West Middlesex Hospital*, Medicine, London*), Ramawad Soobrah (*Ealing Hospital, General Surgery, London*), Andrew Frankel (*London Deanery, Foundation, London*)

Background: Foundation Training is designed for doctors in their first two years post-graduation. The number of foundation doctors (FD) in the UK working nights has reduced because of a perception that clinical supervision at night is unsatisfactory and that minimal training opportunities exist. We hypothesise that removing FDs from nights may be detrimental to training.

Summary of Work: Using a survey, we assessed the number of FDs working nights in London, FDs views on working nights and their supervision at night. We also evaluated whether nights provided opportunities to achieve foundation competencies. A clinical supervision score evaluated FDs perception of supervision at night and the effect of Hospital at Night (HaN) teams.

Summary of Results: 83% (N=2,157/2,593) of FDs completed the survey. Over 90% of FDs who worked night's felt the experience they gained improved their ability to prioritise, make decisions and plan. FDs who worked nights reported higher scores for achieving competencies in history taking (2.67 vs 2.51 p=0.00), examination (2.72 vs 2.59 p=0.01) and resuscitation (2.27 vs 1.96 p=0.00). The majority (65%) felt adequately supervised and this was increased when part of a HaN team.

Conclusions: FDs find working nights a valuable experience, which enhances their ability to achieve foundation competencies. Important training opportunities exist at night, which are additional to those encountered during daytime working. While these experiences are valuable they must be well supervised to ensure patient safety.

Take-home Messages: With the correct supervision, working nights offers valuable training for FDs, enhancing their ability to prioritise and make decisions while providing experience to achieve foundation competencies.

OTT-PB7-07
Assessment of Immediate Reaction to Institutional Goal-Oriented Leadership Training for Chief Residents

Anuradha Lele Mookerjee (*Cooper University Hospital, Cooper Medical School of Rowan University, Internal Medicine, Camden, NJ, USA*), Rajani Sharma (*New York Presbyterian Hospital - Columbia Medical Center, Internal Medicine, New York ,NY*), Shamit Desai (*Albert Einstein Medical Center, Medicine, Philadelphia, PA*), Susan Cavanaugh (Cooper Medical School of Rowan University, *Library, Camden, NJ*), Vijay Rajput (*Cooper University Hospital, Cooper Medical School of Rowan University, Internal Medicine, Camden, NJ*)

Background: Chief residents (CRs) are required to carry out administrative, clinical, managerial, and organizational duties with emotional intelligence. There is a need for adaptive leadership competency framework for CRs who lead and manage near peers.

Summary of Work: A two day workshop with an integrated curriculum was delivered to a multidisciplinary group of CRs to develop effective

targeted key leadership skills in communications, goal setting, empathy, emotional intelligence, management, professional growth and team motivation. A series of case-based interactive workshops were delivered to CRs from eight different departments. Sixteen (n=16) chief residents participated in training. We measured the immediate impact of the leadership development and training program (LDTP). Chief residents completed both a pre-test and post-test with non -validated 15 self assessment statements, grouped by seven key leadership qualities and skills.

Summary of Results: The CRs showed immediate positive change in the "Communicating Goals" between pre-test and post-test (54%, CI 11-69%). The categories "Working with Others" (69.7%, CI 51-84%) and "Managing Services" (72.7%, CI 55-87%) showed no immediate changes. The workshop facilitated a collaborative environment for the exchange of ideas and implementation of solutions to anticipated challenges in organization.

Conclusions: Focused institutional goal oriented leadership workshop can be helpful to chief residents and encourage collaboration between departments. Experiential learning with trained faculty from multiple disciplines enables a positive outcome in developing leadership skills in Chief residents.

Take-home Messages: Chief resident leadership training should provide experiential training in working in teams and managing team on clinical services.

OTT-PB7-08
Value Added General Practice Training Course

Kelly Thresher (*Wessex School of General Practice, Winchester, UK)*, Jonathan Rial (*Wessex School of General Practice, Winchester*)

Background: The introduction of the nMRCGP, and in particular the CSA (Clinical Skills Assessment) component, has seen a shift in teaching on the General Practice Training Course towards preparing trainees for the assessment. This has been compounded by the perceptions of need of those failing the exam. During the Specialist Trainee Year 3 trainees have historically worked within a small group which they remain in throughout the year. It was noted by the facilitators that trainees passing the CSA found themselves in groups with trainees who continued to want to practise skills in preparation for the CSA, creating a mismatch in learning needs.

Summary of Work: After the February CSA results, the groups were rearranged to take account of those who had passed. The trainees who had not passed continued to focus on communication skills. For those passing, five sessions of 'Value Added' small group work were planned. The sessions were resourced and led by the trainees:
• Planning
• Alternative medicine
• A topical debate
• Exploring careers
• Faith in Healthcare

A key aim was to focus on establishing and developing the learning needs of the trainees at the end of their training and to promote self-directedness.

Summary of Results: The sessions were evaluated using feedback from the trainees and facilitators. Findings suggested that trainees were not ready to embrace self directed learning in their post exam 'honeymoon' period.

Conclusions: The sessions did not have the outcome the educators anticipated, as they perceived self-directedness and leadership would be developed. Unfortunately the trainees embraced the clinical learning but missed the 'Value-Added' aspects.

OTT-PB7-09
Evaluating Innovative Electronic CME for Physicians: Actionable Nuggets™ on SkillScribe

Karen M. Smith (*Queen's University, Office of Continuing Professional Development, Kingston, ON, Canada)*, Mary Ann McColl (*Queen's University, Kingston*), Alice Aiken (*Queen's University, Kingston*), Jay Joseph (Queen's University, *Kingston*), Laura McDiarmid (*Queen's University, Kingston*), Danielle Naumann (*Queen's University, Kingston*)

Background: CME for primary care physicians often focuses on issues that are common in primary care practice; however, the literature shows that physicians have expressed anxiety about patients presenting with low prevalence conditions associated with high risk for morbidity. Primary care physicians are often resistant to conventional approaches to CME, and are unlikely to attend CME that will facilitate the care of only a few patients in their practice. Actionable Nuggets™ is an approach that has been developed to address the issue of knowledge translation (KT) in primary care. This innovative KT tool consists of a series of 20 visually attractive postcards that are time released through conventional mail system; participating physicians receive 1 per week and complete brief learning activities in order to gain education and knowledge on the common conditions associated with the featured disorder. Existing Actionable Nuggets™ have been approved for CME credit in Canada; however, broad distribution of this tool has been impeded by the cost associated with production and delivery.

Summary of Work: SkillScribe is an electronic platform that has been designed to deliver Actionable Nuggets through personal electronic devices. This presentation describes the evaluation of learning and knowledge translation using SkillScribe to deliver Actionable Nuggets.

Summary of Results: Pending completion of data collection and analysis.

Conclusions: SkillScribe is an electronic platform that has the potential to replicate the strengths of Actionable Nuggets, using a convenient electronic medium, time-released content, and interactive testing. Moreover, it is not associated with the cost of the traditional approach.

Take-home Messages: Electronic CME is emerging as a valuable knowledge translation tool that can be tailored to meet the unique needs of primary care physicians.

OTT-PB7-10
Written or Typed Note-Taking and the Effect on Memory Retention: Implications for Electronic Continuing Medical Education

Danielle N. Naumann (*Queen's University, Office of Continuing Professional Development, Kingston, Canada),* Karen M. Smith (*Queen's University, Office of Continuing Professional Development, Kingston*), Laura McDiarmid (*Queen's University, Office of Continuing Professional Development, Kingston*)

Background: Electronic Continuing Medical Education (CME) is emerging as a popular option for physicians to conveniently gain knowledge leading to CME credit. Often, these opportunities require participants to engage in electronic note-taking formats and post-test surveys using their electronic devices, with little opportunity or requirements for handwritten notes to supplement their learning. There is very little information available in the scholarly literature on the implications that this approach to learning can have on memory retention for medical knowledge within the context of CME.

Summary of Work: 20 Physician interviews, pre and post-test evaluations, qualitative evaluation of changes to practice following participation in electronic CME requiring either hand-written or type-written note taking

Summary of Results: The CME audience consists of a range of healthcare professionals with varying levels of experience and familiarity with current learning technologies. Results currently pending data analysis.

Conclusions: We expect to conclude that innovative electronic approaches to CME need to consider the effects of note-taking preference on memory retention. Electronic CME is offered to professionals with a range of learning styles, and may not lead to effective practice change if participants are not provided with the opportunity to engage in note-taking procedures that complement their learning style.

Take-home Messages: Considering the widespread movement toward innovative electronic CME and the use of electronic technology in CME, it is important that CME providers understand the strengths and limitations of handwritten and typed note-taking, repetition, and cognitive organization; and the resultant effects on memory retention and changes to practice.

OTT-PB7-11
Assessing physicians' performance in Quebec

Monique Robert (*Collège des médecins du Québec, Practice Enhancement Division, Montreal, Canada),* Marc Billard (*Collège des médecins du Québec, Practice Enhancement Division, Montreal*), Thiffault Johanne (*Collège des médecins du Québec, Practice Enhancement Division, Montreal*), François Goulet (Collège des médecins du Québec*, Practice Enhancement Division, Montreal*)

Background: Since 1974, the Collège des médecins du Québec has been required by law to assess the practice of physicians on a non-voluntary basis. Inspections are made by peer assessments. Goulet F & al (2002) published: > 95% of physicians have satisfactory clinical performance in terms of quality of care.

Summary of Work: In 1997, the Practice Enhancement Division of the Collège developed inspection programs instead of relying on random selection. Evaluations completed between 2001 and 2012 have been analysed to determine if programs enable to establish risk factors affecting quality of care.

Summary of Results: Between 2001 and 2012, 1,918 assessments were made. The main three programs are: Diploma>35 years, Reporting from Inquiry Division and Professional reporting. Unlike random selection, programs highlight significant differences in physicians' performance. Furthermore, senior physicians have a higher probability to be inspected. Protective factors for the maintenance of competency are: continuous professional development, good record keeping and, with regard to specialists, practicing in hospital settings.

Conclusions: We believe our selection programs are efficient and fulfill the objective of identifying physicians at risk regarding quality of care.

Take-home Messages: - Peer-assessments based on selection programs can identify physicians at risk with regard to quality of care.
- Protective factors can be identified.

OTT-PB7-12
Self Assessment items linked to journal articles: A survey of rationale, structure and impact in practice

Douglas L Wooster (*University of Toronto, Surgery, Toronto, Canada)*, Elizabeth M Wooster (*OISE/University of Toronto, Leadership, Higher and Adult Education, Toronto*)

Background: Multiple choice items (MCI) are frequently associated with journal articles as a value-added strategy to promote self assessment, linkage to practice, further reflection on the content of the article and as a means to provide continuing education (CE) credits for journal reading. The nature of such activities and impact as a CE strategy has not been studied.

Aim: The aims of this study were to identify patterns of use and practical impact of MCI associated with journal articles.

Summary of Work: A literature review and survey of CE activities associated with journal articles will be performed. Specific examples will be selected and practitioners in those specialties will be asked to complete a focused survey of the impact of these activities compared to other CE activities to obtain a quantitative data set.

Summary of Results: Based on the limited literature that exists regarding this topic, we are hypothesizing that multiple choice questions associated with journal articles will be poorly constructed, instructions for authors are absent or limited and the associated modules are a poor educational experience. Analysis of

common item flaws may provide cogent guidance to authors and editors to improve the quality of such initiatives. An understanding of practitioners use may guide improved education strategies.

Conclusions: Analysis of common item flaws may provide cogent guidance to authors and editors to improve the quality of such initiatives. An understanding of practitioners use may guide improved education strategies.

Take-home Messages: Analysis of common item flaws may provide cogent guidance to authors and editors to improve the quality of such initiatives. An understanding of practitioners use may guide improved education strategies.

Ottawa 2014 Abstract Book

SUNDAY APRIL 27
1:30PM – 3:00PM

SYMPOSIA

OTT-SB-1
Validity Issues in Medical Education Assessment

Presenters: Katharine Boursicot (*Lee Kong Chian School of Medicine, Singapore)*, Richard Fuller (*University of Leeds, Leeds*), Marjan Govaerts (*University of Maastricht, Maastricht*), Saskia Wools (*CITO*), Chair: Trudie Roberts (*University of Leeds, Leeds*

Summary: The symposium brings medical education testing under scrutiny in relation to more modern argument-based approaches to validity. While the traditional psychometric discourse has been, and is still, dominant in medical education assessment, there are growing concerns that there are limitations to this view, especially in the context of newer assessment tools, such as workplace-based assessments. It is our intention to highlight the wider outlook provided by the unitary concept of validity, with its requirement to consider a range of different factors/evidence when making interpretations of test results, especially in high-stakes situations. The presenters will provide an international perspective of how far the modern views of validity have impacted on medical education testing.

OTT-SB-2
The Non-Medical Expert Roles: Methodological Challenges to Assessment and Evaluation

Presenters: Ayelet Kuper (*Wilson Centre, Toronto)*, Cynthia Whitehead (*University of Toronto, Toronto*), Rachel Ellaway (*Northern Ontario School of Medicine*), Discussant: Brian Hodges (*University of Toronto/Wilson Centre, Toronto*)

Summary: The widespread adoption of role-based competency frameworks, such as CanMEDS, has highlighted the importance of assessing physician roles (often called "non-Medical Expert" or "Intrinsic" roles) that go beyond the performance of medical knowledge and technical skills. This symposium will provide a range of contrasting theoretically-grounded non-psychometric perspectives that challenge concepts such as authenticity and identity that are bound up with the non-Medical Expert roles. We will explore novel approaches to the assessment of these roles and the evaluation of the curricula that support them. Our aim is to draw the audience into a robust and constructive conversation about the assessment of the non-Medical Expert roles in order to explore theoretical, methodological and practical directions for medical educators and researchers to employ in their own practices.

WORKSHOPS

OTT-WB-1
Faculty Development Strategies Focused on Assessment: Not Boring!!

Presenter(s): Deborah Simpson(*Aurora Health Care, Academic Affairs & Family Medicine, Milwaukee, United States*), LuAnn A. Wilkerson(*David Geffen School of Medicine at UCLA, Medical Education, Los Angeles, United States*), Janet P. Hafler (*Yale School of Medicine, Teaching and Learning Center & Pediatrics, New Haven, United States*

Background: The heightened importance of assessment with competency based medical education requires that faculty be facile with basic assessment concepts and able to apply them in classroom, simulated and clinical settings. Yet faculty development strategies often have limited attendance and impact as they are perceived as "dry" and/or not applicable.
Intended Outcomes: To provide participants with a toolbox of proven, engaging faculty development strategies to prepare faculty for their roles as competency-based assessors.
Structure: Following a 10 minute approaches to assessment focused faculty development overview, workshop participants will experience and discuss an array of evidence based yet fun faculty development strategies: (1) Educational Measurement Workshop -A "Sweet Approach" to Understanding the Basic Principles of Educational Measurement; (2) Defining Assessment Terms "In 3 Words"; (3) Crossword Puzzles; (4) PBL Tutor; (5) Building a Penguin – Teacher Assessment; (6) Milestone Theater. Each of the six strategies will be presented as follows with time variation due to differences in strategy complexity: 2-5 minute introduction to strategy (e.g., objectives and lesson plan) 8-15 minute excerpt highlighting the opportunity to practice a skill with a hands on activity, followed by discussion and debriefing, 2-4 minute discussion of key facilitation lessons. Session will close with gathering engaging ideas and strategies from the audience with a brief review of demonstrated strategies with access links (e.g., MedEdPORTAL, POGOe, FMDRL.org) and an assessment sing-a-long.
Who Should Attend: Primary audience: Individuals seeking to obtain faculty development skills and resources to provide just in time to half-day workshops on assessment.
Level of Workshop: Intermediate

OTT-WB-2
Assessing why trainees are struggling or failing

Presenter(s): **Alan** Cook (*Health Education South West, Severn Postgraduate Medical Education, Bristol, United Kingdom*), Davinder Sandhu (*Health Education South West, Severn Postgraduate Medical Education, Bristol, United Kingdom*)

Background: The biggest challenge for Medical Trainers/Educators is how to recognise, analyse and manage performance issues. About 5% of trainees struggle to complete their training and require additional targeted time or an extension of training with some leaving the profession. Many others have periods when they really struggle and need extra support with their training and clinical performance or in managing life pressures. This can be a huge drain of resources and upsetting for faculty who often feel underprepared to deal with such occurrences. Failure to address them properly can lead to bitterness, large remedial costs and legal challenges to educational institutions and employers.

Intended Outcomes: 1. To be able to effectively analyse and assess the causes of poor performance
2. To gain an understanding of why trainees struggle and/or fail
3. To know how to pick up these issues early and skilfully
4. To know how to deal effectively with such matters and get good outcomes.

Structure: An initial presentation and discussion led by Alan Cook will focus on why trainees struggle or fail and how to evaluate the factors involved. This will be followed by group work reviewing case materials and challenging situations. The workshop will conclude with advice based on researched best practice and guidance on how to gauge performance support needs and develop focused action plans with struggling trainees.

Who Should Attend: This workshop is open to all involved with education; undergraduate, postgraduate, interdisciplinary nursing and allied health professionals.

Level of Workshop: Intermediate

OTT-WB-3
SP assessment in a competency-based world

Presenter(s): John Shatzer (*Vanderbilt Universtiy School of Medicine, Nashville, United States*), Sally Santen (*University of Michigan School of Medicine, Medical Education, Ann Arbor, United States*), Gail Furman (*National Board of Medical Examiners, Philadelphia, United States),* Tony Errichetti (*New York Institute of Technology, College of Osteopathic Medicine, New York, United States*)

Background: Competency- based assessment is important to graduating health professionals in a complex health environment. Within this process is programmatic assessment that systematically examines the key performance measures for health professionals in the multiple domains of practice. Standardized Patient (SP) assessments is one method of determining successful performance. As health science education shifts to a competency-based model, SP assessments have measured clinical skills but also are increasingly used to assess professionalism, interpersonal communication, compassion, teamwork, and other skills that are essential to patient-centered care.

Intended Outcomes: Participants completing this workshop will create a framework to help guide the development of competency-based outcomes using SP assessments. Elements will include mapping assessment to curricular outcomes within an assessment program, SP selection for assessment, measurement issues with the competency framework, including milestones and Entrustable Professional Activities (EPAs), and their roles in self-assessment and feedback.

Structure: Workshop activities will be a balance of large and small group work with report outs and lessons learned for take home action plans. Primary sections and discussion to engage learners will center on an introduction to competency-based assessment and various interpretations from the audience. We will connect with participants prior to the workshop to provide them with a pre-workshop exercise to bring to the discussion. The exercise will focus on 2-3 constructed example competencies that could present challenges for an SP exercise, but appears to be appropriate for it. Participants will be asked to respond to questions related the competency and will serve as a basis for workshop activites and outcomes.

Who Should Attend: Program directors, clerkship directors, SP program directors, SP educators, assessment program directors

Level of Workshop: Intermediate

OTT-WB-4
Social Accountability: a logical approach to its assessment

Presenter(s): Trevor Gibbs (*AMEE, Dundee, United Kingdom*), Charles Boelen (*Independent Consultant, Sciez-sur-Léman, France*), Robert Woollard (*University of British Columbia, General Practice, Vancouver, BC, Canada*)

Background: The concept of Social Accountability is now gaining momentum in terms of its understanding and implementation in the curricula of various healthcare learning organisations. As the concept grows it is finding its way into undergraduate and postgraduate healthcare education, continuing professional development, accreditation procedures and research and most recently in the measurement of excellence in teaching within medical schools; the AMEE- ASPIRE initiative.

Although the concept and particularly what it means is gaining momentum, experience, especially from the recent ASPIRE initiative and the most recent AMEE conference held in Prague, suggests that further explanation of its true meaning is required; a firm understanding of the spectrum of social engagement from social responsibility, social responsiveness through to true social accountability.

As important, if not more important however, is an ability of healthcare learning establishments to be able to assess their global approach to social accountability; what has been the effect of this new concept on the community to which they serve, as well as being able to

assess the effect of introducing social accountability on individual programmes or individual students.

Intended Outcomes: - A greater understanding of the concept of Social Accountability

- A realisation that a spectrum of social engagement exists

- A sharing of a validated measuring tool to assess social accountability

- An opportunity to apply this measuring tool to individual programmes

Some Canadian schools lead the way in Social Accountability, and although this is not universal we would hope to use the experience from Canada to benefit others.

Structure: Workshop format, with an early didactic input and sharing of terminology, followed by small group application to curricula presented by the audience.

Who Should Attend: Curriculum developers, educational managers, programme directors and all with an interest in the subject.

Level of Workshop: Intermediate

OTT-WB-5
Overcoming geographic barriers in performance-based assessment: Remote standardized patients (RSPs)

Presenter(s): Elizabeth Kachur(*Medical Education Development, New York, United States*), Erik Langenau (*Philadelphia College of Osteopathic Medicine, Family Medicine, Philadelphia, PA, United States*), Jeanne Sandella (*National Board of Osteopathic Medical Examiners, National Center for Clinical Skills Testing (NCCST), Conshohocken, PA, United States),* Chaoyan Dong (*National University of Singapore, Medical Education Unit, Singapore*)

Background: Sometimes it is logistically difficult and costly to bring trainees and performance-based assessment programs together, into the same place. Learners at all training levels may find themselves at remote sites and, we need to explore new strategies to overcome geographical barriers. The Internet already provides unique opportunities for medical care, conferences and meetings. Telemedicine and distance education are here to stay. Programs such as SKYPE can also be used for performance-based assessments. This workshop will provide some examples and explore the key elements needed to create such programs by engaging with RSPs in the US, Singapore, the Netherlands and South Africa.

Intended Outcomes: By the end of the session participants will be able to:

1. Describe how a remote encounter with a standardized patient can proceed

2. Detail the opportunities and challenges inherent in using RSPs

3. Explore realistic strategies for implementing a remote performance based assessment at their home institution.

Structure: 5 min   Introduction & Orientation

10 min   RSP Projects – brief description of the NBOME and other related projects

10 min   RSP Encounter Set-up – selection of participant volunteers to interview RSPs via SKYPE

40 min   SKYPE Encounters with RSPs in the US, Singapore, the Netherlands and South Africa - volunteers perform interview (5 minutes) and receive instant feedback from the RSP (3 minutes), other participants reflect on the experience using worksheets (ideally connections can be set up with all 4 sites, but if there are technical difficulties fewer will be adequate to provide participants with a general understanding of this assessment tool)

20 min   Opportunities & Challenges – think-pair-share, large group discussion

5 min   Take-Home Points & Closure.

Who Should Attend: faculty & educators.

Additional workshop facilitators: Jan Joost Rethans, University of Maastricht, the Netherlands; Ina Treadwell, University of Limpopo, South Africa.

Level of Workshop: Introductory

OTT-WB-6
Strategic Formative Feedback and Reflection as the Foundation for Effective Learning in a Competency-based Curriculum

Presenter(s): Sharon Krackov (*Albany Medical College, Medical Education, New York, United States*), Antoinette Peters (*Harvard Medical School, Population Medicine and Academy Center for Teaching and Learning, Boston, United States*), Henry Pohl (*Albany Medical College, Medical Education, Albany, United States),* Joan Sargeant (*Dalhousie University, Medical Education, Halifax, Canada*)

Background: Learning is expedited by outcome-based objectives, regular and ongoing assessment, formative feedback, and reflection. The goal of feedback is to reduce the discrepancy between students' current performance and their performance target. Feedback is more effective if the objectives are clear and the student is committed to them. The concept of "deliberate practice", which involves ongoing formative assessment accompanied by strategic feedback, is an accepted method of experiential learning leading to competence. Strategic feedback reinforces learners' achievements and helps them achieve the curricular objectives. Reflection allows learners to focus their learning while sensitizing them to their ongoing learning needs. Incremental learning based on these practices fosters development of a learner's expertise over the course of the educational program.

The level, timing and frequency of feedback affects learners' performance. Levels of feedback include: "task, processing the task, self-regulation, and the self as a person". The first three lead to increasingly deep learning, but the last can inhibit learning. Timing and frequency of feedback relate to the cognitive levels of learning. Feedback should be given promptly and frequently for lower order cognitive tasks, but less often

and after a delay for higher order tasks that require reflection.

Intended Outcomes: Participants will demonstrate teaching/learning cycles that link outcome based objectives, assessment, feedback and reflection to facilitate learning in a competency-based educational program.

Structure: Interactive discussion to produce strategies to link outcome based learning objectives, ongoing feedback and reflection. Opportunities to practice skills.

Who Should Attend: Faculty and curriculum planners who use feedback to help learners achieve outcome-based objectives.

Level of Workshop: Advanced

OTT-WB-7
Contextualising assessment in the basic medical sciences: Designing appropriate assessment strategies

Presenter(s): Dujeepa D. Samarasekera (*National University of Singapore, Medical Education Unit, Singapore*), Matthew C.E. Gwee (*National University of Singapore, MEU, Singapore*), Chay Hoon Tan (*NUS, MEU, Singapore*)

Background: In the past assessment in the basic medical sciences (BMSc) tested mainly in-depth scientific knowledge (i.e. factual content knowledge) of a discipline. The response of students to such an assessment strategy was simply to memorise and regurgitate scientific facts in examinations conducted by the respective disciplines. The assessment strategy therefore drives students to undertake rote-learning an educational outcome which has received much criticism especially because of its negative steering effect. Today, however, there is clear recognition that students need to, not only acquire foundational knowledge, but also to develop knowledge processing skills such as critical thinking and reasoning, as well as problem-solving Student acquisition of knowledge processing skills will be consistent with the application of scientific thinking in approaching problems. There is also now a major shift in assessment paradigm from the assessment of learning to assessment for learning. In fact assessment is now considered a potentially powerful learning tool on the basis that assessment drives student learning behavior. Thus, it is imperative for teachers in the BMSc disciplines to have a clear understanding of the educational implications of the major shift in assessment paradigm, especially the need to contextualize assessment in the BMSc to ensure, not only the relevance of the BMSc disciplines to medical education, but also to contribute to the intellectual skills development of students for future medical practice.

Intended Outcomes: The primary aim of this workshop is to provide participants with:
• insights on the major shift in assessment paradigm and its educational implications
• a pedagogical approach: to contextualizing assessment in the BMSc

• an overview of blueprinting and the key features of various assessment strategies
• opportunities to engage in hands-on small group activities.

Structure: Small group hands-on activities developing written assessment in basic sciences.

Who Should Attend: Basic Science teachers, curriculum designers and administrators.

Level of Workshop: Intermediate

ORALS

OTT-OB-1: ASSESSMENT AND LEARNING

OTT-OB-1-1
Testing to change practice: Does testing facilitate learning and practice change in continuing medical education communities of learning?

Kevin Eva (*University of British Columbia, Medicine, Vancouver, Canada)*, Heather Armson (*University of Calgary, Family Medicine, Calgary*), Stefanie Roder (*Foundation for Medical Practice Education, Research, Hamilton*), Jacqueline Wakefield (McMaster University, *Family Medicine, Hamilton*)

Presenter: Kevin Eva (*University of British Columbia, Medicine, Vancouver, Canada*)

Background: Using assessment to facilitate learning is a well-established priority in medical education. Studies have shown testing can influence knowledge indirectly through motivation and directly by enhancing memory. Whether such strategies are effective for continuing professional development in established practitioners is unknown. We aimed to determine if delivering feedback through a web-based test would influence physicians engaged in maintenance of competence.

Summary of Work: Family physicians belonging to a practice-based learning program studied two educational modules either independently or in small learning groups. Prior to the learning sessions they were assigned the task of completing a pre-test or reading a relevant review article. After the learning session they completed a post-test and documented any practice changes made.

Summary of Results: 141 physicians completed the study, 118 studying in small groups. Average lag between pre- and post-test was 6 weeks. Relative to those given a review article, physicians given a pre-test: (a) spent less time completing the assigned task (16.1vs26.3 minutes); (b) performed better on post-test (69.4%vs64.2%); and (c) reported a greater rate of practice change (63.8%vs49.0%). These effects were more pronounced in independent learners relative to small group learners.

Conclusions: Low stakes formative quizzes can influence the rate at which continuing professional development influences practicing physicians. Such

externally derived tests of knowledge/ability may be particularly important for those striving to maintain competence without the benefit of group interaction.

Take-home Messages: Testing has value for learning even for those who have previously learned all they needed to pass the tests.

OTT-OB-1-2
Impact of a change in assessment rubric on the academic performance of first year medical students at the University of the West Indies (UWI), Barbados

D Cohall (*The University of the West Indies, Cave Hill, Barbados)*, D Skeete (*The University of the West Indies, Cave Hill*)

Background: The Faculty of Medical Sciences, UWI, Cave Hill recently changed the assessment policy for preclinical courses in the undergraduate medical programme. There was a transition of the heavier weight in the coursework component (60%) to the final exam component, formerly 40%, and a stipulation that students had to pass the final exam in order to successfully complete the course. The objective of this study was to determine if the change in the assessment rubric altered the academic performance of students in 'Fundamentals of Disease & Treatment' (FDT), a compulsory, year-long course in the MBBS programme.

Summary of Work: Grades before and after the policy change were recorded and summarized systemically for the first year medical students during the academic years 20011/2012 and 2012/2013. In course and final exam grades were categorised and analysed separately. Student's t-test was used to determine any significant differences between the means of academic performance between the cohorts.

Summary of Results: There were significant improvements in assignments, mid-semester tests, final course exams and overall course grades in the subsequent student cohort after the change in assessment policy.

Conclusions: The results from the overall data analysis would justify some relationship between assessment rubrics and academic performance in the FDT course.

Take-home Messages: Assessment rubrics are critical to the learning process but should not be considered the main determinant of academic performance of students.

OTT-OB-1-3
Knows how, shows how, does? Assessing the impact of facilitated small-group learning on GP trainees' consultation skills

Selina Sawhney (*Wessex School of General Practice, Wessex Deanery, UK, Southampton Patch, GP Education Unit, Southampton)*, Duncan Platt (*Wessex School of General Practice, Wessex Deanery, UK, Southampton Patch, GP Education Unit, Southampton*), Alice Mavrogordato (*Wessex School of General Practice, Wessex Deanery, UK, Southampton Patch, GP Education Unit, Southampton*), Samantha Scallan (Wessex School of General Practice, Wessex Deanery, UK*, Southampton Patch, GP Education Unit, Southampton*), Kelly Thresher (*Wessex School of General Practice, Wessex Deanery, UK, Southampton Patch, GP Education Unit, Southampton*), Johnny Lyon-Maris (*Wessex School of General Practice, Wessex Deanery, UK, Southampton Patch, GP Education Unit, Southampton*)

Background: Preparation for high stakes assessments such as the CSA (RCGP membership exam, UK) can lead trainees to focus on the assessment process and rigid consultation models rather than their knowledge and skills, resulting in them losing sight of the interaction at the heart of the consultation.
This innovative programme of education aimed to help trainees gain insight into their consulting skills whilst preparing for the CSA. Using role-play with facilitated feedback, sessions helped trainees self assess their performance and focus on aspects of the consultation that required further development.

Summary of Work: Trainees worked in small groups facilitated by three newly qualified GPs, the 'Near Peer Educators,' with recent experience of passing the CSA assessment. Each group met for four sessions, during which they role-played CSA-style thumbnail scenarios devised by the facilitators. During the first two sessions they gave and received feedback using the ALOBA approach; in the last two sessions feedback was structured using the generic RCGP marking guidelines. At the end of the programme, the trainees attended a mock CSA circuit comprising cases written by the facilitators. Trainees were assessed using the RCGP marking scheme and received oral and written feedback. Evaluation data was gathered from participants (pre-, mid and post the programme), the group facilitators and the assessors.

Summary of Results: Major benefits reported by the participants were: insight into areas for development, improved feedback skills and refocused learning needs. Facilitators reported observing improved consultation skills.

Conclusions: Approaches to learning that promote shared reflection and constructive feedback can have a positive impact on trainees' learning and consultation skills.

Take-home Messages: Trainee exam performance improved after a structured eductional programme.

OTT-OB-1-4
Fostering Transformative Learning in a Social Pediatrics Research Summer Studentship Through Empowerment and Assessment

Susanna Talarico (*Hospital for Sick Children, University of Toronto, Pediatrics, Toronto, Canada)*, Mohammed Zubairi (*University of Toronto, Pediatrics, Toronto*), Tony Barozzino (*St Michael's Hospital, University of Toronto, Pediatrics, Toronto*), Denis Daneman (Hospital for Sick Children, University of Toronto*, Pediatrics, Toronto*), Angela Punnet (*Hospital for Sick Children, University of Toronto, Pediatrics, Toronto*), Tina Martimianakis (*Hospital for Sick Children, The Wilson Centre, University of Toronto/University of Toronto, Pediatrics, Toronto*)

Background: Medical educators are challenged to integrate new strategies to prepare future pediatricians with the necessary skills to address health disparities (Ford-Jones et al, 2008). In response, the Social Pediatrics Research Summer Studentship (SPReSS) program was developed and implemented for medical students at the University of Toronto.

Summary of Work: Transformative learning principles were applied to the program and curriculum design to facilitate critical reflection and learning, and as an innovative approach to program development and evaluation. The curriculum consisted of research and clinical placements, as well as a formal seminar series. Students were asked to write a reflection describing a situation that challenged their thinking or caused them to re-evaluate previously held attitudes or perceptions. The reflections were assessed with a rubric, and the program was evaluated through thematic analysis of the reflections and an exit survey of faculty and students.

Summary of Results: The analysis revealed that students were preoccupied by their empathic responses to marginalized patients. They described feeling empowered to act as advocates and that these feelings were reinforced through assigned readings or role modeling of faculty. Students found the program both challenging and rewarding, particularly the integration of a clinical and research experience. Faculty found students to be engaged and reflective, making connections between assigned readings and practical experiences.

Conclusions: The theory of transformative learning can be applied to medical education programs and curricula to help students identify with advocacy and develop the necessary critical reflective and advocacy skills to address health disparities.

Take-home Messages: A key strategy to fostering transformative learning is the incorporation of an authentic assessment tool such as the rubric to encourage and evaluate student reflections.

OTT-OB-1-5
Why educators should be evaluating using behaviour-change models: A example in global health

Lucie M T Byrne-Davis (*University of Manchester, Manchester Medical School, Manchester, UK)*, Jo Hart (*University of Manchester, Manchester Medical School, Manchester*), Chris Armitage (*University of Manchester, School of Psychological Sciences, Manchester*), Ged J Byrne (University of Manchester, *Faculty of Medical and Human Sciences, Manchester*)

Background: There is a lack of evidence about whether, how and in what circumstances, education leads to professional practice changes. Health psychologists evaluate interventions using assessments of behavior. Educators typically do not measure behavior, nor identify behaviour change techniques and predictors of behaviour are not explored. Evaluation of education is typically increase in knowledge, skills, satisfaction or patient outcomes. Understanding the efficacy and efficiency of education is even more crucial in low and middle income countries where workforce development is key to improving patient health.

Summary of Work: We tested the feasibility, validity and reliability of assessing predictors of behavior in a multiprofessional group taking part in a one-day course on acute illness management (AIM) in Uganda. Using questionnaires, we measured behavioural intention and predictors of behavior (capability, opportunity and motivation) post-course and behavior 1 month later. We measured knowledge pre and post course.

Summary of Results: 51 health professionals took part in AIM. There were high levels of self-reported capability and motivation and lower levels of opportunity. There was a significant increase in knowledge pre to post course. Behavioural intentions were significantly higher than self-reported behaviours at 1 month.

Conclusions: It is feasible to measure behavioural predictors alongside education. Knowledge was not predictive of self-reported behavior 1 month later.

Take-home Messages: Evaluation of the impact of education by measuring predictors of behaviour allows tailoring of education to increase its efficacy/efficiency. Knowledge is necessary but not sufficient to change professional practice.

OTT-OB-2: ASSESSMENT AND CURRICULUM CHANGE

OTT-OB-2-1
Higher specialist scientist training: Training and assessing the leaders in clinical science

Sue Hill (*Health Education England and NHS England, Modernising Scientific Careers, London, UK)*, Shelley Heard (*Health Education England, Modernising Scientific Careers, London*), Graham Beastall (*Health Education England, Modernising Scientific Careers, London*), Maria Bussey (Royal College of Surgeons*, Education, London*), Winnie Wade (*Royal College of Physicians, Education, London*), Jo Brinklow (*Royal College of Pathologists, Education and Training, London*)

Background: Healthcare science (HCS) is essential to safe/effective patient care. There are about 50,000 employees in the non-medical HCS workforce in the NHS, impacting on approximately 80% of clinical diagnoses. Modernising Scientific Careers (MSC) is a UK wide educational/training framework for HCS and includes higher specialist scientist training (HSST), a formal five-year work based programme supported by a part-time doctoral level award, designed to train clinical scientists to practice at consultant level similar to their medical counterparts.

Summary of Work: Overseen by a national MSC team to ensure consistency of the process and outcomes, the Medical Royal Colleges (MRCs) through the UK Academy of MRCs engaged with senior scientists to develop curricula, assessment processes and clinical scientific

learning programmes, reflecting standards similar to medical specialist training.

Summary of Results: To date, 5 MRCs have been involved in the process, with 3 submitting 16 of the 20 planned curricula through governance systems. Assessment of training fellows is through intensive assessment of workplace practice, a final competence assessment and a doctoral-level programme. Using a common assessment strategy, each MRC working group developed fit-for-purpose assessments documenting progress using the National School of Healthcare Science e-learning portfolio, with universities assessing the academic/research elements.

Conclusions: Co-operation between the MRCs, healthcare science profession and higher education has led to the development of robust curriculum/assessment processes ensuring that clinical scientists are trained to the same standards as medical specialists.

Take-home Messages: Cross professional educational developments can bring benefits to patients by ensuring that standards of clinical and scientific practice/assessment are consistent and of a high quality.

OTT-OB-2-2
Links between the faculty undergraduate assessment and the new licensing Swiss Federal Examination in Human Medicine (FEHM): A retrospective cohort study

Bernard Cerutti (*University of Geneva, Faculty of medicine, Geneva, Switzerland)*, Elisabeth van Gessel (*University of Geneva, Faculty of medicine, Geneva*), Mauro Serafin (*University of Geneva, Faculty of medicine, Geneva*), Nu Vu (University of Geneva*, Faculty of medicine UDREM, Geneva*)

Background: A national revised licensing exam (FEHM), with a 300 vignette-based multiple choice (MC) examination and a 10 stations standardized-patient based clinical exam (CS), was introduced in 2011. The new exam was an opportunity to investigate and compare students' longitudinal performance on the intramural assessments (IA) and their scores on the FEHM administered at the end of medical school.

Summary of Work: Three cohorts (2011 – 2013) of students were included in the study (N=393). The IA included scores on the MC and computer-based (CB) exams and, laboratory (LAB), and CS practical exams taken during the six-year curriculum, and their clerkship ratings. The FEHM scores included the MC score and the CS score.

Summary of Results: Preliminary results, with the 2011 cohort only and without the clerkship ratings, showed that IA MC and CB exams presented significant and moderate correlations between each other (mean 0.44; range 0.18 to 0.75), and very weak with CS (0.16; 0.00 to 0.30). The CS and MC parts of the FEHM showed a significant correlation (0.48). The MC part of the FEHM showed moderate correlation with IA MC or CB (0.43; 0.31 to 0.57) and weak with IA CS (0.23; 0.17 to 0.28). The CS part of the FEHM showed moderate correlation

with IA CS (0.41; 0.38 to 0.48) and weak with IA MC or CB (0.35; 0.21 to 0.45).

Conclusions: The results for the three cohorts will be presented and discussed, as well as the potential importance of other factors (gender, students who repeat a year, predictiveness of clerkship ratings).

OTT-OB-2-3
Assessing Diversity Teaching in Medicine – Leave it to the Students!

Pete Leftwick (*University of Liverpool, Community Studies Unit, Liverpool, UK)*, Sian Alexander-White (*University of Liverpool, Community Studies Unit, Liverpool*)

Background: In 'Tomorrows Doctors' the General Medical Council provided guidance to medical schools in the UK for all students to receive training in diversity to enhance their ability to care for and communicate with their patients. In Liverpool, we have delivered diversity training to all 3rd year students for the past 2 years.

Summary of Work: The lecture and small group tutorial were well received by students. However, there were significant problems with student engagement with the on-line resources and discussion boards through which their overall participation with the course was assessed. The subjective nature of the assessment led to considerable conflict between staff and students and the workload for tutors, supervising over 350 students for the duration of the course, was unsustainable. For the 2013/14 academic year an alternative assessment method was introduced, with students having to produce a 2000 word reflective essay on an area of diversity relevant to medicine for assessment (formative feedback and summative grading) by 2 of their peers. Students were given guidance in both aspects of the peer review process.

Conclusions: The course is currently ongoing and the task of producing a discursive reflective essay is clearly challenging the students. Our presentation will discuss these challenges, student engagement and the overall success (or otherwise!) of the change in assessment format. We will focus on the student response to the peer assessment process, including an analysis of their quantitative grading and a thematic analysis of their formative feedback to their colleagues.

OTT-OB-2-4
Crossing boundaries: The potential for innovative medical education models to challenge traditional assessment practices

Susan Van Schalkwyk (*Stellenbosch University, Centre for Health Professions Education, Tygerberg, South Africa)*, Juanita Bezuidenhout (*Stellenbosch University, Centre for Health Professions Education, Tygerberg*), Hoffie Conradie (*Stellenbosch University, Family Medicine, Tygerberg*), Norma Kok (Stellenbosch University, *SURMEPI, Tygerberg*), Ben Van Heerden (*Stellenbosch University, Centre for Health Professions Education,*

Tygerberg), Marietjie De Villiers (*Stellenbosch University, Dean's Division, Tygerberg*)

Background: The emergence of extended rural clinical placements for medical students creates opportunities to cross curriculum boundaries through the adoption of innovative educational models. Studies attest to largely positive student experiences of these programmes, although concerns regarding assessment emerge as traditional approaches are often not transformed in keeping with the new teaching models.

Summary of Work: A five-year cohort study to investigate the implementation of a year-long rural placement for final year medical students is underway. The exit level results of three successive cohorts at the rural clinical school (RCS) have been analysed and compared with students at the academic hospital. In addition, in-depth interviews have been conducted each year with RCS students to determine their experience of the rural immersion.

Summary of Results: RCS students complete the same exit level assessment as their colleagues at the academic hospital. In-course assessment differs in the form of a patient portfolio which is used both formatively and summatively. While they describe confidence in their clinical skills, the students articulate concerns about their academic preparedness for final assessment. The analysis of exit results, however, indicates that RCS students are not disadvantaged relative to students at the academic hospital.

Conclusions: Assessment remains pivotal to the student experience. The implementation of an educational innovation requires careful consideration of the implications that such innovation holds for assessment. While it creates opportunities to embrace new approaches, these should be nested within a renewed and aligned assessment plan.

Take-home Messages: Innovation in education offers spaces to challenge traditional assessment practices.

OTT-OB-2-5
Cumulative Assessment Programme: A new formative assessment for undergraduates at Dundee Medical School

Vanessa J. Kay (*Dundee Medical School, Obstetrics and Gynaecology, Dundee, UK)*

Background: A new formative assessment (CAP) was introduced in 2013, following revision of the undergraduate curriculum, negative comments in the National Student Survey and GMC inspection and recognised limitation of the previously run progress test.

Summary of Work: A question bank was generated, consisting of EMIs and MCQs, with feedback. Questions were validated, blue printed against the curriculum taught and inputted into an assessment management system, with an exam of 100 questions for each undergraduate year. Students sat the 90 minute exam in a designated week, 2 months before summative assessments on any computer with internet access and received immediate feedback with their results, followed by a ranked report. Student feedback was

requested immediately after the exam and after their summative assessment. Exam performance was analysed, with mean Cronbach's alpha of 0.75 (range 0.70 to 0.80) and mean average discrimination of 0.20 (range 0.18-0.21), with improved performance in 5th year.

Summary of Results: The following positive themes emerged: easy to use exam format, convenience of sitting exam, valued immediate feedback, useful revision aid. Negative issues included: 7% had technical problems, only 40% of poorly performing students attended their mentor, concerns regarding quality of some questions.

Conclusions: The CAP was show to have acceptable reliability and discrimination for a new formative assessment, with many positive features identified. Improvements in question bank writing, quality assurance and assessment management system are planned. .

Take-home Messages: With changes in curriculum and developments in IT, it is important to review formative assessment. The CAP test appears to offer many advantages to the previously used progress test.

OTT-OB-3: WRITTEN ASSESSMENT 1

OTT-OB-3-1
Written assessments using key features cases to assess clinical decision-making, CanMEDS roles and competence

Susan Glover Takahashi (*University of Toronto, Postgraduate Medical Education, Toronto, Canada)*, Jodi Herold (*University of Toronto, Postgraduate Medical Education, Toronto*), Tracy Dignum (*College of Physical Therapists of British Columbia, Vancouver*), Mary Clark (*College of Occupational Therapists of British Columbia, Vancouver*), Chris Corbett (*CSCW Systems Inc, Victoria*), Marla Nayer (*University of Toronto, Postgraduate Medicine, Toronto*)

Background: This study describes the use of key features (KF) cases to assess the clinical knowledge and decision making of postgraduate residents, physical therapists, and occupational therapists. Key features assess the critical elements using patient-related cases and a variety of question/answer formats. For residents, the KF cases are used to assess the mastery of the CanMEDS intrinsic competencies for residents completing mandatory online modules. For the occupational therapists and the physical therapists, their respective regulatory bodies have developed an online KF assessment to assess the continued competence of their registrants.

Summary of Work: This study provides details on the development and implementation online assessment using KF cases to assess the clinical knowledge and reasoning reasoning of occupational therapists, physical therapists and residents.
Using both qualitative and quantitative methods, similarities and differences in the approach to the

development, deployment, program evaluation across the three groups were complete. Document analysis and focused interviews illustrate the developmental processes, lessons learned and tips and strategies for successful use of key features. Three years of outcome data for the online tests is available for the resident group.

Summary of Results: For the resident group, the results show that the flexibility of design and efficiency in case development was valued by faculty and that key features function well for assessing the intrinsic CanMEDS roles. For the occupational therapy and physical therapy regulatory bodies, the key features approach of getting at 'real practice' has resulted in face and content validity with registrants in the development of their continued competence programs.

Conclusions: Online assessments using key features cases are an effective approach to assess clinical knowledge and decision-making of intrinsic CanMEDS competencies and competence in a variety of health professions.

Take-home Messages: Key features cases are effective to assess competence in a variety of health professions.

OTT-OB-3-2
Short answer questions based on Key Features have higher discrimination indices on a certification examination in family medicine

Carlos Brailovsky (*College of Family Physicians of Canada, Academic Family Medicine, Toronto, Canada)*, Tim Allen (*College of Family Physicians of Canada, Academic Family Medicine, Toronto*), Kathy Lawrence (*University of Saskatchewan, Family Medicine, Regina*), Tom Crichton (Northern Ontario School of Medicine, *Family Medicine, Sudbury*), Tom Laughlin (*Dalhousie University, Family Medicine, Moncton*), Theresa Van der Goes (*University of British Columbia, Family Medicine, Vancouver*)

Background: The concept of key features was introduced to improve the validity and effectiveness of assessment. The Evaluation Objectives of the College of Family Physicians of Canada (CFPC) include Priority Topics and their Key Features. For the last four years, new question development for the short-answer written component of the CFPC certification examination is based on Key Features. Any current examination contains new and old questions, so the questions on a topic on any one examination show varying degrees of concordance with Key Features for that topic.

Summary of Work: For three recent certification examinations, each question was coded for its degree of concordance to the key features for its topic (0 = no concordance, 1= minimal concordance, 2= moderate concordance, 3= good concordance). The discrimination indices (item/item total correlations) were calculated for each question on each examination, then were compared to the degree of Key Feature concordance.

Summary of Results: There is a consistent and highly significant relationship between the discrimination index

and the degree of Key Feature concordance for each question. e.g. Concordance question:key feature Average Item/item total correlation

None (0)	0.118
Minimal (1)	0.142
Moderate (2)	0.148
Good (3)	0.162
	$F = 5.1815$ (prob 0.0243 *)

Conclusions: For the CFPC certification examination, written short answer questions based on the key features for their topic discriminate better than do questions that are not based on the key features.

Take-home Messages: Key features provide guidance to question writers and improve face validity. They also improve the statistical performance of the questions, making assessment more effective and efficient.

OTT-OB-3-3
Validity evidence for a key features examination to assess clinical decision making in the internal medicine clerkship

Valerie J Lang (*University of Rochester School of Medicine and Dentistry, Medicine, Rochester, USA)*, Kirk Bronander (*University of Nevada, Medicine, Reno*), Heather Harrell (*University of Florida, Medicine, Gainesville*), Regina Kovach (Southern Illinois University, *Medicine, Springfield*), Sandra Monteiro (*McMaster University, , Hamilton*), Georges Bordage (*University of Illinois at Chicago, Medical Education, Chicago*)

Background: Clerkship directors must assess both students' knowledge and their ability to make clinical decisions. Key features examinations (KFEs) to assess medicine clerkship students and for licensing have been used in other countries, but not in the U.S. An on-line key features exam (KFE) was developed and blueprinted to the national curriculum and Simulated Internal Medicine Patient Learning Experience (SIMPLE) virtual patient cases, which are used at >100 medical schools. A vignette is followed by 1-5 questions to identify and resolve the problem at critical steps in the case.

Summary of Work: To gather validity evidence for the KFE, a study was conducted at nine U.S. medical schools with 786 students over one year. Each student took one of four forms of the 15-case exams, administered over 75 minutes. A repeated measures ANOVA was conducted to determine the relationship between KFE score, NBME examination score, and prior experience. Reliability was determined by Cronbach's alpha.

Summary of Results: 515 (65.5%) students participated. The mean score was 58% (SD 9%). There was a positive correlation between KFE score and NBME subject exam score ($r=0.36$; $p<0.001$) and number of clerkships completed ($r=0.20$, $p<0.001$). Cronbach's alpha for all four exams ranged from 0.44-0.53.

Conclusions: The KFE measured a construct different from, but overlapping with the NBME subject exam. Reliability was similar to other clerkship KFEs of similar length.

Take-home Messages: An on-line KFE may be incorporated into medicine clerkship assessments, but should not be used in isolation for high stakes decisions.

OTT-OB-3-4
Creating plausible distractors in an item generation framework

Hollis Lai (*University of Alberta, Undergraduate Medical Education, Edmonton, Canada)*, Mark Gierl (*University of Alberta, Educational Psychology, Edmonton*)

Background: Medical education requires numerous test items to facilitate student learning and item writing require tremendous efforts from medical educators. Automatic item generation (AIG) integrates the cognitive modelling of knowledge and skills from content experts with technology to systematically generate large numbers of test items. A review of generated multiple-choice items have found although the quality of generated item stem was comparable to written items, quality of distractors were not. The purpose of this presentation is to propose a new distractor modeling approach to improve the quality of generated items in medical education.
Summary of Work: Distractors generated under two existing approaches (random selection from a pool of distractors and distractor generation through a set of common rationale) were problematic, as distractors were not adaptively generated for each item. By modelling the features and related diagnoses of scenarios, we can adapt the generation of distractors by either matching presenting features on each item with scenarios, or selecting distractors based on related diagnoses. Thru this approach of distractor modelling, more plausible distractors can be generated in a systematic manner.
Summary of Results: A cognitive model in the content area of dyspnea was created from two content experts to trial the distractor modelling process. Distractors were generated from three approaches, samples of items from each approach are compared.
Conclusions: The use of distractor modelling allows the collection of information to be used for selectively generating an appropriate set of distractors. By adapt distractors toward the unique features presented in each item, resulting generated options should be more plausible compared with existing methods.
Take-home Messages: Previous approach to AIG was able to model the generation of item stems in a comparable manner. The addition of distractor modelling should provide a new source of information to refine the generation plausible option sets.

OTT-OB-3-5
What matters: item-writing flaws or test content area? – A case study in clinical anatomy

Milton Severo (*University of Porto Medical School, Department of Clinical Epidemiology, Predictive Medicine and Public Health, Porto, Portugal)*, João Pais (*University of Porto Medical School, Center for Medical Education, Porto*), Bruno Guimarães (*University of Porto Medical School, Porto, Portugal, Center for Medical Education, Porto*), Ana Povo (University of Porto Medical School, *Institute of Anatomy, Porto*), Artur Silva (*University of Porto Medical School, Institute of Anatomy, Porto*), Maria Amélia Ferreira (*University of Porto Medical School, Center for Medical Education, Porto*)

Background: Several studies had showed that item-writing flaws negatively affect the psychometric quality of the items, and also, the problematic items were concentrated in clinical anatomy particular content areas.
Summary of Work: 800 multiple-choice test items from 8 examinations were classified as either standard or flawed according to item-written guidelines, and in 8 clinical anatomy content areas (Abdomen, Pelvis and Perineum, Upper Limb, Lower Limb, Neck, Thorax, Head and Imagiology). Four judges (2 teacher/2 students), blinded to all item performance data, independently classified each item.
Summary of Results: Overall, the standard items are 4.1% easier than the flawed items (p=0.004) and the mean biserial correlation is 3.0% higher for the standard vs. the flawed items (p=0.014). The mean difficulty index (p=0.002) ranged from 52% (Imagiology) to 66% (Pelvis and Perineum) and the mean biserial coefficient (p=0.002) ranged from 0.30 (Thorax) to 0.39 (Upper Limb). After adjusting for the examination, item-writing flaws and content areas, the content area explained approximately 2.1% and 2.2% of the total variance for difficulty and discrimination index, while item-writing flaws explained 1.8% and 0.7%, respectively.
Conclusions: The content area showed a stronger relation with both difficulty and discrimination index compared with item-writing flaws. Both factors explained small part of the items quality.
Take-home Messages: The focus when construct good items should be in content and afterwards in item-writing flaws.

OTT-OB-4: INTERNATIONAL DIMENSIONS 1

OTT-OB-4-1
Reforming the high stakes assessment of international medical graduates in Australia

David Prideaux (*Expert Advisory Panel on Assessment, Australian Medical Council/Flinders University, Adelaide, Australia)*, Geoff McColl (*Expert Advisory Panel on Assessment, Australian Medical Council/Melbourne University, Melbourne*), Tim Wilkinson (*Expert Advisory Panel on Assessment, Australian Medical Council/Otago University, Christchurch*), Liz Farmer (Expert Advisory Panel on Assessment*, Australian Medical Council/University of Wollongong, Wollongong*), Gordon Page (*Expert Advisory Panel on Assessment, Australian Medical Council/University of British Columbia, Vancouver*)

Background: The Australian Medical Council (AMC) is responsible for the assessment of more than 2000 international medical graduates for registration in

Australia each year. The majority of candidates undertake MCQ and Clinical Examinations with a smaller number taking work-based assessment as an alternative to the Clinical Examination. The former process has been marked by high failure rates and long waiting lists. It has been difficult to provide sufficient Clinical Examination places across Australia.

Summary of Work: Work has been undertaken to increase the capacity of the assessment system overall and to improve screening for competence prior to the Clinical Examination. The MCQ examination was moved to computer adaptive testing (CAT) to reduce testing time while maintaining reliability. A screening step between the two examinations has been proposed. It involves an innovative, interactive case-based format designed to test clinical reasoning.

Summary of Results: Cases for the new format have been developed by writing teams of clinicians and test developers. Advice has been received from international authorities and experts. Additionally the Council received funding for the construction of a National Test Centre. It was opened in July 2013 and will greatly expand capacity to offer Clinical Examination places.

Conclusions: The move to a CAT format for the MCQ, the addition of the screening step using the new methods and the increased capacity for offering Clinical Examination places will all result in a more efficient and ultimately a fairer examination process.

Take-home Messages: Reforming high stakes examinations is a multi-step process involving changing formats, creating new formats and increasing capacity.

OTT-OB-4-2
The Impact of a Transition Program for International Medical Graduates on Communication and Clinical Skills

Marie Rocchi (*Centre for the Evaluation of Health Professionals Educated Abroad, Toronto, Ontario, Canada)*, Eric Wong (*Centre for the Evaluation of Health Professionals Educated Abroad, Toronto, Ontario*), Shobhan Vachhrajani (*Centre for the Evaluation of Health Professionals Educated Abroad, Toronto, Ontario*), Arthur Rothman (Centre for the Evaluation of Health Professionals Educated Abroad*, Toronto, Ontario*), Murray Urowitz (*Centre for the Evaluation of Health Professionals Educated Abroad, Toronto, Ontario*), Fok-Han Leung (*Centre for the Evaluation of Health Professionals Educated Abroad, Toronto, Ontario*)

Background: The Centre for the Evaluation of Health Professionals Educated Abroad (CEHPEA) delivers brief mandatory transition programs (4 weeks for Family Medicine, 3 weeks for specialties) for International Medical Graduates (IMGs) who have matched to residency training programs in Ontario. Canadians Educated Abroad (CEAs) now constitute the majority of participants in both programs. Concerns surrounding program validity are raised by CEAs despite proven effectiveness in program delivery. For the purpose of participant evaluation, baseline and exit OSCE stations were conducted. Program evaluations were carried out at the programs' conclusions.

Summary of Work: A four-station OSCE was administered on the first and last days of each program. New OSCE stations were developed in 2013 to better reflect the curricula. Trained physician examiners rated participants relative to Canadian medical graduates. Differences in individual participant performance in the OSCEs and cohort performance between IMGs and CEAs were compared at baseline and exit OSCEs using t-tests. Program evaluation data was reviewed.

Summary of Results: "Statistically significant increases were achieved in 14/15 OSCE evaluation criteria for participants in both programs.

There were small differences between CEAs and IMGs at the program outset in communication skills for the specialty group. Although both CEAs and IMGs showed improvement at the end of the programs, CEAs performed slightly better in most measures of communication skills and overall clinical skills. Overall satisfaction increased in both programs."

Conclusions: Two brief transition programs had positive impacts on communication and clinical skills for IMGs, including CEAs. Improved alignment of OSCE stations with the curricula provided motivation for participation, resulting in increased learner satisfaction.

OTT-OB-4-3
Development and implementation of a Canadian pre-bridging competency based assessment for internationally educated physical therapists

Sharon Switzer-McIntyre (*University of Toronto, Toronto, Canada)*, Barbara Norton (*University of Alberta, Edmonton*), Bernadette Martin (*University of Alberta, Edmonton*), Dianne Millette (Physiotherapy Alberta*, Edmonton*)

Background: A standardized, multi-format competency based assessment was collaboratively developed, implemented and evaluated by the Departments of Physical Therapy at the University of Alberta and the University of Toronto. The assessment evaluated whether bridging program applicants demonstrated the required level of foundational physiotherapy knowledge, clinical skills, communication abilities and practice readiness to be successful in a bridging program.

Summary of Work: Following a literature review and benchmarking of current competency based assessments for internationally educated professionals, an assessment model and blueprint were established for the Alberta and Ontario Internationally Educated Physical Therapists (IEPT) Bridging Programs. A panel of educational and physiotherapy experts developed 4 assessment tools including a basic knowledge exam (100 MCQ), an OSCE (6 stations), a Multiple Mini Interview (6 stations) and a written exercise (4 short essay questions).

Summary of Results: A total of 60 IEPT applicants (27 in Alberta, 33 in Ontario) from 17 different source countries completed the one-day assessment. The format and timing were reasonable. The results for each assessment component were analyzed and overall rankings were used for admission decisions.

Conclusions: This multi-faceted competency based assessment provided a comprehensive picture of each applicant's bridging potential. Project collaboration between the bridging programs at 2 universities resulted in high quality assessment tools and portability of applications.

Take-home Messages: Multiple modes of assessment are required to effectively evaluate IEPT bridging suitability. A shared standardized assessment model is beneficial for applicants and programs.

OTT-OB-4-5
National Assessment Collaboration Practice Ready Assessment – Predictive Validity Study

André De Champlain (*Medical Council of Canada, Research and Development, Ottawa, Canada)*, Fang Tian (*Medical Council of Canada, Research and Development, Ottawa*), Sirius Qin (*Medical Council of Canada, Research and Development, Ottawa*), Marguerite Roy (Medical Council of Canada*, Research and Development, Ottawa*)

Background: The National Assessment Collaboration (NAC) is designing a pan-Canadian Practice Ready Assessment (PRA) process to evaluate international medical graduates' practice readiness. It is critical to select candidates who will not only have the highest likelihood of successfully completing the PRA program but will also attain specialty certification. In an effort to identify common selection tools, a study was conducted to determine how well PRA candidates' performance on the Medical Council of Canada Evaluating Examination (MCCEE), Qualifying Examination I (MCCQE Part I) and four demographic variables could predict both the score and pass/fail status on the College of Family Physician's (CFPC) certification examination.

Summary of Work: Data from 132 PRA candidates were analyzed using logistic regression to predict pass/fail status on the CFPC Examination and its two components, i.e., the Simulated Office Orals (SOOS) and Short Answer Management Problems (SAMPS), as a function of the MCCEE, MCCQE Part I and demographic variables. Linear regression was run to predict scores on the SOOS and SAMPS from the same variables.

Summary of Results: Results indicate that MCCQE Part I scores, gender and age were significant predictors of both SAMPS pass/fail status (p<.05) and SAMPS scores (p<.01).

Conclusions: The MCCQE Part I is a good predictor of success on the SAMPS but not the SOOS. This study provides validity evidence for using the MCCQE Part I as a selection tool for PRA. Additional studies to identify tools that can predict success at SOOS are underway.

Take-home Messages: A set of common tools can be used to standardize NAC PRA programs across the country.

OTT-OB-5: CLINICAL ASSESSMENT 1

OTT-OB-5-1

A clarification study of internal scales clinicians use to assess undergraduate medical students

Catherine Hyde (*Keele University, Keele, UK)*, Janet Lefroy (*Keele University, School of Medicine, Keele*), Simon Gay (*Keele University, School of Medicine, Keele*), Sarah Yardley (Keele University*, School of Medicine, Keele*), Robert McKinley (*Keele University, School of Medicine, Keele*)

Background: Clinicians hold internal constructs with which they make intuitive judgements about learners within clinical assessments. Grading scales used in summative assessments which align with these internal constructs enable more reliable assessment. We aim to understand the constructs clinicians use to make judgements of undergraduate medical students' consultation skills and whether we can develop construct aligned scales for their assessment.

Summary of Work: JL and CH conducted semi-structured face-to-face interviews with teaching clinicians with at least 2 years' experience of assessing undergraduates at one English medical school. The clinicians were asked to draw scales for assessment domains and populate them with words and phrases which described the range of student performance. Interviews were audio-recorded and the audio-recording and scales from each interview were analysed by the interviewer and a researcher using framework analysis informed by realist theory. Emerging scales for each construct were reviewed in round-table meetings and are being fed back to subsequent participants. The finalised scales will be reviewed in a focus group with clinicians who participated.

Summary of Results: Preliminary results suggest that clinicians hold internal scales which they can use to describe meaningful scales. The mechanisms by which assessors make judgements are influenced by the context.

Conclusions: Assessment scales aligned to the internal scales clinicians use to assess undergraduate medical students may be feasible.

Take-home Messages: While further work is needed to investigate whether our scales are generalisable to assessment in other schools, we consider them to be a first step in the development of construct aligned scales.

OTT-OB-5-2
Clinical assessment of health professional students – what are the systemic and individual factors that influence assessment in healthcare settings?

Verona du Toit (*University of Western Sydney, School of Medicine, Sydney, Australia)*, Wendy Hu (*University of Western Sydney, School of Medicine, Sydney*), Andrea Bialocerkowski (*Griffith University, School of Rehabilitation Sciences, Gold Coast*), Peter Zelas (University of Western Sydney*, School of Medicine, Sydney*), Lambert Schuwirth (*Flinders University, School of Medicine, Adelaide*), Yenna Salamonson (*University of*

Western Sydney, School of Nursing and Midwifery, Sydney)

Background: Clinical supervision and assessment combines complex clinical skills with the ability to make judgements within highly contextualized environments. Empirical evidence about these environments and how clinical assessments are actually implemented in the workplace is limited.

Summary of Work: Datasets of 25 actual clinical assessments, comprising video and audio recordings, field notes and in-depth interviews were collected in hospital and community based healthcare settings. Participants were expert supervisors and students in the health professions from the western region of Sydney, Australia. An expert panel from five health professions reviewed the data using Delphi techniques to examine the activities and factors influencing the conduct of clinical assessments of students from medicine, nursing, occupational therapy, physiotherapy and podiatry undergraduate programs.

Summary of Results: Common factors that influenced assessment in all professions studied were:
1. Organisational: including documentation, communication, physical facilities, workplace culture, health service, and university governance.
2. Individual: including personal orientation to assessment, knowledge of curriculum, approaches to judgement and feedback, identity and standing as a competent clinician.

Conclusions: Expert assessors were able to develop relationships with students, co-workers and colleagues to effectively navigate the organisational factors influencing the conduct of assessments. Their ability to do this was in turn influenced by individual factors. Unravelling such commonalities between the health professions could improve sharing of training resources for supervisors, and further support interprofessional workplace based assessments.

Take-home Messages: To be effective clinical supervisors need to be role models, show interest in maintaining current evidence based teaching and provide timely and constructive feedback.

OTT-OB-5-3
Assessing medical students' sensitive examination skills: What's working - results of a Clinical Teaching Associate program

Neil Sefton (*University of Tasmania, School of Medicine, Hobart, Australia)*, Prof Richard Tuner (*University of Tasmania, School of Medicine, Hobart*), Dr Mona Loofs-Samorzewski (*University of Tasmania, School of Medicine, Hobart*)

Background: Assessing medical students' sensitive examination technique by trained lay people as Clinical Teaching Associates (CTAs) - men and women from the community specifically trained to teach students appropriate sensitive examination - is increasing being adopted in medical education. The results of our CTA program evaluation has helped guide our program development.

Summary of Work: The University of Tasmania CTA program has evolved to incorporate a program focused at male sensitive examination to complement the successful woman's program. The current results demonstrate improved student confidence and competencies.

Summary of Results: In 2012/13, 128 third year medical students progressed through the program. Evaluation revealed 97% believed learning sensitive examination was very important, and is best delivered through a CTA program. Students also commented on learning examinations as highly valuable in improving their understanding of health issues, and generously parsed the benefits to learning SE with CTAs.

Conclusions: Clinical skills are fundamental to acquire in undergraduate education with many medical schools developing programs specifically targeted at teaching sensitive examinations. Although our finding support the use of CTAs in teach and assess sensitive examination, it is also clear that CTAs need training to support the learning and assessment process appropriately.

Take-home Messages: Innovative programs involving CTAs support the student learning and achievement of difficult competencies. It is important to train CTAs in assessment principles. Students value the assessment feedback from the CTAs.

OTT-OB-5-4
Structured Long Interview and Clinical Examination (SLICE) as a teaching tool in clinical clerkship by providing instant feedback

Rehan Ahmed Khan (*Islamic International Medical College, Riphah International University, Rawalpindi, Pakistan)*, Madiha Sajjad (*Islamic International Medical College, Pathology, Rawalpindi*), Masood Anwar (*Islamic International Medical College, Pathology, Rawalpindi*)

Background: Long case assessment has always been a source of debate in medical education. In many parts of the world, it still forms the mandatory component of the assessment. However students get very little individual opportunity to get feedback on their holistic approach towards management of a patient. To provide students with this opportunity, feedback was provided using the SLICE tool.

Summary of Work: SLICE was developed using the Delphi technique and has been found reliable and valid to assess long case. In our medical school, which has an integrated modular spiral curriculum, students have formative assessment at the end of ward rotation. These students were assessed using SLICE and were provide instant feedback. A questionnaire using Likert scale was then filled by the students to assess the benefits of this feedback.

Summary of Results: 88 students of final year MBBS during their clerkship rotation were provided with the feedback using the SLICE tool. Students felt marked improvement in their approach towards holistic concept of managing a patient, presenting comprehensive history, comprehending weakness in presentation and communication skills and clinical reasoning at higher cognitive level.

Conclusions: Though SLICE was mainly designed to assess undergraduate medical students in clerkship to assess Long case in summative assessment only, it can also be used successfully in providing instant feedback to the students to improve their learning of holistic approach towards managing a patient.

Take-home Messages: Assessment tools can be successfully used to enhance learning of the students by providing them with instant feedback.

OTT-OB-5-5
The perceptions of final year students on Observed Long Case (OLC)

Yokeleong Cheong (*Royal College of Medicine Perak, Univ. Kuala Lumpur, General Surgery, Ipoh, Malaysia)*, Hungloong Yip (*Royal College of Medicine Perak, Univ. Kuala Lumpur, Primary Care, Ipoh*), Osman Ali (*Royal College of Medicine Perak, Univ. Kuala Lumpur, Public Health, Ipoh*)

Background: OLC is used in UniKL Royal College of Medicine as part of the Summative Assessment in the final year. A student has to attempt and pass one OLC during each of these postings (Internal Medicine, General Surgery, Paediatrics and Obstetric & Gynaecology). Cases are selected randomly from the ward before each session. A content specialist will observe and award marks according to a broad checklist. During the discussions, students are given feedbacks on their performance.

Summary of Work: A questionnaire consisting of 14 questions was distributed to the students. This presentation discusses the students' responses to one of the questions. There were 111 students and 88 students returned the questionnaire. There are two parts to this question. The first part asks whether OLC is a useful for learning – the response is a yes/no. The second part asks the students to narrate their reasons.

Summary of Results: 96.6 % of the responders think that OLC is useful for learning, 2.3% disagree and 1.1% uncommitted. Among the positive reasons given are usefulness of immediate feedback, opportunities for remediation, promotes continuous learning, relevant to their career and personal development, motivational, fair and authentic assessment. Negative reasons given include: good for assessment, not good for learning, stressful, not standardised.

Conclusions: The perception of majority of students is very favourable.

Take-home Messages: OLC is a useful assessment and educational tool for students and it is feasible.

OTT-OB-6: OSCE 2

OTT-OB-6-1
Why do the sexes perform differently in a high stakes postgraduate examination?

Lindsey Pope (*University of Glasgow, General Practice and Primary Care, Glasgow, UK)*, Anne Hawkridge

(*Stonehill Medical Centre, Bolton*), Robin Simpson (*Defence Medical Services, The Defence Deanery, Lichfield*), Richard Wakeford (University of Cambridge*, Hughes Hall, Cambridge*)

Background: The MRCGP is an integrated assessment comprising three components that all UK GP trainees must pass to be awarded a licence to practice as a GP. Sex differences in performance in the CSA (Clinical Skills Assessment: an OSCE) exam have been reported and we consider this in detail, looking at performance by sex overall as well as by assessed domain (data-gathering, interpersonal and management skills) and curriculum area (there are 26).

Summary of Work: Examination data were analysed enabling comparative performance by candidate sex. 92 989 encounters had encounters tagged for candidate sex and curriculum area for the case. For each encounter, total score and score for each of the three assessment domains was recorded.

Summary of Results: Female candidates outperformed their male counterparts globally and in all assessment domains and curriculum areas. The biggest difference in mean case score was seen in the areas of 'Women's health' and 'Sexual Health'.

Conclusions: This study provides insight into differences in performance by sex in this OSCE. We discuss factors that could explain these including known baseline performance differences between sexes, differences in patient mix seen by different trainee groups, changing posts included in GP training rotations and sex differences in consulting styles.

Take-home Messages: GP trainees, trainers and programme directors need to recognise that male trainees are more likely to have higher levels of learning needs in all areas of the GP curriculum but particularly in women's health and sexual health. As the CSA is an integrated assessment, learning needs involving both knowledge and skills deficits may need to be addressed more proactively during training.

OTT-OB-6-2
Enhancing the catalytic educational effect of summative assessments: use of skill-domain scoring and feedback for an M4 objective structured clinical examination

Joel Purkiss (*University of Michigan Medical School, Ann Arbor, USA)*, Carrie Bernat (*University of Michigan Medical School, Ann Arbor*), Sally Santen (*University of Michigan Medical School, Ann Arbor*), Brent Stansfield (University of Michigan Medical School*, Ann Arbor*), Caren Stalburg (*University of Michigan Medical School, Ann Arbor*), Mary Ellen Hernandez (*University of Michigan Medical School, Ann Arbor*)

Background: Summative OSCEs often inadequately produce "catalytic educational effects", whereby "assessment provides results and feedback in a fashion that creates, enhances, and supports education [and drives] future learning forward"(Norcini et.al. 2011). Here we describe our recent utilization of skill-domain

scoring/feedback to enhance catalytic educational effects of our M4 OSCE.

Summary of Work: We provide OSCE scoring/feedback at the individual-station level, and the overall-exam level. For the 2012-13 cohort(n=171), we calculated new skill-domains to provide additional scoring/feedback: Data-Gathering, tabulated from History/Physical Examination checklists across seven stations; Communication, from six station checklists; and Clinical-Reasoning, from five post-encounter notes and assessments, an oral case presentation, and written evidence-based medicine and critical values exams.

Summary of Results: Across domains, Mean(SD) were similar. Data-Gathering:83%(5%); Communication:82%(8%); Clinical-Reasoning:83%(5%)). All inter-domain correlations were significant(p<0.01). Communication yielded weak inter-domain correlations with Clinical-Reasoning(r=0.25) and Data-Gathering (r=0.22); Data-Gathering and Clinical-Reasoning were moderately correlated(r=0.45); OSCE-Overall had the strongest correlation with Clinical-Reasoning(r=0.87). Students ≥2SD below average on Communication scored variously on the OSCE-Overall: some poorly, others well. In contrast, students with low Clinical-Reasoning scores fell in the bottom quartile on OSCE-Overall.

Conclusions: Adding domain-specific scoring to our OSCE identified needs for communication and clinical-reasoning curricular enhancements. It also improved remediation for students with domain-related performance patterns. Some who did well on the OSCE overall learned Communication improvements were still warranted; patterns for students with poor Clinical-Reasoning scores revealed some needing data-gathering improvements, while others could data-gather but needed to improve synthesis.

Take-home Messages: Domain-specific scoring/feedback expanded catalytic educational effects of our OSCE, by identifying areas for curricular enhancement, and improving feedback and remediation processes for students.

OTT-OB-6-3
Benchmarking the attainment of clinical competence

Bunmi Malau-Aduli (*James Cook University, Medical Education Unit, School of Medicine and Dentistry, Townsville, Australia)*, Peta-Ann Teague (*James Cook University, School of Medicine and Dentistry, Townsville, Queensland*), Richard Turner (*University of Tasmania, School of Medicine, Hobart, Tasmania*), Karen D'Souza (Deakin University*, School of Medicine, Melbourne*), David Garne (*University of Wollongong, Graduate School of Medicine, Wollongong*), Ian Wilson (*University of Wollongong, Graduate School of Medicine, Wollongong*)

Background: The sharing of co-developed Objective Structured Clinical Examinations (OSCEs) by medical schools for the assessment of clinical competence could serve as a responsive and flexible means of entrenching assurance of high quality assessment processes. It has the added benefit of benchmarking and evaluating comparable competence standards.

Summary of Work: This study was undertaken to explore the feasibility of utilising shared OSCEs to benchmark clinical performance of students in four geographically dispersed Australian medical schools. Four shared OSCE stations were co-developed by the participating medical schools and embedded in the end of year examinations for the assessment of clinical performance in the early clinical phase of the course. Returned checklist, global and total scores from 1670 student results were then analysed using SAS analytical package to compare mean scores and clinical competence levels.

Summary of Results: Data analysis revealed similar clinical competence patterns in the performance of the medical students, indicating comparable standards. The degree of difficulty for the shared OSCE stations was largely similar for participating schools, although mean total student scores varied between schools.

Conclusions: Benefits of benchmarking are available to collaborating medical schools through identification of common curriculum areas requiring specific focus and the sharing of assessment approaches. Similarly, relative underperformance by a school in a particular assessment item may indicate deficiency that can be remediated in order to achieve comparability with its peers.

Take-home Messages: Sharing of assessment materials can provide common defensible, reliable, valid, robust and standardised assessments which in turn, enhance transparency and accountability.

OTT-OB-6-4
The impact of OSCE case information length on examiner markup behavior

Marguerite Roy (*Medical Council of Canada, Ottawa, Canada)*, Caroline Marshall (*Medical Council of Canada, Ottawa*)

Background: Being an OSCE examiner is demanding. Examiners must observe candidate performance in real time, track targeted case information for scoring decisions, complete rating tools correctly, monitor SP performance, and be attuned to the assessment standard. The risk of cognitive overload is high. Cognitive load theory (Sweller, 1983) builds on the premise that working memory has limited capacity and posits that when tasks, materials, or situations impose excessive load on an already limited system, task performance may suffer (Taveres & Eva, 2013). Excessive task-irrelevant load can be effectively offset by redesigning materials or developing strategies to better support task-relevant demands (Chandler & Sweller, 1992; Moreno & Mayer, 1999).

The current study explored the relationship between the complexity of case information sheets provided to aid examiners in using rating scales to score and spontaneous strategies they adopted as they rated candidates.

Summary of Work: Case information sheets for a high stakes OSCE involving 232 examiners rating 548 candidate performances across multiple sites were collected. The length of case information per station was

the count of the distinct information units. All sheets were coded according to whether they contained examiner notes, drawings, checklists, a mix of these, or none.

Summary of Results: The length of case information was significantly and positively related to examiner's tendency to create checklists but was not related to creating notes, or drawings.

Conclusions: Lengthy case information descriptions impose a heavy cognitive load and some examiners spontaneously develop strategies to deal with it.

Take-home Messages: Future research should explore how to mitigate unnecessary cognitive demands of examining.

OTT-OB-6-5
Does rubric improve inter-raters' agreement in a national OSCE?

Gandes Retno Rahayu (*Faculty of Medicine, Universitas Gadjah Mada, Department of Medical Education, Yogyakarta, Indonesia)*, Yoyo Suhoyo (*Faculty of Medicine, Universitas Gadjah Mada, Department of Medical Education, Yogyakarta*), Rachmadya Nur Hidayah (*Faculty of Medicine, Universitas Gadjah Mada, Department of Medical Education, Yogyakarta*), Mohammad Adrian Hasdianda (Faculty of Medicine, Universitas Gadjah Mada*, Department of Medical Education, Yogyakarta*), Datu Respatika (*Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta*)

Background: Since February 2013, Indonesia has been implementing OSCE nationally as competence examination of medical students before they graduate. It consisted of 12 stations, 15 minutes each plus 2 rest stations. Considering the large variation that may emerge among examiners, instead of checklist or rating scale, rubric has been applied. This study identified inter-raters' reliability in each station during one of the 6 pilots done in 25 Indonesian medical schools.

Summary of Work: Among the 12 stations in each school, one chosen station was assessed by two examiners. In 12 schools, the two examiners consist of a specialist that has similar expertise to the assessed case and one other specialist with a different expertise. In the other 13 schools, the two examiners were a specialist that has similar expertise to the assessed case and a general physician.

Summary of Results: The inter-raters' reliability between the specialist that has similar expertise to the assessed case and a specialist that has different expertise to the assessed case in 12 stations were 0.92, 0.73, 0.66, 0.82, 0.91, 0.51, 0.95, 0.96, 0.92, 0.97, 0.83, 0.68. The inter-rater reliability between a specialist that has similar expertise to the assessed case and a general physician were 0.49, 0.62, 0.73, 0.72, 0.76, 0.63, 0.91, 0.42, 0.64, 0.88, 0.72, 0.79.

Conclusions: Inter-raters' agreement using rubric ranges from moderate to almost perfect agreement.

Take-home Messages: Rubric can be used as an alternative of checklist or rating scale in a large-scale OSCE using many examiners.

OTT-OB-6-6
Innovation in evaluation: Transdisciplinary OSCE during clerkship at the University of Sherbrooke

Nathalie Gagnon (*Université de Sherbrooke, études medicales pré-doctorales, Sherbrooke, Canada)*, Carolle Bernier (*Université de Sherbrooke, études medicales pré-doctorales, Sherbrooke*), Caroline Osborne (*Université de Sherbrooke, études medicales pré-doctorales, Sherbrooke*), Marick Beauchamp (Université de Sherbrooke*, études medicales pré-doctorales, Sherbrooke*), Sylvie Bourque (*Université de Sherbrooke, études medicales pré-doctorales, Sherbrooke*), Daniel Gladu (*Université de Sherbrooke, études medicales pré-doctorales, Sherbrooke*)

Background: In 2007, a clerkship reform in the Faculty of medicine and health sciences recommended a transdisciplinary approach, and the development of an array of terminal competencies. 99 relevant clinical situations, aligned with the objectives of the Medical Council of Canada, were selected as learning issues.

Summary of Work: In response to this reform, we developed an integrated evaluation system: an OSCE (objectively structured clinical examination) where selected clinical scenarios are used to develop different stations, following a transdisciplinary approach. These scenarios permit the evaluation of specific medical competencies as well as generalized fundamental skills. Operationally, the student is evaluated, not only with respect to the diagnostic process, but concurrently with respect to communication, organizational ability, professionalism and the quality of patient-doctor relationship. All of the standardized criteria are evaluated by a physician-observer, except for the rubric "patient -doctor relationship", which is evaluated by the standardized patient.

Summary of Results: Although the conceptualization of transdisciplinary stations are more challenging, these stations are more faithful of clinical reality. Indeed, a female patient presenting with abdominal pain could equally have a gynecologic, surgical or metabolic problem. This type of OSCE permits us to evaluate more comprehensively the clinical reasoning skills of our students, as well as other competency domains.

Conclusions: Our transdisciplinary OSCE during clerkship represents a realistic project in developing a tool for a more authentic evaluation of clinical competencies required for our clerks.

Take-home Messages: Our clerkship OSCE, conceptualized in a transdisciplinary fashion approximates closely clinical medical realities, and constitutes an innovative instrument to comprehensively evaluate clerks.

OTT-OB-7: SITUATIONAL JUDGEMENT TEST

OTT-OB-7-1
Validating an Integrity-focused Situational Judgement Test for Pre-Interview Selection into Medical School

Adrian Husbands (*University of Dundee, Division of Clinical & Population Sciences & Education, Dundee, UK)*, Jonathan Dowell (*University of Dundee, Division of Clinical & Population Sciences & Education, Dundee*), Fiona Patterson (*Work Psychology Group, Derby*)

Background: Despite the emerging success of the Multiple-Mini Interviews (MMIs) it is often not practical to interview all applicants to medical school. Situational Judgement Tests (SJTs) present applicants with work-related scenarios accompanied by plausible multiple-choice type responses and hold promise for assessing core personal attributes at the pre-interview level for all medical school applicants. This study aims to investigate the reliability and construct validity of an SJT focused on integrity in the UK context.

Summary of Work: The SJT consisted of fifty effectiveness-type items pertaining to ten integrity-related scenarios, each developed from interviews conducted with faculty. MMI candidates sat the SJTs after their interviews but the result played no part in selection. Relationships were examined between SJT score, pre-admissions measures and the HEXACO personality inventory, with the Honesty-Humility dimension having a clear conceptual link to integrity. Face validity was also determined.

Summary of Results: Two hundred of 553 (36.2%) interviewees participated in the study. Statistically significant correlations were observed between SJT score and the HEXACO personality dimensions of Honesty-Humility (.36), Extraversion (.26) and Conscientiousness (.35), MMI score (.24) and negatively with the academic score (-.23). There were no significant associations with UKCAT. Most applicants strongly agreed that the SJT appeared relevant and valid.

Conclusions: Results suggest the SJT is a valid measure of integrity and other desirable personality traits, with encouraging correlations with MMIs and high face validity. SJTs therefore hold promise as a psychometrically robust pre-interview screening tool.

Take-home Messages: An SJT for medical school admissions demonstrates construct validity and potential utility.

OTT-OB-7-2
Evaluation of a Situational Judgement Test (SJT) to assess the professional attributes of UK junior doctors

Siobhan Fitzpatrick (*Medical Schools Council, London, UK)*, Fiona Patterson (*Work Psychology Group & University of Cambridge*), Helena Murray (*Work Psychology Group*), Katie Petty-Saphon (Medical Schools Council), Paul O'Neil (*University of Manchester*)

Background: Having completed medical school training, students apply for a junior doctor post as part of the UK Foundation Training Programme. It was recommended that alongside an educational performance measure, an SJT was implemented to assess relevant professional attributes (eg integrity, teamwork) to replace the open ended competency-based application form.

Summary of Work: This project describes the design, and analysis of an operational SJT for selection of approximately 8,000 medical students in 2014. It follows on from analysis of the initial piloting and live implementation in 2013.

Summary of Results: Psychometric analysis of the 2013 cohort provided evidence that the SJT is a reliable measurement methodology in this context and overall is able to differentiate between candidates. Early evidence was found for criterion-related validity. Further analysis on the 2014 applicants will be reviewed to evaluate year on year validity and reliability.

Conclusions: The SJT is a reliable and valid methodology to facilitate the appointment of junior doctors to the Foundation Programme.

Take-home Messages: A robustly designed SJT will enhance the predictive validity of the selection process and improve standardisation nationally. SJTs draw on non-cognitive attributes that cannot easily be targeted through traditional exams and reflect the challenging interpersonal context that junior doctors work within.

OTT-OB-7-3
Medical and dental school admissions: Evidence from a Situational Judgement Test to assess non-academic attributes in UK

Stuart Martin (*Work Psychology Group, Derby, UK)*, Safiatu Lopes (*Work Psychology Group*), Fiona Patterson (*Work Psychology Group & University of Cambridge*), Máire Kerrin (Work Psychology Group*)*, Rachel Greatrix (*UKCAT Consortium*), Sandra Nicholson (*UKCAT Consortium*)

Background: The UK Clinical Aptitude Test (UKCAT) is used by a consortium of universities to help them make more informed choices from amongst the many highly qualified applicants who apply for the medical and dental degree programmes. Until 2013, the UKCAT consisted of tests of verbal, quantitative and abstract reasoning, as well as a decision making analysis test. In July 2013 a newly designed situational judgement test (SJT) was used live alongside the sub tests as a selection method
to evaluate important non-academic attributes deemed appropriate in medical and dental students as part of the UKCAT.

Summary of Work: The SJT targets three domains: integrity, perspective taking and team-involvement. In 2012, 18 test forms were piloted, on-line alongside the other UKCAT tests. In 2013, the test was used live (approx N=25,000) and this paper will present back psychometric analysis of the test performance.

Summary of Results: Results from the pilot SJT reported good levels of reliability (α=.75-.85). Initial evidence of criterion-related validity was established as applicant scores on the SJT correlated significantly with the other UKCAT tests. On average, the SJT correlated more strongly with the verbal and decision making analysis test than with the numerical and abstract tests. Similar analysis on the 2013 live data will be conducted and reported.

Conclusions: An SJT is a reliable and valid selection methodology for testing important non-academic attributes for entry to medical and dental school.

Take-home Messages: An SJT is a reliable and valid selection methodology for testing important non academic attributes for entry to medical and dental school.

OTT-OB-7-4
Admissions into Veterinary Medicine in the UK: Evaluation of a pilot to assess professional attributes using a situational judgement test

Máire Kerrin (*Work Psychology Grouo, Derby, UK)*, Liz Mossop (*University of Nottingham*), Helena Murray (*Work Psychology Group*), Michelle Dickson (*University of Nottingham*), Emma Rowett (*Work Psychology Group*)

Background: Admissions into Veterinary Medicine in the UK is highly competitive. For example, the University of Nottingham attracts approximately 1,900 applicants annually for just 100 places. It presents challenges for the delivery of a practical, robust and fair admissions process.

Summary of Work: A role analysis was conducted to identify the professional attributes required for the role of a veterinary student. This was used to inform the design of a Situational Judgement Test (SJT) targeting four key attributes: Empathy & Building Client Relationships, Professional Integrity & Trust, Team Work and Resilience. Scenarios were developed based on best practice (Patterson et al 2012). These were piloted with N=921 applicants alongside live admissions, and evaluated in terms of psychometric properties and candidate reactions.

Summary of Results: The pilot data (N=921) returned a normal distribution and suggests that the SJT is able to differentiate between applicants. Analysis of group differences on the test showed that there were no significant differences in SJT total scores between males and females. Item-level analysis showed that the majority of the SJT items worked satisfactorily. The reliability for the pilot SJT is $\alpha=.601$, which is encouraging for a first pilot process in a new context.

Conclusions: Further development and piloting would be required before an operational test could be implemented. Detailed review and analysis of the items will help to identify trends and patterns that can inform further development of the SJT.

Take-home Messages: The psychometric analysis provides early evidence to suggest that an SJT may be a reliable measurement methodology in this context.

OTT-OB-7-5
Exploring the use of situational judgement tests for selection into anaesthesia training

NA Smith (*Wollongong Hospital, Anaesthesia, Wollongong, NSW, Australia)*, David Ware (*Wollongong Hospital, Anaesthesia, Wollongong, NSW*)

Background: Traditional selection methods are based on academic ability, with applicant letter, curriculum vitaes, interviews and references. There is a growing acceptance that these methods are not robust. Situational judgement tests (SJTs) have been described to have good reliability and validity in improving selection techniques. Their use in anaesthesia selection is not widely described.

Summary of Work: We conducted a job analysis to identify the non-academic attributes that were considered most important in our department. Two years of trainee programme applicants completed 10 SJT questions based on these attributes as part of their application process. Various methods of creating, structuring, and scoring SJTs were trialled in the first year, with lessons learnt from this experience applied in the second year.

Summary of Results: The candidate rankings created by the SJT scores did not display any agreement with the traditional ranking system. This applied regardless of which exact scoring system was used. Each of the three trialled scoring systems gave a different candidate ranking. (Data from the second year are not yet fully analysed but will be by the end of the year).

Conclusions: Using SJTs gave a different candidate ranking compared to the traditional scoring system. We were unable to say which scoring method was best, and whether the use of SJTs led to a better scoring system overall. We raise several questions that need to considered further as to why the ranking results were so varied, including the robustness of the current traditional methods. The SJTs were easy to do although time-intensive to score.

Take-home Messages: This study raises more questions than it answers. SJTs are a potentially useful tool in anaesthesia selection, but the best methods of creation, structuring, and scoring remain to be resolved before widespread use can be considered.

OTT-OB-7-6
Prediction of video-based situational judgment tests for National Licensure scores

K.L. Dore (*McMaster University, Hamilton ON, Canada)*, S. Kreuger (*McMaster University, Hamilton ON*), H.I. Reiter (*McMaster University, Hamilton ON*)

Background: The endpoint of a good admissions decision is performance in clinical practice. National licensure scores act as surrogates for subsequent endpoints in practice. For national licensure, pre-interview screening tests predictive of future cognitive performance exist. This is not true for most measures of personal/professional characteristics. Does Computer-based Assessment for Sampling Personal characteristics (CASPer) (a video-based situational judgment test (SJT)) predict for national licensure scores?

Summary of Work: Applicants to McMaster University in 2006 (n= 110) and 2007 (n=167) participated in CASPer. CASPer includes 8 SJT video-based sections, each with 3 follow-up questions. Applicants have 5 minutes to type responses per section. Five to six years later, CASPer scores were correlated against

participants' subsequent performance on non-clinical decision-making areas of the Medical Council of Canada Qualifying Examination (MCCQE) I and II.

Summary of Results: By 2013, 109 and 65 applicants had attended Canadian medical schools and took MCCQE I and II respectively. After disattenuation correction, correlations of the video SJT with MCCQE subsections most closely related to personal/professional characteristics ranged from r = 0.21, p = 0.237 to r = 0.50, p = 0.025. Correlations between the video SJT and MCCQE subsections most closely related to cognitive skills approached zero.

Conclusions: Video-based situational judgment tests can predict for national licensure scores. Longer-term follow-up with larger groups is anticipated in coming years.

Take-home Messages: The relationship is similar to that found between MCAT performance and National Licensure examination (Julian, 2005) and supports the additional use of CASPer in applicant selection.

OTT-OB-8: WORK BASED ASSESSMENT 1

OTT-OB-8-1
Specialty Specific DOPs -Tailoring the Associated Assessment System to the specialty curriculum: required practical procedures

Eleana Ntatsaki (*University College London, Medical School, London, UK)*, Gavin Clunie (*Cambridge University NHS Hospitals Foundation Trust, Addenbrooke's Hospital, Department of Rheumatology, Cambridge*)

Background: Training in any medical specialty requires competence and skill in the practical procedures that are relevant and specific to the specialty. Direct Observation of Procedure (DOPs) is a workplace based assessment (WBA) utilising relatively broad domains of assessment for practical procedures. For the assessment of specialty specific procedures however, it is reasonable to consider modification of the generic DOPs assessment to assess aspects of competence and skill not covered by the generic domains. We describe a structured, consensus-based, reproducible, stakeholder-relevant approach for designing specialty specific WBAs that enables tailoring assessment tools to the specific specialty's requirements.

Summary of Work: The Rheumatology Specialty Advisory Committee (SAC) identified a need for speciality specific DOPS (ssDOPS) for core curriculum procedures such as joint injections. For the development of procedure-specific assessment indicators for a Rheumatology ssDOPs, a structured opinion-requesting Delphi approach of SAC Committee members was followed. Delphi is a well described systematic and interactive technique used to obtain a consensus from key stakeholders.

Summary of Results: Draft versions of the proposed ssDOPS were circulated to all SAC members who reviewed the specific assessment stems awarding them a score ranging from 1-5. The top scoring stems in each

domain were used to populate the final ssDOPS assessment, before piloting them.

Conclusions: In designing or modifying an assessment tool, a Delphi approach can ensure there is consensus amongst the stakeholders. We suggest this method of generating ssDOPS might be a valid way of generating specialty specific WBAs by SACs and/or other stakeholders.

Take-home Messages: Specialty specific procedures require a relevant and tailored assessment system developed using a structured, consensus-based, reproducible, modifiable, stakeholder-relevant approach.

OTT-OB-8-3
Assessor training to support a bi-national workplace-based assessment initiative

Mary Lawson (*ACEM, Education, Melbourne, Australia)*, Claire Spooner (*ACEM, Education, Melbourne*), Claire Byrne (*ACEM, Education, Melbourne*)

Background: A system of Workplace-based Assessment (WBA) is being implemented across Advanced Training in Emergency Medicine in 2015. To support the implementation and promote standardisation and uptake, a cascade model of assessor training has been developed and implemented.

Summary of Work: Recruitment has been conducted for assessor coordinators at a regional and local level across Australasia. Face-to-face training is being delivered to all coordinators with the expectation that local coordinators take responsibility for oversight of departmental-level training. This locally delivered training is supported by a series of four online modules covering principles and rationale for workplace-based assessment, system and instrument orientation, calibration and frames of reference for assessment and feedback for learning.

Summary of Results: A total of 11 Regional Coordinators and 182 Local Coordinators have been recruited. All Regional Coordinators have been trained and now participate in their own regions as trainers and advocates for the implementation process. The Local Coordinators attend training and subsequently engage in department-based training and orientation. Collectively, they provide coverage of 95% of all ACEM accredited training hospitals bi-nationally. Calibration exercises undertaken at training workshops can be successfully incorporated in online activities to support more general roll-out. Evaluation has been positive and identified further resources required by assessors which can be provided in the online environment.

Conclusions: Assessor training can be successfully implemented on a bi-national basis adopting a cascade approach. A blended approach to training addresses issues of feasibility in a dispersed training model.

Take-home Messages: Assessors are willing to take a leadership role in the implementation and roll-out of a major new assessment initiative.

OTT-OB-8-4

Survey of conducting direct observation of clinical skills on last year medical students of Islamic Azad University of Mashhad

Seyed Mohamad Ali Raisolsadat (*Islamic Azad University, Dept of Surgery, Mashhad Branch, Mashhad, Iran)*, Arezou Farajpour (*Shahid Beheshti University of Medical Sciences, Tehran*), Hamidreza Arshadi (*Islamic Azad University, Department of Psychiatry, mashhad*), Mitra Amini (*Education Development and Research Center, Shiraz University of Medical Sciences, Shiraz, Education Development and Research Center, Shiraz*), Alireza Alehashemi (*Islamic Azad University, mashhad, iran, Department of Neurology, Mashhad*), Elahe Mohamadi (*Shiraz University of Medical Sciences, Education Development Center, Shiraz*)

Background: Assessment is an essential and integral part of medical education. If the goal of education is learning the skills of the work place, the tests should be designed in this regard as well, including the direct observation of clinical skills, mainly DOPS (Direct Observation Of Procedure Skills).

Summary of Work: This is a descriptive study. 60 senior medical students and 23 assessors participated in the emergency department. The questionnaire developed by the researcher and the questions were in areas of satisfaction, acceptability and educational impact on a Likert scale design. Its validity confirmed by a panel of experts and Cronbach's alpha coefficient was 85%. Data were analyzed with the SPSS16 software.

Summary of Results: Of 60 students, 26 were males and 34 were females. Minimum, maximum and mean of age were 23, 28 and 25 years respectively. Of 23 assessors, 15 were male and 8 were female, age range was 30-46 years and the average is 37.39. Educational impact in students and overall satisfaction in assessors were highest(73% and 77% respectively) and other aspects in all areas were above mean average.

Conclusions: DOPS is an acceptable, feasible exam and has a positive educational impact; therefore by implementing DOPS, educational objectives can be achieved.

Take-home Messages: DOPS is an efficient assessment tool for assessing workplace skills.

OTT-OB-9: COMPETENCY BASED ASSESSMENT 2

OTT-OB-9-1
CanMEDS 2015: The environmental scan on directions for the future

Elizabeth Wooster (*Royal College of Physicians and Surgeons of Canada, Ottawa, Canada)*, Elaine Van Melle (*Queens University, Kingston*), Linda Snell (*McGill University, Medicine, Montreal*), Jonathan Sherbino (*McMaster University, Medicine, Hamilton*), Ming-Ka Chan (*University of Manitoba, Pediatrics, Winnipeg*), Jason Frank (*Royal College of Physicians and Surgeons of Canada, Ottawa*)

Background: The CanMEDS framework has guided medical education since 1996. The Royal College of Physicians and Surgeons of Canada is systematically designing an update to be called CanMEDS 2015. We wanted to begin by understanding current medical educator and stakeholder opinions regarding the current CanMEDS framework.

Summary of Work: A Delphi-informed environmental scan targeted program directors, deans, specialty committees, Royal College educators, fellows, trainees and continuing professional developers in Canada and around the world through email, Royal College notification and specific invitations. Four medical educators coded the results and produced themes using qualitative analysis.

Summary of Results: 1074 respondents identified minor revisions to the current framework from which we drew 4 major themes: emerging concepts; simplification; clarity, and; practical application to medical education.

Conclusions: Stakeholders believe that the CanMEDS framework will require strengthening in certain areas, updating of some concepts that are in flux in the intrinsic roles, and better engagement of programs and faculty in the delivery of medical education using CanMEDS.

Take-home Messages: In general, medical educators and other stakeholders endorse the current CanMEDS framework with minimal changes recommended for CanMEDS 2015. Enhanced knowledge translation of CanMEDS concepts for front-line educators is needed.

OTT-OB-9-2
Priority topics for the assessment of competence in family medicine

Tim Allen (*College of Family Physicians of Canada, Academic Family Medicine, Toronto, Canada)*, Tom Crichton (*Northern Ontario School of Medicine, Family Medicine, Sudbury*), Carlos Brailovsky (*College of Family Physicians of Canada, Academic Family Medicine, Toronto*), Kathy Lawrence (*University of Saskatchewan, Family Medicine, Regina*), Cheri Bethune (*Memorial University, Family Medicine, St John's*), Michel Donoff (*University of Alberta, Family Medicine, Edmonton*)

Background: Assessment in any field should be based on a well-defined domain of competence, establishing clear expectations and limits for the learners and the assessment. The College of Family Physicians of Canada identified 99 Priority Topics for the purposes of assessment at the certification level. The purpose of this study is to determine whether the topics are the same when determined by different groups of practitioners, and whether they stable over time.

Summary of Work: 99 topics originally generated in 1998, using modified Delphi technique. Validation survey was done in 2004 to a stratified sample of 1104 family physicians across Canada. Repeat survey was done in 2013, with the same stratification criteria, to a different group of family physicians

Summary of Results: Results from 2013 are pending. Results from 2004: 37% response, 25% participation, 17% completion rate (172 responders), same demographic distribution as population. Very high correlation with initial 99 topics, both in selection and ranking: only 8% of suggested topics not in the 99; 24 topics were common amongst the top 30 citations in both lists; Spearman rank correlations 0.438, p<0.01, and 0.740, p< 0.000, respectively for matching and write-in formats.

Conclusions: The topics selected as priority were stable over a 5-year period, and over two different groups of family physicians. Hypothesis: the differences will not be substantive after another 10-year interval.

Take-home Messages: Management of patients may change greatly over time, but the basic topics that comprehensive family physicians must deal with change slowly. A stable list of priority topics can used for assessing that competence.

OTT-OB-9-3
A multicenter validation study of a Gastrointestinal Endoscopy Competency Assessment Tool for pediatric colonoscopy (GiECAT-KIDS)

Catharine M Walsh (*The Wilson Centre and The Hospital for Sick Children, University of Toronto, Department of Paediatrics, Toronto, Canada)*, Simon C Ling (*The Hospital for Sick Children, University of Toronto, Department of Paediatrics, Toronto*), Petar Mamula (*The Children's Hospital of Philadelphia, Department of Pediatrics, Philadelphia*), Jenifer R Lightdale (*Boston Children's Hospital, Department of Pediatrics, Boston*), Jeffrey Yu (*The Wilson Centre, University of Toronto, Toronto*), Heather Carnahan (*The Wilson Centre and Centre for Ambulatory Care Education, Women's College Hospital, University of Toronto, Toronto*)

Background: Validated assessment tools are required to support competency-based education and demonstrate trainees are progressing. We aimed to assess the validity and reliability of the Gastrointestinal Endoscopy Competency Assessment Tool for pediatric colonoscopy (GiECAT-KIDS); an instrument developed by 41 North American experts using Delphi methodology.

Summary of Work: GiECAT-KIDS consists of a 7-item global rating scale (GRS) and 18-item checklist (CL). 85 colonoscopies performed by 45 endoscopists at 3 North American teaching hospitals were evaluated by an attending physician: 22 novices (<50 procedures), 17 intermediates (50-250), and 6 experts (>500). Another observer rated procedures to assess inter-rater reliability using intraclass correlation coefficient (ICC). Construct validity was examined by comparing level of experience with GiECAT-KIDS scores. Concurrent validity was assessed by correlating scores with (a) colonoscopy experience; (b) cecal and (c) terminal ileal intubation rates; and (d) physician global assessment (PGA) of skill. Test-retest reliability was measured by correlating endoscopists' first and second procedure scores.

Summary of Results: GiECAT-KIDS GRS and CL scores differed significantly between endoscopists based on level of experience (p<0.001). There was a significant positive correlation (p<0.001) between scores and (a) number of previous colonoscopies (GRS: r=0.81, CL: r=0.73); (b) cecal intubation rate (GRS: r=0.84, CL: r=0.73); (c) terminal ileal intubation rate (GRS: r=0.85, CL: r=0.76), and (d) PGA (GRS: r =0.94, CL: r=0.84). Inter-rater reliability was good (GRS ICC: 076, CL ICC: 0.71). Test-retest reliability was high (GRS: r=0.94, p<0.05, CL: r=0.83).

Take-home Messages: GiECAT-KIDS is a feasible, valid and reliable measure of performance of pediatric colonoscopy that can be used to support training and evaluation.

OTT-OB-9-4
Evaluation of an innovative learner-centred assessment program for family medicine residency training

Maria Palacios (*University of Calgary, Department of Family Medicine, Calgary, Canada)*, Keith Wycliffe-Jones (*University of Calgary, Department of Family Medicine, Calgary*), Tyrone Donnon (*University of Calgary, Department of Community Health Sciences, Calgary*), Vishal Bhella (University of Calgary, *Department of Family Medicine, Calgary*)

Background: As one of the first in Canada to meet the new College of Family Physicians of Canada (CFPC) standards for Family Medicine Residency training, the Calgary Family Medicine Residency Program introduced its new "Triple-C" competency-based curriculum in 2012 and concomitantly developed and implemented an innovative competency-based assessment program based on current best-practice recommendations. This new assessment program, utilizing multiple assessment data, includes the use of field notes, progress reviews, self-assessments and Entrustable Professional Activities ("EPA's").
This 2-phase project studies the impact of the implementation of this new assessment program at both Resident and Preceptor levels (Phase I) and also the evidence for the reliability, validity and feasibility of the assessment methods chosen (Phase 2).

Summary of Work: In Phase I of the study, a number of Preceptors and Residents will be interviewed using a semi-structured questionnaire to investigate their experiences of the new assessment program. Data from these interviews will then been used to generate a program-wide survey instrument for use by all Preceptors and Residents in the Program.

Summary of Results: Qualitative & Quantitative data will be presented from Phase 1 of the study in relation to;
i) Implementation issues –barriers and facilitators
ii) Preceptors and Residents perceptions around educational benefits

Conclusions: The results of this present study will help our understanding of how a multi-method, work-placed based assessment program impacts learners and

preceptors and to what extent both learners and teachers accept the legitimacy of these processes.

Take-home Messages: What works and what doesn't in relation to the implementation of a competency-based assessment program.

OTT-OB-9-5
Is it time to validate the National Competency Profile for Physician Assistants?

Maureen Gottesman (*University of Toronto, Department of Family and Community Medicine, Toronto, Canada*)

Background: The Canadian Association of Physician Assistants released the National Competency Profile (NCP) in 2009, which includes an appendix list of Diseases and Conditions, to specify "the abilities of the Generalist entry-to-practice level PA in recognizing, diagnosing and treating specific conditions and diseases" (CAPA, 2009). The appendix list was generated by a consensus of expert educators, clinicians and administrators. Accredited by the Canadian Medical Association, PA education programs are expected to provide "verifiable data on student learning outcomes"(CMA, 2008) that supports the attainment of competencies, including the appendix list of the NCP.

Summary of Work: The University of Toronto's Physician Assistant program requires learners in their clinical year to electronically log procedures and diagnoses. The template includes both 193 required diagnoses and 36 required procedures, as well an additional 819 diagnoses and an additional 155 procedures that students are expected to encounter in their clinical training.

Summary of Results: The completion rates of the 44 student logs from the first 3 years of clinical placements at the University of Toronto will be reported. Although online, the logging systems and templates have been barriers for students to complete their logs efficiently and accurately.

Conclusions: The use of the self-reporting logs to document clinical experiences may be an inaccurate and misleading body of data by which to measure the true competency profile of our PA graduates.

Take-home Messages: The "required" diagnoses and procedures listed in the CAPA NCP should be validated in the clinical field, with practicing PA professionals, PA students and supervising physicians.

POSTERS

OTT-PAB8: SELECTION

OTT-PAB8-01
Personality Assessment: Can It Help to Select the Medical Student?

Anupong Suthamnirand (*Chonburi Medical Education Center, Chonburi, Thailand)*, Nattinee Shinajitpun (*Chonburi Medical Education Center, Chonburi*)

Background: Selection for admission to medical school in Thailand mostly relied on knowledge score (cognitive attribute). But to train medical students to become a good doctor needs more than that, and perhaps personality and coping style could help. This study was to determine the personality of the medical student and the grade score in medical school.

Summary of Work: Ninety-eight (forty-three male and fifty-five female), 4th to 6th year medical students, academic year 2013, Chonburi Medical Education Center, were enrolled. The Sixteen Personality Factor Questionnaire (16 PF, Thai version), was used to measure the big five secondary traits; extroversion, anxiety, self-control, independence, and receptivity of all students. The cumulative grades of medical students were compared with five secondary traits using Pearson's r correlation coefficient.

Summary of Results: In all students, there was statistic significant correlation between cumulative grade and extroversion. For male students, there was no statistically significant correlation between grade and personality, but in female students, there was significant correlation with extroversion and independence.

Conclusions: Some personality traits can predict the learning achievement in medical students.

Take-home Messages: Good selection builds great doctor.

OTT-PAB8-02
Removing the selection interview leads to gender bias amongst enrolled medical students

Mavourneen Casey (*The University of Queensland, School of Medicine, Brisbane, Australia)*, Diann Eley (*The University of Queensland, School of Medicine, Brisbane*), David Wilkinson (*Macquarie University, The Chancellery, Sydney*)

Background: Selection processes for entry into medical school are continually under scrutiny, and in 2009, the selection interview was discontinued at The University of Queensland (UQ). The student cohort that had originally participated in an interview, were the domestic graduate entry students.

Summary of Work: A retrospective analysis of student records (n= 4051) for nine consecutive cohorts (2004-2012) was conducted to examine the impact of changes in the admissions selection criteria on medical student characteristics. Main outcome measures included change in number (%) of male and female students over time and gender-specific admissions test section scores (GAMSAT).

Summary of Results: Between 2004-2008, 891 (51.4%) enrolled students were male, whereas between 2009-2012, 1134 (57.7%, p< 0.0001) were male. This change in gender ratio was limited to domestic direct graduate entry students and the male proportion in this group rose from 50.9% (705) when an interview was part of the selection process to 64.0% (514, p< 0.0001) after removal of the interview (reaching 73.8% in 2012). Between 2004 and 2012 males consistently performed

better on the GAMSAT section III (mean score 71.5 vs. 68.5, p< 0.0001).

Conclusions: The proportion of males enrolled in the medical program at this university increased markedly following removal of the interview from the selection process. This change was limited to domestic direct graduate entry students, the sub-cohort who had originally participated in an interview, and seems to be due to a gender bias, favouring males, in the GAMSAT.

Take-home Messages: The selection interview may play an important role in ensuring gender equity in selection and medical schools should carefully monitor the consequences of changes to selection policy.

OTT-PAB8-03
Assessment of academic preparation and motivation to create a standard of entry behavior for International Dental Course program at Faculty of Dentistry Hiroshima University

Ketut Suardita (*Institute of Biomedical & Health Sciences, Hiroshima University/Faculty of Dentistry, Airlangga University, Department of International Collaboration Development for Dentistry/ Department of Conservative Dentistry, Hiroshima/Surabaya, Indonesia)*, Hiroko Oka (*Institute of Biomedical & Health Sciences, Hiroshima University, Department of International Collaboration Development for Dentistry, Hiroshima*), Takashi Takata (*Institute of Biomedical & Health Sciences, Hiroshima University, Department of International Collaboration Development for Dentistry / Department of Oral and Maxillofacial Pathobiology, Hiroshima*), Darmawan Setijanto (Faculty of Dentistry, Airlangga University*, Department of Dental Public Health, Surabaya*), Motoyuki Sugai (*Institute of Biomedical & Health Sciences, Hiroshima University, Department of Bacteriology, Hiroshima*)

Background: To develop Asia-based global collaboration in dental education and research, Faculty of Dentistry Hiroshima University established an innovative dental program, International Dental Course (IDC) program. After finishing the first year of study in their country, the international students will study for 4 years as IDC students with regular program students. We developed a Japanese-English dual linguistic for our education system.

Until this second year of the program, we received six students from three sister schools (Indonesia, Vietnam and Cambodia). We used to evaluate our program to provide feedback that will improve the program better. For the first step, we try to find the standard of entry behavior of the international students therefore we can make recommendations on how to prepare prospective students that join IDC.

Summary of Work: We assessed the academic preparation including curriculum, grade point average (GPA) and motivations of the students before joining the IDC program. Relationship between the academic preparation and motivation of the students before joining IDC with their academic achievement after joining IDC program was examined.

Summary of Results: The results showed that curriculum and motivation should be essential as standard of entry behavior but in this study did not play a crucial role on the academic achievement because all of the students had strong motivations and the curriculums were not so different among the schools. High GPA before joining the program had a significant effect on excellent academic achievement.

Conclusions: High GPA was recommended as standard of entry behavior for IDC Program at Faculty of Dentistry Hiroshima University, though motivation and adjustment of the curriculum seem important too.

Take-home Messages: Standard of entry behavior is very important to get prospective students who are expected to gain excellent academic achievement.

OTT-PAB8-04
Evaluation of machine-marked tests for selection into core medical training: Evidence from a longitudinal validation study

Safiatu Lopes (*Work Psychology Group, Derby, UK)*, Fiona Patterson (*Work Psychology Group & University of Cambridge*), Liz Berkin (*Royal College of Physicians*), Stephen Harding (Royal College of Physicians), Bill Burr (*Royal College of Physicians*)

Background: With over 2,500 applicants being interviewed annually for UK Core Medical Training (CMT), a major challenge is ensuring calibration and standardisation of assessments. This study builds on previous research which examined the incremental validity of two machine-marked tests (MMT), a clinical problem-solving test (CPST) and a situational judgement test (SJT), in predicting in-training performance over current interview methods.

Summary of Work: The study tracked candidates' performance over an extended time interval and assessed whether MMT scores are able to predict speciality training (ST3) selection outcomes. A longitudinal design was used to examine the predictive validity of the MMT. The correlation between MMT scores at selection and subsequent performance in the College licensure examination (MRCP) approximately 4 years into training (N=1,733) was analysed. The MRCP has 3 components including applied-knowledge tests (Parts 1 & 2) and a high-fidelity patient simulation (PACES).

Summary of Results: The MMT showed good reliability (α=.85). Both the MMT and interview scores were significantly positively correlated with all elements of the MRCP (p<.001), indicating good predictive validity. For MRCP Part 1 and 2, the strongest predictor was the CPST (r=.53-.66, p<.001), reflecting similar content and style of these methods. For PACES, the strongest predictor was the interview (r=.45, p<.001), followed by the SJT (r=.37, p<.001). Further analyses with ST3 selection scores as a criterion will be conducted.

Conclusions: The MMTs offer an appropriate measurement methodology for selection into core medical training, and add significant incremental validity to the process.

Take-home Messages: MMTs can be used to more accurately calibrate high volume nationally coordinated selection practices.

OTT-PAB8-05
Reliability of Standardized Interviews at Schulich Medicine

Lorne Gula (*University of Western Ontario, London, Ontario, Canada)*, MaryAnn Kennard (*University of Western Ontario, London, Ontario*), Kathy Sadler (*University of Western Ontario, London, Ontario*), Elana Whelan (University of Western Ontario*, London, Ontario*), Francis Chan (*University of Western Ontario, London, Ontario*), Bertha Garcia (*University of Western Ontario, London, Ontario,* Presenter: <u>Robert Hammond</u>)

Background: Interviews are a traditional element of the Medical school admissions process and have intuitive value in the assessment of an applicant's ability to interact with patients and colleagues. However, to be a fair and functional tool, logistical challenges for applicants should be minimized and the atmosphere of the interview made conducive to a representative assessment of a variety of individuals. Similarly, sources of potential bias should be minimized and reliability maximized

Summary of Work: Schulich Medicine interviews are conducted by teams of 3 individuals (community member, senior medical student, faculty member) who receive annual on-site training to follow a standardized format. Annual post-interview surveys and analyses are conducted to confirm and strengthen best practices from applicant, interviewer and administrative perspectives.

Summary of Results: The interview atmosphere and format have been overwhelmingly positively received by applicants and interviewers who express a high degree of confidence in the interview format, content and the ability of the interview to assess candidates in a consistent and thorough manner. These sentiments are echoed in analyses of the interviews. Daily and annual score distributions are highly consistent. Reliability is strong between and within interviewer constituencies. Factor analysis has consistently demonstrated significant covariance of scoring clusters within CanMEDs domains, further validating the use of these domains as a framework for the interview.

Conclusions: Standardized interviews at Schulich Medicine have proven to be reliable and free of significant bias. This selection tool will continue to be informed and refined on the basis of annual stakeholder feedback, statistical analysis and in studies of validity based on longitudinal outcomes.

OTT-PAB8-06
If At First You Don't Succeed. . . Try MCAT Again!

<u>Sarah Wright</u> (*University of Toronto, Toronto, Ontario, Canada)*, Lindsay Jackowetz (*University of Toronto, Toronto, Ontario*), Kulamakan Kulasegaram (*University of Toronto, Toronto, Ontario*), Mark Hanson (University of Toronto*, Toronto, Ontario*)

Background: Medical schools often consider MCAT as a stable measure of a student's cognitive ability, despite evidence that MCAT scores fluctuate from one attempt to another. While the extent of MCAT score change has been previously documented, the proportion of applicants taking the MCAT multiple times has not been addressed. Furthermore, gains beyond the second attempt at MCAT have not been documented. This study looked at the MCAT test taking behaviour of applicants to University of Toronto, including the number of times MCAT was taken and the extent repeat test taking affected scores.

Summary of Work: The MCAT history of two recent cohorts U of T applicants were collected (N=6205). The proportion of repeat test takers was identified. For this subset, MCAT total score and subtest score changes were analyzed using repeated measures ANOVA.

Summary of Results: "40% of all applicants took the MCAT at least twice. Of the population of MCAT repeaters, 37% had taken it 3 or more times. Mean total score increased significantly when MCAT was taken a second time (T1 = 26.99, T2= 29.12, p<.01, Cohen's d = .64). For the subsample that took the test 3 times (N=792), there was a progressive increase in scores (T1= 25.82, T2 = 27.13, T3 = 28.85, p<.01, d= 0.76), with a similar significant pattern for those who took the test 4 times (N= 229, d= 0.96). Increases were uniform over the four subscales."

Conclusions: This study indicates that a substantial proportion of U of T applicants take the MCAT more than once. Repeat test taking increases scores for those who repeat the test more than once, with no diminishing returns. More research is needed to clarify the cause of sustained score increases with repeated test taking and how repeat MCAT scores should be used by medical schools.

OTT-PAB8-07
The Cost of MCAT Preparation: Findings from a Medical Student Survey

<u>Jesse Kancir</u> (*University of Toronto, Toronto, Ontario, Canada),* Sarah Wight (*University of Toronto, Toronto, Ontario*), Lindsay Jackowetz (*University of Toronto, Toronto, Ontario*), Kulamakan Kulasegaram (University of Toronto*, Toronto, Ontario*), Mark Hanson (*University of Toronto, Toronto, Ontario*)

Background: The (MCAT) is variously but commonly used for admission to Canadian medical programs. There are a variety of preparation activities available to students, but little is known about the proportion of students who actually engage in such activities or their associated costs. The purpose of this study was to investigate student preparation methods for the MCAT.

Summary of Work: Incoming first-year medical students at the University of Toronto were asked a series of questions about their experiences preparing for MCAT. A total of 199 students (76% response rate) responded

to questions related to preparation time, resources utilized, and estimations of associated costs.

Summary of Results: Nearly all students had taken a practice MCAT, with 48% reporting having taken more than 6 practice tests. 52% of students enrolled in general preparatory courses for the MCAT. Of the course takers, 88% reported spending $1000 or more and 32% reported that parents had funded their preparation course. 67% of students reported preparing for MCAT over the summer. Of those who typically held a summer job, 54% reported having to reduce hours or not seeking employment at all with the sole purpose of preparing for the MCAT.

Conclusions: This study demonstrates that a large proportion of medical students have engaged with various forms of MCAT preparation, indicating a perception that significant preparation for MCAT is necessary. This perception coupled with costs associated with preparation courses, opportunity cost of not taking summer employment and costs of writing the test may have an impact on the socioeconomic demographic of the medical school applicant pool.

OTT-PAB8-08
Selection into specialty training: assessing behaviour and values based recruitment to improve quality of patient care

Alison Carr (*Health Education England, London, UK),* Fiona Patterson (*Work Psychology Group & University of Cambridge*), S Lopes (*Work Psychology Group*), Bill Irish (Health Education South West), Selena Gray (*Health Education South West*)

Background: Having completed a two year Foundation training post, doctors in the UK apply for specialty training (ST1/CT1). Currently, each specialty has their own recruitment process that seeks to assess the attributes and competencies required for that specialty. With an aim to improve way doctors are selected in the future for ST1/CT1 recruitment, a single Speciality Selection Test (SST) is being piloted with all specialties in the UK.

Summary of Work: This project describes the design, piloting and planned evaluation of a Clinical Problem Solving test and Situational Judgement Test for selection of over 11,000 trainee doctors.

Summary of Results: Evaluation of the pilot phase is due to take place in February 2014. Psychometric analysis will look at the appropriateness of the test within this context and its use across all specialties. In particular, relationships between the SST scores and the current selection processes for each specialty will also be investigated. Feedback will be gathered from candidates to evaluate whether they felt that the SST was relevant and fair.

Conclusions: This project has been endorsed by Health Education England and is part of the Medical and Dental Recruitment and Selection (MDRS) Programme whose aim is to ensure that the UK are recruiting the right people with the right skills, values and behaviours, which will, in turn, improve the quality of patient care.

Take-home Messages: A robustly designed SST will ensure that selection into NHS funded training posts incorporates the testing of clinical problem solving, situational judgement testing and values and behaviours based recruitment.

OTT-PAB8-09
Is Performance at GP training selection related to Myers Briggs personality type?

Richard Elliott (*Wessex School of General Practice, Wessex Deanery, UK, Southampton GP Education Unit, Southampton, UK)*, Sandy Miles (*Wessex School of General Practice, Wessex Deanery, UK, Southampton GP Education Unit, Southampton*)

Background: Myers Briggs personality indices are metrics that enable individuals to better understand their preferences in four domains, and thus how they are most comfortable thinking and acting. Our research considered whether Myers Briggs personality indices are related to trainee performance in the Stage two and three assessments for admission to GP training, and thus the thinking and learning preferences of trainees recruited to GP specialty training.

Summary of Work: Performance in stages two and three of selection for a cohort of 28 Southampton GP trainees was retrospectively collected from deanery records. Myers Briggs data was available for 23 of the 28 trainees. Performance in two assessment exercises was compared to their Myers Briggs results. Non-parametric T testing was performed on each data set, to determine whether they were significantly different to the cohort as a whole.

Summary of Results: Trainees with an 'Introverted' personality demonstrated superior performance than those with an 'extraverted' personality in both stages of assessment (p=0.055 for stage two, p=0.07 for stage three).

Trainees with a 'Thinking' personality performed better than those with a 'Feeling' personality in stage two of interview (p=0.047).

In all other areas, no correlation between personality type and performance was demonstrated.

Conclusions: These results were surprising; those found to be 'thinkers' were not anticipated to outperform 'feelers' in the assessments due to the need for GPs to demonstrate a deep understanding of their patients' feelings and emotions. Further research is warranted to determine whether trainees type alters during training.

OTT-PB5: COMMUNICATION SKILLS / WORK-BASED ASSESSMENT

OTT-PB5-01
Evaluation of Communication Skills Training in the Undergraduate Dental Curriculum

Dixuan Zhang (*University of Toronto, Faculty of Dentistry, Toronto, Ontario, Canada)*, Laura Dempster

(*University of Toronto, Faculty of Dentistry, Toronto, Ontario*)

Background: The Faculty of Dentistry, University of Toronto developed a new Communication Skills program to give students simulated clinical experience practicing communication skills.

Summary of Work: Second year dental students were randomly selected to play the role of "dentist" in communication scenarios common to clinical practice. A cross-over design compared Standardized Patient (SP) and Role-Playing (RP) sessions. Seven elements of communication were scored (1/poor–5/excellent) using the validated Kalamazoo Essential Elements Communication Checklist. Research investigated students' skill improvement; agreement among self, Faculty, and educator raters; and the effect of SP vs RP formats.

Summary of Results: One element (gathering information) was significantly improved pre- and post-program ($F=17.91$, $p<0.01$). In session 2, post-hoc analysis identified significant differences ($p<0.01$-0.04) in mean(SE) scores between self (3.25(0.17)-3.35(0.13)) and Faculty/educators (2.62(0.20)-2.65(0.29)) in 3 elements. Faculty and educators reported moderate-high correlations (ICC: 0.57-0.84); self and Faculty reported low correlations (ICC: 0.04-0.53). No effect of SP or RP format was reported although students' reported higher scores for session 2 (irrespective of SP or RP format order), despite similar Faculty ratings between sessions.

Conclusions: Students' communication skills in gathering information improved at the completion of the program. Differences between self and Faculty/educator scores were noted with low correlations between students' and Faculty/educator ratings. Students' reported improved performance in session 2, irrespective of order of SP or RP format; however, Faculty ratings remained similar.

Take-home Messages: Interest lies in differences in student learning of communication elements; the need to improve agreement between student and faculty ratings; and why higher sessions 2 scores were reported by students (confidence vs performance).

OTT-PB5-02
The Audio-COT (Clinically Observed Tool) – a friend or foe? Assessment of GP trainers' interest in and use of this clinical assessment tool

Bryony Sales (*Wessex School of General Practice, Wessex Deanery, Portsmouth, UK)*, Samantha Scallan (*Wessex School of General Practice*), Sue Crane (*Wessex School of General Practice, Wessex Deanery*), Johnny Lyon-Maris (Wessex School of General Practice*, Wessex Deanery*)

Background: The use of telephone triage and consultations in healthcare has increased in recent years, requiring GP trainees to not only develop face-to-face communication skills, but those using the telephone. It can be challenging for trainers to find ways to teach and assess these skills in an authentic way. This project piloted and evaluated an audio-COT (clinically

observed tool) which aimed to increase trainers' ability and confidence in using this assessment tool with trainees.

Summary of Work: Trainers completed a pre-session questionnaire to ascertain their awareness of the audio-COT, its perceived relevance for GP training and confidence in teaching telephone consultations skills. Trainers then attended a facilitated session, which included sharing telephone consultation teaching experience, guidance on useful clinical models to use, potential equipment to support learning and tips to facilitate trainee engagement. Examples of audio-COTs were assessed and scores discussed within the group, to consolidate learning and facilitate assessment benchmarking. After the session trainers provided feedback and information about their subsequent use of audio-COTs.

Summary of Results: This work is in progress. It is anticipated that GP trainers will increase their awareness of and willingness to use the audio-COT following the session; trainers will be more prepared and confident in facilitating trainee learning and assessment of telephone consultations.

Conclusions: The use of audio-COTs allows trainers an additional supervised learning episode to formally assess and develop the clinical competence of trainees' telephone consultation skills.

OTT-PB5-03
What do patients and caregivers say about how clinicians communicate a diagnosis of schizophrenia?

Carmel Loughland (*University of Newcastle / Hunter New England Health Service, Psychiatry / Priority Research Centre for Translational Neuroscience and Mental Health, Newcastle, Australia)*, Sue Outram (*University of Newcastle, Priority Research Centre for Health Behaviour, Newcastle*), Brian Kelly (*University of Newcastle / Hunter New England Health Service, Psychiatry / Priority Research Centre for Translational Neuroscience and Mental Health, Newcastle*), Harsimrat Sandhu (Hunter New England Health Service*, HNET, Newcastle*), Marina Vamos (*Hunter New England Health Service, Psychiatry / HNET, Newcastle*), Tomer Levin (*Memorial Sloan Kettering Cancer Center, Department of Psychiatry and Behavioral Science, New York*)

Background: Despite widespread acceptance of the principle that patients should be informed about their diagnosis, many clinicians are reluctant to provide a diagnosis of schizophrenia. This study examines patients and their caregivers' experience of the communication of a schizophrenia diagnosis and related information.

Summary of Work: A generic qualitative methodological approach was used. Thirteen family caregivers were recruited in regional New South Wales, Australia. Semi-structured interviews were used to explore their perceptions and experiences of discussing the diagnosis, prognosis and treatment of schizophrenia with mental health professionals. Interviews were recorded, transcribed, codes generated and thematic analysis undertaken.

Summary of Results: Patients and caregivers described long and difficult pathways to being given a diagnosis, haphazard means of finding out the diagnosis, high unmet needs for information, exclusion from the medical care process, and problematic communication and general interactions with mental health clinicians. Patients and caregivers were unanimous about the importance of receiving a timely diagnosis, with over half the patients interviewed reporting that receiving a diagnosis of schizophrenia was beneficial to their well being and recovery as it relieved confusion and provided a 'name' for their experiences. One-third of patients reported negative impacts when diagnosis was delayed or not provided and many felt a lack of information was disempowering.

Conclusions: This research provides evidence for the need to improve doctorl-patient communication through communication skills training in psychiatry to empower patients in their own health recovery process.

Take-home Messages: The perspectives of patients and caregivers on communicating a diagnosis of schizophrenia provide important information for communication skills training of psychiatrists and other mental health professionals.

OTT-PB5-04
What are the barriers to good practitioner-patient communication in psychiatry: A qualitative study

Sue Outram (*University of Newcastle, Priority Research Centre for Health Behaviour, Newcastle, Australia*)

Background: Mental health clinicians are often criticized for their failure to communicate well and share diagnostic information despite evidence that people with schizophrenia and their families want this. A generic qualitative methodological approach was used to examine clinicians' reasons for giving, or not giving, a diagnosis of schizophrenia, and to document the process of communicating a diagnosis, so as to better understand the complexities facing clinicians in communicating diagnostic news.

Summary of Work: Sixteen mental health clinicians were recruited from a public mental health service in regional Australia. Semi-structured interviews were used to explore their perceptions and experiences of communication with patients and their family concerning the diagnosis of schizophrenia. Interviews were recorded, transcribed and thematic analysis undertaken.

Summary of Results: Over half the clinicians interviewed believed in giving a diagnosis, but a large proportion also felt reluctant to do so in practice. The reasons stated for not giving a diagnosis included diagnostic uncertainty, loss of hope, stigma, and ineffective treatment options. Communicating diagnostic news to patients and families was generally not planned for, and was often described by clinicians as chaotic and ad hoc with little consistency regarding leadership approaches, how the team operated with regard to management, or the way diagnostic information was communicated to the patient and family. This contributed to tensions between different team members.

Conclusions: This study provides insight into the barriers that prevent clinicians from communicating a diagnosis of schizophrenia and highlights the need for specialist communication skills training in psychiatry.

Take-home Messages: Specialist communication skills training in psychiatry is needed to assist practitioners to communicate diagnostic information effectively, to lead and manage multi-discipline teams to improve health care outcomes for patients.

OTT-PB5-06
Assessment of physicians' communication skills through a physician-patient questionnaire

François Goulet (*Collège des médecins du Québec, Practice enhancement division, Montréal, Canada)*, Johanne Thiffault (*Collège des médecins du Québec, Practice enhancement division, Montréal*)

Background: Doctor-patient communication is considered a core clinical skill at the heart of health care delivery. Good patient-doctor communication has been shown to have positive influence on patients' recall, understanding and adherence, symptom resolution, clinical outcomes, physicians' and patients' satisfaction, as well as frequency of malpractice claims.

Summary of Work: A 19-item physician-patient questionnaire developed and validated by Suzuki-Laidlaw and al. in 2001 based on the Calgary-Cambridge Guide and on other existing instruments, was translated in French and submitted to 90 physicians practicing in Quebec and to a sample of their patients. Participants: 30 family physicians seeing their own patients, 30 family physicians seeing patients for the first time at a walk-in clinic and 30 specialists seeing new patients and patients for follow-up. Each physician completed the self-evaluation questionnaire and a minimum of 15 of their patients completed the same 19-item questionnaire. The questionnaire was completed by the patient at the end of the consultation and the physician self-assessment questionnaire was completed only once at the end of the day.

Summary of Results: Of the 290 physicians who were randomly selected, 90 agreed to participate (31 family physicians in walk-in settings, 26 family physicians with their own clientele and 33 specialists (7 dermatologists, 10 orthopedists, 5 rheumatologists and 11 cardiologists)).

A total of 1 372 patients were interviewed. Completed questionnaires were coded and statistical analyses were performed using t-tests, chi squares, item analysis and generalizability analysis.

Conclusions: Younger patients are more critical in the assessment of their physician's communication skills. Physicians tend to be more critical of themselves, compared to their patients' assessment.

Take-home Messages: Younger patients are more critical in the assessment of their physician's communication skills. Physicians tend to be more critical of themselves, compared to their patients' assessment.

OTT-PB5-07

Evaluating the baseline communication skills of Family Medicine residents through a benchmark evaluation

Alim Nagji (*University of Alberta, Department of Family Medicine, Edmonton, Canada)*, Doug Klein (*University of Alberta, Department of Family Medicine, Edmonton*)

Background: Effective communication is the cornerstone of the patient-physician relationship, especially in Family Medicine. The diverse background of residents creates a potential for unequal footing. Communication OSCEs are widely used and form a core component of resident evaluation. However, opportunities for formative, rather than summative, feedback are limited.

Summary of Work: New residents in the University of Alberta Family Medicine program underwent a 20 minute communication OSCE and a 12 question self-administered content quiz at the start of their residency. Each resident was directly observed by a Family Physician and provided with 10 minutes of structured feedback, with an opportunity for questions and discussion. The entire process remained private and did not affect summative evaluations.

Summary of Results: 65 Residents completed the skills assessment (54 Canadian Graduates, 11 International Graduates). The mean score for the content quiz was 20.60 of a total score of 24. Resident scores ranged from 8 to 24. The mean score on the practice interview was 20.9 out of 30 with a range of 13 to 29. Learner feedback indicated that the skills assessment was useful (4.68/5) and would lead to a change in practice (4.43/5).

Conclusions: An introductory communication assessment offers new residents an opportunity to gauge their baseline skill level, become aware of program expectations early in their training and garner specific suggestions in a non-threatening environment. This tailored approach helps orient residents while acknowledging their previous experiences.

Take-home Messages: A baseline, formative communication evaluation can help ensure equal footing for new residents.

OTT-PB5-08

Mail from the Heart: Simulated E-mail to Assess Knowledge and Communication Skill in Medical Students

Wasana Hongkan (*Chonburi Medical Education Center, Chonburi, Thailand*)

Background: The teaching hours for congestive heart failure and acyanotic congenital heart disease in the classroom were limited. There is an increasing trend of using e-mail for communication between doctors and patients and communication skills via email should be taught in medical students. Would Simulated E-mail help them to learn?

Summary of Work: In academic year 2010-2012, two hours of lectures about congestive heart failure and acyanotic congenital heart disease was given to fourth year medical students at the fourth week of 10 weeks rotation. At the fifth week, the students would received simulated e-mails. Simulated e-mails contained questions about caring of ventricular septal defect, patent ductus arteriosus pediatric patients from concerned parents. Students had to reply to the first email at the sixth week then received feedback from the teacher. The final email was sent to students in the seventh week and students replied before ending the rotation. Rubric score was used to assess score of knowledge and communication skill of responses to first and final e-mails. Focus group was used to assess student satisfaction.

Summary of Results: Thirty students completed the study. Clinical performance score was statistically significant improvement from the first to final email response, overall score from 6.76 to 10.2 (p 0.001), knowledge score from 1.7 to 3.2 (p 0.001) and communication skill score from 5 to 7 (p 0.001). Most students were satisfied with this study method for increasing knowledge and communication skill.

Conclusions: Simulated E-mail is an effective assessment tool for knowledge and communication skill.

Take-home Messages: Self-directed learning and feedback are the keys to improve knowledge and communication skill via email.

OTT-PB5-09

Re-evaluating educational assessment: Using ethnography to assess the learning of community health workers in Ecuador

Peter Nugus (*McGill University, Montreal, Canada)*, Alison Doucet (*McGill University, Montreal*), Helene Rousseau (*McGill University, Montreal*), Andrea Evans (University of Toronto*, Toronto*), Alexander Caudarella (*University of Toronto, Toronto*), Ann Macaulay (*McGill University, Montreal*)

Background: The few studies that have used ethnography to investigate learning tend to use it merely for description rather than assessment. Although observation is crucial to judgments of clinician and student performance, it is usually ad hoc. Observing professional training in context might help to provide guidelines to standardize practical learning and competence at work.

Summary of Work: This mixed method study drew on ethnographic observations, surveys, interviews and document analysis in relation to a two-week workshop in which Canadian family medicine doctors trained rural, indigenous people in Ecuador to be community health workers. The workshop was conducted in April 2013. Survey data were analyzed using descriptive statistics and qualitative data were analyzed using iterative cycles of thematic analysis in which all variation in the data was incorporated into existing themes or presented as additional themes.

Summary of Results: Systematic observation by a trained ethnographer showed that developments in learning were able to be traced to particular interventions. Changes in practical competence and

knowledge acquisition were evident in the strategic use of open and closed questions and practical exercises in the form of role play.

Conclusions: On a daily basis, clinical supervisors rely on observation to assess the practical skills of residents. The principles and strategies discerned in this study can provide a roadmap to document these to formalize and systematize work-based assessment in health care. Practical exercises and questions are amenable to independent assessment, and summative, as well as formative evaluation.

Take-home Messages: Ethnographic observation is an ideal and under-exploited tool to assess learning in the workplace.

OTT-PB5-10
Mini-MAS, a workbased assessment tool to assess milestone stages

Moyez Ladhani (*McMaster University, Pediatrics, Hamilton, Canada*)

Background: The success of competency-based education relies on frequent assessment and feedback. Much work remains to be done on how one will measure the defined competencies. Intrinsic skills such as communication, collaboration etc. remains difficult to measure yet will be integral in the desired competencies. The learner's advancement will rely on effective assessment methods.

Summary of Work: The American Board of Pediatrics Milestones project pieced together the developmental progression of knowledge, skills and attitudes for each sub-competency and then translated this background information into milestones.

Using these milestones, we have developed a tool to assess these milestones called the mini-MAS (Mini Milestones Assessment). A separate tool was created for 6 competencies; data gathering, physical exam, clinical reasoning, communication with families, communication with physicians and other health professional and collaboration. The tool has descriptors and is meant to differentiate the novice from the expert. The tool has been implement for our competency-based curriculum in our PGY 1 year in pediatrics.

Conclusions: Reliability, validity, feasibility and educational impact will be measured to assess the tool.

Take-home Messages: The mini-MAS will help faculty assess key competencies and milestones to assess resident progress through a competency-based curriculum.

OTT-PB5-11
Needs assessment of teachers conducting workplace based assessments of professional behavior

Hester Daelmans (*VU University medical center, Amsterdam, Netherlands)*, Rashmi Kusurkar (*VU University medical center, Amsterdam*), Jean Bronzwaer (*VU University medical center, Amsterdam*), Martina Cornel (VU University medical center*, Amsterdam*),
Ariadne Meiboom (*VU University medical center, Amsterdam*), Gerda Croiset (*VU University medical center, Amsterdam*)

Background: Assessment of professional behavior is embedded in the assessment of our medical school (VU University Medical Center) and in the clerkships executed by doctors who have a teaching function (further called teachers). They received information and guidelines for assessment of professional behavior.

Summary of Work: A team of 4 persons visited the peripheral hospitals (n=13), where the student clerkships are arranged, to address the experiences with workplace based assessments and especially with professional behavior. These visits were executed as discussions requesting specific inputs from teachers and education coordinators in the hospitals. Reports of the discussions were analyzed for recurring themes reported below.

Summary of Results: Teachers reported specific difficulties in assessment of professional behavior of students. They strongly stated that they needed information about the students' professional behavior in earlier clerkships to be able to guide students effectively and evaluate them fairly in their own clerkship. This is at odds with current practice. The teachers also wanted information on the remediation trajectory after an unprofessional behavior judgment. They wanted stricter action for students with a judgment of unprofessional behavior, like not being allowed to continue in the next clerkship. Teachers evaluating clerkships of shorter durations (4 weeks) found the available time too short to be able to make fair judgments of unprofessional behavior.

Conclusions: Providing guidelines and an occasional training is not sufficient to master assessment of professional behavior of students.

Take-home Messages: Teachers need longitudinal support for enabling them to evaluate professional behavior of students in the workplace. The availability of need-based expert consultation aids this process.

OTT-PB5-12
Residents' competency assessment by Direct Observation of Procedural Skills and Global Rating Scale: Its Correlation and Agreement

Maysam Asadilari (*Tehran University of Medical Sciences, Emergency Medicine, Tehran, Iran)*, Amir Nejati (*Tehran University of Medical Sciences, Emergency Medicine, Tehran*), Mohammad Jalili (Tehran University of Medical Sciences*, Emergency Medicine, Tehran*), Mahvash Alizadeh Naini (*Shiraz University of Medical Sciences, Internal Medicine, Shiraz*), Ali Ardalan (*Tehran University of Medical Sciences, Emergency Medicine, Tehran*)

Background: Core competency achievement is the main aim of each specialist training course. Assessment methods should be able to predict future specialist performance in a good manner; to be valid, reliable, feasible and efficient.

For competency based evaluation, observation of trainees' performing activities is the best.
The role of diagnostic & therapeutic procedural technical skills is clear in patients' care. Direct Observation of performing Procedural Skills (DOPs). Another method is GRs, which rates trainees' performance globally.

Summary of Work: To assess the resident's procedural skills by two methods (GRs & DOPs) and evaluate its agreement.
All post graduate (n=41) EM residents of Imam Khomeini hospital in TUMS completed 4 DOPS and 4 GRs assessment, the scores of each procedure, each level and totally were analyzed and their agreement was assessed.

Summary of Results: 143 DOPS Scores (85% of expected scores were collected). The Global rating scores of residents who participated in EM rotation 2 montsh each trimester (n=85) was analyzed. Nearly all residents achieved satisfactory scores by both tools. Reliability of both methods was statistically significant(alpha Cronbach 0.7). The inter method agreement assessed by Bland Altman method was acceptable as statistically significant>95%between 2 SD.

Conclusions: Both methods were feasible for skills assessment. Both were reliable but both had some limitation. Although it was more difficult to implement DOPs but missing was seen more by GRs.

Take-home Messages: DOPs & GRs are two known good WPBA methods. Both had some limitations, so it is advisable to implement both methods in core competency assessments.

OTT-PB5-13
Use of Shift Card as a Learning Tool in the Emergency Department

BYG Leong (*Changi General Hospital, Emergency Department, Singapore)*, GG Sim (*Changi General Hospital, Emergency Department, Singapore*)

Background: Evaluation and feedback on trainee performance during shifts are integral parts of resident development and learning. Traditional methods of informal debrief at the end of each case/shift was found to be inconsistent and inadequate for training purposes. Thus, our Emergency Department started a "shift card" system in an attempt to formalize and standardize evaluation and feedback on trainee performance and to enhance learning experience.

Summary of Work: All residents were required to fill up minimum 2 shift cards per month comprising of 2 components – resident self-evaluation and faculty evaluation. Residents were to identify what went well during the shift and what can be improved upon; faculty was required to choose a specific learning issue to discuss. After introducing this for 7 months, residents' views were polled in an anonymous survey.

Summary of Results: 86% felt shift cards help to point out deficits while 86% felt shift cards help in identifying and building on their strengths. Majority agreed minimum 2 shift cards per month were adequate, however views were split on shift card being a good

learning tool overall, citing difficulty implementing on the ground.

Conclusions: As a concept the shift card is helpful in pin-pointing deficiencies and strengths but translating it to a practical learning tool will require more work to develop it to meet all expectations.

Take-home Messages: Moving forward, a faculty survey to assess the impact of the shift card on residency learning and feedback process should be done.

OTT-PB5-14
The development and training of mini-CEX central assessors to evaluate students in undergraduate medical school

Wyn Harris (*Swansea University, College of Medicine, Swansea, UK)*, Steven Capey (*Swansea University, College of Medicine, Swansea*), Jess Murtagh (*Swansea University, College of Medicine, Swansea*)

Background: Workplace based assessments have been implemented in postgraduate medical education as formative assessment tools for many years. We have implemented a assessment model that utilises clinicians, both placement assessors and central assessors, to deliver a reliable and valid assessment which is used as a summative assessment for undergraduate medical students.

Summary of Work: We evaluated the training of the assessors utilising the mini-CEX assessment as a summative tool in undergraduate assessment. We ran training sessions for central assessors supported by web-based instructional material. Central assessors were also involved in the development of the mini-CEX documentation through a number of training seminars.

Summary of Results: The training in the use of the mini-CEX was evaluated as extremely useful by the central assessors. There was good agreement between central assessors about the relative weighting of the different domains within the mini-CEX assessment. An interactive booking system was developed to enable students placed in different geographical areas to book appointments with central assessors.

Conclusions: The mini-CEX training and booking system has demonstrated that it is possible to deliver an effective and reliable summative mini-CEX assessment in a geographically distributed undergraduate setting.

Take-home Messages: It is possible to deliver an effective and valid assessment of undergraduate medical students using the mini-CEX.

OTT-PB5-15
Queen's Ob/Gyn clerkship ambulatory clinic rubric: making the focus of assessment and standard for performance explicit

Susan Chamberlain (*Queen's University, Obstetrics and Gynecology, Kingston, Canada)*, Laura McEwen (*Queen's University*), Ulemu Luhanga (*Queen's University*), Richard Thomas (Queen's University*, Obstetrics and Gynecology*)

Ottawa 2014 Abstract Book

Background: Assessment of clinical clerks in the ambulatory clinic has been done using a daily encounter card. Faculty assessed clerks' Knowledge, Clinical Assessment, Fundamental skills (communications, record-keeping, technical skills, problem-solving) and Attitudes on a 4-point rating scale. Known problems with these encounter cards include leniency bias, halo effect, confusion about the focus of assessment, and a lack of agreed upon standards for performance.

Summary of Work: A rubric was developed that took into consideration learning objectives for ambulatory Obstetrics and Gynecology. Items were separated into discrete observable competencies (attributes). Three levels of performance were defined for each attribute to clarify expectations for faculty and students. Consultation with faculty ensured agreed upon standards and the feasibility of direct observation for defined attributes. Faculty were asked to complete a rubric for every ambulatory clinic. Clerks were required to self-assess at entry, mid-point, and end of rotation.

Summary of Results: An evaluation of this pilot project is currently underway to explore ease of use, quality of feedback, consensus between faculty and clerks' self-assessment, and the utility of assessment data. On-line surveys will be administered to faculty and students to elicit perceptions about ease of use and quality of feedback. Statistical analysis of faculty and clerk generated data will be conducted to examine consensus between groups. Finally, an interview with the course director will explore the utility of assessment data generated.

Conclusions: in process

Sunday April 27
3:30PM – 5:00PM

SYMPOSIA

OTT-SC-1
What is best practice in the selection of medical students?

Presenters: Jennifer Cleland (*University of Aberdeen, Aberdeen*), Sandra Nicholson (*Barts and the London, London*), Fiona Patterson (*Cambridge University, Cambridge*), Jonathan Dowell (*University of Dundee, Dundee*)

Summary: Selection can be seen as the first assessment in the medical education and training pathway. Admission to medical school has traditionally used educational attainment as a primarily hurdle, increasingly in conjunction with aptitude test of some sort. Non-academic abilities are then usually considered by interview and/or other sources such as personal statements or even personality tests. However, these approaches have been criticised heavily on the basis of poor reliability as well as dubious validity and it is also clear they are not infallible: with regulators concerned about some of those entering medicine. And rarely is the major influence of self-selection considered. This symposium will explore the question "What is best practice in the selection of medical students?" from a number of angles including "evidential weight" and supporting Widening Participation. Emerging selection tools and their evidence-base will be reviewed, including Situational Judgement tests and Multiple Mini Interviews. Finally, the view of the regulator will be considered.

OTT-SC-2
Exploring Rater Cognition in Workplace-Based Assessment from Three Different Research Perspectives

Presenters: Eric Holmboe (*American Board of Internal Medicine, Philadelphia)*, Andrea Gingerich (*Northern Medical Program (UBC Medicine), BC*), Jennifer Kogan (*Perelman School of Medicine, University of Pennsylvania*), Peter Yeates (*University of Manchester, Manchester*), Marjan Govaerts *(Maastricht University, Netherlands)*

Summary: Workplace-based assessments are an integral part of our assessment systems. In efforts to improve the defensibility of assessment decisions and our accountability to patient safety, researchers have begun investigating raters' cognitive processes. Although a relatively new domain of inquiry, there appear to be three distinct (though not exclusive) perspectives on rater cognition. One considers raters' cognitive processing to be conscious and controllable, and seeks tangible training solutions. A second

acknowledges the automatic and unavoidable limitations of human cognition and will ultimately seek to provide design solutions to minimize such weaknesses. The third casts the rater as a valuable source of information whose expertise is squandered in current practices but could be harnessed in radically different assessment approaches. This symposium features a group of international rater cognition researchers representing the current understanding of rater cognition. We see this symposium as an important tool for stimulating a discussion about prevailing assumptions and conceptual gaps as well as potential implications for improving assessment.

WORKSHOPS

OTT-WC-1
Using Rubrics to Increase Reliability and Decrease Anxiety for Trainees and Faculty

Presenter(s): Susanna Talarico (*Hospital for Sick Children, University of Toronto, Pediatrics, Toronto, Ontario, Canada*), Shirley Yee (*Mount Sinai Hospital, University of Toronto, Family and Community Medicine*), Helen Batty (*Women's College Hospital, University of Toronto, Family and Community Medicine*)

Background: The use of portfolios in medical education follows a move to competency-based education and the formal emphasis on non-medical expert roles. This move occurred without a parallel shift in assessment tools that formatively facilitate the development of competence and summatively assess it. The portfolio is considered to be this instrument (Driessen et al 2007). However, the challenge remains in how to assess portfolios in an authentic manner, which considers the context.
Intended Outcomes: This workshop will describe the evidence behind portfolios as valid and reliable tools to assess the development of competence in medical education. It will introduce the rubric as a tool that increases the reliability of portfolio assessments by incorporating "standardized narratives", or descriptors of various criteria and levels of competence (Regehr et al 2012). The development and use of a rubric to assess resident and faculty teaching and reflective portfolios, both based in a constructivist epistemology, will be described.
Structure: During this interactive workshop, participants will have the opportunity to practice using a rubric to assess both electronic and paper portfolios and create their own rubric as an assessment tool for their program.
Who Should Attend: Educators involved in undergraduate or postgraduate education, faculty development or curriculum development are invited and encouraged to attend
Level of Workshop: Intermediate

OTT-WC-2

Validating entrustable professional activities for the assessment of early medical students

Presenter(s): H. Carrie Chen (*University of California San Francisco, Pediatrics, San Francisco, United States*), Margaret McNamara (*University of California San Francisco, Pediatrics, San Francisco, United States*)

Background: Despite growing emphasis on early clinical experiences, preceptors continue to have difficulty integrating and assessing early students in their practices. Entrustable professional activities (EPAs), which frame competencies in the context of workplace activities, may provide more explicit guidance for students' clinical roles and activities and better alignment with how preceptors assess performance. We developed five EPAs for preclerkship students using medical student focus groups, preceptor interviews, preclerkship course objectives, graduation competencies, and available resident-level EPAs. This workshop is designed to engage participants in the validation process of EPAs for early medical students.
Intended Outcomes: Upon completion of this workshop, participants will be able to:
• Identify characteristics of an entrustable professional activity (EPA)
• Describe potential EPAs for preclerkship medical students
• Describe an assessment scale for levels of entrustment/supervision
• Determine expected levels of entrustment/supervision for entering clerkship students
Structure: 1) Large group (15 min):
• Introduction to workshop
• Background on EPAs and local EPA project
• Introduction to proposed preclerkship student EPAs
2) Small group break-out – address content validity of proposed EPAs (25 min):
• Each small group reviews one EPA
• Critique and discussion of EPA
3) Large group (10 min):
• Small group debrief
• Introduction to assessment scale for levels of entrustment/supervision
4) Small group break-out – address response format validity of proposed assessment scale (20 min):
• Application of scale to group's EPA
• Determine expected level of entrustment/supervision of EPA at beginning of clerkship
5) Large group (15 min):
• Small group reports
• Summary and wrap-up
Who Should Attend: Medical student educators, directors of clinical skills courses and clerkships
Level of Workshop: Intermediate

OTT-WC-3
The Uses and Misuses of Cronbach's Alpha with OSCEs: How to Estimate Reliability Appropriately with Generalizability Theory

Presenter(s): Andre De Champlain (*Medical Council of Canada, Research & Development, Ottawa, Canada*), Andrea Gotzmann (*Medical Council of Canada, Research & Development, Ottawa, Canada*)

Background: Reliability is a central psychometric property that needs to be assessed with all OSCEs. Given its inclusion in common statistical packages, Cronbach's coefficient alpha is the most popular reliability index. However, Cronbach's alpha cannot provide an appropriate description of the quality of our OSCE scores given that measurement error, i.e., the sources that detract from reliability, is multifaceted. That is, Cronbach's alpha does not allow the quantification of all sources of measurement error nor does it permit the correct calculation of reliability for OSCEs. It can only account for one omnibus source of measurement error. Generalizability theory does, however, enable the computation of an appropriate reliability estimate for OSCE scores and also allows the researcher to gauge the importance of all sources of measurement error in the examination. This can provide useful information not only for scoring purposes but also for training of examiners and other test development issues.
Intended Outcomes: The aim of this workshop is to outline the pitfalls of Cronbach's alpha with common OSCE data sets and illustrate how generalizability theory can properly address these limitations
Structure: Part 1: Overview of classical test theory reliability and limitations for OSCE scores.
Part 2: Outline of G-theory and how it can address these limitations.
Who Should Attend: Medical educators who may be responsible for OSCEs at their institutions would benefit from this workshop. Some knowledge of reliability and assessment is needed.
Level of Workshop: Intermediate

OTT-WC-4
Implementing best practices for in training assessment - establishing and implementing institution wide improvements

Presenter(s): Susan Glover Takahashi (*University of Toronto, Postgraduate Medical Education, Toronto, Canada*), Marla Nayer (*University of Toronto, Postgraduate Medical Education, Toronto, Canada*)

Background: In-training assessment is the foundation of the workplace educational model of clinical medical education. A good system relies not only on principled design of assessment instruments but also on the local context in which they are deployed. Numerous psychometric, logistic and socio-cultural challenges impede successful implementation. The literature provides direction in all of these areas. This workshop outlines a comprehensive evidence-informed approach to developing best practices for in-training assessment with a focus on institutional change management and broad uptake.
Intended Outcomes: Participants will be able to: a) Identify key issues that can undermine workplace-based

in-training assessment and the literature base supporting these b) Systematically design a process for developing and implementing an effective workplace in-training assessment system, and c) Describe various models for monitoring and oversight of such a system.
Structure: The workshop will involve a brief review of the literature around in-training assessment, including original research from the local context of the presenters. Participants will then work in small groups to review case studies showing how to integrate evidence with local context to design an in-training assessment system. The workshop will finish with a general discussion of principles, questions and declared next steps for participants interested in changing assessment systems in their local environment.
Who Should Attend: Who Should Attend: Program directors, education coordinators, decanal members, and other education leaders interested in improving intraining assessment strategies.
Level of Workshop: Intermediate

OTT-WC-5
Learning to train SPs to simulate acute life-threatening medical problems for teaching and assessment

Presenter(s): James Parle (*University of Birmingham, Health and Population Sciences, Birmingham, United Kingdom*)

Background: Clinical students rarely have the opportunity to examine patients with severe acute illness, and are never examined on such patients. This sends a clear although hidden message that faculty do not expect students to acquire experience seeing patients with such clinical problems nor competence in their assessment, especially examination. This is dissonant with our expectations of them on qualification. We propose therefore that such clinical scenarios need to be simulated (as is already done routinely with cardiac arrests) and that such simulations can be sufficiently realistic with the use of experienced trained actors and simulation of symptoms and signs, the latter using both make-up and acting skills.
Intended Outcomes: By the end of this workshop participants will understand the mechanics of designing a complex simulation with backstory, current presentation and simulation of physical signs.
They will:
Experience (by 'role playing learners') a complex human simulation ('one we prepared earlier')
Learn how to develop a new simulation with scenario and how to train an SP to simulate signs for teaching purposes
Structure: After brief introductions the workshop leader will 'teach' using an SP who has been trained to simulate physical signs of an acute severe life-threatening illness, with a backstory and presenting symptoms. Participants will then develop a new scenario and take part in the training of simulation of signs.

Who Should Attend: All teachers of clinical scenarios, but particularly those who work with human based simulation
Level of Workshop: Introductory

OTT-WC-7
"What we hope ever to do with ease, we must learn first to do with diligence": measuring conscientiousness in health care settings

Presenter(s): Marina Sawdon (*Durham University, School of Medicine, Pharmacy & Health, Cleveland, United Kingdom*), Andrew Chaytor (*Durham University, School of Medicine, Pharmacy & Health, Cleveland, United Kingdom*), Gabrielle Finn (*Durham University, School of Medicine, Pharmacy & Health, Cleveland, United Kingdom*), John McLachlan (*Durham University, School of Medicine, Pharmacy & Health, Cleveland, United Kingdom*)

Background: Robust measures of professionalism continue to be elusive. This may be attributed to the complex nature of professionalism, which goes beyond the application of knowledge and skills to encompass humanism, accountability, altruism and the pursuit of excellence.
There is evidence that concerns about undergraduate performance are a risk factor for subsequent disciplinary action. We have previously developed an objective, scalar measure of diligence or conscientiousness; a proxy measure of the trait of professionalism, which is being implemented in health care settings both nationally and internationally. Conscientiousness measures have been shown to be valid and reliable, and cost effective, and correlate to staff and student views of professionalism, as well as academic performance.
Intended Outcomes: To enable colleagues to develop and implement a conscientiousness measure suitable to their environment, and demonstrate how to subsequently evaluate its reliability and validity, as well as how to develop a related cohort study in their institution.
Structure: The workshop will start with a brief introduction to the data in settings such as medical undergraduate students, foundation level undergraduate students, anaesthetic trainees, paramedics and podiatrists. Participants will be provided with exemplars of conscientiousness measures and divided into discipline groups to develop conscientiousness instruments suitable to their own particular health care setting. As a group we will consider these individual instruments, and then explore the necessary data collection required to determine validity and reliability, and develop a cohort study.
Who Should Attend: Colleagues interested in measuring professionalism in a variety of health settings
Level of Workshop: Intermediate

ORALS

OTT-OC-1: PEER AND SELF ASSESSMENT

OTT-OC-1-1
Redesigning peer assessment with experienced physician assessors

Rhoda Reardon (*College of Physicians and Surgeons of Ontario, Research and Evaluation, Toronto, Canada*), Craig Nathanson (*College of Physicians and Surgeons of Ontario, Research and Evaluation, Toronto*), Nanci Harris (*College of Physicians and Surgeons of Ontario, Research and Evaluation, Toronto*)

Background: CPSO has a 30 year history of in-situ physician assessments (chart review, physician interview) purposed to promote continuous quality improvement by providing feedback to validate appropriate care and show improvement opportunities. Program revision was initiated to create an assessment process that used a common framework but allowed for discipline specificity. The primary goals of the revision are to:
• respond to evaluator feedback to make assessments more discipline-specific
• better align peer assessment with its purpose
• improve consistency in decision-making.
Summary of Work: Eight diverse medical specialities were selected to begin the revision process. Small subgroups of evaluators from each discipline worked independently to propose discipline-specific assessment topics and define the corresponding elements of quality to guide assessment decisions.
Summary of Results: The groups' independent outputs were analysed to determine commonality. Eight common topic areas emerged and were grouped using the SOAP framework: Subjective (History), Objective (Examination, Investigation), Assessment (Diagnosis), Plan (Management, Medication, Follow-up & Monitoring, Documentation & Communication). A common, four-point scale was agreed upon that captures need for improvement opportunities by topic. Using a consensus-building approach, assessor guides of discipline-specific elements of quality were developed for each topic.
Conclusions: The output of the eight subgroups was consolidated into a common tool aimed at promoting quality improvement that is usable across disciplines while maintaining discipline specificity. Creation of such a tool requires collaboration with experienced, physician peer evaluators across medical disciplines.
Take-home Messages: Building an effective quality improvement peer assessment process requires on-going, collaborative work with practicing physicians.

OTT-OC-1-2
Validity of Peer Assessment in Team Based Learning

Asty Amalia (*Faculty of Medicine Hasanuddin University, Medical Education, Makassar, Indonesia)*, Irwin Aras (*Faculty of Medicine Hasanuddin University, Medical Education, Makassar*), Alfian Zainuddin (*Faculty of Medicine Hasanuddin University, Public Health, Makassar*)

Background: Team-based learning (TBL) is a learning strategy that can deepen students' learning and promote the development of high-performance learning teams. It enhances students' problem-solving skills and reduces lecture time.
One important step in TBL is to ensure that members contribute time and effort to group work and peer-assessment. This is essential because team members are typically the only ones who have enough information to accurately assess one another's contributions. The problem related with peer assessment is the students' ability to evaluate their peers objectively and that will influence the validity. This research is performed to find out predictive validity of peer assessment in TBL compared to the final test.
Summary of Work: This study is performed on 203 first year students who are undertaking Immunology and Hematology block. The students were asked to assess their peers from one group by ranking them based on their contribution to group work. The results were analyzed and compared to the final test result using SPSS 17.
Summary of Results: From the analysis using linear regression method, it was found that the result of the peer assessment has a positive linear correlation with final test. The peer assessment then has been proved to be valid as one assessment tool.
Conclusions: Peer assessment using ranking system has predictive validity and can be used as one component assessment in TBL.
Take-home Messages: Students should be prepared and have orientation on peer assessment and technique to assess their peers. Further research on other peer assessment methods should be performed.

OTT-OC-1-4
Culturally competent healthcare in Emergency Medicine: An Australasian approach

Holly Donaldson (*ACEM, Continuing Professional Development, Melbourne, Australia)*, Alyssa Vass (*ACEM, Continuing Professional Development, Melbourne*), Natalie Johnson (*ACEM, Continuing Professional Development, Melbourne*), Mary Lawson (ACEM*, Education, Melbourne*)

Background: Using culturally competent and patient-centred care have been promoted as means to improve healthcare delivery and improve patient safety. Achieving these goals can be particularly challenging in the frenetic and noisy environment of the emergency department (ED).
Summary of Work: The Australasian College for Emergency Medicine (ACEM) has created a suite of educational resources targeting knowledge and skill acquisition of cultural competency principles to improve care delivered to Indigenous Australians, New Zealanders and other culturally and linguistically diverse

patients. Resources have been developed in collaboration with a diverse group of experts to ensure their cultural appropriateness, accuracy and relevance to the ED. The project was formed in late 2012 through funding by the Australian federal government. Key deliverables include ten self-paced, self-assessed online modules and a podcast series highlighting case studies, research and initiatives occurring in this arena. Recognising cultural competence (CC) as a key attribute of an emergency medicine specialist, has enabled learning outcomes to be embedded across the range of curriculum from training through to continuing professional development.

Summary of Results: Online self-directed modules address multiple domains of the ACEM curriculum framework including health advocacy, professionalism, communication, team work and collaboration.

Conclusions: A vertically integrated approach to CC has been developed demonstrating ACEM's commitment to promoting high standards of care for all Australasians in ED. CC will be a component of our revised accreditation standards requiring a systematic approach to teaching and assessing in this area to be developed.

Take-home Messages: High quality care in the Emergency setting can be supported by providing an holistic approach to cultural competence.

OTT-OC-1-5
Queen's multisource feedback rubric: Frames of reference to enhance self-assessment

Jane Griffiths (*Queen's University, Family Medicine, Kingston)*, Karen Schultz, Laura McEwen, <u>Ulemu Luhanga</u> (*Queen's University, Faculty of Education, Kingston, Canada*)

Background: The assumption that learners are less than accurate self-assessors has been around for a long time (see Kruger & Dunning 1999). These ideas endure in spite of questions raised about the manner in which this topic has been researched (see Eva & Regehr, 2005). Pangaro and ten Cate (2013) advocate the need for assessment frameworks. They hypothesize these to function as frames of reference that can be used to benchmark relevant facets of performance to be taken into account and define superior, adequate and unacceptable performance.

Summary of Work: A group at Queen's School of Medicine developed a multisource feedback (MSF) rubric designed to collect information about residents' performance in the four CanMEDs domains of Communicator, Collaborator, Manager, and Professional. Unacceptable, adequate, and superior performance levels were explicitly described for facets of performance across the four domains. Preliminary MSF rubric data were collected for 1st year Family Medicine residents between April and June 2013. Administration, Allied Health Professionals, Nurses, Peers, and Physicians assessed resident performance with electronically distributed MSF rubrics. Residents completed the MSF rubric as a self-assessment prior to viewing assessment results.

Summary of Results: Wilcoxon Signed Ranks tests were conducted to examine how self-assessments compared with raters. 232 MSF rubrics were completed for a subset of 36 residents. 60 - 85% of residents achieved consensus with their various rater groups. Consensus was defined as selecting the same pattern of performance across all rubric items. The highest rate of consensus was between residents and nurses.

Conclusions: Self-assessment accuracy was positively influenced by the provision of a frame of reference in the form of a rubric.

Take-home Messages: • Rubrics provide a valuable frame of reference to guide the focus of assessment and make standards for performance explicit.
• Explicit standards for performance appear to support learners to more accurately self-assess.

OTT-OC-2: SIMULATORS AND VIRTUAL PATIENTS

OTT-OC-2-1
Establishing validity from multiple evidence sources for a simulation-based assessment of procedural skills

<u>Rodrigo B. Cavalcanti</u> (*University of Toronto, Department of Medicine, Toronto, ON, Canada)*, Ryan Brydges (*University of Toronto, Department of Medicine, Toronto, ON*), Laura Naismith (*Toronto Western Hospital, Toronto, ON*), Graham Slaughter (University of Toronto*, Department of Medicine, Toronto, ON*), Khalil Sivjee (*University of Toronto, Department of Medicine, Toronto, ON*), Lynfa Stroud (*University of Toronto, Department of Medicine, Toronto, ON*)

Background: Simulation-based assessments of procedural skills are increasingly common because they provide a reproducible platform that safeguards patient safety. To date, however, validity evidence for the scores of simulation-based assessments and associated decisions (e.g., pass/fail) have been sparsely collected. We sought to collect, organize and interpret the validity evidence for a structured simulation-based assessment of procedural skill competence.

Summary of Work: Second-year internal medicine residents' clinical judgement and procedural skills in ultrasound-guided paracentesis and thoracentesis were assessed using task trainers. From examiners (N=17) and residents (N=43), we collected four sources of evidence: content, response process, internal structure, and relations to other variables.

Summary of Results: Content: Residents and examiners rated the expert-designed simulations as accurate predictors of clinical performance (MR=3.61; ME=3.94; 3=neutral; 5=strongly agree). Response process: Residents rated their own performance as representing clinical performance (M=3.29; 3=somewhat; 5=exactly). Internal structure: Inter-rater reliability (intra-class correlation coefficient) ranged from .78 (checklist) to .92 (global ratings). Relations to other variables: Prior experience with ultrasound-guided paracentesis

correlated with performance, for both checklist, rs(41)=.32, p=.04, and global ratings, rs(27)=.43, p=.02. Also, checklist scores and global ratings correlated for both procedures (rs-PARACENTESIS(27) =.56, p=.001; rsTHORACENTESIS(12) =.71, p=.004).

Conclusions: Multiple sources of evidence support this exercise as a valid assessment of clinical competence in procedural skills. Further study will examine consequences, including training and retesting for underperforming residents.

Take-home Messages: We used a validity framework to collect and organize multiple sources of evidence supporting our inferences of competence in procedural skills measured by simulation-based assessment.

OTT-OC-2-2
The Development and Validation of Simulation-Based Resuscitation OSCE Stations for Emergency Medicine Trainees

Andrew K. Hall (*Queen's University, Emergency Medicine, Kingston, Canada)*, J. Damon Dagnone (*Queen's University, Emergency Medicine, Kingston*), Lauren Lacroix (*Queen's University, Emergency Medicine, Kingston*), William Pickett (Queen's University*, Emergency Medicine, Kingston*)

Background: The use of high-fidelity simulation is emerging as an effective method for competency-based assessment in postgraduate medical education. We aimed to develop and validate simulation-based Objective Structured Clinical Examination (OSCE) stations to assess post-graduate emergency medicine trainee competency in resuscitation.

Summary of Work: An expert panel of emergency physicians developed and revised 10 OSCE stations including the scenario and corresponding assessment tool. Assessment tools comprised four domain-specific anchored scores (primary assessment, diagnostic work-up, therapeutic action, and communication) and one global assessment score. Trainee performances were independently reviewed by three trained clinician experts blinded to the residents' identity and level of training. Assessment correlation and variance were calculated.

Summary of Results: Emergency Medicine postgraduate trainees (FRCP & CCFP-EM) from Queen's University (N = 19-26 per OSCE station) participated in the study. Inter-rater reliability showed substantial agreement for each scenario (mean Spearman's rho = 0.75 [0.63-0.87]). Discriminatory validity was also strong, with Senior FRCP-EM residents outperforming Junior FRCP-EM and CCFP-EM residents in all 10 scenarios (p=<0.001-0.11).

Conclusions: This study demonstrates the development and validation of 10 OSCE stations using high-fidelity mannequins for the evaluation of resuscitation competence in post-graduate emergency medicine trainees. The assessment tools have strong inter-rater reliability and discriminatory capabilities.

Take-home Messages: We have developed and evaluated 10 simulation-based OSCE stations utilizing domain-specific and anchored global assessment scoring as tools to assess post-graduate emergency medicine trainee competency in resuscitation. These assessment tools have strong inter-rater reliability, discriminatory capabilities, and have distinct potential for high-stakes competency-based assessment.

OTT-OC-2-3
Development of a new tool to assess clinical performance of anesthesiologists: The structured oral interview with high definition simulator

Pierre Champagne (*Collège des médecins du Québec, Practice enhancement division, Montréal, Canada)*, Marguerite Dupré (*Collège des médecins du Québec, Practice enhancement division, Montréal*), François Goulet (*Collège des médecins du Québec, Practice enhancement division, Montréal*), Johanne Thiffault (Collège des médecins du Québec*, Practice enhancement division, Montréal*)

Background: The Collège des médecins du Québec (CMQ) is the medical licensing authority in Québec whose mission is to monitor the quality of practice of all Quebec physicians. In 2006, the coroner recommended that the CMQ evaluate more specifically the practicing anesthesiologists' capacity to manage emergencies.

Summary of Work: Since during direct observation emergencies seldom occur, we opted for a structured oral interview (SOI), a case-based assessment tool where cases are presented according to the key features approach. We also use hands-on anesthesia high definition simulator that allows presentation of similar standardized crisis situations to anesthesiologists under evaluation. Actually, we present to the assessed practicing physician 12 cases that include 5 to 6 high definition simulation cases. A secondary benefit of the anesthesia simulation context is that it allows the evaluators to assess the communication skills of anesthesiologists and their technical skills.

Summary of Results: The results of a pilot study to test the simulator with 2 practicing anesthesiologists will be presented. We will also present the results of all candidates submitted to the SOI with to high definition simulation since 2011 (n=9). Results indicate that the SOI in anesthesiology is a valid and reliable tool, albeit costly, to assess practicing anesthesiologists' performance and management of critical incidents. Cases where the patient is in need of emergency management are the cases most often failed, especially by older practicing anesthesiologists working in a low-risk environment.

Conclusions: The SOI with the use of high definition simulator to assess the clinical performance of practicing anesthesiologists is valid, reliable and useful to determine the remedial programs best adapted to their educational needs. However, cost remains an important consideration.

Take-home Messages: High definition simulation is a promising tool to assess the clinical performance of anesthesiologists.

OTT-OC-2-4

Integrated assessment of basic science knowledge and clinical skills: CardioSim

Robin DeMuth (*Michigan State University College of Human Medicine, Family Medicine, East Lansing, MI, USA)*, Dianne Wagner (*Michigan State University College of Human Medicine, Medicine, East Lansing, MI*)

Background: Our College's second-year curriculum has ten problem-based learning (PBL) domains. We wanted to better integrate the students' clinical skills and basic science learning. CS already required the students to demonstrate competency in a focused encounter with a patient having cardiac complaints. The Cardiovascular PBL domain wanted to increase formative feedback to students and to place basic science within a patient-care context.

Summary of Work: We created a joint objective structured clinical evaluation between the cardiovascular PBL domain and clinical skills course, "CardioSim," to assess student's necessary science knowledge and clinical skills.

Communication skills, history-taking, focused physical examination in two of the stations and creation of a progress note served as a summative assessment of CS competencies. Students answered relevant basic science questions (pathophysiology, pathology, pharmacology) after each case and accessed correct answers to questions in preparation for their final domain examination.

Summary of Results: When asked at the end of the two-year clinical skills sequence "what type of experience would you like more of?" the majority of students answered "CardioSim." Overall satisfaction with the PBL course improved.

Conclusions: Formative and summative assessment can be successfully combined without student confusion. Integration between clinical skills and basic science competencies is well received by students.

Take-home Messages: Integrating formative with summative assessment is feasible. Assessment across clinical skills and basic science competencies is also feasible and highly desired by students.

OTT-OC-2-5
High fidelity and virtual patient simulation as mixed method needs assessment tools

Aimee Sarti (*The Ottawa Hospital, Critical Care, Ottawa, Canada)*, Angele Landriault (*Royal College of Physicians and Surgeons of Canada, Practice, Performance and Innovations unit, Ottawa*), Stephanie Sutherland (*University of Ottawa, Academy for Innovation in Medical Education, Ottawa*), John Kim (The Ottawa Hospital*, Critical Care, Ottawa*), Susan Brien (*Royal College of Physicians and Surgeons of Canada, Practice, Performance and Innovations unit, Ottawa*), Pierre Cardinal (*The Ottawa Hospital, Critical Care, Ottawa*)

Background: As part of a variety of data collection strategies for a mixed methods critical care needs assessment, insitu, high fidelity simulations (HFS) and virtual patient simulations (VPS) were conducted to explore whether qualitative data collected from the debrief session combined with quantitative performance data could provide more in-depth exploration of critical care resuscitation in a community hospital context.

Summary of Work: Teams of practitioners completed two 10-minute HFS cases and two VPS cases. The interprofessional simulations were recorded then scored using the Ottawa Global Rating Scale and the Team Emergency Assessment Measure. Debrief sessions were recorded, transcribed and analyzed using an inductive coding technique.

Summary of Results: Simulations provided unique data and debriefings permitted clarification of observed performance gaps and corroborated findings from other data sources. Analyses of team performance objectively identified three primary issues needing improvement: Use of personal protective equipment, overall crisis resource management, and treatment of shock. Qualitative data contributed to 15 of the 18 themes that emerged from all data sources, and elicited information about the community context, how patients navigate the system at different times of day, the roles and responsibilities of various professionals, and thought processes underlying behaviours during the simulations.

Conclusions: Simulations and debriefings can be useful for identifying perceived and unperceived educational needs and eliciting information about a clinical context and system function. Simulations and debriefings may be suitable substitutes to the typical focus group.

Take-home Messages: Simulations can be used as part of a needs assessment and may be a suitable substitute to the typical focus group.

OTT-OC-2-6
Assessing progression of clinical reasoning through Virtual Patients

Elenita Forsberg (*Halmstad University, School of Social and Health Sciences, Halmstad, Sweden)*, Kristina Ziegert (*Halmstad University, School of Social and Health Sciences, Halmstad*), Håkan Hult (*Karolinska Institutet, Clinical Sciences, Intervention and Technology, Stockholm*), Uno Fors (*Stockholm University, Computer and Systems Sciences, Stockholm*)

Background: There have been discussions to use more formative assessments in health care education to contribute to students' deep learning. Feedback is important, but a lot of student's do not do anything with it. Thus, interventions which force students to reflect the new knowledge need to be introduced.

In order to explore if formative VP- based exams had an impact on development of clinical reasoning ability and achievement of clinical decision making, we let postgraduate pediatric nurse students complete self-evaluation forms in connection with three VP- based assessments to follow their progress.

Summary of Work: After performed assessment and before answering the self-evaluation form, the students' were asked to take part of the feedback section of the VP-system and the recommended interactions in the VP system. Data was conducted using content analysis with

a deductive approach. Kolb's' model of Learning Cycle guided the analysis.

Summary of Results: The result showed a perceived progression of clinical reasoning skills by the students. After the first assessment the students described uncertainty and that knowledge gaps were exposed, at the second exam the awareness of clinical reasoning was obvious and the students were more certain of knowing how. Finally, self-efficacy in patient solving was expressed.

Conclusions: VP-based assessments with self-evaluation early in the education resulted in a gain of students' own identification of the concept of clinical reasoning, awareness of what to focus on, and pay attention to during clinical practice.

Take-home Messages: VP with reflective tools is excellent to use in formative assessments to identify progress and to visualize the expected clinical competence.

OTT-OC-3: WRITTEN ASSESSMENT 2

OTT-OC-3-1
Insight, foresight and risk-taking: response analysis from an MCQ

Mike Tweed (*University of Otago Wellington, Medicine, Wellington, New Zealand)*, Tim Wilkinson (*University of Otago Christchurch, Medical Education, Christchurch*), Sarah Stein (*University of Otago, Higher Education Development Centre, Dunedin*), Jeff Smith (University of Otago*, Education, Dunedin*)

Background: Insight (estimated by appropriate certainty in MCQ responses), foresight (estimated by safeness of incorrect responses) and their interaction, and risk-taking (estimated from responses to MCQs with correct option removed) can provide information about students.

Summary of Work: 89 undergraduate medical students sat an optional 75 MCQ examination covering core presentations and conditions. Responses included certainty. Incorrect responses were categorised for unsafeness. Correct option had been removed from 6 questions.

Summary of Results: The cohort demonstrated insight: the mean proportion of correct responses increasing from 0.31 with low to 0.73 with high certainty (p<0.001); the proportion of incorrect not unsafe responses decreasing from 0.17 with low to 0.10 with high certainty (p<0.001). A degree of foresight was demonstrated: the mean proportion of unsafe responses decreased from 0.53 for low to 0.16 for high certainty (p<0.001); the proportion of incorrect responses that were unsafe decreased from 0.76 for low to 0.61 for high certainty (p<0.001). Individual students could be identified as having insight and/or foresight, or not, by a significant change in responses with increasing certainty: increase in the proportion of correct responses; a decrease in the proportions of incorrect, unsafe, and proportion of incorrect that are unsafe.

Individual students were also identified with potentially limited foresight by multiple unsafe responses with high certainty, and/or risk-taking response patterns.

Conclusions: In addition to ability (number correct), information on insight, foresight and risk-taking was available for the cohort and individuals.

Take-home Messages: This MCQ format adds information. Most have a degree of insight. Some have a degree of foresight. Some take risks.

OTT-OC-3-2
Negatively-worded MCQs – what does the evidence tell us?

Neville Chiavaroli (*University of Melbourne, Melbourne, Australia)*

Background: While most MCQ drafting guidelines explicitly instruct writers to 'word stems positively', many also leave the door open for the use of negatively-worded questions, with qualified instructions such as to 'minimise' the use of negative questions, or to use them 'only if necessary'. This effectively endorses their use, and indeed it seems that the use of negatively-worded MCQs remains relatively commonplace in medical, bioscience and health professional courses (Downing 2005, Jozefowicz et al 2002).

Summary of Work: This presentation is based on an extensive review of the literature on the issue, experience with faculty development workshops on written assessment, and statistical analysis of the performance of negatively-worded questions in local examinations.

Summary of Results: Despite a number of studies, statistical data remains ambiguous. The issues surrounding the use of negatively-worded MCQs are more complex than is often acknowledged, involving some conflicting evidence and other factors which may actually encourage teachers to utilise them.

Conclusions: In the context of ambiguous evidence, it is unlikely that a pedagogical principle will be enough to dissuade writers from using negatively-worded MCQs. Nor should we, as educators, be content to simply 'mandate' against their use. Rather, we need to offer cogent validity-based arguments justifying why negatively-worded should be avoided, rather than just 'minimised', and why statistical evidence may not be sufficient to prove the point.

Take-home Messages: What may seem as a self-evident pedagogical principle to educationalists can seem unnecessarily pedantic to content experts. In the absence of conclusive empirical evidence, we need to provide stronger validity-based arguments.

OTT-OC-3-3
Use of Candidate Feeback for Improving Examinations

Gary Cole (*Royal College of Physicians and Surgeons of Canada, Ottawa, Canada)*, Farhan Bhanji, Jonathan Dupre, Stacey Brzezina

Background: The Royal College of Physicians and Surgeons of Canada develops and administers high stakes exit examinations for 67 medical specialties. Following each exam candidates respond to a survey regarding their perception of the exam. The survey asks a number of questions regarding perceived validity, amount of time, etc.

Summary of Work: The study focuses on the relationship between various factors and the candidates' perceptions of the short answer question (SAQ) exams used for Royal College certification. Factors include, type of questions, number of questions, number of responses, overall number of words in the exam, quality of the questions, time allotted for the exam, and relationship of perceptions to other exams (eg., OSCE).

Summary of Results: Results were obtained for 53 examinations. Overall, the short answer question exams are less well perceived than the oral or OSCE exams. The metric for determining the amount of time that is appropriate for the short answer question exam seems best related to both the number of answers (not questions) as well as the number of words in the exam.

Conclusions: Performance based exams, such as Orals or OSCES, are more likely to be better perceived than written SAQ exams. It is the number of answers that most influences the time that candidates require to write the SAQ exam. The length of the scenarios is also an influence. The overall perception of the SAQ exams is related to a number of factors.

Take-home Messages: Candidates' post examination responses to a survey can provide useful feedback to help understand and improve examinations.

OTT-OC-3-4
Application of Item Response Theory for Analysing the Psychometrics of Multiple Choice Questions Used in MD Certifying Exams at the University of Calgary

Mona Nasir (*University of Calgary, Calgary, Canada)*, Claudio Violato (*University of Calgary, Calgary*)

Background: Multiple choice questions (MCQs) are used worldwide for summative assessment in undergraduate medical education. Very few studies have looked at their stability over the years using item response theory (IRT). The main purpose of this study was to assess the temporal stability of multiple choice questions over a period of three years at the University of Calgary for MD Certifying exams using a two-parameter logistic model of IRT.

Summary of Work: Three courses were chosen for running the item analyses. Thirty MCQs from each course over a period of three years were scrutinised for stability. Item response theory utilizing Xcalibre 4.0.2 was employed to carry out the analyses. Test Response Function (TRF) was generated year-wise for the MCQs for a 2 parameter logistic model. These TRFs were then compared across the years.

Summary of Results: Standardized thetas and discrimination indices were stable for the MCQs across the three year period when compared for alternative

year indicating the fact that TRFs were equivalent for the 2-PL model of IRT.

Conclusions: Multiple choice questions used by the University of Calgary over a period of three years have shown to be internally consistent and have equivalent test response function. This indicates that the MCQs can be employed for MD certifying exams and may be used repeatedly without affecting their psychometric properties.

OTT-OC-3-5
Don't know option in progress testing

Cecile J Ravesloot (*UMC Utrecht, Utrecht, Netherlands)*, Marieke F Van der Schaaf (*Utrecht University, Utrecht*), Jan P J Van Schaik (*UMC Utrecht, Utrecht*), Cas LLJ Kruitwagen (UMC Utrecht*, Utrecht*), Olle ThJ Ten Cate (*UMC Utrecht, Utrecht*)

Background: Using a "don't know" option (DKO) with true/false items in progress tests is subject of discussion. Its use could create construct irrelevant variance in test results, but, on the other hand, can positively impact test reliability, especially with junior participants. We aimed to empirically evaluate the impact of the DKO on construct validity and reliability.

Summary of Work: A 200 item progress test for radiology residents was divided in two 100 parallel items tests (A and B). Each test had a version with DKOs and a version without. Participants were randomized in two groups. One group took test A with DKO first followed by test B without DKO, and the other group did the reverse. Differences in scores between both tests were calculated for each participant. The regression coefficient of the use of the DKO on difference scores was calculated to evaluate construct validity. Group, gender, length of training and knowledge level were also included as predictors. Cronbach's alphas were calculated to estimate reliabilities.

Summary of Results: The participants' use of DKOs and knowledge level were significant, positive predictors of variance in difference score (β.38, SE.05, p<.001; β.24, SE.06, p<.001). Cronbach's alphas of the tests with DKO were .83 (A) and .82 (B), and without DKO .68 (A) and .80 (B).

Conclusions: The results support our hypothesis that using a DKO weakens the construct validity of the test due to construct-irrelevant variance. Participants with low DKO use seem to be disadvantaged. Impact of DKO on test reliability differed between tests.

Take-home Messages: DKO should not be used in TF progress tests in postgraduate medical education.

OTT-OC-4: INTERNATIONAL DIMENSIONS 2

OTT-OC-4-1
Assessing the preparedness of clinical facilities in Kazakhstan to participate in US-style medical student education

John Mahoney (*University of Pittsburgh School of Medicine, Office of Medical Education, Pittsburgh, Pennsylvania, USA)*, Nurgul Khamzina (*Nazarbayev University School of Medicine, Astana*), Marat Tavassov (*Nazarbayev University School of Medicine, Astana*), Arman Saparov (*Nazarbayev University School of Medicine, Astana*), Margaret McDonald (*University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania*), Steven Kanter (*University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania*)

Background: As part of creating a new medical school at Nazarbayev University in Kazakhstan, it has been necessary to assess the readiness of clinical sites and their physicians to participate in US-style medical education.

Summary of Work: A comprehensive assessment instrument was developed and applied using a combination of self-assessment and evaluator observation, much like the process used in US medical school accreditation.

Summary of Results: The instrument was used by each hospital to collect and organize data on the full spectrum of issues relevant to clinical medical education, including: facility resources, patient types and availability, faculty and resident issues (clinical and teaching experience, English proficiency, availability), educational environment, and existing programs. Experienced medical educators completed in-person site surveys in a targeted and accelerated manner, primed by the data in the self-generated information profile.

Conclusions: By clearly articulating the needed information, it was possible to rapidly collect and organize a comprehensive view of the preparedness of multiple hospitals. The instrument was readily completed with minimal guidance. The self-assessment component of the process gave the individual hospitals an investment in the assessment, and also provided a clear introduction to what was going to be expected from them when students arrived. The method and instrument can readily be generalized for use in any locale, and to other health professions.

Take-home Messages: A systematic, accreditation-like approach can be highly effective in assembling a complete picture of institutional readiness for medical education, and help overcome complexities of differing language and culture.

OTT-OC-4-2
Competence Assessments of Internationally Educated Nurses (IENs) Successfully Evaluate Eligibility for Canadian Licensure Examination

Debra Sibbald (*Centre for the Evaluation of Health Professionals Educated Abroad (CEHPEA), Assessment Services, Toronto, Canada)*, Arthur Rothman (*Centre for the Evaluation of Health Professionals Educated Abroad, Assessment Services, Toronto*), Tammie McParland (*Centre for the Evaluation of Health Professionals Educated Abroad, Assessment Services, Toronto*), Heather Scott (*Centre for the Evaluation of Health Professionals Educated Abroad, Assessment Services, Toronto*), Murray Urowitz (*Centre for the Evaluation of Health Professionals Educated Abroad, Assessment Services, Toronto*)

Background: These assessments, based on positive pilot tests, screen IENs for eligibility to write the Canadian Registered Nursing Examination (CRNE) by validating the scores generated.

Summary of Work: Three assessments with 82, 138 and 136 participants effectively evaluated IENs in an appraisal which combines written MCQ testing and a 12 station OSCE rated by nurse examiners. Seventy percent of candidates were 10 years or less post-graduation, from diverse institutions primarily in the Philippines and India.

Summary of Results: Valid multiple choice items and performance cases were developed by content experts conforming to an Entry-to-Practice competency blueprint. Trained examiners pre-rated the items, with criterion anchored tools validating content and competency scoring schemes. Validation: OSCE station and total test score psychometrics were acceptable. The total OSCE test score α-reliability value for each assessment was 0.8, which is higher than values reported with similar tests. Test results discriminated performance between and within candidates. There were no significant track and/or session effects.

Conclusions: The tests performed exceptionally well in terms of the psychometric quality of results, and its ability to differentiate between and within large numbers of candidates.

Take-home Messages: This clinical examination is offered to IEN candidates in lieu of traditional credential screening. The test design incorporated an understanding of parameters for the licensing exam and the appropriate measures to validate competency-based performance assessment. The results provide evidence of this test for screening IENs for eligibility for the CRNE as a valid, reliable, acceptable and sustainable assessment process, improving recognition of overseas qualifications.

OTT-OC-4-3
International Consortium for Assessment Networks (ICAN): facing the challenges of competency-based assessment

Jana Jünger (*ICAN-International Consortium for Assessment Networks, Heidelberg, Germany)*, Konstantin Brass (*ICAN-International Consortium for Assessment Networks, Heidelberg*), Andreas Möltner (*ICAN-International Consortium for Assessment Networks, Heidelberg*), Achim Hochlehnert (*ICAN-International Consortium for Assessment Networks, Heidelberg*), Jörn Heid (*ICAN-International Consortium for Assessment Networks, Heidelberg*) (Presenter: W Kurtz)

Background: The teaching and assessment culture in medical education is currently undergoing extensive changes: instead of assessing knowledge, teaching should be aligned with competencies needed in medical practice. Accordingly, assessment should be carried out with appropriate competency-based procedures. At the same time such exams are costly: content has to be

produced and validated with scarce resources. Also the options for standardization, validation and for quality assurance are limited. Furthermore, in order to adequately assess competencies, a programmatic structure of exams is needed - as outlined in a curriculum based on the principles of constructive alignment.

Summary of Work: ICAN is a non-profit organization for different assessment networks. 57 schools, boards and councils from 7 countries use a common platform for the exam preparation, exchange, delivery and evaluation: the ItemManagementSystem. Within close cooperation with several partners ICAN is steady developing actual needed formats, contents, procedures and tools.

Summary of Results: ICAN designed new formats for the ItemManagementSystem, such as OSCE, EncounterCards and DOPS. These standardized formats can be used and exchanged by partners independent from specific requirements.

For the delivery, several apps, such as tOSCE for tablet-based OSCEs, were developed.

A competency-based student feedback system was introduced, giving students feedback on their strengths and weaknesses in different competencies.

Currently a competency-oriented progress test comprising MCQs and other exam formats is composed.

Conclusions: To address the current challenges, a cooperative approach towards competency-based assessment is needed.

Take-home Messages: With ICAN, a comprehensive platform for such cooperation exists and partners can use the new tools and innovations in the field of competency-based assessments.

OTT-OC-4-4
A global medical education challenge: keeping cultural values while developing international assessment standards

Cynthia Whitehead (*University of Toronto, Faculty of Medicine, Family and Community Medicine, Women's College Hospital, Toronto, Canada)*, Ming Ho (*National Taiwan University College of Medicine, Taipei*), Maria Athina (Tina) Martimianakis (*University of Toronto, Faculty of Medicine, Pediatrics, Toronto*), Jessie Norris (National Taiwan University College of Medicine*, Taipei*), Brian Hodges (*University of Toronto, Faculty of Medicine, Psychiatry, Toronto*)

Background: Globalization is a well-recognized phenomenon in 21st Century medical education. Standardization of assessment measures aims to ensure high quality education globally. The current focus on standardized outcomes, developed in large part in Western contexts, may unintentionally limit attention to values and cultural concerns. Medical education researchers from one Western and one Asian country have begun to work together to explore how values and cultural issues can better be incorporated into international assessment approaches.

Summary of Work: Taiwanese and Canadian medical education researchers gathered for an international symposium in Taipei in January 2013 to share their globalization research findings. Participants engaged in extensive group discussions about guiding principles and approaches that might help to ensure cultural sensitivity and equity in international medical education assessment initiatives.

Summary of Results: Two key systems level themes were identified: the need to be clear about the specific rationale for internationalizing certain assessment standards, and the need to clarify which aspects of medical education are best conceptualized (and then standardized) at the international level. In addition, participants emphasized the need to acknowledge the costs associated with supporting globalization in medical education, and the need to ensure that equitable approaches are considered.

Conclusions: Cross-cultural conversations amongst medical educators and medical education researchers are essential to achieve high quality international assessment standards that respect cultural values. Building collaborative networks with researchers and educators who bring multiple perspectives will further advance and enrich this work.

Take-home Messages: While international assessment standards can help to ensure high quality medical education globally, these standards must incorporate cultural values.

OTT-OC-4-5
Reforming the criteria used to assess the spoken English clinical communication skills of international medical graduates

Tim McNamara (*The University of Melbourne, School of Languages and Linguistics, Melbourne, Australia)*, John Pill (*The University of Melbourne, School of Languages and Linguistics, Melbourne*)

Background: Globalization of the medical workforce has led to concerns about adequate clinical communication skills in immigrant medical professionals, potentially impacting on patient safety. The Occupational English Test (OET) is used in Australia, NZ and Singapore to assess the readiness of international medical graduates (IMGs) to communicate with patients in English via performance on simulated clinical communication tasks. But evaluating spoken performance on the test against linguistic criteria does not always identify those who subsequently experience difficulty in communicating in the clinical setting.

Summary of Work: Feedback from medical educators and supervisors on examples of clinical communication with patients involving IMGs was analysed thematically and yielded a model of the assessment criteria actually oriented to by these medical professionals.

Summary of Results: Two new criteria, Clinician Engagement and Discourse Management, were developed on the basis of the model to supplement the existing language criteria used in the OET speaking sub-test. Current language assessors, trained in the use of the new criteria, were found to be able to apply them reliably. Standard-setting involving judgements of performances on OET tasks by medical professionals

enabled a subsequent comparison of passes and failures using the old and new criteria, and found differences.

Conclusions: Using the assessment criteria indigenous to the clinical setting proved to be operational within a language test setting, and yielded more valid conclusions about individuals' readiness for clinical communication.

Take-home Messages: Incorporating the criteria used by medical professionals to judge effective communication with patients in a language test admitting IMGs to practice can lead to more defensible admission and exclusion decisions.

OTT-OC-5: CLINICAL ASSESSMENT 2

OTT-OC-5-1
Clinical audit as a component of MBBS curriculum and assessment modality: a critical evaluation of validity and reliability

Elina Tor (*The University of Notre Dame Australia, Fremantle, Medical Education Support Unit, Fremantle, Australia)*, Donna Mak (*The University of Notre Dame Australia, Fremantle, Population and Preventive Health, Fremantle*), Carole Steketee (*The University of Notre Dame Australia, Fremantle, Medical Education Support Unit, Fremantle*)

Background: The teaching and learning regarding quality and safety of medical practice across the MBBS curriculum at the University of Notre Dame, Fremantle, Australia, culminates in the final year in which all students are required to undertake a real clinical audit for a health service. Students are assessed both formatively and summatively to achieve the goals of 'assessment for learning', 'assessment as learning' and 'assessment of learning'. This paper examines the validity and reliability of scores interpretation and use from this assessment component.

Summary of Work: A critical appraisal of the clinical audit project (termed project) as an assessment modality was conducted based on a unified validity framework (Messick, 1989) for evidence of content, substantive, structural, external, generalizability and consequential validity.

Summary of Results: Qualitative analysis showed alignment between the project's conceptual underpinning and the MBBS curriculum, that students' learning was well scaffolded and the project's utility to students and participating health services. Quantitative analysis of the summative assessment scores demonstrated high overall internal consistency reliability (Cronbach's alpha >0.8); internal structure consistency between scores from main sections of rubric (Cronbach's alpha > 0.7); inter-rater consistency reliability (intraclass correlation coefficient 0.8 – 0.9) and criterion validity. More procedural and empirical evidence for validity will be discussed at the conference.

Conclusions: The project has demonstrated compelling evidence of validity as a component in the MBBS curriculum and assessment program.

Take-home Messages: The principle and quality assurance practices underpinning this project could be adapted for other student-driven performance assessment modalities involving real work in health services.

OTT-OC-5-2
Problematizing the assessment of simulated patient cases

Geneviève Gauthier (*University of Alberta, Department of Educational Psychology, Edmonton, Canada)*, Jeannine Conway (*University of Minnesota, College of Pharmacy, Minneapolis*)

Background: Variation in the resolution of complex cases by clinical experts is well documented. Yet, it is assumed that in the context of teaching cases, a single content expert can provide the criteria by which learners will be assessed. This research questions this assumption, which has important implications for the assessment of learners.

Summary of Work: This case study investigates the validation of pre-planned answers and arguments guiding the scoring of learner's performance in pharmacy education. Three instructors, considered to be experts at both pedagogical and clinical levels, solved five teaching cases twice. Solving these simulated cases required experts to submit a final hypothesis regarding the medical condition(s) affecting each patient, along with a ranked evidence list that support their hypothesis and treatment plan. Logs, and answers compiled and compared to the proposed "good" answer as defined by the case creator who was one of the experts.

Summary of Results: Experts, including the case creator, could neither reproduce the proposed "good" answer for the cases, nor upon replication of the cases, repeat their first answer to the cases. However, in retrospect participants deemed all of these answers to be acceptable resolutions to the given cases.

Conclusions: The within and between variability in expert answers and paths to these answers suggest that the assessment of simulated patient cases is more complex than simply identifying a "good" answer to a problem.

Take-home Messages: The feedback given to future practitioners, and the assessment of their performance, should reflect the inherent variability in the resolution of teaching cases.

OTT-OC-5-3
Assessing the learning of medical students after spending time in general practice

Heidi Penrose (*Wessex School of General Practice, Wessex Deanery, UK, Primary Healthcare Education, Winchester, UK)*, Samantha Scallan (*Wessex School of General Practice, Wessex Deanery, UK, GP Education Unit, Southampton*), Simon Newton (*Wessex School of General Practice, Wessex Deanery, UK, Primary Healthcare Education, Winchester*), Sharon Kibble (*Winchester*)

Background: There is much debate as to the role and relevance of time spent in general practice to the learning of medical students. Much of what is known is drawn from the feedback of trainees regarding the usefulness of attachments with little demonstrating actual change in thinking.

Summary of Work: The aim of this exploratory investigation was to identify changes in the student's thinking about a clinical case scenario before and after time spent in general practice. Participants 'mind mapped' a clinical case prior to going into general practice and again after. They were required to record their thoughts on the given case scenario: what they would do (the consultation process) and the issues they would consider (the case and primary care context). The maps were thematically analysed, taking into account the content and structure, drawing on the approach used in Kibble et al.

Summary of Results: The thematic analysis demonstrated change between the pre and post maps across a range of items and in structure. Participants demonstrated a more structured approach to thinking about the case and took greater account of context. The mind maps provided a useful method of assessing the learners' starting points prior to going into general practice and how their thinking changed. For the tutor the maps highlighted area to develop in follow up discussion.

Conclusions: Development of clinical thinking is a complex and continuous process which, for the study participants, was demonstrated to benefit greatly by exposure to real cases from general practice.

Take-home Messages: Mind maps provide a novel and innovative way to assess and demonstrate learning.

OTT-OC-5-4
Disrupting diagnostic reasoning: The effect of interruptions on the diagnostic performance of residents and emergency physicians

Sandra Monteiro (*McMaster University, Psychology, Hamilton, Canada*), Jonathan Sherbino (*McMaster University, Medicine, Hamilton*), Jonathan Ilgen (*University of Washington, Medicine, Seattle*), Kelly Dore (**McMaster University, *Medicine,* Hamilton**), Timothy Wood (*University of Ottawa, Medicine, Ottawa*), Meredith Young (*McGill University, Medicine, Montreal*)

Background: Diagnostic errors in emergency medicine are presumed to occur more frequently with increased time pressure and interruptions. Previous studies have reported that increased time pressure did not result in increased diagnostic error. The effect of interruptions has not been tested. As well, it is unclear whether experience will modulate the combined effects of time pressure and interruptions.

Summary of Work: Residents (N = 152) were recruited from several Medical Council of Canada (MCC) Qualifying Examination, Part II test sites (McMaster, Toronto, McGill, Ottawa and Queens). Emergency physicians (N = 46) were recruited from health centers in two cities (Hamilton and Seattle). Participants were randomly assigned to two instructional conditions (fast (time pressure) or slow (no time pressure)). Case interruptions were manipulated as a within subject factor (present or absent).

Summary of Results: Accuracy was not affected by interruptions or time pressure. Only experience level was related to diagnostic accuracy; emergency physicians were more accurate (70%) than residents (43%), F = 235.0, p < 0.0001. Response time (RT) was shorter in the fast condition (55s) than in the slow condition (73s), F = 21.8, p < 0.0001.

Conclusions: Concerns that diagnostic errors are more frequent with increased time pressure and frequent interruptions were not substantiated.

Take-home Messages: As the present study demonstrated that experience determined performance, education programs should focus on improving the quality and quantity of experience during residency.

OTT-OC-5-5
Do peer tutors perform better in examinations? An analysis of medical school finals

Kazuya Iwata (*University College London Medical School, London, UK)*, Daniel Furmedge (*University College London Medical School, London*), Alison Sturrock (*University College London Medical School, London*), Deborah Gill (*University College London Medical School, London*)

Background: Peer-assisted learning (PAL) is recognized as an effective learning tool in medical education for learners. PAL tutors themselves also report improved clinical skills and confidence through tutoring. The objective of our study was to examine whether final year medical students who have been PAL tutors performed better in medical school final examinations than non-PAL students.

Summary of Work: We conducted a retrospective analysis of final year examinations results (written and OSCE) from 2010-2012 at UCL Medical School, and compared examination performance between PAL tutors and non-PAL students. Scores from all years were compared by calculating composite Z-scores. Year 4 examination results were used as indicators of previous academic achievement, and this was controlled for in the final analysis.

Summary of Results: PAL tutors outperformed non-PAL students in all components of the final examinations for all three years analysed, and the composite scores showed statistically significant difference between the two groups. However, when previous academic achievement was controlled for, the difference was no longer significant.

Conclusions: Although PAL tutors perform better than non-PAL students in final year examinations, this difference is likely to reflect their background academic abilities; high performing students appear to be self-selecting to become peer tutors.

Take-home Messages: Peer tutoring in itself may not necessarily lead to better examination performance, but PAL tutors probably are better prepared for high stake

examinations as they gain confidence, knowledge, and improvement in clinical and presentation skills through peer tutoring.

OTT-OC-6: OSCE 3

OTT-OC-6-1
Comparing OSCE global rating scores to candidate performance

Dwight Harley (*University of Alberta, School of Dentistry, Edmonton, Canada*), Margaret Dennett (*Vancouver Community College, School of Health Sciences (Dental), Vancouver*)

Background: The OSCE is a commonly used performance-based assessment tool that provides objective measures of clinical competency. OSCE's strength comes from the objectivity inherant in its checklists, which in concert with evaluator calibration, are intended to maximize objectivity.
The OSCE provides an opportunity to collect a subjective global rating provided by the examiners. This allows the examiner to record his professional, expert opinion regarding the candidate's performance. This rating is central to some standard-setting methods.
Examiner calibration generally focusses on the checklist rubric, a question arose as to whether or not the global rating reflects a similar view as to candidate performance as does the checklist score. This study examines the relationship between the examiner-based global rating score and the checklist score.
Summary of Work: An eight-station OSCE was administered to the 88 fourth-year medical students. Checklist and global ratings scores were determined for each candidate on each station.
Pearson correlations were calculated for each pair of scores. In addition their respective scatter plots were examined. The scatter plots were most enlightening.
Summary of Results: The global ratings correlated strongly with candidate performance on the checklist items (0.67 to 0.81). Scatter plots supported the assumption of linearity although a hesitation on the part of the examiners to award the lowest global rating score was evident.
Conclusions: Global ratings strongly reflect performance recorded by the checklists. Further research needs to be done to see if examiner calibration on global rating would improve the correlations
Take-home Messages: Correlations between global ratings and checklist score should be calculated routinely.

OTT-OC-6-2
Understanding assessor judgements in OSCEs: can item weighting reconcile assessor mismatch between global and checklist ratings?

Godfrey Pell (*Leeds Institute of Medical Education, University of Leeds, Leeds, UK*), Richard Fuller (*Leeds Institute of Medical Education*), Matthew Homer (*Leeds Institute of Medical Education*)

Background: Constructivist views of assessor behaviour highlight many of the challenges assessors face as they make assessment judgements. In OSCE formats, an indicator of this is the degree of mismatch between station-level global judgements and checklist scores. Differential item-weighting is often applied within stations, but literature suggests little impact on overall reliability. However, no work has explored the impact at station level.
Summary of Work: As part of a broader research programme looking at borderline and failing students, we investigated whether varying item weightings away from those set by faculty during the development of stations affected the level of misclassification between checklist scores and global ratings.
Summary of Results: Moving weightings away from those set by faculty to a simple, equal item-weighting model usually increases the degree of assessor misclassification within a station, impacting on station level quality metrics.
Conclusions: When determining global ratings, assessors are likely to implicitly 'weight' checklist items differently based on their clinical judgement of the relative importance of the particular item. Differential weighting provides a way of investigating these behaviours, understanding assessor misclassification and improving assessment quality.
Take-home Messages: Investigation of item weighting techniques within OSCE marking schema provide key opportunities to understand the values assessors place on individual items within stations. Thoughtful application of differential weighting may help bridge the gap between psychometric determinants of quality and more 'rounded' views of high-stakes assessments.

OTT-OC-6-3
Assessing rater quality using a simple descriptive approach

Ilona Bartman (*Medical Council of Canada, Research & Development, Ottawa, Canada*), Fang Tian (*Medical Council of Canada, Research & Development, Ottawa*)

Background: A variety of methods exist for assessing the quality of raters; for example, generalizability theory, IRT, or the dove and hawk analyses (Bartman, Roy and Smee, 2010). However, these methods require multiple raters per station per candidate or a large sample of raters. What if you have neither? If a single rater is used or a sample size makes it untenable to use these methods, then consider a descriptive approach, as explained in this presentation. The method outlined below is employable in situations with a single rater and a small number of raters.
Summary of Work: This method correlates scores assigned by one judge with an average score assigned by nine other judges. The Medical Council of Canada developed this method to identify the most consistent raters for the National Assessment Collaboration (NAC) OSCE; it has three-steps:

1) Correlate station score with total score for all raters
2) Correlate station score with total score for each rater
3) Subtract 2) from 1)
Raters with large absolute differences (Difference >= 0.1) are judged less consistent; those with smaller absolute differences (Difference < 0.1) are judged more consistent.

Summary of Results: Using 2013 NAC data we ran these analyses on 208 raters (across 17 stations and 547 candidates). Only 31% of raters were consistent.

Conclusions: For NAC, an increased effort to train raters is occurring and rater selection criteria are being reviewed using a variety of analyses.

Take-home Messages: These analyses are quick and easy to run, provide useful information to OSCE administrators, and can be applied in many contexts eg. at a school level.

OTT-OC-6-4
A Novel Approach for Developing Rating Scales for a Practice Ready OSCE

Bruce Holmes (*Dalhousie University, Division of Medical Education, Halifax, Canada*), Saad Chahine (*Mount Saint Vincent University, Halifax*)

Background: The Clinician Assessment for Practice Program uses an OSCE to assess IMG readiness for practice. Eight competencies each have a 6-point rating scale: History Taking, Counseling, Problem Definition & Diagnosis (PDD), and Investigation & Management (IM), Physical Exam, Professional Behavior, Quality of Spoken English, and Communication. Although each is defined, the rating form also includes more precise descriptors (Elements) of candidate performance to guide the examiner ratings. The Elements for each competency are the same for all stations. Our research question is, "To what extent are the ratings of Elements related to the competencies?"

Summary of Work: The study contained 36 IMG examinees and in total 432 observations of their performance. Each set of Elements was correlated with the competency. For example, the Elements that made up the history taking competency are: Breadth & Depth, Relevance, Focus, Sequencing, and Chronology.

Summary of Results: A correlational analysis showed a high positive correlation r>0.7 for the History Taking, Counseling, and IM with their respective Elements. Additionally PDD had high r>0.7 correlations with the Elements except for Alternate Diagnosis which was slightly lower r=0.6. Similarly for the Physical Exam there was high correlations r>0.7, except for Sensitivity r=0.6.

Conclusions:

Take-home Messages: Overall the results show a strong positive relationship between Elements and overall competency rating. The PDD Elements of Alternate Diagnosis and Sensitivity show less of a relationship; however these may be due to case specificity. The results also show us that a follow up qualitative study through focus groups with examiners may help identify the utility of the Elements.

OTT-OC-6-6

The reliability of domain based assessment in a high stakes OSCE examination

Robert McKinley (*Keele University, School of Medicine, Newcastle-under-Lyme, UK)*, Jim Crossley (*Sheffield University, School of Medicine, Sheffield*), K Hartley (*Keele University, School of Medicine, Newcastle-under-Lyme*), Mark Cowling (Keele University, *School of Medicine, Newcastle-under-Lyme*)

Background: All assessments of consultation skills at Keele University School of Medicine use GeCoS - a Generic Consultation Skills assessment tool, which contains nine consultation skills domains. Assessors make judgements of students' competence in each domain and an overall judgement at each station.

Summary of Work: Students' scores were analysed using G theory. We have Research Ethics committee approval for research using anonymised student data.

Summary of Results: We have data from 4 years (544 candidates) for year 3, 3 years (392 candidates) for year 4, and 2 years (241 candidates) for year 5. Using median scores across the years, and taking G= 0.8 as the cut off for acceptable reliability, the D studies indicated that assessment of Opening would be reliable with 27 stations, History with 26, Examination could not be estimated, Management with 24, Clinical reasoning with 21, Organisation with 27, and Building and maintaining the relationship with 19.

Conclusions: Performance depended more on the variable challenges of the case than the skill domain being assessed. Overall global and structured scores were more reliable than domain scores.

Take-home Messages: The skill domains are helpful for assessment blueprinting, examiner training and feedback structuring, but the domains did not exhibit the psychometric stability of 'traits' or improve the reliability of the assessment.

OTT-OC-7: PROGRAMME EVALUATION 1: ACCREDITATION AND CURRICULUM

OTT-OC-7-1
Understanding the Severe Action Decisions of the Liaison Committtee on Medical Education over an Eight Year Period

Dan Hunt (*Association of American Medical Colleges, Washington, DC, USA*), Barbara Barzansky (*American Medical Association, Chicago, Washington, DC*)

Background: The Liaison Committee on Medical Education (LCME) provides accreditation for medical education programs leading to the MD in the US and Canada. This study identifies variables associated with severe action decisions over an eight year period.

Summary of Work: All 143 accreditation full survey decisions made by the LCME between 11/2004 and 6/2012 were categorized into either non-severe actions (n=102) or severe actions (n=41) and tested for

explanatory variables that characterized the latter group.

Summary of Results: Variables associated with severe action decisions were, 1. Insufficient response in the medical education database/self-study materials (Odds Ratio (OR), 7.30; 95% CI, 2.38-22.46; P = .001), 2. Chronic noncompliance with 1 or more standards (OR, 12.18; 95% CI, 1.91-77.55; P = .01), 3. Noncompliance with ED-8 (comparability across instructional sites) (OR, 6.73; 95% CI, 2.32-19.47; P = .000), 4. Noncompliance with ED-33 (curriculum management) (OR, 5.40; 95% CI, 1.98-14.76; P = .001)2.4-22.5), and 5. Noncompliance with a large number of standards (rpb = .62; P < .001).

Conclusions: Medical schools with larger than average number of noncompliant standards, had less than optimal preparation for the visit, and had problems in the same areas as the previous survey are at higher risk for a severe action decision. Also significantly associated with a severe action decision is the lack of evidence for comparability of educational opportunity and the absence of a strong central committee capable of integrating content across the educational program.

Take-home Messages: These findings may be useful for medical education programs efforts to avoid severe action decisions and ensure the quality of the educational offering.

OTT-OC-7-3
Assessing cultural immersion experiences at the John Burns School of Medicine in Hawaii

Martina Kamaka (*University of Hawaii John A. Burns School of Medicine, Native Hawaiian Health, Honolulu, USA*)

Background: The Department of Native Hawaiian Health at the John A. Burns School of Medicine Medicine has been offering cultural immersion for students and residents as part of a comprehensive cultural competency curriculum focusing on provider training in addressing the large health disparities suffered by the indigenous, Native Hawaiian population. The curriculum is strengths based focusing on experiential learning and native culture. A major challenge for any cultural competency curriculum is assessment.

Summary of Work: In 2004, the Cultural Competency Curriculum (C3) Project began searching for assessment tools for our immersion curriculum. The MAKSS (Multicultural Awareness Knowledge and Skills Survey) developed at the University of Hawaii for counselors looked promising. It was piloted in 2006 and then modified, with the help of one of the developers of the MAKSS survey, to reflect both medical and Native Hawaiian perspectives. Qualitative information was also gathered.

Summary of Results: Preliminary analysis of the modified survey tool utilizing pre/posttest including factor analysis showed significant differences in all three subscales (awareness, knowledge and skills), however results were limited due to small numbers. Several years later, we are undergoing final data analysis of the use of this tool as well as of our qualitative data. Results will be shared with conference participants.

Conclusions: Preliminary analysis is encouraging. Upon final analysis, we hope this pre/posttest tool will be reliable and valid and also transferable, with modifications, to other institutions. Combining this type of quantitative information with qualitative data should demonstrate meaningful impact.

Take-home Messages: Don't be afraid to use pre-existing tools and modify them.

OTT-OC-7-5
Alignment of Readiness for Clerkship and Residency Surveys for comprehensive evaluation of undergraduate medical education programs

Linda Peterson (*University of British Columbia, Evaluation Studies Unit, Vancouver, Canada*), Shayna Rusticus (*University of British Columbia, Evaluation Studies Unit, Vancouver*), Kevin Eva (*University of British Columbia, Centre for Health Education Scholarship, Vancouver*), Derek Wilson (University of British Columbia*, Evaluation Studies Unit, Vancouver*), Chris Lovato (*University of British Columbia, Evaluation Studies Unit, Vancouver*)

Background: We recently published the development and performance of a novel competency-based evaluation instrument ("Readiness for Clerkship Survey") to measure the effectiveness of the pre-clerkship of an undergraduate medical education program (UME). The purpose of the present study was to further this effort to develop a Readiness for Residency Survey to evaluate the curriculum as a whole.

Summary of Work: Building on the results of our earlier study, the Readiness for Residency Survey was designed to align with the Readiness for Clerkship Survey but to reflect a higher level of competency where appropriate. The 42-item survey was administered electronically six months after the start of residency and completed by 81 UBC MD graduates. Descriptive statistics and decision studies were conducted.

Summary of Results: The reliability of a single rater to discriminate between the competencies was G=0.34. However, when averaged across all raters, G=0.97 was obtained. A reliability of G=0.8 could be obtained with as few as 9 raters. Whereas highest rated competencies in the clerkship survey were in the domains of professional, communicator and manager; three medical expert competences were identified on the residency survey as being strengths of the program.

Conclusions: The survey provides reliable ratings of the relative effectiveness of various aspects of training with small numbers of respondents.

Take-home Messages: The Readiness for Residency Survey provides reliable data that can help guide improvement of the UME program and anticipation of the needs of new residents while reducing survey burden.

OTT-OC-8: FEEDBACK UNDERGRADUATE 2

OTT-OC-8-2

Custom Software Facilitates Individualised Test Feedback

Anna T Ryan (*The University of Melbourne, Medical Education Unit, Melbourne, Australia)*, Terry Judd (*The University of Melbourne, Medical Education Unit, Melbourne*)

Background: A major challenge in medical education is the provision of specific and timely student feedback. Although computer-based assessment has significant utility in meeting this challenge, paper-based examination formats are frequently considered incompatible with expedient feedback.

Summary of Work: This study demonstrates how purpose-built software applications were used to provide rapid and individualised test feedback to medical students following progress testing. Test items used in this study were from a protected bank and could not be released to students. Each item was coded according to the core presentation(s) and clinical rotation of relevance. Students sat paper-based tests and submitted their answers on scannable MCQ answer sheets. Software applications were developed using an open-source rapid application development tool and were designed to:

1. Simultaneously mark the various test forms (using MCQ scanner output data)
2. Create the individualised feedback document for each student, and
3. Email all students their individualised feedback report.

Summary of Results: Students sat the paper-based tests at nine different clinical school sites. Use of the software allowed students to receive feedback within one day of receipt of all answer sheets. Students were provided with three different feedback types (comparative data; correct and incorrect answers related to core presentations; or core presentations related to certainty) depending on their randomisation within the study.

Conclusions: Coding of test items and development of purpose-built software allows rapid delivery of individualised student feedback. Coding during initial item development would further decrease the feedback provision workload.

Take-home Messages: Custom software applications can facilitate swift and individualised feedback after paper-based examinations.

OTT-OC-8-3
Understanding pre-clinical medical student choices in the use of formative assessment

Ancil Abney (*University of Central Florida, College of Medicine, Orlando, USA)*, Jonathan Kibble (*University of Central Florida*)

Background: Our previous studies have consistently demonstrated that student performance and participation rates in online formative assessment correlate with summative assessment outcomes.

Summary of Work: This study used in-depth interviews to investigate the decision making process used by medical students as they choose to participate in formative assessment. A grounded theory approach was used in which student interviews were recorded, transcribed and coded to identify key themes.

Summary of Results: Sixteen in-depth interviews were needed to saturate emerging themes. Four major themes were found describing what was most important to students when choosing when, and how often, to participate in online formative assessment: 1) effectiveness of feedback, 2) accuracy and relevance of information presented, 3)time constraints (curricular and extra-curricular), 4) fear of judgment (self, peer and faculty).

Conclusions: Several best practice points were identified to maximize student engagement in online formative assessment: 1) provide elaborated feedback, 2) pay scrupulous attention to information accuracy [or the hidden curriculum is promoted], 3) ensure non-judgmental learning environment, 4) provide student academic advising about the importance of self-regulated learning and use of feedback.

Take-home Messages: Great care is needed in the design and deployment of online self-assessments to maximize student engagement.

OTT-OC-8-4
A Paradox explained? A cautionary tale concerning student satisfaction with teaching

Janet Lefroy (*Keele University, School of Medicine, Keele, UK)*, Ashley Hawarden (*Keele University, School of Medicine, Keele*), Maggie Bartlett (*Keele University, School of Medicine, Keele*), Simon Gay (Keele University*, School of Medicine, Keele*), Robert McKinley (*Keele University, School of Medicine, Keele*)

Background: At Keele, we perform serial formative workplace based assessment (WPBA) of students' consultation skills with written feedback. However, student evaluation consistently shows greater satisfaction with verbal feedback and, while our assessments of the quality of the written feedback are correlated with student satisfaction with teaching (Kendal tau 0.38 p=0.01), there is no correlation with satisfaction with verbal and written feedback (tau = 0.075 and 0.092). We now explore this paradox.

Summary of Work: A secondary analysis of the transcripts from semi-structured interviews with students who had recently had three WPBAs with verbal and written feedback in a four week period. Comparative analysis was made of all references to verbal/informal feedback and to written/formal feedback.

Summary of Results: 24 students participated in the study. Students reported that their assessors gave verbal feedback on multiple occasions before completing each formal assessment. This was seen by students as supportive, immediate, and desirable. They felt that the written feedback was superfluous. What was notable was the correlation of their reported

informal feedback with the wording of their written feedback.

Conclusions: Although the WPBA was formative, students and their assessors wanted positive outcomes. Assessors also became familiar with language for assessing and giving feedback on consultation skills. Consequently, students had more and richer verbal feedback than we expected. This at least in part explains the disparity in student satisfaction with verbal and written feedback. However, we consider this a positive unintended consequence of the requirement to complete three formal WPBAs.

Take-home Messages: Negative student feedback can reflect unseen benefits.

OTT-OC-9: COMPETENCY BASED ASSESSMENT UNDERGRADUATE

OTT-OC-9-1
Basic science instruction: Preparing medical students for future learning

Maria Mylopoulos (*University of Toronto, Toronto, Canada)*, Nikki Woods (*University of Toronto, Toronto*)

Background: 'Preparation for future learning' (PFL) is the ability to learn new information from available resources, relate new learning to past experiences and be innovative and flexible when problem solving. PFL has been proposed as a key competence of adaptive expertise. Thus, there is a need to ensure that instruction supports the development of the PFL ability and that assessments accurately measure its progress The objective of this research was to compare the impact of basic science instruction versus clinical-focused instruction on performance on a PFL assessment (PFLA).

Summary of Work: This study used a "double transfer design." Fifty-one pre-clerkship students were randomly assigned to either basic science instruction or clinical-focused instruction . After completing a 'sequestered problem solving' assessment (SPSA)on the learned material, all participants then received new, related clinical instruction and completed a PFLA on the new material. The data from the SPSA and PFLA were submitted to independent sample t-tests.

Summary of Results: Scores on the SPSA for the basic science (M=.65 SD =.11) and clinical learning (M=.62 SD=.11) conditions were not significantly different; $t(42)=.90$, $p =.37$, $d=.27$. Analysis of the PFLA revealed significantly higher scores for participants in the basic science learning condition (M= .72 , SD= .14) compared to those in the clinical learning condition (M= .63 SD =.15); $t(42)=2.02$, $p=.05$, $d=.62$.

Conclusions: Our results show that basic science instruction enabled better learning of novel, related content.

Take-home Messages: Critically, the PFLA revealed the differential impact of different forms of instruction on learning. We discuss the implications of our results on assessment for learning.

OTT-OC-9-2
Clinical assessment in Australian and New Zealand medical schools: An overview and the development of a suite of national assessment blueprints

Monique Hourn (*Medical Deans Australia and New Zealand, Sydney, Australia)*, Richard Hays (*Bond University, Health Science and Medicine, Pro Vice Chancellor Learning & Teaching, Gold Coast*)

Background: Australian and New Zealand medical schools face increasing pressure to demonstrate that their assessment processes are measuring agreed graduate outcomes, thus ensuring public safety in the health system.

Over the last four years, Medical Deans Australia and New Zealand has conducted a series of projects to increase understanding of clinical assessment in Australian and New Zealand medical schools and to provide clinical education resources to assist medical schools develop a national approach to clinical assessment. The latest stage has produced a suite of clinical assessment blueprints based on accreditation standards.

Summary of Work: An extensive consultation process with all Australian and New Zealand medical schools was undertaken to collect data on clinical assessment, assessment blueprints, Workplace Based Assessments and standard setting for clinical exams.

Summary of Results: The project has provided a unique snapshot of clinical assessment occurring in 2012/13 and used this information to develop a suite of clinical assessment blueprints.

Conclusions: This project has increased medical schools' awareness of the technical aspects of assessment, has promoted collaboration between the schools and has encouraged a more consistent approach to assessment standards across Australian and New Zealand medical schools.

Take-home Messages: Overall Australian and New Zealand medical schools clinical assessment programs are robust. Recommendations have been developed to enhance clinical assessment programs and assessment blueprints have been developed based on national standards. This is useful for medical schools, accreditation agencies and health services and should increase confidence that graduates across Australia and New Zealand are assessed against the agreed outcomes, ready to commence postgraduate employment and training.

OTT-OC-9-3
Validity study of workplace portfolios using a recently developed validated inventory (VIP-A)

Nele R.M. Michels (*University of Antwerp, Faculty of Medicine and Health Sciences, Antwerp, Belgium)*, Marijke Avonts (*University of Antwerp, Faculty of Medicine and Health Sciences, Antwerp*), Griet Peeraer

(*University of Antwerp, Faculty of Medicine and Health Sciences, Antwerp*), Kris Ulenaers (*University of Antwerp, Faculty of Medicine and Health Sciences, Antwerp*), Erik W. Driessen (*Maastricht University, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht*), Benedicte Y. De Winter (*University of Antwerp, Faculty of Medicine and Health Sciences, Antwerp*)

Background: Portfolios, designed for mentoring and assessing students at the workplace, are often based on competencies. The purpose of this validity study was to investigate whether the intended CanMEDS-based competencies are actually measured using the Validating Inventory for Portfolio-Assessment/Antwerp (VIP-A), a shortened version of the CLIMAS (Michels et al. 2012).

Summary of Work: The VIP-A was used to validate 120 workplace portfolios of medical students during internships at three different universities in Belgium and The Netherlands (Antwerp, Maastricht, Utrecht). Each portfolio was evaluated by two raters who indicated for each competency whether the information in the portfolio was sufficient to enable the assessment of the competency. Percentage Agreement between the raters was calculated. A descriptive analysis of the validation data was performed for each role and competency. Information on the context and portfolio-content was gathered and correlated with the different roles.

Summary of Results: For almost all competencies, interrater agreements were high to very high (>87%). All the involved internship portfolios, were embedded in learning-oriented programs. The descriptive analyses showed that almost all roles and competencies could be assessed sufficiently or partially by one of the portfolios. Overall, the communicator (90,2%) and professional role (87,1%) were best covered, whereas the health advocate role (59,1%) was least covered by the different portfolio-tasks.

Conclusions: This study showed that portfolios could be used successfully for assessing the CanMEDS-roles and their competencies.

Take-home Messages: To improve content validity of portfolio-data, we confirmed the importance of clear aims and blueprints, learning-oriented contexts and an appropriate selection of portfolio-content.

OTT-OC-9-4
Designing Competency-Based Learning Objectives to Ensure Uniform Evaluations of Clinical Rotations Delivered at Multiple Off-Campus Institutions

Carmen Fuentealba (*Ross University School of Veterinary Medicine, Programs Assessment, North Brunswick, USA*)

Background: Schools and Colleges of Veterinary Medicine are faced with increased scrutiny regarding the quality of the educational program. Educational institutions are accountable to numerous stakeholders and are required to provide evidence of fair and unbiased evaluation of students throughout the DVM program. Development of competency-based learning objectives that align with professional competencies, and design of adequate assessment instruments are some of the many challenges faced by educators, and individual responsible for the program's quality assurance.

Summary of Work: Students enrolled in Ross University School of Veterinary Medicine (RUSVM) complete the clinical year at 25 accredited veterinary schools(Affiliates) in the US and Canada. To ensure uniformity in evaluation, RUSVM defined the competencies and learning objectives for all students completing the program at Affiliate institutions. The documents were reviewed by rotation coordinators and their feed back was incorporated into a consensus document.

Summary of Results: The presentation will include an overview of evaluation methods currently used to assess student performance and will describe best practices used in the design of competency-based grading rubrics.

Conclusions: The process followed to implement changes, the challenges encountered, and strategies used to achieve consensus across institutions will be discussed.

Take-home Messages: Design of Compentency-Based Learning Objectives is the cornerstone in the development of fair an unbaissed grading rubrics. Collaboration across institutions facilitates the process of achieving intramural concensus.

OTT-OC-9-5
Comparison of a novel CanMEDS-based assessment tool to a traditional approach for formative assessment of clinical clerks

Siddhartha Srivastava (*Queen's University, Internal Medicine, Kingston, Canada)*, Wilma Hopman (*Kingston General Hospital, Clinical Research Centre, Kingston*), David Taylor (*Queen's University, Internal Medicine, Kingston*)

Background: Queen's University School of Medicine has a competency based curricular framework built on the CanMEDS roles. The current midblock formative assessment tool for clinical clerks on internal medicine uses a traditional checklist approach, with limited effectiveness at helping students meet curricular milestones. We wanted to determine if a descriptor-based assessment tool built from local curricular competencies can help undergraduate medical students progress through educational milestones.

Summary of Work: Queen's undergraduate curricular competencies (built on the CanMEDS competency framework) were reviewed and divided into graduated stages of competence. Collectively, these stages represent the expected maturation process, or milestones, of a clinical clerk. A competency-based assessment tool (CBAT) based on these stages was developed. During the 2012-2013 academic year, clinical clerks received mid-block formative assessment using the traditional model. In addition, the clerkship course director applied the CBAT. The perceived usefulness of the CBAT was assessed by an anonymous student survey

comparing it to the traditional midblock assessment method.

Summary of Results: Of 70 students, 66 enrolled with a 100% survey completion rate. 95% (63) of students reported that CBAT effectively identified deficiencies, compared to 53% (35) for the traditional method (p < 0.001). 98% (65) of students reported that CBAT was useful or very useful in directing improvements, compared to 42% (28) for the traditional method (p<0.001).

Conclusions: The CBAT was perceived to be more effective at identifying deficiencies and directing improvements for clinical clerks on internal medicine compared to the traditional approach.

Take-home Messages: Using a descriptor-based formative assessment tool can help clerks progress through milestones and attain curricular competencies.

OTT-OC-9-6
Assessing Competencies and Combating Grade Inflation Through a New Evaluation Form

Machelle Linsenmeyer (*Oklahoma State University Center for Health Sciences, Tulsa, USA*), Johnathan Franklin (*Oklahoma State University Center for Health Sciences, Tulsa*)

Background: As accreditation of medical schools focuses more on student achievement of competencies, the need to assess students based on national competencies has increased. This can be especially challenging in clinical clerkships where assessment and evaluation of student competencies is more difficult. This, coupled with grading inconsistencies (fair and reliable scoring of students and the avoidance of grade inflation), can be an issue. Limited literature on this topic currently exists.

Summary of Work: As our institution struggled with grading issues (grade inflation or giving a passing grade when it wasn't deserved just to avoid conflict) as well as the need to assess competencies of students in a valid, reliable and consistent manner, we searched for a new assessment tool. To begin this process, we did an exhaustive review of instruments being used at other medical schools around the world. From this, we developed an instrument to meet the objectives we had for the tool.

Summary of Results: We have seen more accurate grading, more of a bell curve in scoring----where before we saw mostly As, more useful and constructive comments, and greater satisfaction with the instrument.

Conclusions: The new instrument has taken the subjectivity out of the scoring. It has provided a mechanism to ensure consistency across scoring because the descriptors provide details that can be paired with actual performance and observance of a specific behavior or activity.

Take-home Messages: This new instrument provides a mechanism for faculty to give fair, consistent, and reliable scores based on competencies.

POSTERS

OTT-PA2: ASSESSMENT FOR LEARNING

OTT-PA2-01
Collecting patient feedback to assess resident and staff physician patient care at a family medicine teaching unit: The Feedback From Patients on Physician Care (FFPPC) tool

John Brewer (*University of Ottawa, Ottawa, Canada)*, Eric Wooltorton (*University of Ottawa, Ottawa*), Madeleine Montpetit (*University of Ottawa, Ottawa*), Farhad Motamedi (University of Ottawa*, Ottawa*), Peter Brewer (*University of Ottawa, Ottawa*), Erica Battram (*University of Ottawa, Ottawa*)

Background: In our clinic, we have not previously solicited feedback from patients on individual physicians' strengths and weaknesses. Our innovation was piloted at a large university academic teaching site (>8000 patients) to more systematically collect feedback from patients on the care delivered by our trainees and staff physicians.

Summary of Work: We developed and piloted a paper-based tool seeking anonymous qualitative and quantitative patient feedback. No existing tool identified in the literature was deemed adequate by experienced family medicine staff physicians at a postgraduate education retreat (n=12, Feb 2013) nor at two academic units (n= 18, April 2013). Staff and resident physicians provided input on the elements they deemed most important to assess. A new tool was creasted (the FFPPC tool) consisting of ten aspects of physician care during the visit. Nurses selected a convenience sample of patients to pilot the form at our family medicine teaching unit in 2013.

Summary of Results: The tool is being assessed in terms of acceptability (to patients and physicians) in terms of content, process of administration, and value of its results. This work in progress allows feedback to this individual, with normative comparisons (PGY1, PGY2, staff physicians) to offer better context for the comments.

Conclusions: Patient feedback using the FFPPC tool allows more complete formative assessments of individual physicians (staff and residents) because it includes patient feedback on performance. It is general enough to be used in other hospital settings (in-patient, out-patient).

OTT-PA2-02
Final clerkship grades: do they relate to written feedback in earlier clerkships?

Franciska Koens (*VUmc School of Medical Sciences, Institute for education and training, Amsterdam, Netherlands)*, Floor Mulder (*VUmc School of Medical Sciences, Institute for education and training, Amsterdam*), Hester Daelmans (*VUmc School of Medical Sciences, Institute for education and training,*

Amsterdam), Gerda Croiset (VUmc School of Medical Sciences*, Institute for education and training, Amsterdam*)

Background: Student performance in clerkships in the Netherlands is rated on a 1-10 scale (10 being the highest grade given, students pass with grade 6 or higher) accompanied by written feedback. The final clerkship in our medical school (VU University Medical Center, Amsterdam, The Netherlands) consists of a 16-week clerkship in which students integrate their gained competences in a near professional work setting. We were interested whether the grades in the final clerkship and the written feedback in previous clerkships correlated.

Summary of Work: We randomly selected 6 to 10 student files for each grade (5 to 9) for the final clerkship. We counted the number of positive and negative written feedback in all clerkships and calculated the positive/negative feedback ratio. We defined positive feedback as encouraging remarks and negative feedback as discouraging remarks on, or disapproval of observed behavior.

Summary of Results: The positive/negative feedback ratio increased with increasing grades. Students who failed the final clerkship (grade 5) received mainly and consistent negative feedback during previous clerkships. Students with higher final clerkship grades (8 or 9) received more positive feedback then with grades 6 or 7. In addition, students who scored a 9 were given less negative feedback during their clerkships than those with grades 8.

Conclusions: High achieving students received more positive feedback whereas students who failed the final clerkship mainly received negative feedback and hardly showed improvement.

Take-home Messages: Written feedback in earlier clerkships seems to correlate with the grades of students in a 16-week final clerkship.

OTT-PA2-03
Is feedback getting S.M.A.R.T.er?

Bridget Langham (*Health Education East Midlands, Postgraduate Medical Education, Ruddington, UK*), Heidi Breed (*Health Education East Midlands, Postgraduate Medical Education, Ruddington*)

Background: The design of formative assessment tools in the UK Foundation Programme has changed to refocus on feedback and promote action planning for development needs. Our audit looks objectively at the impact of this change on the quality of feedback given to trainees in one formative assessment; the Case Based Discussion (CbD).

Summary of Work: One hundred FY1 trainee CbDs were analysed pre and post change; 50 from the period August 2011 to December 2011 and 50 from the corresponding period in 2012. The quality of feedback in each assessment was allocated a rating using the SMART objective setting mnemonic. The score allocated was dependent on the number of SMART objectives identified.

Summary of Results: Prior to the change the average number of CbDs completed were 2 per trainee of which 27% had no SMART objectives. Post change this has increased to 4 per trainee of which 50% had no SMART objectives. The number containing three or more objectives was 38% and 37% respectively.

Conclusions: Whilst the removal of a scoring system on the CbD form has encouraged increased engagement with formative assessment, there is a decrease in the quality of feedback received. Investigation is required to determine the factors leading to this behaviour but a modification to the formative assessment record may be the root cause of this undesired effect as prompts present in 2011 are absent in 2012.

Take-home Messages: When considering changes to the design of formative assessments, the effect of a change to the recording mechanism should be considered, otherwise the proposed benefit of the change may be cancelled out.

OTT-PA2-04
Encouraging residents' feedback seeking behavior

Janine van der Rijt (*Erasmus MC, Institute of Medical Education Research Rotterdam, Rotterdam, Netherlands)*, Susanna M. Lucieer (*Erasmus MC, Institute of Medical Education Research Rotterdam, Rotterdam*), Matthijs de Hoog (*Erasmus MC, Department of Pediatrics, Division of Pediatric Intensive Care, Rotterdam*), Wim H. Gijselaers (Maastricht University*, Department of Educational Research and Development, Maastricht*)

Background: Developing a life-long-learning attitude is an essential competency for residents to become proficient medical specialists. Research suggests that a majority of learning is gained informally on the job from colleagues. Feedback is a key component of informal learning, and is considered as a crucial resource in clinical medical education. Traditionally, feedback is seen as a top-down process where supervisors provide feedback. Nowadays, research suggests that individuals should not wait for feedback, but should proactively seek feedback during daily work interactions. However, this requires creation of safe and supportive clinical environments, including supervisors, in which residents feel free to seek feedback to support their expertise development.

Summary of Work: A series of empirical studies in accountancy and in medical residency investigated conditions within work environments that encourage (or inhibit) feedback and help seeking behavior. The effect on career development and (expert) performance was studied. We will present the main findings of our studies and the implications for the medical profession.

Summary of Results: Important elements to develop expertise are; support for learning, psychological safety, high quality feedback of colleagues and in particular of supervisors, and the quality of relationships within the work environment.

Conclusions: A supportive supervisor and a work environment in which employees feel safe and

supported to seek feedback or help, is imperative to continue to develop.

Take-home Messages: In order to create effective clinical learning environments, it is vital to enhance understanding of conditions that encourage residents' feedback seeking and the quality of the provided feedback.

OTT-PA2-07
Self-directed learning readiness among undergraduate physiotherapy students at University of Ghana

Jonathan Quartey (*School of Allied Health Sciences, College of Health Sciences, University of Ghana, Department of Physiotherapy, Accra, Ghana*), Carina Eksteen (*University of Pretoria, Department of Physiotherapy, Pretoria*), Glynis Pickworth (*University of Pretoria, Department of Physiotherapy, Pretoria*)

Background: The core characteristic of a problem based teaching approach is that students' learning is facilitated by challenging them to solve demanding artificial key professional, clinical and real life problems. It is structured with the aim that they become self-directed, lifelong learners.

The aim of this study was to determine physiotherapy students' readiness to engage in self-directed learning.

Summary of Work: This 'repeated pre-test post-test study was carried out over two academic years involving 119 third and fourth year physiotherapy students. The students' readiness was assessed using the self-directed learning readiness scale (SDLRS) questionnaire before and after a problem-based teaching approach. Pearson's correlation coefficient, Pearson's Chi-square and ANOVA were used to determine the level of significance between variables with $p < 0.05$ using SPSS version 20.0.

Summary of Results: Participants included 58 males and 60 females (50.4 % third and 49.6 % fourth years). The mean scores of 60 % of all participants were 203.17 (SD 20.08) before and 203.96 (SD 22.58) after the module (which is interpreted as having average, above average or high readiness on the self-directed learning readiness scale) with no significant difference between scores ($p=0.776$). There was no significant difference between age and total scores ($p=0.757$); total scores between first and second year of study ($p=0.227$), third and fourth years ($p=0.683$) and between males and females ($p=0.074$).

Conclusions: The module presented did not affect students' readiness for self-directed learning.

Take-home Messages: Self-directed learning should be assessed with a battery of tools instead of only one questionnaire.

OTT-PA2-08
The impact of curriculum rearrangement in the lead-up to barrier exams

Teresa M Treweek (*University of Wollongong, Graduate School of Medicine, Wollongong, Australia)*, Jessica R Hughes (*University of Wollongong, Graduate School of

Medicine, Wollongong), Kylie J Mansfield (*University of Wollongong, Graduate School of Medicine, Wollongong*), Alistair Lethbridge (University of Wollongong*, Graduate School of Medicine, Wollongong*)

Background: The Graduate School of Medicine (GSM) MBBS is a four year programme preparing students for the practice of medicine regional, rural and remote Australia. In response to student and staff feedback, we recently implemented curriculum change prior to the barrier summative exam in Year 2 of the course. The Nervous System (NS) Block, arguably the most challenging subject for students, was previously delivered immediately prior to the exam but was relocated to allow students greater opportunity for consolidation.

Summary of Work: Mean student exam scores for each Year 2 Block (Haemopoietic/Immune; HI, Musculoskeletal; MS and NS) and student survey responses were compared between different student cohorts encompassing the time of the curriculum change.

Summary of Results: When this change was implemented (2012), student exam results for NS Block content were showing a declining trend. Survey data revealed that students had low levels of confidence in their knowledge of the content, and considerable anxiety related to delivery of the Block immediately prior to the exam. Following curriculum change, this decline in student performance halted and students reported reduced anxiety. Conversely, a statistically significant decline in exam scores for HI content was evident, highlighting the impact of placing this Block last.

Conclusions: The change had a positive effect on student results for NS Block, accompanied by increased confidence in material delivered in this Block. The decline in HI performance was ameliorated post-2012, with the improvement likely influenced by additional factors.

Take-home Messages: Timing of content delivery is critical to student performance and confidence in complex topics.

OTT-PA2-09
Assessment of competencies in medical students: The challenges of large-scale diagnostic assessment in Mexico

Melchor Sánchez-Mendiola (*UNAM Faculty of Medicine, Medical Education, Mexico City, Mexico*), Adrián Martínez-González (*UNAM Faculty of Medicine, Medical Education, Mexico City*), Andrés Trejo-Mejía (*UNAM Faculty of Medicine, Medical Education, Mexico City*), Fernando Flores-Hernández (UNAM Faculty of Medicine*, Medical Education, Mexico City*), Iwin Leenen (*UNAM Faculty of Medicine, Medical Education, Mexico City*), Teresa Fortoul van der Goes (*UNAM Faculty of Medicine, Coordination of Basic Sciences, Mexico City*)

Background: In Latin America there is little published information about knowledge retention and formative assessment of competencies in medical students, during medical school training and curricular changes. The

objective of this study was to assess knowledge level and clinical competencies, in medical students at the end of the second year in a new curriculum.

Summary of Work: Observational, cross-sectional study in UNAM Faculty of Medicine students in Mexico city. A diagnostic evaluation was performed in the first class of the "Plan of Studies 2010" curriculum, as part of an integral program evaluation strategy. The assessment had two components: a multiple-choice question test and a practical Objective Structured Clinical Examination (OSCE).

Summary of Results: 456 (87%) of the 524 students that successfully completed second year of Plan 2010 were assessed. The written test had 211 items, mean percent-correct score of 60±14.5, mean difficulty index of 0.60, Cronbach's alpha of 0.85. The OSCE mean global score was 58±9, with a G-coefficient of 0.48, results were provided for each station. Results by area of knowledge, course and station were reported to the Institution authorities, academic departments and individual students.

Conclusions: The results in general are acceptable, compared with previous written evaluations at the end of the second year, suggesting that the new program is achieving its educational goals. Competencies were formally assessed, for the first time in our Institution, in all students using a standardized instrument, establishing a starting point for follow-up. The study provided useful information to the institution, teachers and students.

Take-home Messages: Diagnostic assessment during the implementation of major curricular changes is extremely important for all the stakeholders.

OTT-PA2-12
Pictorial representation of assessments used at exit level of a medical training programme: What does this tell us?

Christina Tan (*Stellenbosch University, Centre for Health Professions Education, Cape Town, South Africa)*, Susan van Schalkwyk (*Stellenbosch University, Centre for Health Professions Education, Cape Town*), Francois Cilliers (*University of Cape Town, Educational Development Unit, Cape Town*), Juanita Bezuidenhout (Stellenbosch University*, Anatomical Pathology, Cape Town*)

Background: Assessment is an essential component of any medical curriculum and is central to public accountability. In high stakes exit level examinations, assessments should move to higher levels of Miller's pyramid, to allow decisions to be made about fitness to practise.

Summary of Work: An investigation was undertaken to map what current exit level assessment takes place at one South African medical school. Document analysis of student study guides for information relating to assessment was conducted. Clarificatory interviews with discipline coordinators verified the content of the study guides. Assessment methods were mapped with consideration to Miller's pyramid and the proportion which each method contributed to the final overall

discipline mark. This provided a pictorial summary of assessment for all disciplines at exit level.

Summary of Results: Assessment practices varied across disciplines based on the: 1) number of assessments, 2) weighting of individual components, and 3) range of assessment methods used.

Conclusions: This assessment mapping has highlighted the range and diversity in existing assessment practices and has provided a useful reference overview for discipline and programme coordinators.

Take-home Messages: A pictorial representation of the proportional use of different assessment methods is a useful tool to enable stakeholders to easily identify the range of methods used and the extent to which assessment at exit level is addressing the higher levels of Miller's pyramid. It also highlights potential areas where shifts to better practice should be encouraged.
This tool can be adopted for any medical training programme.

OTT-PA2-14
Reflective Diary and Dialogue: Critical Reflection Tools for Transformative Learning

Nipon Uppamanoraset (*Queen Sawang Vadhana Memorial Hospital, Thai Red Cross affiliated with Burapa University, Medical education Center / Faculty of medicine, Chonburi, Thailand)*, Manavika Punnan (*Queen Sawang Vadhana Memorial Hospital, Thai Red Cross affiliated with Burapa University, Medical education Center / Faculty of medicine, Chonburi*)

Background: Psychological problems and psychiatric patient care are sophisticated and require a deep understanding. Critical reflection tools contribute to understanding. These could result in holistic patient care and understanding human nature.

Summary of Work: These activities were organized in the academic year 2013 for 32 of the 5th medical students (MDs), Burapa University, during the community psychiatry rotation at Queen Savang Vadhana Memorial Hospital, Thai Red Cross. Prior to the apprenticeship, the dialogue group was organized to practice emotional recognition and self assessment, and entrusted the MDs to write a reflective diary by posting in a private Facebook group which was supervised by a facilitator during the apprenticeship.

Summary of Results: Most reflective thinking was on a behavior level. Forty two percent of the dialogue group had medium level satisfaction. Some of them thought that this period was useful, they had a chance to assess their own minds. Hopefully the meditation and positive talking should be retained. Most of the MDs (37%) had low level satisfaction, felt it cumbersome to write every day and suggested that they should write once a week but 42 percents found that the reflective diary activity gave them a more positive attitude and social ability. In both activities, they requested clear objectives and that moderators introduce themselves and share their feelings first which might encourage them to reveal their own feelings at the beginning.

Conclusions: Reflective practice and reflection on practice are part of the applied Mezirow's theory. Self

assessment gave them a better start, helped their critical thinking and gave them a better perspective on holistic patient care.

Take-home Messages: The delicate emotion is hard to comprehend. Recognition and following clinical reflection tools could eventually transform learning.

OTT-PA2-15
The viewpoints of medical graduates toward their achievement of expected competencies at Kerman Medical School, Iran, 2013

Habibeh Ahmadipour (*Kerman University of Medical Science, Community Medicine Department, Kerman, Iran*), Ameneh Mozafari (*Kerman University of Medical Science, Kerman*)

Background: Medical education systems need to be assured their graduates are competent in the tasks that are appropriate for their roles. The aim of this study was self assessment of the graduates' competency in Kerman Medical School.

Summary of Work: This research was a cross sectional study carried out in the medical school, Kerman, Iran. All medical graduates who graduated in 2010 and 2011 were selected. We analyzed self-assessment of graduates' competency. The instrument used was a questionnaire including 77 skills. The level of each competency was assessed on a five-point likert scale (no skill = 1, full skill =5). The validity of the questionnaire was confirmed by a number of experts and its reliability determined using Cronbach alpha 0.8. Data was analyzed by SPSS version 19.

Summary of Results: Fifty-seven graduates completed the questionnaire. Overall mean of competency (±SD) was 51.25(±12.19). They assessed their skills as intermediate. There was no difference in the overall mean between male and female. Performing suture procedure, taking accurate history, respect for patient rights and familiarity with medical websites had the highest scores. Calculation of drug doses, application of biostatistics, familiarity with regulation of sex change and performing venous cutdown procedure had the lowest scores.

Conclusions: Our results can be used by educational planners for curriculum reforms that should address the identified areas of competency that need further improvement.

Take-home Messages: Medical education systems need to be assured their graduates are competent in the tasks that are appropriate for their roles.

OTT-PB6: THE STUDENT/THE TEACHER

OTT-PB6-02
Comparison of clinical science comprehensive examination performance in different groups of medical students based on their cumulative GPA at 3rd year

C Ngarmukos (*Faculty of Medicine, Ramathibodi hospital, Mahidol University, Bangkok, Thailand*), S Liawwanich (*Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok*), N Plungpongpan (*Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok*), P Phuapradit (Faculty of Medicine, Ramathibodi Hospital, Mahidol University, *Bangkok*), S Wanvarie (*Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok*)

Background: Ramathibodi provides a 6-year medical curriculum, emphasizing preclinical science during years 1-3 and clinical science during years 4-6. In year 6, students are required to pass clinical science examinations (CS) including MCQ, Radiology SAQ, OSCE and MEQ. This study aimed to explore students' performance in preclinical years based on cumulative GPA at 3rd year (cGPA3) when compared to clinical science performance represented by scores of each CS.

Summary of Work: Graduates of year 2012 (N=150) were divided into 3 tertile groups according to cGPA3: high (H), middle (M) and lower (L). The scores of each CS were compared among the 3 groups using ANOVA. Data were shown as mean ± SD for MCQ, SAQ and OSCE and median for MEQ.

Summary of Results: Each group had 50 students (mean cGPA3 of H 3.56±0.14, M3.21±0.11, L 2.88±0.19). There were differences in mean CS score among the 3 groups. MCQ score: H 61.7± 3.9, M 56.8±3.8, L 54.64± 4.5 (p< 0.01); SAQ score: H 71.5± 5.5, M 67.5± 5.3, L 64.6 ± 5.8 (p<0.01), OSCE: H 68.5 ±4.7, M 65.6±4.8,L 61.3 ± 5.0 (p< 0.01). MEQ median score: H 65.5±3.7, M60.8±3.8, L56.8± 5.0 (p< 0.01.) Independent t-test also showed differences in score for all CS comparing H vs. M and M vs. L (p< 0.01).

Conclusions: Among the 3 performance groups of preclinical years, all show differences in scores of all clinical science examinations.

Take-home Messages: The findings suggest that students with lower performance in preclinical years may need more attention and support to improve their performance in later years.

OTT-PB6-03
Emotional Intelligence and correlation to anxiety and stress in Medical students studying in an All-English Curriculum in China

Sima Sarvari (*Shantou University Medical College, Medical Department, Shantou, People's Republic of China*)

Background: In today's society much emphasis is placed on 'academic intelligence' and less consideration is given to 'emotional intelligence' in the domains of learning. Language Anxiety presents potential problems "because it can interfere with the acquisition, retention and production of the new language" (MacIntyre & Gardner, 1991) by inhibiting the ability to process incoming language and hindering the process of knowledge acquisition.

Summary of Work: If young students are not emotionally developed to handle anxiety, this can have adverse effects on their learning ability, further paving the way for problems with self-esteem, self-confidence, and risk-taking ability (Crookall and Oxford,1991). Consequently, anxiety affects cognitive function causing anxious students to learn less and not be able to demonstrate what they have learned. Thus, experiencing more failure furthermore escalating their anxiety.

Summary of Results: The findings of our quantitative surveys shed some light onto students' emotional intelligence and their coping mechanisms with stress while studying in an All-English-Based Medical curriculum in China.

Conclusions: Emotional intelligence can be trained and taught to individuals low in emotional competencies to improve their abilities to better recognize, express and regulate their feelings.

Take-home Messages: Heads of departments and Deans of programmes can be encouraged to include training programs, for both staff and students alike, to identify and raise the emotional competencies of their students hence facilitating their learning.

OTT-PB6-05
Aligning assessment structures and support systems: A method for early flagging of struggling students

Claire L. Vogan (*Swansea University, College of Medicine, Swansea, UK)*, Ana L. Da Silva (*Swansea University, College of Medicine, Swansea*), Judy McKimm (*Swansea University, College of Medicine, Swansea*), Alice Hoon (Swansea University, *College of Medicine, Swansea*), Steve Capey (*Swansea University, College of Medicine, Swansea*)

Background: Medical students have been shown to be more vulnerable to depression, anxiety and psychological distress than the general population (Dyrbye et al 2006). Stressors are numerous, but final year examinations periods with a direct impact on progression have been identified as contributing towards exhaustion and burnout (Law, 2007). Furthermore such "high-stakes" examinations have been criticised due to the limited opportunity for feedback and remediation (Van der Vleuten,2000). Traditional approaches to curriculum alignment include curriculum design, learning and assessment methods and evaluation, but little attention has been paid to aligning support mechanisms as a key component of the curriculum.

Summary of Work: At Swansea (UK) Graduate-Entry Medicine the curriculum and assessment structures were designed to provide information on students' progress throughout the programme while maintaining validity and reliability of examinations. Integral to the design was an integration with support and remediation systems to maximise students' chances of progression.

Summary of Results: Each year consists of two exam periods with each period having mirrored assessment components that contribute to one final end-of-year mark for each component. This mark determines whether the student needs to undertake resit exams. After the first exam period, all students underwent a mid-year appraisal that helped them identify and define an action plan to address areas of weakness. Students that fell below the cut-scores are given advice on remediation from our assessment lead. Results from two cohorts will be presented and discussed.

Conclusions: Assessments structured to enable opportunities for feedback and remediation aligned with structured formalised support mechanisms can have a positive effect on students' wellbeing and progression without compromising the validity or reliability of assessments.

Take-home Messages: Curriculum alignment should include support systems. Structured opportunities for the early-flagging and remediation of struggling students requires close alignment between learning and teaching, assessment and support systems.

OTT-PB6-06
Meeting community needs: career paths of the first James Cook University medical school graduates

Aileen Traves (*James Cook University, General Practice and Rural Medicine, Cairns, Australia)*, Sarah Larkins (*James Cook University, General Practice and Rural Medicine, Townsville*), Tarun Sen Gupta (*James Cook University, General Practice and Rural Medicine, Townsville*), Richard Hays (Bond University*, Health Sciences and Medicine, Gold Coast*)

Background: The James Cook University (JCU) medical school was established in North Queensland, Australia in 2000 with a focus on rural, remote, Indigenous and tropical health. Over 60% of the first cohort of students selected had a local or rural background. The first graduates entered the workforce in 2006. Their career paths during the first five postgraduate years were studied to determine what factors influenced postgraduate careers.

Summary of Work: A longitudinal mixed methods study with data collection from the first cohort of JCU medical school graduates (n=58). Quantitative data was collected via an online survey and analysed using simple bivariate statistics. A subgroup of participants was interviewed to further explore their experiences. Interviews were transcribed in full and analysed with an iterative thematic analysis.

Summary of Results: Most graduates have worked in rural, remote, Indigenous or tropical health related employment, with 65% currently working in regional, rural or remote areas and 43% remaining in North Queensland. Only 18% of students intended to work in general practice when selected, yet 45% of the graduates have undertaken general practice training. More than 80% of graduates participate in teaching, and nearly 50% intend to continue participation in research.

Conclusions: The focus on rural, remote, Indigenous and tropical health and strong promotion of general practice throughout the JCU medical course encouraged graduates to pursue careers in these fields.

Take-home Messages: Selecting local and rural background students and providing adequate local postgraduate training opportunities improves local retention.

OTT-PB6-07
Australian Medical Schools Outcomes Database and Longitudinal Tracking (MSOD) Project

Angela Carberry (*Medical Deans Australia and New Zealand Inc., Sydney, Australia),* Don Roberton (*Medical Deans Australia and New Zealand Inc., Sydney*)

Background: The Medical Schools Outcomes Database and Longitudinal Tracking (MSOD) Project is a longitudinal project which aims to determine career objectives of all medical students in Australia.
Summary of Work: The Project enrolled its first commencing medical students in 2005. From 2005 – 2011, 27,403 questionnaire responses were provided.
Summary of Results: Graduate entry programme commencing medical student numbers were 9,113, and direct entry commencing student numbers were 9,548. Gender profiles for commencing students were 51.6% female: 48.4% male for graduate entry programmes, and 55.1% female: 44.9% male for undergraduate entry programmes (chi2 p<0.001).
For commencing medical students, the three most prevalent career choices were Surgery (27%), Paediatrics and Child Health (16%), and General Practice (14%).
For completing medical students the three most prevalent career choices were Surgery (19%), Adult Medicine (17%), and General Practice (13%).
There were significant differences in profiles of the top five potential career choices for commencing male and female students (chi2 661, 4 dof, p<0.001). There were also significant differences in the top five potential career choices for commencing students in graduate entry and undergraduate entry programmes (chi2 223, 4 dof, p<0.001)
High proportions of commencing, and completing, medical students stated that they wished to practice in future in capital cities.
Conclusions: Differences are evident in career choices according to gender and to type of medical school programme
Take-home Messages: Longitudinal linkage of medical student career intentions provide important national information for medical workforce planning and for monitoring of evolving specialist workforce provision.

OTT-PB6-08
Right task, wrong time. Influences on prescribing in an undergraduate OSCE

Jennifer Illingworth (*Royal Brompton Hospital, Anaesthesia, London, UK*), Stephanie Strachan (*King's College London, Critical Care, London*)

Background: Prescribing medication is routine practice for new doctors. However when studied an error rate of 8.4% was found, with time pressure an identified contributor (GMC EQUIP, 2009). Learning safe prescribing practice is an expected outcome of medical curricula, so could errors be promoted or prevented during undergraduate training?
King's College London medical undergraduate final year Objective Structured Clinical Examination (OSCE) includes assessment of prescription writing.
Performance in these stations is historically poor so we asked students their views on the prescribing stations.
Summary of Work: Students completed an anonymised questionnaire immediately following their OSCE.
Open questions designed to encourage student free text were used.
Students asked to consider factors influencing their performance.
Answers were subjected to thematic analysis.
Summary of Results: 181/224 (81%) students completed the questionnaire.
Themes emerging:
133/179 (74%) felt the stations were realistic.
75/181 (41%) students commented that time pressure influenced their prescribing, regardless of whether the station had been used previously.
"Expected station but still found time pressure issue, guessed dose".
Additional themes included fairness, fear of 'trick questions' and panic.
Conclusions: Time pressure was a key theme to emerge.
Students were aware of the safety implications of rushed prescribing but prioritized attempting to gain knowledge marks.
Professional behavior and patient safety need to be adequately rewarded in these assessments and students need to be aware of this.
Allowing insufficient time to complete station may encourage poor prescribing behavior.
Take-home Messages: The educational impact of assessments needs to be considered when limiting access to resources, including time, when creating assessment tools.

OTT-PB6-09
An evaluation of senior medical students as problem based learning (PBL) facilitators: Are they as good as experienced faculty?

Ashley W Newton (*University of Liverpool, Liverpool, UK),* Jennifer L Whiteley (*University of Liverpool, Liverpool*), David C M Taylor (*University of Liverpool, Liverpool*)

Background: Problem based learning (PBL) is used widely as a core component of medical school curriculums internationally. Existing studies explore the qualities of a good PBL facilitator, and highlight that the PBL facilitator can "make or break a session." Senior students within the same curriculum have an excellent working knowledge of it, and a broad understanding of the topics being delivered in the early years, which have been noted as very desirable characteristics for a good facilitator.

Summary of Work: Four senior medical students facilitated four groups of first year medical students. Four focus groups were held, covering 24 students at the end of the academic year after they had been exposed to both a student and faculty facilitator. Results were analysed using thematic analysis.

Summary of Results: Attitudes towards senior student facilitators were very positive. Key factors included their familiarity with the curriculum, active facilitating style including use of questioning, approachability, appreciation and adherence to the PBL process, and highlighting useful resources. There were also overarching references to the adoption of a mentoring style. The subtle disadvantages and direct comparisons with experienced faculty will also be presented.

Conclusions: Senior students can make excellent PBL facilitators - curriculum familiarity and approachability were key to success. Their potential utility is likely to be limited by their availability in terms of time and provision of appropriate supervision, rather than their effectiveness as facilitators.

Take-home Messages: Using senior students as PBL facilitators can be a fruitful experience for both the students and facilitators, provided appropriate training and monitoring systems are in place.

OTT-PB6-11
Developing a tool to measure and inform changes in clinical faculty engagement in a Canadian medical school

Tanya Matheson (*Dalhousie University, Faculty of Medicine, Operations and Policy, Halifax, Canada)*, Anne Weeden (*Dalhousie University, Faculty of Medicine, Operations, Halifax*), Diane LeBlanc (*Saint Mary's University, Department of Psychology, Halifax*), Kate Calnan (Saint Mary's University, *Department of Psychology, Halifax*), Thomas Marrie (*Dalhousie University, Faculty of Medicine, Halifax*), Diane Gorsky (*Dalhousie University, Faculty of Medicine, Operations and Policy, Halifax*)

Background: Faculty engagement drives success through increased dedication, retention, and productivity. Seeking direction to increase faculty engagement, we developed a tool to measure and diagnose specific elements of engagement for clinical faculty in Canada.

Summary of Work: We defined clinical faculty engagement as encompassing affective, cognitive, and behavioural aspects of involvement in the institution. Using this definition, a model of potential predictors and outcomes for faculty engagement was created and used to select and modify potential items from existing validated measures. Following a pilot run, a final survey with 52 measurement items and 6 demographic items was used to collect data from 368 full- and part-time clinical faculty (a response rate of 23.5%).

Summary of Results: Ten predictor variables accounted for 46% of the variance in participants' reported behavioural involvement, and fourteen predictor variables accounted for 63.3% of the variance in psychological well-being. The results also identified some interesting patterns in responses based on gender, years of appointment, role, and primary appointment.

Conclusions: Results suggest that strategy penetration, involvement in decision making, and the quality of working relationships are key to advancing faculty engagement. Using these results evidence-based intervention strategies will be developed that are expected to yield desired results using a strengths-based approach to organizational development.

Take-home Messages: The development of a Canadian-centric clinical faculty engagement survey is a first step to organizational change in developing a culture which supports and engages clinical faculty. Sharing and refining this tool and other strategies to improve faculty engagement should be a collaborative venture to benefit all medical schools.

OTT-PB6-12
Factors Affecting Resident Evaluation of Faculty Entrustment of Autonomy

M Wolff (*University of Michigan Medical School, Ann Arbor, USA),* B Bassin, L Hopson, K Saxon, N Juneja, W Brent Stansfield, Presenter: Sally A Santen (*University of Michigan Medical School, Ann Arbor, MI*)

Background: The role of supervising faculty is to entrust residents with progressive autonomy as residents advance to independent practice. The objective of this study was to identify faculty factors associated with resident perception of increased autonomy.

Summary of Work: Trainees anonymously completed a 9-item faculty evaluations. Performance domains included: resident entrustment, communication, clinical skills, feedback, availability. Each item was completed using a 5-point scale from poor to excellent. Pearson correlations between entrustment and faculty performance factors were determined. Cronbach's alpha and factor analysis were conducted to determine the dimensions of faculty factors.

Summary of Results: 52 residents completed 2738 evaluations of 104 faculty. Aggregate data showed very high correlations between all nine domains ($r=0.83$-$0.94, p<0.001$). There were strong correlations across all distinct characteristics surveyed with responses generally being all favorable or unfavorable. Residents did not view faculty as more or less entrusting compared to other characteristics. Cronbach's alpha was 0.99. Using factor analysis, all of the items loaded on a single dimension that predicted 91.5% of the variance.

Conclusions: Residents appear to perceive faculty performance and skills as either in a positive or negative context and do not differentiate between an individual faculty's performance measures such as entrustment or clinical skills.

Take-home Messages: Evaluation of faculty needs to be further explored to better provide feedback to faculty

OTT-PB6-13
Building Learning Communities in a Changing Curriculum

Dana Sayre-Stanhope (*Emory University School of Medicine, Family and Preventive Medicine, Atlanta, USA)*, M Mayfield (*Emory University School of Medicine, Family and Preventive Medicine, Atlanta*)

Background: In 2011 the Emory University School of Medicine's Physician Assistant program implemented a major curriculum change which included the addition of learning communities. Initially these communities focused on two main goals: promoting team based care and through case discussions, enhancing clinical reasoning skills.

Summary of Work: Students are carefully assigned to societies to achieve diversity of background, age and prior health care experience. Each society has two faculty mentors who serve as advisors to the students for the duration of their educational experience. Meeting weekly, students and faculty engaged in case-based discussions primarily focused on the solution of a specific clinical problem designed to correlate with current didactic studies.

Summary of Results: Students provide quarterly evaluations of their peers, mentors and the society's function overall. Metrics focus on student professionalism, faculty responsiveness, relevance of case problems, and overall satisfaction. Students report a high degree of satisfaction with the learning community model, with the amount of faculty guidance received and with the team-based activities. They were less satisfied with the construct of the cases which they believed were too constraining and did not provide sufficient opportunity for exploration.

Conclusions: The learning community model provides an elegant way to address multiple curriculum issues. Building on the didactic and clinical strength of the program, the learning communities enhance team-based practice through shared learning activities while reinforcing clinical reasoning and self-directed learning. Lessons learned include redesigning cases to provide more fluid discussion and expanding opportunities to discuss humanistic aspects of medical decision making.

Monday April 28
8:30AM – 10:00AM

PLENARY: MAKING TEACHING AND ASSESSMENT RELEVANT

Cheating in Assessments: Doing, Detecting and Deterring

Speaker: *Professor Trudie E. Roberts BSc., MB.ChB, PhD, FRCP, FHEA (Director – Leeds Institute of Medical Education, University of Leeds, England)*

Cheating is commonly defined as breaking the rules to gain advantage. How common is cheating in medical school examinations? It probably occurs more frequently than we would like to think. Why do students do it and how do they justify it when found out? Is cheating more morally wrong in would-be doctors than in other students? Are some types of cheating worse than others? How can we expose this type of deception and how can we deter students from deciding to cheat in assessments. In this presentation I will explore these issues and look at the dilemma posed by medical students and trainee doctors who are academically dishonest.

What if they aren't playing our game? : Education theories, curriculum intent and learner goals re-examined

Speaker: *Glenn Regehr, PhD*
Professor, Department of Surgery
Associate Director (Research), Centre for Health Education Scholarship, University of British Columbia

As education theorists and innovators we carefully construct curricular strategies and practices based on our goals regarding what we want students to acquire and on our understanding of what will maximize that acquisition. Most of these strategies and practices are designed with the assumption that students appreciate our goals and adopt them as their own ... that students are willing participants in our plans for them and play along. It is likely, however, that several social and environmental factors of our own making lead our students to a different set of goals that are focused around efficiently and effectively accomplishing the tasks they perceive as necessary to succeed in the system. If we are trying to make teaching and assessment more relevant to our goals for the students, therefore, perhaps the question we should be challenging ourselves with is not what educational strategies would ideally maximize the acquisition of these goals, but rather how do we create willing partners in our students, and what do our grand educational strategies reduce to if they don't play along?

Monday April 28
10:30AM – 12:00PM

SYMPOSIA

OTT-SD-1
Faculty Development and Learner Assessment: The Missing Link

Presenters: Yvonne Steinert, Beth-Ann M Cummings, Robert Sternszus (*Centre for Medical Education, Faculty of Medicine, McGill University, Montreal*)

Summary: The assessment of learners at all levels of the educational continuum is the focus of much debate and research, as are specific aspects of assessment including standard setting, psychometric properties of assessment methods, and the value of an assessment program. However, the role of clinical teachers in assessing students and residents, and the need to prepare faculty members to observe critically, question effectively, and judge appropriately, is often neglected. The goal of this symposium is to highlight the role that faculty development can play in promoting reliable, valid, and fair learner assessments. The symposium will highlight common approaches to preparing faculty for their role as assessors as well as the proposed content of a faculty development curriculum that includes the goals and principles of learner assessment, an overview of diverse assessment methods (including their strengths and limitations), standard setting and 'inter-rater' reliability, and the role of contextual factors in assessment. It has been said that the lack of agreement among faculty members – and the difficulty of assessing learners in a meaningful way – is a threat to the reliability and validity of decisions made about learner competence. The goal of this symposium is to address how faculty development can help to overcome this challenge.

OTT-SD-2
What is Excellence in Assessment?

Presenters: Trudie Roberts (*University of Leeds, Leeds*), David Wilkinson (*Macquarie University, Sydney*), Ronald Harden (*AMEE, Dundee*)

Summary: There is increasing acceptance of the need to recognise, alongside excellence in research in medical schools, international excellence in education. The symposium examines the concept of excellence in education, in particular excellence in assessment. The ASPIRE programme was established to go beyond the traditional accreditation process and to recognise that the education programme in a medical school can be subjected to peer review against an agreed set of standards or benchmarks that identify world-class excellence in education. The criteria for excellence in assessment established by an international panel as part of the ASPIRE initiative will be explored. The ASPIRE initiative encourages and promotes outstanding performance and excellence in education and recognises that the characteristics of excellence will vary according to local contexts.

WORKSHOPS

OTT-WD-1
Simulation based assessments of competency & performance to provide a capable healthcare workforce

Presenter(s): Thomas Gale (*Plymouth University Peninsula Schools of Medicine and Dentistry, Director of Assessment, Plymouth, United Kingdom*), Martin Roberts (*Plymouth University Peninsula Schools of Medicine and Dentistry, Dept of Assessment, Plymouth, United Kingdom*), Ian Anderson (*Plymouth Hospitals NHS Trust, Director of Peninsula Simulation Suite, Plymouth*)

Background: Many undergraduate & postgraduate curricula focus on competency based assessments to define learning outcomes such as those listed in GMC Tomorrows Doctors 2009. However, graduates entering the complex world of healthcare need to be able to perform at higher levels of skill acquisition than mere competence. Simulation is increasingly used to measure proficiency of technical & non-technical skills in various formats plus a host of metrics have been validated for this purpose.

Intended Outcomes:
1. Familiarise attendees with validated metrics used for simulation based assessment of technical & non-technical skills
2. Increase understanding of how simulation can be used to assess competence, proficiency & higher levels of performance
3. Understand differences in using checklist versus domain specific scoring in different contexts
4. Examine ways of standardising simulation scenarios

Structure: Short presentations will cover important concepts to consider when using simulation in formative & summative assessment within medical curricula and in high stakes scenarios such as recruitment.

Video demonstrations will be used to highlight strengths & limitations of specific examples of simulation based assessments.

Facilitated group discussions will follow the presentations / video demonstrations to cover:
1) Appropriate use of checklists versus domain specific scoring in different contexts.
2) Techniques to standardise scenarios including linear programming.
3) Use of participant versus observer raters in high stakes assessments.
4) Advantages / disadvantages of simulation based assessments versus other methods.

Who Should Attend: Undergraduate / postgraduate clinical tutors
Simulation providers / experts

Directors of education / simulation / assessment
Interviewers for undergraduate & postgraduate
programmes
Level of Workshop: Intermediate

OTT-WD-2
Using a TOSCE (or Team OSCE) to Evaluate Interprofessional Education

Presenter(s): Dawn Schocken (*University of South Florida Health, Morsani College of Medicine, Tampa, FL, United States*)

Background: A team of IPE educators developed modules for early clinical healthcare professionals to work in teams. Using the IPEC Competencies, a total of five modules were developed. Each module included simulation to reinforce the educational message. The culminating event was iCare: Integrity, Compassion, Accountability, Respect and Excellence. This module was a hospital-based clinical event for early clinical learners.
Following the didactic portion of iCare, a TOSCE, a Team-based OSCE, was run as a formative evaluation tool to help document students' performance. The TOSCE evaluated key benchmarks for identification. These benchmarks included defined roles and responsibilities, interprofessional communication skills, teams and teamwork and values based practice demonstrating patient centered care.
Intended Outcomes: This workshop is developed to demonstrate the TOSCE, allowing course participants to develop their own TOSCE. The presenter will assist the course participants in tailoring a TOSCE for their use.
At the end of this session, the participants will be able to:
1. Develop a case scenario for the TOSCE for early clinical learners.
2. Demonstrate clear objectives for evaluation of students during the TOSCE.
3. Discuss methods to train SPs to address the needs of clinical scenarios.
4. Analyze methods of evaluation to determine the effectiveness to meet the IPEC Competencies.
Structure: The participants will:
Discuss the TOSCE and demonstrate a TOSCE
Round table discussion of TOSCE development
Large group demonstration of table's TOSCE event
Discussion of evaluation methods
Wrap up
Who Should Attend: Course participants who work with early clinical and/or preclinical learners.
Level of Workshop: Introductory

OTT-WD-3
How might TQM enhance your workplace-based assessment?

Presenter(s): Simon Gay (*Keele University School of Medicine, , Keele, United Kingdom*), Maggie Bartlett (*Keele University School of Medicine, Keele, United*

Kingdom), Janet Lefroy (*Keele University School of Medicine, Keele, United Kingdom),* Robert McKinley (*Keele University School of Medicine, Keele, United Kingdom*)

Background: Keele Medical School has a Total Quality Management (TQM) approach to delivering 1400+ WBAs per annum. Facilitators will use this experience to assist workshop participants in the exploration of the possible benefits of TQM for WBA within their own organisation.
Intended Outcomes: Following the workshop participants will be able to:
• Recognize opportunities to strengthen the performance of WBA within their organisation.
• Identify a wide range of strategies to promote the quality of WBA within their clinical context.
• Appreciate the value of the synergistic application of simultaneous continuous quality improvement strategies to WBA.
Structure: 1. Introduction: to orientate delegates to the principles of TQM to ensure a shared starting point.
2. Clarifying Challenges: In small groups, delegates will have opportunity to discuss the challenges to WBA within their own curriculum.
3. Presentation: brief outline of the TQM approach to WBAs at Keele.
4. Practicalities: In small groups, delegates will be able to explore how the strategies contained within Keele's TQM approach might offer solutions to the previously identified challenges to WBA within their own context.
5. Aspiration to Inspiration: closing discussion will bring together the reflections of attendees and focus on the challenges of WBA and the potential solutions TQM offers.
Who Should Attend: Those with an interest in introducing or increasing the WBA contribution to assessment within their institution.
Level of Workshop: Intermediate

OTT-WD-4
Reflection, Integration and Assessment (RIA) week in an integrated competency-based curriculum model

Presenter(s): Alice Fornari (*Hofstra North Shore-LIJ School of Medicine, Science Education, Hempstead, United States*), Judith Brenner (*Hofstra North Shore-LIJ School of Medicine, Internal Medicine, Hempstead, United States*), Maria Barilla-Labarca (*Hofstra North Shore-LIJ School of Medicine, Internal Medicine, Hempstead, United States),* Lauren Block (*Hofstra North Shore-LIJ School of Medicine, Internal Medicine, Hempstead, United States*)

Background: It is known that medical students value that which is assessed, therefore the connection between curricula and assessment must be seamless. How to achieve and maintain the ideal that "assessment drives learning", using diverse assessment modalities, is a challenge. As a new School of Medicine we have developed an innovative curriculum that is supported by integrated assessment and reflection weeks. The goal of

these weeks, which occur eight times over two years, is developmental formative/ summative assessment individualized for each learner and aligned to core competencies of patient care, medical knowledge, interpersonal/communications skills and professionalism. The outcome data of formative assessment activities is part of a longitudinal feedback process by "core" faculty assigned to individual students. Feedback from summative assessments is also reported back to students.

Intended Outcomes: 1) Engage in discussion about the core principle: "assessment drives learning".

2) Describe an assessment week that is: formative/summative, developmental, aligned with integrated competency-based curriculum strategies, involves collaborative leadership, and diverse faculty resources.

3) Brainstorm about development of assessment systems that meet educational principles connected to core competencies.

Structure: Small table discussions on how to achieve "assessment drives learning" as a core principle and challenges to overcome, with report out (30 minutes) Overview/Q &A of our assessment weeks, including resource allocation (15 minutes)
Video clip of a student progressing in clinical skills (30 minutes: 8 minute clips each with structured discussions)
Take home points/next steps (15 minutes)
Who Should Attend: Faculty involved with assessment of early medical students
Level of Workshop: Intermediate

OTT-WD-5
Uniform diversity: How to achieve standardization in live clinical assessments while still retaining local autonomy

Presenter(s): Richard Turner (*University of Tasmania, School of Medicine, Hobart, Australia*), Peta-Ann Teague (*James Cook University, School of Medicine and Dentistry, Townsville, Australia*), Bunmi Malau-Aduli (*James Cook University, School of Medicine and Dentistry, Townsville, Australia),* Karen D'Souza (*Deakin University, School of Medicine, Geelong, Australia*)

Background: Objective Structured Clinical Examinations (OSCEs) are integral to assessment in medical education. Creating and delivering quality OSCE items can be time-consuming and intellectually demanding. Sharing items among like-minded institutions alleviates some of this burden, while also addressing the desire for standardization of assessment at a national level. This often must be harmonized with certain content, process and procedural specifications imposed at a local level. The ACCLAiM project was founded by a number of Australian Medical Schools with these goals in mind. Those who have been involved in the project are now in a position to share their acquired expertise.
Intended Outcomes: Participants will emerge with a framework for developing, implementing and evaluating collaborative OSCE items. They will also acquire

experience in skills required by examiners and item creators.
Structure: Presenters will facilitate discussion around various topics relating to inter-institutional sharing of live assessment items. They will also provide illustration from their own experiences as part of a collaborative demonstration project and evidence gleaned from the literature. Topics covered will include inter-examiner calibration, objective scoring, standard setting, curriculum alignment, and iterative item improvement. The session will be interactive, with practical exercises in inter-examiner calibration and consensus-driven item improvement.
Who Should Attend: This workshop is intended for all medical educators who engage in live clinical assessments, and wish to share assessment items in order to maximise intellectual capital and continuously improve item performance. Familiarity with the OSCE paradigm is a p
Level of Workshop: Intermediate

OTT-WD-7
Learner-Centered Remediation for Clinical Skills: Diagnosing, Reflecting and Monitoring

Presenter(s): Norma Saks (*Rutgers Robert Wood Johnson Medical School, Piscataway, United States*), Carol Capello (*Weill Cornell Medical College, New York, United States*), Felise Milan (*Albert Einstein College of Medicine of Yeshiva University, New York, United States),* Eilzabeth Kachur (*Medical Education Development, New York, United States*)

Background: The OSCEs have been a mainstay for teaching and evaluating clinical skills across the medical education continuum. The format is useful for identifying learners with deficits in various clinical domains, including communication skills, medical knowledge and clinical reasoning, and professionalism. A robust remediation curriculum requires time and resources. How to best remediate learners who fail the OSCEs remains challenging.
Intended Outcomes: By the end of this workshop, participants will be able to
1. Diagnose domains where learners require remediation
2. Provide targeted feedback to learners using learner-centered approach
3. Plan a monitoring system for learners undergoing remediation.
Structure: This interactive workshop will harness the expertise of medical educators from various institutions who design and teach clinical skills and have implemented remediation programs. The workshop will include large and small group activities, and will consist of 1) Diagnosis: After introduction, faculty will discuss best practices for diagnosing domains where learners demonstrate deficits, using the ACGME & CanMEDS frameworks. Large group discussion will center around videotaped OSCEs from workshop faculty; 2) Reflection: Through guided reflection, faculty will demonstrate a learner-centered approach for providing constructive

feedback and formulating collaborative action steps. Small group role-plays will provide participants with skills that apply important concepts; 3) Monitoring: Faculty will discuss "prescription for action" for learners and strategies for monitoring improvement. Faculty will then lead small group discussions on feasibility and challenges of implementation at individual institutions, and explore best strategies for next steps. Participants will leave with a tool kit of materials from institutions with successful remediation programs.

Who Should Attend: Directors and faculty of clinical skills centers, and faculty involved in teaching, assessing and remediating clinical skills.

Level of Workshop: Introductory

ORALS

OTT-OD-1: POSTGRADUATE ASSESSMENT

OTT-OD-1-1
General Practice Out of Programme Study Sessions

Aurelia Butcher (*Wessex School of General Practice, Winchester, UK)*, <u>Jonathan Rial</u> (*Wessex School of General Practice, Winchester*), Johnny Lyon-Maris (*Southampton GPEU, Wessex Deanery, Southampton*)

Background: Failure to complete specialty training within the proscribed timeframe is a growing problem for those managing specialty training for general practice.

Seven trainees from Wessex did not obtain their certificate of completion of training, and were placed out of programme for six months, as they could not be offered extensions. Once out of programme, it is difficult for such trainees to maintain continuity in their learning and assessment.

Summary of Work: A short-term programme of bi-weekly study sessions was set up and facilitated by recently qualified GPs with an interest in education ('Near Peers'). The programme included Clinical Skills Assessment (CSA) practice with feedback, as well as careers advice. Feedback data was gathered over the programme by the facilitators.

Summary of Results: All trainees attended the full programme. Facilitators found that they improved their consultation skills and in particular gained insight into the purpose of the CSA assessment and their professional development as doctors.

Conclusions: The programme allowed trainees to keep in touch with general practice and provided a support network for trainees OOP. For them the experience prompted insight and a reassessment of their learning.

Take-home Messages: Regular facilitated study sessions are helpful for trainees who are out of programme. They can maintain continuity of learning as well as promote insight. Although short term extra funding is required for sessions, in the long run it is anticipated

that the outcome for trainees will be positive and cost effective.

OTT-OD-1-2
Implementing best practices for in-training evaluation reports (ITERs)

<u>Marla Nayer</u> (*University of Toronto, Post Graduate Medical Education, Toronto, Canada)*, Susan Glover Takahashi (*University of Toronto, Post Graduate Medical Education, Toronto*), Khushnoor Adatia (*University of Toronto, Post Graduate Medical Education, Toronto*), Caroline Abrams (*University of Toronto, Post Graduate Medical Education, Toronto*), Glen Bandiera (*University of Toronto, Post Graduate Medical Education, Toronto*)

Background: In-training evaluation reports (ITERs) are used to systematically document the performance of learners in a clinical setting. A post-graduate medical education advisory committee sets standards for the ITERs. The Postgraduate Medical Education (PGME) office provides support to the programs in complying with the standards.

Summary of Work: In 2012 the PGME advisory committee developed new standards for ITERs. A key operational decision was implemented that required residency program directors (PDs) to submit new ITERs to the PGME office for review by an educational consultant. A compliance template was created and PDs were provided with feedback. ITERs were then re-submitted and only approved for posting once they met the approved guidelines.

Summary of Results: In the first year of this process 181 ITERs were reviewed. All ITERs were linked to the goals and objectives (previously none had been so linked). The length of the submitted ITERs went from 12 to 133 ratings, median of 26. Following revisions the length went from 8 to 36 ratings, median of 19. Significant PGME staff time was devoted to providing one-on-one support to PDs in the revision of their ITERs.

Conclusions: Significant improvements to ITERs occurred during this review and revision process. This included providing greater clarity to the ratings, providing greater clarity of the level of expected performance, focusing on the rotation goals and objectives, and shortening the ITERs to a more appropriate length.

Take-home Messages: ITERs can be significantly focused and shortened with the support of educational leadership and clear standards.

OTT-OD-1-3
'Telemedicine' video-links for supervised learning events with junior doctors in the clinical environment

<u>Jim Price</u> (*Brighton and Sussex Medical School, Division of Medical Education, Brighton, UK)*, Susie Calderbank (*Western Sussex Hospitals NHS Foundation Trust, Department of Anaesthetics, Worthing*), Lyndsey Forbes (*Western Sussex Hospitals NHS Foundation Trust, Department of Anaesthetics, Chichester*), Richard Venn

(*Western Sussex Hospitals NHS Foundation Trust, Department of Anaesthetics, Worthing*), Chris Smith (*Western Sussex Hospitals NHS Foundation Trust, Department of Anaesthetics, Chichester*)

Background: Supervised learning events (SLEs) are an important aspect of the curriculum for junior doctors. In the UK, getting the required assessments carried out can be difficult for learners due to accessibility of the assessors, poor planning by the trainees, and frequently delayed and less satisfactory feedback from the assessors.

Summary of Work: Secure video-conferencing was used as part of the Health Education England national pilot scheme 'Better Training, Better Care'. The 15 core competencies required by Foundation Year 1 doctors were assessed in 42 SLEs at a distance using the video-link, both in a simulated environment, and subsequently in the real clinical environment, using a variety of hardware.

Summary of Results: All core competencies were deemed assessable in both clinical environments. The technology was generally reliable, and when fully functional, the learning experience was deemed as good or excellent. Patient data were accessible on-line, as was the learner's eportfolio, which was completed during the SLE. Feedback was given immediately in a timely and interactive way, something highlighted by learners and assessors to be an improvement on the current process.

Conclusions: The use of such a video-link offers a practicable, efficient, convenient and secure method for assessing SLEs. Planning the SLE may be more efficient for both assessors and learners. The quality of the educational interaction is usually as good as a face-to-face encounter, and offers other significant advantages.

Take-home Messages: Using a secure video-link for SLEs in a planned way is efficient and may provide an improved educational interaction for both learner and assessor.

OTT-OD-1-4
Towards development and validation of an intraoperative objective assessment tool for evaluation of robot-assisted laparoscopic radical prostatectomy

Jen Hoogenes (*McMaster University, Department of Surgery, Division of Urology, Hamilton)*, Christopher T. Morris (*McMaster University, Department of Surgery, Division of Urology, Hamilton, Canada*), Bobby Shayegan (*McMaster University, Department of Surgery, Division of Urology, Hamilton*), Ranil Sonnadara (*McMaster University, Department of Surgery, Hamilton*), Edward D. Matsumoto (McMaster University, *Department of Surgery, Hamilton*)

Background: No intraoperative assessment tool currently exists for robotic-assisted laparoscopic radical prostatectomy (RARP), and the development, validation, and implementation of such a tool is timely. As a first step in this process, we conducted a Delphi study to establish an inventory of critical procedural steps for RARP to provide a basis for an assessment instrument for new urologic robotic surgeons.

Summary of Work: Thirteen expert RARP surgeons from North America were recruited for our Delphi panel. Starting with 13 critical RARP procedural steps and 60 sub-steps derived from our institutional modified Delphi process, international experts iteratively rated each of these steps on a 5-point Likert scale of agreement for inclusion in the final assessment tool. Qualitative feedback was elicited for each item and a video of a de-identified RARP case was provided for reference as required. Responses were compiled and the inventory was edited through three iterations, after which 100% consensus was achieved among the expert panel.

Summary of Results: Throughout three iterations, the inventory was reduced by 10 sub-steps, a skip pattern was incorporated, and edits were made as necessary. No attrition occurred until the final round (N=12). The final inventory comprised 13 primary steps and 50 sub-steps.

Conclusions: This comprehensive inventory of RARP steps will be used to develop a psychometrically sound intraoperative RARP assessment tool for training future urologic robotic surgeons. The validation phase is in progress via expert RARP practitioners as raters.

Take-home Messages: Development and validation of this intraoperative RARP assessment instrument will lead the way for curriculum change in international urological robotic training.

OTT-OD-1-5
Specialty-Specific Clinical Performance Milestones Evaluation in an Introductory Anesthesiology Program: Correlation between Composite Independence Scores & Recommended Supervisory Levels

Michael G. Richardson (*Vanderbilt University School of Medicine, Anesthesiology, Nashville, USA)*

Background: This 1st month program ensures that novice anesthesia residents progress from novice to proficient (capable of functioning safely with routine supervision) in 20 clinical days.

Summary of Work: 24 specialty-specific core clinical performance milestones were developed & incorporated into a web-based REDCap evaluation tool. Resident clinical preceptors reported level of independent performance for each milestone at the end of 2-4 days of direct observation. Milestones performance ratings included three levels of independence, as well as the option "not observed/not tried". Each rating was assigned a numerical score (1, 2, 3) in ascending order of independent performance (not observed = 0). Numerical score were absent from evaluation forms. A "composite independence score" (CIP) was calculated for each evaluation (max, 24x3=72). The evaluation solicited one judgment (recommended supervision level, RSL): "Based on your observations, what level of supervision do you recommend for coming days?" Three options: 1) Continue 1:1, 2) "Semi-solo" (may be left alone for brief / appropriate periods); 3) Routine (2 trainees:1 supervisor).

Summary of Results: 25 preceptors completed 93 evaluations (100%) for 15 trainees. CIP progressed over 4 weeks, with good inter-rater reliability. There was strong correlation between RSL & CIP. The cohort's rate of development of proficiency varied for various core elements—easier, less complex skills and those addressed early in the curriculum developed quickly, while more complex skills developed more slowly.

Conclusions: Inter-rater reliability, correlation between RSL & CIP, and cohort proficiency variation with skill sets evidence tool validity.

Take-home Messages: This clinical milestone evaluation tool is easily implemented and yields valid performance evaluation data.

OTT-OD-1-6
New school, new assessments: A psychometric analysis of the assessment programme at the National School of Healthcare Science, England

Suzanne Chamberlain (*National School of Healthcare Science, Birmingham, UK)*, Jennie Bell (*National School of Healthcare Science, Birmingham*), Chris Gibson (*National School of Healthcare Science, Birmingham*), Val Davison (National School of Healthcare Science*, Birmingham*), Sue Hill (*National Health Service*)

Background: The National School of Healthcare Science was established in 2009 as part of a government initiative for modernising healthcare science education and training in England. Trainees on the pre-registration postgraduate training programme for clinical scientists in genetics complete an electronic portfolio of competencies, 28 workplace based assessments and a 12-station Objective Structured Final Assessment (OSFA).

Summary of Work: Assessment data were analysed using classical test theory and univariate analysis techniques in order to produce early indicators of the psychometric properties of each tool and explore the performance profiles of the genetics trainees.

Summary of Results: The OSFA produced reliability coefficients that indicated sound internal consistency reliability within and across stations (within-station Cronbach's α average = 0.73; OSFA Cronbach's α = 0.70). The trainees' workplace based assessments tended towards the higher end of the performance grade scale.

Conclusions: First analysis of the school's assessment data suggests that there is appropriate coverage of domains, balance between formative and summative assessment, and reassuring indicators of psychometric performance. The explanation for the clustering of trainees' workplace grades may lie within the recruitment process and the selection of highly competent trainees.

Take-home Messages: First analysis of the assessment programme at the National School of Healthcare Science suggests that the assessments are adequately blueprinted to the curriculum and working as intended.

OTT-OD-2: RESEARCH AND ASSESSMENT

OTT-OD-2-1
Changing Faces, future research methods training in medical education courses: views from the coal face

Lesley Pugsley (*Cardiff University, School of Postgraduate Medical and Dental Education, Cardiff, uk)*, Lynne Allery (*Cardiff University, School of Postgraduate Medical and Dental Education, Cardiff*), Janet Macdonald (*Cardiff University, School of Postgraduate Medical and Dental Education, Cardiff*)

Background: In 2005 our ASME funded project scoping research training on UK MSc & PhD programmes in Medical Education found limited numbers of courses, with wide variations between each.
Since then there has been increased recognition for medical education career pathway, resulting in an exponential increase in available programmes. It seemed timely to undertake a second iteration of the study to provide a comparative account of research methods training available then and now.

Summary of Work: In the early summer 2013, we undertook a   review of online databases to determine the nature and provision of current research methods training and support and then surveyed delegates at the ASME Conference for grassroots reactions. To enable us to develop more in-depth viewpoints, in autumn 2013 one to one semi structured interviews were conducted with course directors.

Summary of Results: Now, as then, similar patterns of variance have emerged in relation to the nature and provision of research training and support offered across the UK. Grassroots responses indicated support for a rigorous grounding in research methods training with experienced and appropriately qualified dissertation supervisors; declared contact hours for dissertation supervision and summatively assessed research proposals. Analysis of the interview data will allow further exploration of these themes.

Conclusions: Core standards in educational research must be identified, supported and assessed in medical education qualiifications.

Take-home Messages: Robust research in medical education demands a firm grounding in educational research and robust supervisory support.

OTT-OD-2-2
Assessment of primary research study quality in medical education systematic reviews

J.E.F. Fitzgerald (*Barnet Hospital, London, UK)*, A.G. Care (*Liverpool Women's Hospital, Liverpool*)

Background: Medical education research is increasingly looking to evidence-based practice and one result has been the increase in systematic literature reviews. Robust quality assessment can be more challenging for educational literature compared to clinical studies, particularly for novice researchers, as criterion must

accommodate for qualitative and quantitative data across a wide breadth of areas. This work investigated the frequency of primary study quality appraisal through structured criteria in medical education systematic reviews, the criteria used, and the reporting of this.

Summary of Work: Thirty systematic reviews published across 6 medical education journals were examined for evidence of an appraisal process for the primary literature reviewed, details of the specific process and inclusion of the results of this.

Summary of Results: The literature search yielded 19,043 results, 83 of which reported a systematic review (0.4%). Only 1 was published in 2000; in 2011 that had risen to 16. Of the 30 most recent, 24 (80%) described using one or more appraisal processes. Eight (33%) reported quality grading scores at an individual paper level. Sixteen utilised previously published quality-appraisal instruments, or variations thereof. The most commonly cited was the Medical Education Research Quality Instrument (MERSQI).

Conclusions: The appraisal of primary evidence in medical education systematic reviews remains variable. Appraisal instruments are heterogeneous and the clear reporting of findings is infrequent. The lack of clear framework for this creates difficulty in interpreting the strength of review findings.

Take-home Messages: Greater focus is required on the assessment of primary evidence quality in the synthesis of medical education reviews.

OTT-OD-2-3
Latent variables are dead, long live latent variables! On the relation between case specificity and psychometric models

Stefan K Schauber (*Charité - Universitätsmedizin Berlin, Institute of Medical Sociology, Berlin, Germany)*, Zineb M Nouns (*Charité - Universitätsmedizin Berlin, Berlin*), Sebastian Schubert (*Charité - Universitätsmedizin Berlin, Berlin*), Susanne Dettmer (Charité - Universitätsmedizin Berlin*, Institute of Medical Sociology, Berlin*)

Background: Case specificity in the form of high residual variances are reported as a common phenomenon in medical competence measurement. Starting from this empirical evidence, we explored the appropriateness of latent variable modelling in medical education, since these high residual variances are deemed to be problematic from various perspectives.

Summary of Work: We reviewed existing conceptions on the relation between case specificity and latent variables. On the other hand we introduce an alternative viewpoint on this issue. To this end, we systematized which efforts have been made to bring case specificity and the psychometric theory of latent variables into accordance. Critically, we highlight how different conceptions of measurement error are associated with different interpretations of large residual variances and thus case specificity.

Summary of Results: From a psychometric perspective, the discussion on case specificity can be interpreted as suggesting different forms of dimensionality of medical competence. This incorporates assumptions on uni-

dimensionality versus multi-dimensionality on the one hand. By contrast, case specificity has also been deemed an indicator of the inappropriateness of the concept of dimensionality. However we point at the critical possibility that this interpretation of case specificity largely depended on the concept of measurement precision. This concept differs critically between classical test theory and generalizability on the one hand and item response theory on the other hand.

Conclusions: The interpretation of case specificity seems to be dependent on different concepts of measurement precision. Consequently, implications for future directions of psychometric application in medical education are discussed.

Take-home Messages: Is case specificity a "bug" or a "feature"?

OTT-OD-2-4
Validity and Validation of Assessment: A survey of researchers in health sciences education

Christina St-Onge (*Université de Sherbrooke, Médecine, Sherbrooke, Canada)*, Meredith Young (*McGill University, Center for Medical Education, Sherbrooke*)

Background: Validity could be considered a « god term » when discussing assessment, as a label of 'valid' can be used to endorse a given assessment strategy. However, recent views stipulate validity is a continual process rather than a destination. These differences led us to explore how researchers view validity, and whether views aligned with findings of a recent literature analysis (validity conceptualized as a test characteristic, an argument or a social responsibility).

Summary of Work: Researchers in health sciences education were invited to complete a survey with two open-ended questions (definition of validity and validity evidences), 14 items presenting conceptions of validity and four items regarding involvement in assessment (6-point Likert scale; 1=totally disagree to 6=totally agree). Validity definitions were coded using identified validity conceptions. Descriptive analysis was conducted on participants' involvement in assessment. Rated items were categorized by validity conception, and these subscores were compared.

Summary of Results: 35 participants (64%) completed the survey. Open-ended definitions of validity aligned with all three previously reported validity conceptions. Bi-modal distributions were observed for involvement in assessment. Repeated measure ANOVA showed that participants least agreed with the conception of validity as a test characteristic (M=50.75%, vs. 73.40% and 79.04% for validity as argument and social responsibility conceptions, $p<.001$).

Conclusions: Results from this study corroborate findings from a recent discourse analysis regarding identified validity conceptions in the literature.

Take-home Messages: Although participants endorsed the conception of validity as a test characteristic least frequently, results from a recent literature review suggest that it may answer specific needs and thus merits consideration.

OTT-OD-2-5
One measurement approach, two functions: applying the Rasch model for program evaluation and learner assessment for a simulation-based lumbar puncture training program

Deborah Rooney (*University of Michigan, Medical Education, Ann Arbor, MI, USA)*, Shawna Shafer (*University of Michigan, Pediatrics and Communicable Diseases, Ann Arbor, MI*), Joseph House (*University of Michigan, Pediatrics and Communicable Diseases, Ann Arbor, MI*)

Background: Medical training programs require a robust assessment framework and sustainable program evaluation strategies. Traditionally, these requirements are met by applying different techniques in independent processes. We propose that a many-facet Rasch measurement model can be dually employed for learner assessment and program evaluation. We demonstrate the model's value in a simulation-based lumbar puncture (LP) training program at an academic Level 4 NICU.
Summary of Work: A total of 45 residents (14 interns, 19 PGY2, 10 PGY3) engaged in a simulation-based LP program. In a pre-post design over 6 months, residents were assessed while performing a neonate LP in simulation and clinical environments. We captured two primary measures— dichotomous scores from an 11-item MCQ to assess knowledge, and ratings from a three-point, 6-item OSATS-type instrument (Iyer, 2012) to assess psychomotor skills via live observation by one of three faculty raters. We applied the Rasch model to a) identify specific gaps in learners' knowledge and performance, and b) evaluate quality of the program by reviewing learner improvement over time and rater bias.
Summary of Results: Learner assessment: Following intervention, residents continued to have difficulty with anatomy-relevant MCQ item 10 (OA=.5), and OSATS item 5-CSF return (OA=2.0).
Program evaluation: Rasch analyses indicated learner improvement in knowledge (pre- OA=.6, post-OA=.9, p<.001), and psychomotor skill (pre-OA=2.2, post-OA=2.7, p<.01). Bias analyses indicated rater differences (p<.001) and suggested Rater 2 was more severe (OA = 2.18) than other raters (OA=2.44) for item 1-Prep (p=.004).
Conclusions: The many-facet Rasch model may be applied effectively for the dual purpose of learner assessment and program evaluation in simulation-based education.

OTT-OD-3: POT POURRI

OTT-OD-3-1
How hard are you thinking? Objective measures of cognitive load in experts and novices during test taking

Adam Szulewski (*Queen's University, Kingston, Ontario, Canada)*, Nathan Roth (*Queen's University, Kingston, Ontario*), Daniel Howes (*Queen's University, Kingston, Ontario*)

Background: Changes in pupil size have been shown to be associated with cognitive processing demands. The magnitude of this change has been found to be a reliable marker of cognitive load in non-medical settings. The purpose of this study was to compare cognitive load and confidence using task-evoked pupillary responses (TEPR) among expert and novice medical trainees when answering questions of varying difficulty.
Summary of Work: Twenty participants were recruited into two experimental groups based on level of medical training—novice (junior medical students) and expert (PGY-4, PGY-5 and first year attending emergency physicians). Mobile eye-tracking glasses were used to assess pupil diameter change from baseline as participants answered a variety of emergency medicine questions.
Summary of Results: The mean pupil diameter change from baseline, a surrogate for cognitive processing demand, was significantly higher in novices compared to experts for all question types (p<0.0002). Mean pupil diameter change was significantly greater for difficult compared to easy questions in the expert group (p=0.023), however this difference was not observed in the novice group (p=0.148). Moreover, the time to reach peak pupillary response, and answer the question, occurred significantly faster in the expert group (p=0.042), and was shorter for easy compared to difficult questions across all groups (p=0.046). The novice group was also, on average, overconfident compared to the expert group.
Conclusions: TEPRs show systematic and quantitative differences in the cognitive loading effects of different examination questions and are able to reliably differentiate levels of medical expertise. Together with traditional modes of examination, TEPR has the potential to provide the basis for a new, more objective assessment of medical competence.
Take-home Messages:

OTT-OD-3-2
Addressing Filter Failure in Medical Education: Finding the Right Filter

Marcel D'Eon (*University of Saskatchewan, Saskatoon, Canada)*, Krista Trinder (*University of Saskatchewan, Saskatoon*)

Background: Medical schools are mired in content overload due partly to "filter failure", the failure to regulate (or filter) the flow of information according to the needs of the audience. This study set out to compare common variations within three elements of potential filters: (1) reference populations, (2) data collection methods, and (3) indicators of relevance.
Summary of Work: We compared how the various forms of each of the three elements filtered the same bio-medical content as expressed in actual examination questions: multiple regression analysis was used to

detect any differences. Students and physicians (reference populations) were surveyed or participated in a Delphi small group (data collection method) with sets of examination questions and asked to indicate for each one (a) the correct score and (b) the level of relevance to general clinical practice as well as (c) an Angoff cut-score (indicators of relevance).

Summary of Results: There were no significant differences among the various filters created by combining reference populations, data collection methods, and indicators of relevance.

Conclusions: All filter elements were approximately equivalent. For convenience we recommend surveying senior medical students for Angoff cut scores: students are more willing and able to participate than busy clinicians and one-time surveys spread the time and effort out among many contributors. Angoff cut-scores may be used prospectively to determine course content and to create better competency-based examinations.

Take-home Messages: Content of medical school curriculum needs to be filtered. There are valid filters available. It would be best to survey senior medical students using an Angoff method.

OTT-OD-3-3
Assessment on the highest Bloom-levels in a discipline-based integration program challenges students and teachers

Herman J.M. van Rossum (*University of Antwerp, Faculty of Medicine and Health Sciences, Antwerp, Belgium*), Nele R.M. Michels (*University of Antwerp, Faculty of Medicine and Health Sciences, Antwerp*), Guy Hubens (*University of Antwerp, Faculty of Medicine and Health Sciences, Antwerp*)

Background: Students must be able to solve medical problems in an integrated way at the highest - analytical and creative - levels of the pyramid of Bloom. Assessment at these levels has a high fidelity but requires much effort to organize. Therefore, the Antwerp medical school (Belgium) constructed an integration program across the discipline-based undergraduate curriculum. This is a report about the first experiences.

Summary of Work: In 2012-2013, 127 bachelor-1 students started a 2-week end-of-year module. Students worked in 12 groups around clinical vignettes. End products were a three minutes instructional film, a poster (group presentation), and an individual presentation. Tutors offered coaching in formative sessions half way to provide feedback and to challenge students to perform creatively at high cognitive levels. Summative assessment of the end products was based on criteria related to both form and content. Additionally, students' essential basic knowledge was measured by a pre-coded high-stakes test (90%).

Summary of Results: All students passed. Moreover, the performance levels exceeded expectations of all participating teachers. All groups scored sufficiently (> 60%) as regards the poster. 124 students (97.6%) passed the individual presentation. 94.5% and 100% of the

students passed the knowledge test after a first and second testing respectively.

Conclusions: Through a combination of sharply defined tasks, formative and summative sessions and a mix of evaluation forms, students performed surprisingly well at higher order-tasks.

Take-home Messages: Challenging students to perform on the higher cognitive levels of Bloom is inspiring and rewarding for both students and their teachers.

OTT-OD-3-4
The influence of spaced testing on test preparation time: a randomized controlled experiment

Rene Tio (*University Medical Center Groningen, Cardiology, Groningen, Netherlands*), Wouter Kerdijk (*University Medical Center Groningen, Center for innovation and research in medical education, Gronigen*), Janke Cohen-Schotanus (*University Medical Center Groningen, Center for innovation and research in medical education, Groningen*

Background: The average student procrastinates. It is known that test preparation is delayed till 2-4 weeks before the test date. In many curricula, tests are planned at the end of a course and courses usually take up more than 4 weeks. We wondered how spaced testing influences procrastination and performed a randomized controlled experiment in which end of course testing was compared to spaced testing.

Summary of Work: Second year medical students (n=395) were invited to participate in the experiment during a 10-week block with an end of course test as regular assessment. After informed consent was obtained, participants (n=78) were randomly assigned to one of the following two conditions: (1) one test in week 10 (control group) and (2) tests in week 4, 8 and 10 (experimental group). Students recorded home study time on-line weekly. They also filled out a questionnaire about spaced versus end of course tests.

Summary of Results: In the control group with only a test after 10 weeks, students spent 112± 54 hours on test preparation. Students in the experimental group with spaced test condition spent 50% more time on test preparation (176± 78 hours; p< 0.05). Students preferred spaced testing over end of course testing. They perceived spaced testing to stimulate preparation and deeper learning and prevent procrastination more than end of course assessment.

Conclusions: Regular testing enhances test preparation and students prefer this over end of course testing.

Take-home Messages:

OTT-OD-3-5
Home and learning environment effects on assessment performance

Tim Wilkinson (*University of Otago, Christchurch, Christchurch, New Zealand*), Joy Rudland (*University of Otago, Faculty of Medicine, Dunedin*), Anthony Ali (*University of Otago, Christchurch, Christchurch*), Jan

McKenzie (University of Otago, Christchurch, *Christchurch*), Caroline Bell (*University of Otago, Christchurch, Christchurch*), Frances Carter (*University of Otago, Christchurch, Christchurch*)

Background: The 2010 and 2011 earthquakes in Christchurch, New Zealand, had variable effects on the circumstances of our medical students, thus providing a case study by which to explore the effects of the home and work learning environments on assessment performance.

Summary of Work: A questionnaire survey was administered to all medical students prior to their high stakes examination asking about impacts on home and work environments. We used performance on examinations in previous years to account for prior ability. We then compared expected performance with actual performance on the final examinations and correlated this with questionnaire results.

Summary of Results: The greatest adverse effect was seen for those who were recent immigrants or who had a greater disruption on their living arrangements. There was a surprising beneficial association between exam performance and markers of stress or feeling unsupported by university. Home study place disruption and library disruption adversely affected performance on the written exam while the OSCE was adversely affected by not feeling safe, and lack of social support.

Conclusions: A positive learning environment should also consider the home environment and social support. The interaction between 'stress' and support has relevance in situations where disruptions are not evident.

Take-home Messages: A secure physical and emotional base, but with some stress, contributes to assessment performance.

OTT-OD-4: PROFESSIONALISM 1

OTT-OD-4-1
A Simple Behavioral Measure of Professionalism

Betsy Williams (*Rush University Medical Center, Behavioral Sciences, Chicago, USA)*, Michael Williams (*Wales Behavioral Assessment, Research, Lawrence*)

Background: In the measurement of professionalism, indirect measures range from paper and pencil self-report measures through simulations. More recently there has been an effort to develop measures of small professionalism actions as prognostic of larger lapses in behaviors (eg. McLachlan and colleagues.)

Among the efforts to bridge the gap are the multisource feedback instruments. Evidence suggests that such an approach is valid, but it can also be a somewhat weighty methodology.

We report an effort to combine the strengths of both the direct observation approach and multisource feedback.

Summary of Work: Multisource feedback assessments were gathered on 59 individuals in an intensivist

department in a large multisite health care system. All members of the microsystem were queried and their responses serve as our database.

In addition the respondent individuals were asked to estimate the frequency of their consultations with the individuals under assessment.

Summary of Results: There was a striking correlation between assessor's estimation of the individual's professionalism measured by the MSF instrument and their self-reported frequency of consultation.

Conclusions: Frequency of consultation is a direct behavioral measure of the degree of an individual's participation in the team process. Frequency of consultation appears to be directly correlated with an explicit assessment of professionalism. This relationship has implications for both assessment and education.

Take-home Messages: A single item measure that requires no more than a few seconds to obtain was heavily correlated with a more complex measure of professionalism. Further exploration of social connection indices may be warranted.

OTT-OD-4-2
Relationship between Medical Students' Performance of Routine Tasks and their Performance

S. Beth Bierer (*Cleveland Clinic, Cleveland Clinic Lerner College of Medicine, Cleveland, USA)*, Elaine F. Dannefer (*Cleveland Clinic, Cleveland Clinic Lerner College of Medicine, Cleveland*)

Background: Research suggests that not completing routine, administrative tasks during medical school predicts later academic deficiencies or egregious behavior. We explore if proxy measures of conscientiousness (i.e., meet deadlines) relate to student performance on USMLE Step 1 and/or promotion decisions during medical school.

Summary of Work: With IRB approval, we used school records to document students' late/incomplete weekly medical knowledge assessments (essays and multiple-choice questions), missing course evaluations, and focus group attendance for six student cohorts (n=191) at the Cleveland Clinic Lerner College of Medicine over a 2-year period. These variables were combined into scales and compared to the following student performance measures: (1) overall promotion decisions in Years 1-2, (2) portfolio-based promotion decisions for professionalism and medical knowledge in Years 1-2, and (3) first-time performance on USMLE Step 1. We examined descriptive statistics, evaluated scale reliability, conducted nonparametric comparisons, and calculated effect sizes.

Summary of Results: Students identified for professionalism in Years 1-2 were more likely to score < 200 on USMLE Step 1, not complete course evaluations, or submit late/incomplete assessments (p < .01). We did not identify statistically significant relationships between students with medical knowledge deficiencies and their performance on USMLE Step 1 or submission of weekly assessments (p > .05, effect sizes < .20).

Conclusions: We demonstrate a relationship between completion of routine tasks and USMLE Step 1 performance not reported elsewhere.

Take-home Messages: Monitoring routine tasks gave us a low-cost, feasible approach to identify at-risk students and discuss "habits of mind" with students and faculty advisors.

OTT-OD-4-3
Correlates of a behavioural measure of conscientiousness in medical school

Don Munro (*University of Newcastle, Australia, School of Psychology, Callaghan, NSW, Australia)*, Gabrielle Finn (*University of Durham, UK, School of Medicine and Health, Stockton-on-Tees*), David Powis (*University of Newcastle, Australia, School of Psychology, Callaghan, NSW*), Bore Miles (University of Newcastle, Australia*, School of Psychology, Callaghan, NSW*)

Background: In order to track the genesis of professional behaviour in medical students, the medical school of the University of Durham (McLachlan, Finn & Macnaughton, 2009) introduced a behavioural measure of conscientiousness, the Conscientiousness Index (CIndex). This involved routinely recording a number of student behaviours, including the timely completion of forms required by the school, attendance generally and at particular sessions, and volunteering for extra experiences. Reference: McLachlan JC, Finn, G and Macnaughton, J. (2009). The Conscientiousness Index: A Novel Tool to Explore Students' Professionalism. Academic Medicine 84(5), 559-564).

Summary of Work: First and second year students also completed the IPIP Big5 scales, including all facets of Conscientiousness (n=163), and/or the University of Newcastle PQA instruments (Munro, Bore & Powis, 2009; Bore, Munro & Powis, 2008; see www.pqa.net.au), which measure Involvement, Resilience and Control, and their facets (n=120). Results of routine academic and clinical assessments were also available.

Summary of Results: The CIndex showed moderate internal consistency and good construct validity, correlating moderately with aspects of Conscientiousness from the IPIP tests and a similar variable named Control from the PQA tests, but not with the other personality variables (Extraversion, Stability, Agreeableness, Openness, Resilience, Involvement and Moral Orientation). Moderately strong predictive validities by CI for 1st. year examination performance was also evident, but not for 2nd year performance.

Conclusions: The results will be discussed in relation to the possibility of introducing into the assessed curriculum more comprehensive behavioural measures of personality relevant to professionalism.

Take-home Messages: It is possible to measure conscientiousness in medical students using observations of student behaviour, and this technique may be applicable to other aspects of professional behaviour.

OTT-OD-4-4

Understanding the Social Network of Professionalism in the Medical School Environment

Varun Shahi (*Mayo Clinic, Mayo Medical School, Rochester, USA)*, Trey Mullikin (*Mayo Clinic, Mayo Medical School, Rochester*), Frederic Hafferty (*Mayo Clinic, Program in Professionalism and Ethics, Rochester*), Douglas Grbic (Association of American Medical Colleges*, Policy Research Studies, Washington, D.C.*), Wojciech Pawlina (*Mayo Clinic, Department of Anatomy, Rochester*)

Background: Ideals of professionalism transmitted among medical students via the hidden curriculum are important in forming informal student-based networks. In recent research, social network analysis was revealed as an important tool in understanding the dynamics and structure of informal peer learning, including professionalism. However, the qualities students actually value in their peer-based professionalism role models remains unknown.

Summary of Work: We sought to identify qualities influencing the perception of peer-based professionalism. First year students at Mayo Medical School were surveyed every two weeks during a 13-week Histology and Gross Anatomy course. They identified an individual, among their peers whom they saw as a model for professionalism, and answered questions regarding the qualities of that chosen individual.

Summary of Results: Preliminary results reveal a somewhat stable and small number of nodes (role model stars) within an evolving network of support in which the qualities used to identify these professionalism hubs are changing over time. Factors such as leadership and academic performance are emerging over time as criteria in the selection of these role models. Other qualities, such as personal appearance, remain relatively constant and highly ranked.

Conclusions: We are seeing the emergence of spatially distinct sub-communities of students who support a small number of professionalism role models. We identified several characteristics that differentiate these individuals from their peers in the attribution of professionalism.

Take-home Messages: Undergraduate medical students' understanding of professionalism evolves over the course of their medical education. The qualities that determine their role models for professionalism are dynamic and should be considered when developing professionalism curriculum in medical schools.

OTT-OD-4-5
Student opinion toward First Code of Conduct for Thai Medical Students: Preclinical vs. Clinical years

Amnuayporn Apiraksakorn (*Khon Kaen Medical Education Center, Khon Kaen, Thailand)*

Background: In November 2012, the Thai Medical Students Association Committee and the Thai Medical Council announced a 13-section Code of Conduct for Medical Students, created by consensus. Our research Student focused on the students' evaluation regarding practicability of the code as applied to preclinical (first to third year) vs. clinical year (fourth to sixth) students.

Summary of Work: A self-administered questionnaire was distributed in 2013 to 139 preclinical and 140 clinical year students practicing at Khon Kaen Hospital in Northeast Thailand. Student opinion was rated using a 5-point Likert scale and each section was ranked from easiest to hardest with respect to practicability.

Summary of Results: Respectively the response rate by the pre-clinical and clinical groups was 79% and 84%. Most students "strongly agreed" or "agreed" with the practicability of the code. Preclinical vs. clinical year students, respectively, indicated the easiest to do was "being helpful, supportive and thoughtful toward friends and colleagues" vs. "not gaining from patients except education". Both groups ranked "efficiently using technology whenever economical" as the most difficult to achieve. Some clinical year students reported difficulty in following the code with respect to communication skills and patient confidentiality.

Conclusions: Student opinion with respect to practicability of the Thai Medical Students Code of Conduct was comparable. Overall preclinical and clinical year students are able to implement the Code, but all find adopting and adapting technology efficiently and economically very difficult.

Take-home Messages: Timely introduction and implementation of the Code of Conduct will facilitate training and help to overcome student apprehensions and limitations.

OTT-OD-5: SCRIPT CONCORDANCE TEST

OTT-OD-5-1
A SCT on cardiovascular disease in an undergraduate curriculum: who should be member of the expert panel?

Annick Dermine (*KU Leuven, OnderwijsOndersteuning, Leuven, Belgium*), Inge Fourneau (*UZ Leuven, Vascular Surgery, Leuven*), Lisa De Jonghe (*KU Leuven, OnderwijsOndersteuning, Leuven*), Katrien Bosselaers (KU Leuven*, OnderwijsOndersteuning, Leuven*), Geraldine Clarebout (*KU Leuven, OnderwijsOndersteuning, Leuven*)

Background: In their 5th year, KU Leuven medical students have a course on cardiovascular diseases. Given the importance of assessing clinical reasoning in ill-defined cases in this course, an SCT was developed. Previous research (1-3) demonstrated the influence of expert panel composition on the absolute scores but not on student ranking. In this study we seek for the ideal composition of our expert panel.

Summary of Work: 264 students took this 90-item SCT voluntarily after the official examination of this course.

Different panels (teaching staff of the course (P1), general practitioners (P2), specialists and clinical teachers working in the associated hospitals (P3), final year students (ready for residency) (P4)) were composed. Student results, reliability and item analysis were compared with the scoring keys of the different panels.

Summary of Results: Student results: mean scores were significantly different for P2 and P4 (unless GP final year students were eliminated).

Item analysis: item-total correlations were comparable between all panels; the panel modus answer that correlated best with the domain experts' answer was that of P1 (r = 0.8).

Reliability – analysis revealed acceptable ICC scores within juries (0.95-0.85) and good reliability of student scores for all juries (α: 0.86-0.87).

Conclusions: Although more analyses are ongoing, our first findings suggest that the most appropriate jury experts are members of the teaching staff, which is probably the one with the highest face validity.

Take-home Messages: When choosing SCT panels, teachers of the course material generate the highest absolute scores.

OTT-OD-5-2
Dynia – combining of script concordance testing, virtual patients, webinars and forums to assess knowledge translation in chronic pain

David Topps (*University of Calgary, Family Medicine, Calgary, Canada)*, Heather Armson (*University of Calgary, Family Medicine, Calgary*), Cathlin Mutch (*University of Calgary, Family Medicine, Calgary*), Eloise Carr (University of Calgary*, Faculty of Nursing, Calgary*), Paul Taenzer (*University of Calgary, Psychology, Calgary*), Ashi Mehta (*University of Calgary, Physician Learning Program, Calgary*)

Background: The translation of evidence into clinical practice and clinical practice guidelines (CPGs) have not shown strong uptake. Strategies to enhance update are difficult to assess. Combining Webinars, online forums and virtual patients and integrating script concordance testing (SCT) as both needs assessment and outcome measure appears to overcome some of these issues.

Summary of Work: We used a combination of webinars, forums and virtual patients incorporating a script-concordance approach to test behaviour change in experienced clinicians. Participants tackled virtual patients before, during and after webinar sessions on chronic pain CPGs. Case design emphasized areas where field pragmatism conflicts with CPG principles

Summary of Results: Clinicians strongly favored this approach as being more relevant to their concerns in practice, avoiding the classic best-of-five MCQ style, and allowing them to explore investigative and treatment options with their peers.

Immediate reporting of decision pathways using simple graphics enhanced discussions of clinical controversies and enriched ensuing forum debate, following the webinars.

Comparative metrics across similar case structures showed a change in both thinking and how experienced practitioners tackle cases grounded in their own areas of concern.

Conclusions: Blending webinars, online forums and virtual patients significantly enhances learner engagement and enriches case discussions

An SCT approach fits better with how experienced clinicians like to be challenged, if placed within a realistic clinical context

Take-home Messages: SCT-based virtual patients provide a rich assessment tool to evaluate learning and behaviour change in experienced practitioners

OTT-OD-5-3
CROSSBAR – combining script concordance testing, virtual patients and scenario based assessment of learner suitability for residency training

Jean Rawling (*University of Calgary, Family Medicine, Calgary, Canada)*, David Topps (*University of Calgary, Family Medicine, Calgary*), Heather Armson (*University of Calgary, Family Medicine, Calgary*), Robert Gagnon (University of Montreal*, Center of Pedagogy Applied to Health Sciences (CPASS), Montreal*), Paul Duggan (*University of Adelaide, Discipline of Obstetrics and Gynaecology, Adelaide*), Maureen Topps (*University of Calgary, Family Medicine, Calgary*)

Background: Our program assesses a wide variety of candidates for acceptance to residency training, all of whom are international medical graduates (IMGs). Clinical decision making (CDM), and the Canmeds competencies of professional, manager and health advocate are areas identified in the literature, as being of greatest concern. Script concordance testing (SCT) has shown the best correlation with CDM but current test formats are somewhat inflexible.

Summary of Work: We developed a series of SCT-based virtual patients that explore candidates' clinical reasoning skills, in combination with other scenario-based assessments of key competencies. We used a reference panel of peer learners and successful participants in the target residency programs for benchmarking of acceptable behaviours

Summary of Results: Program designers found that this approach to case-based SCT designs afforded a more practical means of authoring realistic scenarios. The detailed metrics provided by our open-source scenario-based learning design platform, OpenLabyrinth, enabled program evaluators to rapidly assess learner competencies and program weaknesses

Conclusions: Combining a flexible scenario-based learning design tool with strong principles from SCT provided our program with a more objective, standardized and repeatable means of assessing candidates from a wide variety of clinical educational backgrounds

Take-home Messages: An SCT approach to case authoring is a more effective way of assessing key competencies, when using scenario-based assessment designs

OTT-OD-5-4
Implementation and evaluation of the Script Concordance Test in osteopathy

Jorge E Esteves (*British School of Osteopathy, London, UK)*, Marianne Bennison (*Oxford Brookes University, Department of Sport and Health Sciences, Oxford*), Oliver P Thomson (*British School of Osteopathy, London*)

Background: Osteopathic pre-registration education aims to produce competent autonomous practitioners who are capable of dealing with and mastering the uncertainty of clinical practice. The students' preparedness for clinical practice is typically assessed using high-fidelity long case exams, i.e., clinical competence assessments using real patients in a real clinical setting.

Summary of Work: In this presentation, we describe the development, implementation and initial critical evaluation of the Script Concordance Test (SCT) in the osteopathy programme at Oxford Brookes University. The SCT was used to complement the long case exam in assessing final year students' ability to effectively operate in situations of clinical and professional uncertainty. To our knowledge, this was the first use of the SCT in osteopathy. The development and implementation of this assessment strategy was informed by available research literature. Student feedback was used to evaluate the effectiveness and usefulness of the newly implemented clinical reasoning assessment.

Summary of Results: From currently available data, the reliability and validity of the SCT in osteopathy cannot yet be established. Notwithstanding this, students consider the SCT to be a useful and effective way to assessing clinical reasoning.

Conclusions: The SCT is a useful addition to assessing clinical reasoning in osteopathy, particularly in situations of clinical uncertainty. It has the potential to effectively assess the students' diagnostic reasoning, evaluation of risk and patient safety, and ethical aspects of osteopathic care; however, its reliability and validity needs to be established.

Take-home Messages: The SCT may provide an important vehicle to assess osteopathy students' preparedness for autonomous clinical practice using a standardised format.

OTT-OD-5-5
Comparison of geographically separated expert panels for a modified script concordance test

Edward Simanton (*University of South Dakota, Sanford School of Medicine, Office of Medical Education, Sioux Falls, USA)*, Denese Playford (*The Rural Clinical School of WA, Medical Education, Crawley*)

Background: Clinical reasoning is strongly case and context specific, however, script concordance tests show

that experts generally arrive at final clinical decisions with some degree of consistency. In order to examine diagnostic reasoning among students, a series of diagnostic reasoning tests were developed by a group of medical school faculty in South Dakota, USA. This test was based on general principles of script concordance testing and grading was based on data from an expert panel of primary care physicians in South Dakota. As other medical schools thousands of kilometers from South Dakota began using the same test, questions arose regarding the validity of grading from a South Dakota expert panel for diagnostic reasoning in other geographical areas with different culture and patient populations.

Summary of Work: An expert panel of rural clinicians representing a diverse range of rural subspecialties in Australia was recruited and their expert panel results were compared with the results of a South Dakota expert panel. Responses of groups were classified as identical grade provided, similar grade provided, and dissimilar grade provided. Comparisons were made by total, case, and diagnosis.

Summary of Results: The two expert panels provided identical results to 35% of items and similar results for 73% of items. The two groups were much most similar in an abdominal pain case (diverticulitis) and least similar in a cough case (drug side-effect).

Conclusions: There are definitely differences in findings from geographically separated expert panels used for modified script concordance tests. Additional study is needed to determine factors relating to differences that were identified.

Take-home Messages: It is best to use an expert panel from the same geographic region as the examinees.

OTT-OD-6: OSCE 4

OTT-OD-6-1
Neurology examination on video as an OSCE

Peter Watson (*Queen's University Belfast, Centre for Medical Education, Belfast, UK)*, Michael Stevenson (*Queen's University Belfast, Centre for Medical Education, Belfast*), Stanley Hawkins (*Queen's University Belfast, Centre for Medical Education, Belfast*)

Background: Due to the difficulty of supplying sufficient numbers of standardised neurology patients for the final MB OSCE examination we introduced a video station for neurology in 2010.

Summary of Work: Each station consists of a series of several short clips played on a lap top of a neurological examination of a patient by an experienced clinician.The station lasts 8 minutes. The student is asked to comment on what abnormalities they see and ultimately to offer a differential diagnosis. We have presented cases of Parkinson's disease, spinocerebellar degeneration and multiple sclerosis causing spastic paraparesis. In the design of stations we aimed to present clear signs readily interpreted by an observer, not just the person performing the exam e.g.

assessment of reflexes and involuntary movements. Demonstration of tone and power loss are more challenging if the changes are subtle. With good editing the clip duration could be kept under 30 seconds and repeated if requested by the student.

Summary of Results: It is logistically easier to organise these stations than with multiple real patients and the scenarios are highly standardised. It is possible to test students on their recognition of neurological signs and diagnosis of complex neurology. They have been well received by examiners and psychometrically have performed well with the pass rate of 90% similar to other stations, and examiner variation less than 20%

Conclusions: A video neurology station is a viable way of testing neurological examination skills as an OSCE.

Take-home Messages: A video OSCE is a good way of testing neurological skills.

OTT-OD-6-2
Innovative assessment using an automated OSCE system

Alia Zawawi (*King Saud Ben Abdulaziz University for Health Sciences, Family Medicine, Riyadh, Saudi Arabia)*, Muneera Baraja (*King Saud Ben Abdulaziz University for Health Sciences, Family Medicine, Riyadh*)

Background: Probably for the first time in the Middle East an automated OSCE [objective structured clinical examinations] system for postgraduate training was implemented as part of a family medicine residency assessment. Difficulties were noted previously in gathering and reporting data, and also in merging OSCE and OSPE (objective structure practical examination) in the same exam. So how can we utilize an automated OSCE system to overcome these problems?

Summary of Work: Mock exams were conducted to test the system, feedback was sent to coordinators and technical staff. Adjustments were made in the system before the actual OSCE exam. An automated OSCE system was implemented for first year residents' end of rotation exam in January 2013.

Summary of Results: Completed evaluations were digitally submitted and compiled by the software with accompanying video for each participant of the OSCE. Evaluations were automatically calculated and each participant received a grade report based on the overall points received during the OSCE. Three types of report were provided. Score reports, ranking reports; and Evaluation reports.

Conclusions: Having an automated OSCE system helps residents to react as if they are in a real clinical situation. Data generated gives instant and accurate results about participants' performance. Additionally, it can be modified to conduct both OSCE and OSPE at the same time.

Take-home Messages: Automated OSCE systems created an interest in assessment researches and auditing .Further development in both assessment and feedback will lead to an improvement in this teaching innovation.

OTT-OD-6-3

The UCLMS experience of using integrated professional skills OSCE stations during a final year examination

Alison Sturrock (*UCL Medical School, London, UK)*, Sarah Bennett (*UCL Medical School, London*), Laura-Jane Smith (*UCL Medical School, London*), Jayne Kavanagh (UCL Medical School, *London*), Lorraine Noble (*UCL Medical School, London*), Aroon Lal (*UCL Medical School, London*)

Background: Recent graduates of UCL medical school reported confidence in clinical knowledge, communication skills and professional skills, however found it difficult to integrate these elements. To help meet this unmet need we introduced an online module called 'Case of the Month' and altered our final year assessment.

Summary of Work: In our 2013 final year long station OSCE we replaced the communication skills and ethics and law station and evaluation of evidence viva with three 'professional skills stations'. These stations integrated testing clinical knowledge and other professional skills including communicating skills, evaluation of evidence, ethics and law and health promotion. The stations were a self discharge request by a homeless man (Station a), a medication review in a GP surgery (Station b), and an end of life care discussion (Station c). We reviewed the results in terms of ability to discriminate. We also asked each examiner whether the station was able to discriminate, was relevant and realistic for a FY1 doctor.

Summary of Results: Each station was out of 15. The range of scores was wide (Station a:1.75-15; station b:2.25-15, station c: 2.5-15) with the median being around 11. There was a good correlation between station scores and OSCE result.
Examiner feedback suggested the stations were discriminating, relevant and realistic, however dependent on the quality of the simulated patient.

Conclusions: Although writing integrated OSCE stations is challenging, we have shown that they can be discriminating. Examiner feedback suggests they are useful and should be developed further.

Take-home Messages: The use of OSCE stations that integrate clinical skills, communication skills and professionalism challenges students. Their inclusion in a final year exam will encourage students to acquire the relevant skills prior to qualification.

OTT-OD-6-4
Identification of Students with Clinical Skills Deficiencies: The iOSCE

Rajesh Gupta (*University of Toronto, Medicine, Toronto, Canada)*, Mahan Kulasegaram (*University of Toronto, Toronto*), Richard Pittini (*University of Toronto, Obstetrics and Gynecology, Toronto*)

Background: While some accrediting bodies have used a comprehensive OSCE to assess clinical skills, few undergraduate programs have reported utilizing or published the reliability and validity of a comprehensive integrated OSCE.

Summary of Work: We designed a two part integrated OSCE (iOSCE) with the goal of assessing medical students progress towards becoming integrated medical graduates during clerkship. Each traditional station was based on a number of specific disciplines (eg post partum depression) while broader concepts were assessed in specific "integrated" stations in the final OSCE, eg. end-of-life care. A unique global rating scale was devised utilizing CanMEDS roles and a transparent pass mark (meets expectations or 3.0/5 for each role) was used.

Summary of Results: The reliability of the iOSCE examinations ranged from 0.37 to 0.62 for the last six iterations. Item-total correlations ranged from 0.08 to 0.64. The reliability of the integrated stations was not different from the traditional stations. Factor analysis showed that integrated stations correlated with the same three factors as non-integrated stations. This examination identified more students in academic difficulty (failure or borderline pass) than other OSCEs in the undergraduate program (up to 15%), and correlated with performance in other courses or course director evaluations.

Conclusions: An integrated OSCE utilizing a global rating scale with transparent pass marks is a useful tool for assessing global clinical skills during the clerkship. Our global rating scale performed similarly on traditional and integrated stations but identified students with deficient clinical skills that had not been identified in other courses.

Take-home Messages: A comprehensive clerkship OSCE utilizing a global evaluation scale with a transparent pass mark is feasible and identifies students with significant clinical skills deficiencies.

OTT-OD-6-5
Can a Likert-type scale integrated into an OSCE station be useful for detection of clinical skill improvement?

Diem Quyen Nguyen (*University of Montreal, Medicine, Montreal, Canada)*, Benoit Deligne (*University of Montreal, Medicine, Montreal*), Jean Victor Patenaude (*University of Montreal, Medicine, Montreal*), Robert Gagnon (University of Montreal*, Bureau d'évaluation, Montreal*)

Background: OSCE is considered the gold standard method for clinical skill assessment. Although global rating scales best assess clinical competence, checklists are used whenever evaluators' training cannot ensure standardization. It is unclear what type of OSCE rating scale could best assess progress in clinical skill development.

Summary of Work: We examine two types of rating scale (checklist and global rating) used to evaluate two student cohorts for the same OSCE station, to verify if an OSCE-integrated global rating scale can assess clinical skill progress for medical students advancing to second year.

In 2012, the musculoskeletal OSCE station, using a checklist as a rating scale, was designed to assess the first- and second- year students' clinical skills. In 2013, a global rating scale was created and incorporated into the same OSCE station for both first- and second-year students; however, the first-year students only performed the clinical interview.

Summary of Results: 570 students took the 2012 exam, and 584 took it in 2013. In 2012, no significant improvement was observed between the first- and second-year student total score (76.6 ± 8.9 and 59.9 ± 9.0). The score for first- and second-year students in the interview sub-section was 78.5 ± 12.1 and 64.8 respectively in 2012.In 2013, it is 81.6 ± 8.6 and 75.6 ± 9.0 respectively. However, the progress of the same student cohort by their second year was significant, advancing from 67.4 to 75.2.

Conclusions: These preliminary results based on one OSCE station, suggest the use of OSCE with a rating scale may be considered whenever OSCE is used to assess clinical skill progress.

Take-home Messages: Global rating scale integrated into OSCE could be a mean to detect progress in clinical skill.

OTT-OD-7: PROGRAMME EVALUATION 2: CURRICULUM 1

OTT-OD-7-1
Then vs now: 'So what?'. Evaluating the aftermath of whole-programme change in a medical curriculum. (16-year follow-up interviews with inaugural problem-based learning tutors still involved)

Gillian Maudsley (*The University of Liverpool, Department of Public Health & Policy, Liverpool, UK)*

Background: When the Liverpool undergraduate medical curriculum underwent a major change in its educational philosophy and design/implementation in 1996, interviews with the 34 inaugural problem-based learning (PBL) tutors (by the 35th tutor) showed a positive approach overall, despite much upheaval. They were particularly bothered at their own fallibility about intervening without 'teaching'. Facilitating active learning of clinically relevant, integrated knowledge (while tolerating uncertainty) is a continuing, worthy focus. What lessons do the remaining tutors from that first-ever cohort offer though, from their considerable experience of implementing and maintaining the original major shift in educational philosophy?

Summary of Work: Aim---What insights about 'So what?' do the inaugural PBL tutors provide when evaluating their continued and extensive experience as educators in an undergraduate medical curriculum that long since underwent major problem-based transformation?
~~~Setting: Liverpool MBChB curriculum. Participants: The ten of the first-ever problem-based learning tutors who remained actively involved as educators in the

curriculum to-date. ~~~Method: Semi-structured interviews (a 16-year follow-up, by the eleventh remaining tutor) about 'then vs now', inductively analysed for themes, within the pragmatism paradigm.

Summary of Results: ~~~The tutors revealed how their educational outlook had developed, and reacted constructively to their own contemporaneous comments about the curriculum change from their original interviews.

Conclusions: Tutors who remained actively involved in promoting medical students' active learning years after a major curriculum change gave valuable insights about the 'So what?' of experiencing that upheaval, and appeared upbeat despite considerable continuing challenges.

Take-home Messages: Long-term evaluative follow-up of educators' experience of the aftermath of a major curriculum change is feasible and productive.

OTT-OD-7-2
Assessment of training programmes: is there concordance between tools?

Taruna Bindal (*Alexandra Hospital, Worcestershire Acute Hospitals NHS Trust, Department of Paediatrics, Redditch, UK)*, David Wall (*Tayport, Fife*), Helen Goodyear (*Health Education West Midlands, Birmingham*)

Background: In the West Midlands Deanery (WMD) UK, trainees assess the education and training environment by completing a JEST (Job Evaluation Survey Tool) questionnaire about each training post. In addition face to face Quality Assurance (QA) visits take place at hospitals by an external panel of consultants and Deanery staff on a 4 yearly cycle. A meeting takes place with trainees, trainers and managerial staff to assess whether General Medical Council standards are being met.

Summary of Work: Analysis of whether the JEST questionnaires and QA visit reports gave comparable assessment scores for the 17 hospital Trusts in the WMD paediatric training programme.

Summary of Results: Overall, the JEST scores were higher than the QA scores with six Hospital Trusts showing significant differences in JEST and QA scores ($p = 0.001 - 0.033$). Of note, four of these hospitals were given significantly different assessment scores by both junior (years 1-3) and senior (years 4-8) trainees. These score differences were attributed to trainees concerns around patient safety, induction, quality of teaching provided, rota design and availability of facilities.

Conclusions: Issues about education and training at QA visits led to lower scores in comparison to JEST. We postulate that trainees felt more at ease in talking rather than writing about their concerns despite JEST being an anonymous questionnaire.

Take-home Messages: A site visit is essential to assess clinical training programmes as questionnaires will not capture all relevant information.

OTT-OD-7-4

Evaluating meaningful outcomes to enhance healthcare provider competencies

Sean Hayes (*AXDEV Group, Brossard, Canada)*, Mila Kostic (*University of Pennsylvania, Philadelphia*), Laura Young (*University of North Carolina, Chapel Hill*)

Background: Outcomes assessment methodologies are evolving beyond assessment of participants' satisfaction with program content, and self-reported acquisition of knowledge, to tap into the impact of the program upon participants' competencies, clinical performance, and in some cases, the impact on clinical practice and patient outcomes. A collaborative sought to explore if CME/CPD designed to improve patient-provider interactions in type 2 diabetes (T2DM) could facilitate patient behavior change. The mixed-methods evaluation approach for this program will be presented as it relates to the evolving nature of outcomes assessments.
Summary of Work: A mixed-methods, repeated-measures, time-series, evaluation was conducted. The evaluation consisted of three types of data collection: (1) learner's pre and post self-assessments questionnaires, (2) IRB-approved pre/post evaluation including self-report questionnaires and qualitative interviews with a sampling of participants and their patients, and (3) pre/post audit of patient medical charts. The research design included methodological, and source triangulation to strengthen the validity of the evidence.
Summary of Results: Results from the program evaluation showed that behavioral change at the health care provider (HCP) level was achieved, positively impacting key clinical markers related to treatment, management, and adherence in patients with T2DM. Evidence of lasting change was present at three- and six-months post-workshop.
Conclusions: This study describes a mixed-methods evaluation approach that was used to determine a causal link between the educational program and improvements in HCP behaviors and patient outcomes, which can be generalized to future program evaluations.
Take-home Messages: The proposed mixed-methods evaluation could be used to inform future educational assessments.

OTT-OD-7-5
Use of a developmental evaluation model to assess curriculum change

Matt Vassar (*Oklahoma State University Center for Health Sciences, Tulsa, USA)*, Machelle Linsenmeyer (*Oklahoma State University Center for Health Sciences, Tulsa*), Johnathan Franklin (*Oklahoma State University Center for Health Sciences, Tulsa*)

Background: Developmental evaluation utilizes a model that supports the development of innovation and adaptation in dynamic environments, and provides a way to obtain rapid, real-time feedback that can be used to make program changes (Patton, 2006; 2010). Our institution has adopted a developmental evaluation approach in the assessment of our new medical school curriculum in order to take advantage of the real-time feedback that can inform accelerated curriculum change.
Summary of Work: Focus groups were conducted with 1st year medical students selected by stratification of exam performance and other characteristics. Online reflection methodologies were also used by a separate group of the student body. Both student groups provided curriculum feedback which was communicated to the Curriculum Oversight Committee and Course Coordinators.   Faculty members participated in interviews conducted by the Office of Educational Development to obtain feedback reflective of their experiences in course delivery. Support staff participated in online surveys to discuss logistical/coordination issues involved with courses. The COM also utilized periodic course evaluations and student-written executive summaries of evaluation feedback.
Summary of Results: Results from these evaluations have been incorporated for rapid improvement of the curriculum. In some cases, course content was clarified quickly or additional time was given to particular concepts. Additional training was provided to small group student facilitators. These are just a few examples of improvements.
Conclusions: Developmental evaluation is a useful model for evaluating curriculum change and is unique in its ability to expedite changes.
Take-home Messages: Medical schools should consider developmental evaluation for curriculum assessment.

OTT-OD-8: WORK BASED ASSESSMENT 2

OTT-OD-8-1
Using Unannounced Standardized Patient visits to assess OSCE practice effects on clinical performance

Colleen Gillespie (*New York University School of Medicine, New York, USA)*, Angela Burgess (*New York University School of Medicine*), Sondra Zabar (*New York University School of Medicine*)

Background: Completing an 10 Objective Structured Clinical Examination (OSCE) could be associated with a "practice effect" leading to enhanced clinical skills. We compared residents' performance in Unannounced Standardized Patient (USP) visits by time elapsed since participation in a 10-station OSCE to explore possible impact on clinical performance.
Summary of Work: 81 USP visits were included (1-4 visits; 38 residents). Both USP visits and OSCE stations reflected a common set of primary care scenarios. SPs (USP and OSCE) used a behaviorally anchored checklist (not done, partly done, well done) to assess Communication (12 items), Treatment Plan (4 items), Patient Education (4-6 items) practices as well as their satisfaction (4 items) and activation (4 items).

Cronbach's alpha > .78. Scores=% well done. Mean USP visit scores were compared by time elapsed since the OSCE, by immediate impact (OSCE < 2 weeks), and by month (up to 6 months; ANOVA).

Summary of Results: Clinical performance, as measured by USPs, did not differ significantly by time elapsed since OSCE (Pearson's r .10 to -.12; p>.05). Mean communication scores did not differ by whether occurred within (75%, SD 38%) or more than 2 weeks (70%, SD 22%) of the OSCE (p=.63). Similar results were found for all clinical skills and also when the interval was months.

Conclusions: In our sample, with a small number of USP visits, an OSCE was not associated with better clinical performance.

Take-home Messages: Concluding that OSCEs do not lead to practice effects will require additional studies.

OTT-OD-8-2
Using Unannounced Standardized Patients (USP) to Assess Quality of Care: Charting and Outpatient Safety

Sondra Zabar (*NYU School of Medicine, Medicine - Division of General Internal Medicine, New York, USA)*, Angela Burgess (*NYU School of Medicine, Medicine - Division of General Internal Medicine, New York*), Kathleen Hanley (*NYU School of Medicine, Medicine - Division of General Internal Medicine, New York*), David Stevens (New York City Health and Hospitals Corporation*, New York*), Adina Kalet (*NYU School of Medicine, Medicine - Division of General Internal Medicine, New York*), Colleen Gillespie (*NYU School of Medicine, Medicine - Division of General Internal Medicine, New York*)

Background: Accurate charting is critical to patient safety yet little is known about residents' documentation of the care they provide. We piloted the use of Unannounced Standardized Patient (USP) visits for assessing documentation skills and practices.

Summary of Work: 15 medicine residents each saw four USP over 6-months in their outpatient practice. After each visit, USPs completed checklists and residents entered their notes believing these to be actual patients. Core variables were abstracted from 60 notes (including correspondence between the chart and USP-report of 12 physical exam elements (over- or under-documentation); documentation of allergies and residents' ordering of labs.

Summary of Results: There was complete or almost complete (>90%) agreement between the chart and the USP-report for 7/12 physical exam (PE) elements. 5 PE elements were over-documented (extremities 27%, neurological 20%, vitals 13%, heart 20%, abdomen 20%) and two were both over- and under-documented (extremities, vitals). 88% of the residents charted the PCN allergy, but only 29% in the correct chartfield. On average, residents ordered 78% of essential labs with substantial variation among residents (SD 9%) (range 22 -100%).

Conclusions: USP methodology is a feasible way to assess resident clinical practice. We found that residents tend to over-document PE elements and to fail to document allergies in the chartfield that triggers a patient safety response. Variation in ordering of labs suggests differences in clinical reasoning skills or system-based practice patterns.

Take-home Messages: USP visits provide a unique method for assessing charting skills and provide essential information for feedback for improving patient safety and outcomes.

OTT-OD-8-3
Assessing Clinical Competence Through a Patient Safety Lens

Beverley Bird (*Monash University, Faculty of Medicine, Nursing & Health Sciences, Clayton, Australia)*, Tangerine Holt (*Australian-American Fulbright Commission, Executive Director, Canberra*), Brian Jolly (*University of Newcastle, School of Medicine & Public Health, Newcastle, NSW*)

Background: Workplace Based Assessment (WBA) methods for interns and residents have been widely adopted. There has, however, been little practical guidance on assessing and predicting future clinical competence through the lens of patient safety.

Summary of Work: An evidence-based Patient Safety Education Model was developed from a comprehensive review of the relevant curricular literature. A mapping exercise identified common competency areas, competency items and competency indicators for the structure of a Patient Safety mini-CEX tool (PSM). The draft tool was tested and modified through OSCE, High Fidelity Simulation and WBA testing, and expert panel review. Reliability was achieved through standardization of clinical scenarios and facilitator and assessor training.   Patient safety focussed formative feedback was an integral component of each testing session.

Summary of Results: One hundred and fifty participants participated in 2 eight –station OSCE rotations, 4 four-scenario high fidelity simulation encounters, and 24 WBA encounters in acute clinical settings. Reliability modelling and expert review of the OSCE and Simulation tools identified that they needed 13-18 items to adequately sample the domain with a Cronbach's Alpha of .937.

Areas where participants consistently performed at an unsatisfactory or borderline standard were Clinical Judgement and Decision Making, Communication, and Managing Information.

All participants valued the patient safety focussed formative feedback. Residents and assessors in workplace settings considered the patient safety focus of feedback to be valuable and non-threatening.

Conclusions: Workplace Based Assessments of clinical competence through the lens of patient safety provided a broader focus.

Take-home Messages: Patient Safety focussed competency assessments in early post-graduate years may engender patient safety orientated practice.

OTT-OD-8-4

Workplace-based Assessments: Using an e-Portfolio system to improve learner feedback

Luke Woodham (*St. George's, University of London, e-Learning Unit, London, UK)*, Trupti Jivram (*St. George's, University of London, e-Learning Unit, London*), Megan Wilkins (*Kingston University and St George's, University of London, School of Radiography, London*), Terry Poulton (St. George's, University of London*, e-Learning Unit, London*)

Background: Workplace-based learning is key to education and training in medicine and healthcare disciplines, but assessing this represents a substantial challenge; practical issues such as printing, sending and storing easily lost forms, and infrequent access to clinical tutors, severely limit the quality of feedback provided to students.

Summary of Work: St George's, University of London (SGUL) introduced an e-Portfolio system, ReallyManaging Assessment, aiming to improve feedback for workplace-based assessments. Data entry is performed "on-the-go" using mobile devices, without need for PC equipment, largely unavailable in a clinical environment. Pilot groups of tutors and students in Radiography and Medicine have been used to evaluate the impact of the system from a practical and educational perspective.

Summary of Results: Both tutors and students have responded favourably to the system. Student feedback was gathered from volunteers on the medicine course, and a wider trial in Radiography has resulted in assessments across their range of competencies being implemented and tailored to suit mobile data entry.

Conclusions: The tool provides functionality that allows students to actively monitor their performance and completion rates, formatively assessing their achievements against relevant competency frameworks. This is not possible using paper-based assessment portfolios, and the improved feedback facilitates just-in-time responses to under-performance, rather than requiring remedial action after summative failures. The use of mobile devices for data entry is essential for this approach to work in a clinical environment.

Take-home Messages: An e-Portfolio system providing easily accessible, high quality feedback can increase the learning value of formative workplace-based assessments.

OTT-OD-8-5
Improving the quality of written feedback from workplace based assessments: The utility index in action

M.H. Bartlett (*Keele University School of Medicine, Stoke on Trent, UK)*, R.K. McKinley (*Keele University School of Medicine, Stoke on Trent*)

Background: Keele's undergraduate medical students spend 113 days on placement in general practice over the 5 year course, during which they have eleven workplace based assessments (WBA) of their consultation skills. A written summary of each assessment is generated for the student.

Summary of Work: These written summaries are scored for quality and the results fed back to the practices, used in tutor development and for quality assurance. We tested a process for assessing the quality of the written summaries for its utility (Van der Vleuten 1996) by testing its reliability, validity and feasibility. Current work tests its educational impact.

Summary of Results: Reliability : G theory with a D study showed the reliability of a single assessor scoring 6 summaries was 0.82.
Validity: The quality score of the written summaries was related positively to students' satisfaction with the teaching in the practices (Kendall's Tau 0.38, p=0.01), which is used as a proxy for the quality of teaching. The students' perceptions of the value of the written summaries were related positively to their satisfaction with the verbal feedback given during the WBAs (Kendall's Tau 0.7, p=0.01).

Conclusions: We introduced this process in the academic year 2011-2012. We are currently assessing its impact on the quality of the WBAs(552 summaries from 99 practices) in 2012-2013 following a year of tutor development informed by this work.
We will describe the development of the process and present the results of the current work on its educational impact.

OTT-OD-9: FEEDBACK POSTGRADUATE 1

OTT-OD-9-1
Trainees' perceptions of the feedback they receive during case based discussions

Lucy Flanders (*North Middlesex University Hospital, London, UK)*, Aruchuna Mohanaruban (*North Middlesex University Hospital, London*)

Background: Workplace based assessments have been incorporated into postgraduate medical training as a useful tool for providing feedback to trainees on their performance. However a disparity exists between what assessors and trainees perceive as useful feedback. We analyzed trainees' perceptions of the feedback they received from case based discussions (CbDs).

Summary of Work: 78 medical trainees at a London hospital completed a 12 item questionnaire rating their last CbD experience including whether it was preplanned or adhoc, the duration spent receiving feedback and how they responded to feedback they received. Focus groups were held to discuss common themes that arose from the questionnaire.

Summary of Results: Only 19% of assessments were preplanned. The average duration of assessments was 6-10mins with time allocation to giving feedback less than 5mins. 76% of trainees responded to the feedback they received by completing self-directed learning or by addressing the specific action points. The focus groups highlighted the barriers to incorporating assessments into every day practice. Trainees felt that feedback was

often too generalized and quality varied according to the seniority of the assessor.

Conclusions: The majority of CbDs were performed on an adhoc basis dedicating little time for feedback. Despite this, most trainees showed a positive engagement in the process by reflecting on cases discussed or completing self-directed learning.

Take-home Messages: This study demonstrates that trainees appreciate the educational value of CbDs but greater emphasis is required in planning these assessments and in providing feedback that is both more specific and actionable.

OTT-OD-9-2
Exploring experienced feedback facilitators' performance in multi-source feedback in specialist training

Gitte Eriksen (*Center of Medical Education, Aarhus University & Aarhus University Hospital, Denmark)*, Gitte Bjørg (*PeoplePartner, Aarhus*), Hans Ehlert (*Business Psychology, Knebel*), Bente Malling (Center of Medical Education, Aarhus University & Aarhus University Hospital, Denmark*, Department of Human Resources, Aarhus*)

Background: Personal feedback in multi-source feedback (MSF) processes in specialist training entails medical doctors to act as feedback facilitators (FF). Novice FFs seem to prefer a deductive rather than facilitating style in the dialogue, thus seem to have difficulties in fostering trainees' reflection. It is not known how experienced FFs perform.

Summary of Work: Video-recordings of 18 feedback sessions were analysed and scored by two independent coaches using a previously developed criteria list for FF-performance. Performance was compared to novice FFs' performance. Participating FFs had performed at least 10 feedback dialogues.

Summary of Results: Overall the skilled FFs obtained a higher global mean score than novice FFs. The experienced FFs had acquired a more facilitating style, compared to novices. They asked various types of questions and potential learning emerged during the dialogue. The trainees were given more opportunities to reflect and to formulate focus areas for further development. However, several learning opportunities were overlooked as it seemed difficult for the FFs to foster elaboration of the learning opportunities. Even for skilled FFs it seemed challenging to guide trainees to work out a developmental plan.

Conclusions: Feedback facilitators develop a more coaching style with experience. Experienced FFs managed to shift focus from data presentation to data analysis and thus create reflection in the dialogue. However, there are still difficulties in making out a developmental plan.

Take-home Messages: Feedback facilitators acquired guiding and coaching styles with experience, however further training in coaching techniques and formulation of developmental plans might be advisable.

OTT-OD-9-3
How surgeon teachers use feedback to assess residents in the Operating Room

Deepak Dath (*McMaster University, Surgery, Hamilton, Canada)*, Jennifer Hoogenes (*McMaster University, Surgery, Hamilton*), Edward Matsumoto (*McMaster University, Surgery, Hamilton*), David Szalay (McMaster University*, Surgery, Hamilton*)

Background: Feedback is absolutely central to teaching and assessment in the operating room (OR). However, recent literature suggests several reasons why feedback is not well utilized in the OR. This qualitative study examines how surgeon teachers use feedback as a teaching and assessment technique in an effort to provide a foundational understanding of the current practice of giving feedback to residents in the OR.

Summary of Work: In this qualitative study, feedback was identified as a theme for exploration. A qualitative methodologist ensured adherence to rigorous process. Surgeon teachers from various disciplines in a single tertiary care centre participated in one of ten surgeon-facilitated focus groups. Transcripts were analyzed using qualitative content analysis to develop a codebook. Themes were derived from further discussion and consensus.

Summary of Results: Fifty-four surgeon teachers discussed 17 concepts around feedback. These concepts fell into 4 main themes: an understanding of feedback; the technique of giving feedback; the timing of feedback, and; the affective component of giving feedback. Although these themes emerged from the surgeons' descriptions, discussion was not rich in specific techniques about how to give feedback and the level of understanding was not uniform.

Conclusions: Surgeon teachers describe feedback and consider it an important part of their intraoperative teaching and assessment strategy, but their discussions demonstrate only a rudimentary understanding of the concepts and techniques of feedback.

Take-home Messages: To raise the quality of teaching significantly for our surgical residents we must design and deliver faculty development events that teach and train surgeons to give feedback appropriately in the OR.

OTT-OD-9-4
Theory, evidence and testing: Development of a model for facilitating performance feedback

Joan Sargeant (*Dalhousie University, Halifax, Canada)*, Heather Armson (*University Calgary, Calgary*), Eric Holmboe (*ABIM, Philadelphia*), Jocelyn Lockyer (University of Calgary*, Calgary*), Karen Mann (*Dalhousie University, Halifax*), Ivan Silver (*CAMH, Toronto*)

Background: Studies of physicians' responses to formal performance feedback show that it is not always accepted or used for practice improvement. This occurs for various reasons - inconsistency of feedback with self-perceptions, concerns regarding credibility, and perceived barriers to data use. The purpose of this

research was to develop an evidence and theory-informed model for facilitating physicians' performance feedback and practice improvement.

Summary of Work: We reviewed the theory and evidence informing approaches for enhancing feedback acceptance and use. Three approaches appeared especially applicable: person-centred approaches (Rogers 1969), integration of external feedback with one's self-assessment (Sargeant et al 2010), and cognitive domains influencing knowledge translation (Cane, 2012). Drawing upon these, we developed a 4-stage model for facilitating feedback. We prepared 12 physician facilitators from 3 regulatory programs in model use. Each conducted a feedback interview with a physician using his/her formal performance report. Data collection included: initial videotaped feedback and debrief sessions (n=4); subsequent audiotaped feedback and debrief sessions (n=8). We used content and interpretive analysis of the videos and transcripts to understand the model and its components.

Summary of Results: Evidence, theory and testing support the model for facilitating feedback. The 4 stages include: 1. build rapport and relationships, 2. explore reactions, 3. explore content, 4. coach for change.

Conclusions: The stages of the model are iterative and each is important. Next research steps will include testing the model in additional settings and its impact upon performance change.

Take-home Messages: Both facilitators and recipients found the model useful in understanding the performance data and considering change.

OTT-OD-9-5
Moving from numbers to words: assessing the quality of a written narrative leadership feedback tool

Susan Kennedy (*East Kent Hospital University NHS Foundation Trust, Medical Education Directorate / Leadership Faculty, Margate, UK)*, Bohaira Geyoushi (*East Kent Hospital University NHS Foundation Trust, Medical Education Directorate/Leadership Faculty, Ashford*), Rags Subramaniam (*East Kent Hospital University NHS Foundation Trust, Medical Education Directorate / Leadership Faculty, Margate*)

Background: Increasingly, educational theories around assessment focus on the use of narrative, qualitative feedback rather than more quantitative methods. Research points to the effectiveness of narrative feedback, reflection and action planning. We wanted to investigate the quality of a relatively new narrative leadership assessment tool. LEADER is used as a record of a professional conversation between trainer and trainee.

Summary of Work: Using specificity of feedback, evidence of reflection and the recording of developmental goals as our primary indicators of quality, we analysed a sample of LEADER tools. Each LEADER was analysed for the three indicators of quality with a 4 point scale for each indicator ranging from 0 to

4 e.g 0, no specific feedback through to 3, very specific feedback.

Summary of Results: 80% of assessments provided specific/very specific feedback. 75% indicated very clear/clear developmental goals being set. Evidence of reflection was least in evidence with 50% of LEADER tools analysed providing no evidence of trainee reflection. This may have been due to poor recording of this aspect of the conversation.

Conclusions: The LEADER tool is a new, qualitative assessment tool that seeks to build a complex picture of the trainee as emerging leader, formed with words, that arises out of a professional conversation. This local, small-scale analysis evidences valuable learning conversations that involve specific feedback and developmental goals. Less evident is trainee reflection on their leadership experiences and competencies.

Take-home Messages: In terms of specificity of feedback and the setting of developmental goals, LEADER as an instrument is helpful in shaping rich narrative assessment.

POSTERS

OTT-PAB1: WRITTEN ASSESSMENT

OTT-PAB1-01
Do gender and candidate instructions impact self-confidence of candidates sitting a multiple-choice assessment?

Danny Mathysen (*Antwerp University Hospital, Department of Ophthalmology, Edegem, Belgium)*, Christina Grupcheva (*Specialized Eye Hospital, Department of Ophthalmology, Varna*)

Background: Intuitively it may be expected that as the candidate ability level increases, the level of confidence of the candidate should increase as well. However, little is known about the effect on the level of confidence of gender and candidate instructions prior to the assessment.

Summary of Work: The level of confidence of candidates participating at the European Board of Ophthalmology Diploma (EBOD) examination was objectively assessed by evaluation of the amount of use of the "don't know" answer option for multiple-choice test items. A statistical analysis of the level of self-confidence in relation to gender of candidates and in relation to the instructions provided prior to the EBOD examination was performed.

Summary of Results: For the EBOD 2012 and EBOD 2013 examination the most able candidates have shown to be more self-confident (choice of the "don't know" answer option for a median of respectively 10 and 15 test items) compared to the least able candidates (choice of the "don't know" answer option for a median of respectively 46 and 49 test items). Over all editions of the EBOD examination, it was observed that male

candidates had significantly higher self-confidence levels.

Conclusions: Results obtained from not less than 1,439 candidates over 4 EBOD examinations have shown that as candidate ability level increases, the level of self-confidence of the candidates also increases, although in general female candidates appear to be less self-confident compared to their male colleagues. Further research will be needed to explore whether or not highly able female candidates are more self-confident. Furthermore it has been shown that assessment of the level of self-confidence can be enhanced by providing more detailed candidate instructions prior to the assessment.

Take-home Messages: Detailed candidate instructions are instrumental to optimize candidates to trust on their self-confidence

OTT-PAB1-02
Formative MEQ: Can it be the challenging tool for preclinical medical students in encouraging their own self-directed learning and improving their own scores?

Panadda Rojpibulstit (*Faculty of Medicine, Thammasat University, Biochemistry, Pathumthani, Thailand*), Nuchanart Suealek (*Faculty of Medicine, Thammasat University, Pharmacology, Pathumthani*)

Background: The purpose of this study was to determine whether formative MEQ could trigger medical students' SDL and stimulate the improvement of their scores in PBL curriculum or not.

Summary of Work: How to develop problem solving self-directed learning skills for formative MEQ examination (at the 3rd week of the 4-week endocrinology block) was informed week by week since the orientation session. After finishing formative exam, 3rd- year preclinical students at the Faulty of Medicine, TU (n=173), were assigned to complete thirty-eight items with a five rating scale questionnaire. Eleven and six were its SDL-triggering efficiency (A) and the problems due to inexperienced in formative (B), respectively. The last twenty-two was the satisfaction on the instruction (C), content (D), answering processes (E). Once the entire questionnaire was completed, data was analyzed using mean ± SD and the pair-t test comparing between formative and summative MEQ was also examined.

Summary of Results: From 98.84% response rate, mean scores on the (A), (B), (C), (D) and (E) questionnaires were 4.3, 4.3, 4.2, 4.2, 4.2, respectively. Average subscales scores varied from M=4.0 to M=4.7 (5: totally agree, 1: totally disagree). Interestingly, the most significant improvement scores from formative to summative MEQ were the top and the last thirty learner students (p-value <0.05) while insignificant difference was detected in the middle score students.

Conclusions: This study demonstrated the potential of formative MEQ in triggering SDL behaviors and improving students' ability for both the "top-thirty" and "last-thirty" learners.

Take-home Messages: To continuously stimulate SDL behaviors, implementation of formative MEQ in other blocks should be considered.

OTT-PAB1-03
Investigating item parameter drift in a multiple-choice based clerkship examination: a multiple case study

Meredith Young (*McGill University, Medicine, Centre for Medical Education, Montreal, Canada*), Beth-Ann Cummings (*McGill University, Medicine, Centre for Medical Education, Montreal*), Christina St-Onge (*Universite de Sherbrooke, Medicine, Sherbrooke*)

Background: Written examinations are often used during clerkship rotations. Unfortunately, their psychometric properties may be less stable (i.e. item parameters may drift across cohorts) since clerkship cohorts are often small.. Since item parameters underlie assessment decisions (e.g. items excluded from the final score), variations due to many small clerkship cohorts raises concerns for assessment equivalency. We set out to examine parameter drift across cohorts of students completing an MCQ-based clerkship exam.

Summary of Work: Exam results for a 32-item MCQ-exam were used. Discrimination and difficulty indices were calculated for 22 cohorts (n=10-15 students). Discrimination coefficients were categorized according to Ebel. Difficulty coefficients were categorized according to three guidelines proposed by Laveault & Gregoire. Descriptive analyses examined parameter drift for item discrimination and difficulty.

Summary of Results: Discrimination coefficients for individual items varied greatly across cohorts, with all items occurring in both Ebel's "poor" and "excellent" categories and 16/32 items occurring in all 5 categories. For item difficulty, the application of different guidelines resulted in a wide range of items removed (from 1 to 20 items), and in widely different predicted exam difficulty (from .61 to .85).

Conclusions: Large amounts of parameter drift was observed across cohorts, for item discrimination and difficulty.

Take-home Messages: The findings of this study relate primarily to the application of assessment guidelines and standards within medical education. Understanding the applicability of standards of item characteristics, with consideration to assessment equivalency, is of utmost importance, particularly as we move towards a competency-based health professions education model.

OTT-PAB1-04
Using a multiple choice examination to assess and enhance Internal Medicine residents' knowledge of evidence-based clinical examination

Rebecca Jarvis (*University of Western Ontario, London, Ontario, Canada*), Elaine Zibrowski (*University of Western Ontario, London, Ontario*), Kathryn Myers (*University of Western Ontario, London, Ontario*)

Background: Interpreting findings from the history and physical examination is a critical skill for residents as these findings guide investigations and patient management. These skills are primarily acquired through bedside case review. Residents consistently express to their teachers a desire for more bedside clinical teaching, but time constraints often intervene. Multiple choice questions provide a reliable, efficient tool for assessing knowledge. We developed the Evidence-based Examination MCQ (EEMCQ) to allow residents to assess their knowledge in applying evidence to examination interpretation with a goal of providing them with formative feedback about these skills.

Summary of Work: The EEMCQ is a 50 question, case-based multiple choice examination developed to assess Internal Medicine residents' interpretation of the clinical examination. Items were developed using standard references in evidence-based examination. We pilot-tested the questions with faculty and revised them based on their feedback. The EEMCQ was then administered to a cohort of PGY-1 through PGY-3 Internal Medicine residents (n=70) at an Academic Half Day. Some of the questions were reviewed with the residents after the administration of the examination as a formative exercise. A satisfaction survey was completed by the residents at the end of the examination.

Summary of Results: The overall mean score on the EEMCQ was 27/50 (min 12, max 40). Internal consistency was moderate for items with an item-item correlation above 0.2 (alpha=0.64). While a main effect of postgraduate year on exam performance was not detected, residents reported that the EEMCQ was of appropriate difficulty and content. They perceived that the examination was relevant to their learning and would stimulate interest in studying evidence around the clinical examination. They expressed a desire to have the examination available and linked to the evidence to enhance their learning in this area.

Conclusions: Resident feedback suggests that the EEMCQ provided them with useful formative feedback about their ability to interpret clinical examination findings. Low overall scores suggest that applying evidence to the interpretation of the history and physical examination remains challenging for residents. The development of a learning module based on the EEMCQ that the residents can access, and that is linked to the evidence, is underway.

OTT-PAB1-05
The development and piloting of a key features exam based on a national clerkship curriculum

Kirk Bronander (*University of Nevada School of Medicine, Internal Medicine, Reno, NV, USA)*, Valerie Lang (*University of Rochester School of Medicine and Dentistry, Internal Medicine, Rochester, NY*), James Nixon (*University of Minnesota School of Medicine, Internal Medicine, Minneapolis, MN*), Norman Berman (Geisel School of Medicine*, Pediatrics, Hanover, NH*)

Background: A key feature exam (KFE) based on a US national clerkship curriculum has not previously been developed or pilot tested.

Summary of Work: In consultation with an expert in KFEs, nine internal medicine educators from varying US medical schools convened over four days and worked online to develop a KFE. They began by defining the key features (KFs) - the critical steps in the identification and resolution of a problem with a focus on steps where students are likely to go wrong. Seventy-two KFs based on the Clerkship Directors in Internal Medicine (CDIM) core curriculum were refined until consensus was achieved. Case vignettes and questions were written for these KFs. Each KF and subsequent case was peer reviewed within the working group. The KFE was pilot tested in an electronic format with 109 students at six medical schools. An electronic survey assessed students' attitudes with a 5-point Likert scale.

Summary of Results: Seventy-six students (69.7%) completed the survey. The portion of students who agreed or strongly agreed with each question included: KFE cases resembled problems from clinical practice (63; 82.9%); the KFE accurately represented clerkship content (46; 60.5%); the KFE accurately assessed ability to apply what was learned in the clerkship (41; 53.9%); and the format more accurately assessed ability than a MCQ exam (35; 46%).

Conclusions: 1) Development of a KFE is a labor intensive process. 2) Although students found the exam a more authentic representation of clinical problems, they were more comfortable with a MCQ format.

OTT-PAB1-06
Development of applicable knowledge assesed by confidence-based testing

Daniela Kampmeyer (*University of Cologne, Pharmacology, Cologne, Germany)*, Jan Matthes (*University of Cologne, Pharmacology, Cologne*), Presenter: Stefan Herzig (*University of Cologne, Pharmacology, Cologne, Germany*)

Background: Medical "knowledge" as assessed in multiple-choice (MC) examinations is biased by guessing. However, to be safely applied in practice, a correct self-judgment is also required ("I know it, and I know that I know it"). This led to "confidence-based testing", where students have to state their level of confidence for each answer. We investigated whether "applicable knowledge" improves during undergraduate education.

Summary of Work: Teaching by pharmacologists is embedded longitudinally in the University of Cologne Medical Curriculum, with two disciplinary courses (3rd and 5th year), plus many interdisciplinary lectures and small-group sessions across years 1-6. We firmly assumed that students will improve over years with respect to "knowledge", and tested whether their confidence would develop congruently. A formative exam offered to 3rd-and 5th-year students (40 and 30 MC items, respectively) was deliberately taken by 129 (of 160 course participants) and 87 (of 145), respectively. MC-items were scored binary (1-correct; 0-

wrong), and confidence using a 4-point scale (0-guess; 0.33-quite unsure; 0.66-quite sure; 1-very sure).

Summary of Results: 5th-year students had higher knowledge (mean 0.73+/-0.16 versus 0.49 $\square$0.13) and confidence (0.61 $\square$0.18 versus 0.46+/0.13) The congruence between the two parameters was analyzed according to Bland-Altman.

Conclusions: Astonishingly, confidence develops in parallel but not congruently with "knowledge". Advanced students display lower confidence at any given level of knowledge. Yet, the proportion of 5th-year students giving mostly "informed answers" was higher (41% versus 5%), along with lower inter-item variance of congruence.

Take-home Messages: Confidence-based testing requires extensive statistics to demonstrate an increase in "applicable knowledge".

OTT-PAB1-07

What is the technical quality of the written assessment used to qualify medical education in Brazil?

Fábio Antonio Perecim Volpe (*Barretos School of Health Sciences, Dr. Paulo Prata - FACISB, Barretos, Brazil)*, José Henrique Benedetti Piccoli Ferreira (*Barretos School of Health Sciences, Dr. Paulo Prata - FACISB, Barretos*), Céline Pinheiro (*Barretos School of Health Sciences, Dr. Paulo Prata - FACISB, Barretos*), Lucas Tadeu Bidinotto (Barretos School of Health Sciences, Dr. Paulo Prata - FACISB*, Barretos*), Maurício Gomes Meirelles (*Barretos School of Health Sciences, Dr. Paulo Prata - FACISB, Barretos*), Eduardo Anselmo Garcia (*Barretos School of Health Sciences, Dr. Paulo Prata - FACISB, Barretos*)

Background: One of the measurements to evaluate higher education institutions in Brazil is a written exam, applied every three years to students of the first and last years. We analyzed all assessments applied so far to medical students as regards the relevance and technical quality of the questions.

Summary of Work: Six faculty members rated the quality of the questions of the written assessments applied in 2004, 2007 and 2010. MCQs were analyzed according to the criteria stipulated by Jozefowiz RF et al. 2002. Scores varied from 1-5, 5 being awarded to items that included clinical vignette, one-best-answer format and no item flaws. Items were scored as 1 if presenting at least two fails, such as involving recall and significant item flaws. Open-ended questions were evaluated considering the presence of a clinical vignette, Bloom's taxonomy, and the clarity of the lead-in. We also classified all items according to the medical relevance based on the national curricular recommendations.

Summary of Results: When analyzing the three assessments together, MCQs quality average score was 2.54. Open-ended items percentage with clinical vignette was 78.43%, items with at least analysis skill order in Bloom's taxonomy were 37.26% and questions with lead-in clarity were 84.31%. 71.1% of the items were considered relevant.

Conclusions: Despite being relevant and directed to medical practice, this assessment shows fragilities in the technical quality of their MCQs and Open-Ended items. Considering its purpose, this format of assessment contains important limitations.

Take-home Messages: The format of the assessment should be better related to its purpose so it can be successfully reached.

OTT-PAB1-08
Is MEQ better than MCQ to assess problem-solving skill in preclinical medical students?

Nampet Ngodngamthaweesuk (*Thammasat University, Faculty of Medicine, Klonglaung, Thailand)*, Panadda Rojpibulstit (*Thammasat University, Faculty of Medicine, Klonglaung*)

Background: Though the advantages of MCQ examination including justice, objectivity and ease of use supported it as the most widely tool to evaluate medical students in my institute, assessment of the level III cognitive domain or problem-solving skills were still controversial especially comparing with MEQ. Hence, this study was designed to compare the effectiveness of MEQ versus MCQ in assessing problem-solving skills.

Summary of Work: The twenty high achievers and twenty low achievers of the third year medical students are grouped by previous performance in three sessional examinations. MEQs and MCQs were formed in endocrine block by the specialists in each discipline and subsequently ensured by the block committee. Data was analyzed using mean ± SD and pair-t test.

Summary of Results: Mean score of MCQ (76.17) was significant difference ($p < 0.05$) with MEQ score (70.90). Additionally, when individual scores from MCQ and MEQ were compared, the most statistical difference was between only those of the top twenty achievers ($p <0.05$)but not in the last twenty achievers($p>0.05$).

Conclusions: Assessment by MEQ and MCQ can differentiate the level III cognitive domain especially among high achievers but not in low achievers.

Take-home Messages: Definite proportion between MCQ and MEQ should be considered for assessing the cognitive domain of preclinical medical students rather than using only one assessment tool.

OTT-PAB1-09
The effect of asking for explanation for choosing an option on an MCQ test on students' scores in a Biochemistry module test

Nadia Al Wardy (*Sultan Qaboos University, Medical Education Unit / Biochemistry, Al Khod, Oman)*, Marit Delghandi (*Sultan Qaboos University, Biochemistry, Al Khod*), Syed Rizvi (*Sultan Qaboos University, Family Medicine & Public Health/ Epidemiology and Statistics Unit, Al Khod*)

Background: MCQs are selected response formats of questions in which there is no provision for explaining why an option was selected but rather it is assumed

that students know why they are selecting the correct answer. In this study we wanted to:
a. explore the effect of asking students to explain their answer on their overall score,
b. compare that score with the one they would have obtained if no explanation was required, and,
c. explore if there was a correlation between the scores of students who provided or did not provide explanations and their performance on the overall course.

Summary of Work: Five MCQs were included in an open-book test in a Biochemistry module. For each question, students were asked to explain why they chose a particular option. Each question was allocated 2 marks; 1 for selecting the correct option and 1 or 0.5 for the correct explanation. Marks were aggregated and analysed. Eighty eight students took the test giving a total number of 440 questions to be analyzed.

Summary of Results: The number of questions answered correctly but without providing an adequate explanation were 47 (10.7%). Difference between the test averages, calculated with and without consideration for explanation, was not found to be statistically significant. However, students who scored highly on the test when calculated with marks for explanation scored higher in other tests in the module compared to students who scored highly on the test when calculated without marks for explanation.

Conclusions: Students can select correct answers to MCQs without being able to provide explanation. However, asking to provide explanation for selecting a particular option can be advantageous for the better performers.

Take-home Messages: Some students can select the correct answer to an MCQ without knowing why.

OTT-PAB1-10
Pharmacy students' perceptions and performance in open and closed book examinations

Srinivasan Ramamurthy (*International Medical University, Kuala Lumpur, Malaysia)*, Hui Meng Er (*International Medical University, Kuala Lumpur*), Vishna Devi Nadarajah (*International Medical University, Kuala Lumpur*), Peter CK Pook (International Medical University*, Kuala Lumpur*)

Background: There is a huge body of knowledge that is constantly changing and growing in every field. It is important that students are able to use and manage knowledge effectively. Open-book examination can be incorporated to meet this change. This study aimed to investigate pharmacy students' perception, reasons for their preference for open or closed book examinations, and their performance in the examinations.

Summary of Work: The study was conducted among Year 1 and Year 2 pharmacy students of International Medical University, Kuala Lumpur. The students in each cohort undertook both open and closed book online formative examinations. At the end of the open book examination, students filled in an online questionnaire consisting of questions related to their experience, number and usefulness of online resources and references referred to during the examination, and their preference with reasons on the types of examinations. In addition, the difference in the mean scores between open and closed book examinations were analysed using t-test.

Summary of Results: Majority of the students preferred open-book examinations, with the main reasons being less anxiety, no memorisation, more room for logical thinking and more problem solving.Students scored significantly higher in open book examination than closed book examination.

Conclusions: Overall students' examination experience and their positive perception of open book examination provide reassurance that the use of the open book examination will complement existing assessment methods.

Take-home Messages: Openbook examinations promote critical thinking rather than memorisation, reduce stress and encourage students to self-monitor their learning.

OTT-PAB1-11
Standard-setting: Retrospective validation of item difficulty and importance judgments

George Pachev (*University of British Columbia, Faculty of Medicine, Vancouver, Canada)*, Cicely Bryce (*University of British Columbia, Faculty of Medicine, Vancouver*), Andrea Busse (*University of British Columbia, Faculty of Medicine, Vancouver*), Niyati Kasana (University of British Columbia*, Faculty of Medicine, Vancouver*)

Background: The purpose of this study, a part of a continuous quality improvement process, was to validate key judgments made during a modified Ebel's standard-setting procedure for the written examinations of two consecutive iterations of a Clinical Skills course. During the standard-setting session, a content expert classified each item according to its difficulty (three levels) and importance (four-level definition for one exam and three-level definition – for the other). Expected performance-levels of "borderline examinees" were then assigned to the each of the categories formed by the combinations of the importance and difficulty subcategories.

Summary of Work: The approach to validation involved comparisons of expected performance levels of "borderline examinees" to the actual performance levels of the lowest scoring 15 % of each class, for items in each of the categories of item difficulty, importance, and the combinations of their levels. In addition, the discrepancies between actual and expected performance, as generated by three-level and four-level "importance" definitions were compared, to evaluate the effectiveness of each definition for setting the passing standard.

Summary of Results: The content expert underestimated the performance-levels of examinees on items classified as most difficult and least important. The three-level definition of "importance" yielded less discrepancy between expected and actual performance.

Conclusions: Recommendations for improvement of the process include: refinement of categories' definitions

and use of three-level importance definition. When available, actual performance levels from previous administrations of items should be used to guide the standard-setting.

OTT-PAB1-12
Using progress tests for cross-institutional comparisons of growth of knowledge: making the context count

Sebastian Schubert (*Charité - Universitätsmedizin Berlin, Dieter Scheffner Center for Medical Teaching and Educational Research, Berlin, Germany)*, Zineb M Nouns (*Charité - Universitätsmedizin Berlin, Dieter Scheffner Center for Medical Teaching and Educational Research, Berlin*), Stefan K Schauber (*Charité - Universitätsmedizin Berlin, Dieter Scheffner Center for Medical Teaching and Educational Research, Berlin*)

Background: Cross-institutional comparisons have been of interest in medical education frequently. A common problem in this context is that the result of such a comparison might merely reflect differences in the familiarity with certain content and not differences in academic achievement. A possible way to approach this issue has been proposed within the framework of explanatory item response modeling. Critically, this approach views the particular items in a given test as a random sample from a larger universe of items. It therefore offers a way to explore between-school differences while at the same time quantifying the degree to which results are dependent on the educational context.
Summary of Work: Data from six consecutive progress tests obtained in two different medical curricula was analyzed. We conducted two types of analysis: First, we conducted comparisons based on grand means by the use of t-tests. Second, we analyzed cross-institutional differences incorporating a more complex model within the framework of explanatory item response modeling.
Summary of Results: The analysis indicates crucial differences in term of the results between the two approaches. While comparisons based on t-tests revealed statistically significant differences between these learning environments, an explanatory item response model did not. These findings suggest that in this particular case the result of comparisons may be more dependent on the particular content administered than on factual difference in achievement.
Conclusions: Although being no panacea, explanatory item response theory offers a comprehensive and feasible approach to address particular concerns in cross-institutional comparisons.
Take-home Messages: The appropriateness of approaches to cross-institutional comparisons has to be considered carefully.

OTT-PAB1-13
Progress testing: two countries divided by a common language

Barbara O'Connor (*University of Auckland, Medical Programme Directorate, Auckland, New Zealand)*, Steven Lillis (*University of Auckland, Medical Programme Directorate, Auckland*), Kimberley Buckley (*University of Auckland, Medical Programme Directorate, Auckland*), Adrian Freeman (University of Exeter, *Medical School, Exeter*), Warwick Bagg (*University of Auckland, Medical Programme Directorate, Auckland*)

Background: The internationalisation of progress testing (PT) has anticipated benefits, but also challenges. We report on the transferability of items from an 'experienced' to a 'novice' university, who share a common first language.
Summary of Work: Two people analysed the changes required to customise a PT of 125 questions from a UK University. Initial modifications were agreed with the partner university through joint moderation. Analysis is underway for three additional PTs following a common blueprint. The fourth makes use of access to a UK question bank and a locally-adapted blueprint, instead of customising an individual test.
Summary of Results: 10 questions for the first PT, sourced from another UK question bank were omitted for propriety reasons and replaced with local questions. All questions were modified to include ethnicity, aligning with local curriculum developments. After these two changes, 46% (58/125) questions required modifications, of which 98% were regarded as minor. Five categories of change were identified: national workforce titles (28), medical terminology (28), drug names (5), changes in units (8), cultural context (8). Changes to four questions were regarded as major.
Conclusions: While sharing PTs and items is an efficient and effective method of constructing tests, a significant amount of work is required to customise test items and tests for a different country, even with a common language.
Take-home Messages: A locally-based moderating group is important, even for pre-moderated questions, and minimal governance is needed for minor and major changes. Consider whether a collaboration will share tests or test items.

OTT-PAB1-14
A growing literature on key feature questions for assessment of clinical reasoning

Patricia Hrynchak (*University of Waterloo, Optometry and Vision Science, Waterloo, Ontario, Canada)*, Susan Glover Takahashi (*University of Toronto, Faculty of Medicine, Toronto, Ontario*), Marla Nayer (*University of Toronto, Faculty of Medicine, Toronto, Ontario*)

Background: The purpose of this presentation is to review the evidence on the reliability and validity of key feature exams (KFEs) to determine if they are an appropriate assessment method for clinical reasoning in health professions education and certification processes.
Summary of Work: A scoping literature review was conducted in Medline (1946 to 2012) and Embase (1980 to 2012). Search terms included: key-feature, question

or test or exam, or assess or evaluation, case-based or case-specific. Articles not in English were eliminated.

Summary of Results: The literature search resulted in 560 articles. Duplicates were eliminated, as well as the articles that were not relevant, leaving 15 articles. A review of the references and a citation review from these articles resulted in an additional 12 articles for a total of 27 for this review. Format, language and scoring of KFEs have been studied and modified to maximize reliability. Internal consistency reliability has been demonstrated to be between 0.49-0.95. Case specificity has not been shown to be a consistent problem. Face and content validity have been shown to be moderate to high. Construct validity has been shown to be good using vector thinking processes, novice vs. expert paradigms and only moderate correlations with knowledge-based examinations. The results of KFEs have been shown to successfully predict future physician performance, demonstrate the effect of fatigue on performance and study the impact of teaching methods on clinical reasoning.

Conclusions: Considering the evidence on the reliability and validity of KFEs presented, KFEs used to assess clinical reasoning can achieve good reliability scores and have good validity properties when properly designed.

Take-home Messages: Consideration should be given to increasing the use of KFEs in assessment systems for health professions education.

OTT-PA5: TECHNOLOGY

OTT-PA5-01
Enhancement of formative assessment in blended learning of histology

Arnaud Van Hecke (*Université Catholique de Louvain, Louvain Medical School, Brussels, Belgium)*, Marie-Christine Many (*Université Catholique de Louvain, Louvain Medical School, Brussels*), Boucquey Marie (*Université Catholique de Louvain, Louvain Medical School, Brussels*), Jean-François Denef (Université Catholique de Louvain*, Louvain Medical School, Brussels*)

Background: Formative self-assessment exercises were developed in the teaching of general histology for 1rst year students in medical science Brussels using the Learning Management System Claroline®. This allowed to manage large number of students. A team of ten teaching assistants in charge of the practical classes was supervised by a single teacher.

Summary of Work: The classical quiz initial system was enhanced with two new types of questions : a)"hotspot" exercises, based on the identification of specific structures in complex images as it is a basic competency expected at the final year examination; b) individual scenarios, i.e. tree-branched succession of questions, depending on success or failure of the previous ones. Feedbacks depending on answers were proposed immediately after each question as well at the end of the whole exercise.

At the end of the year, a survey containing of 66 Likert scale-based items, and 6 identification items was proposed to students.

Summary of Results: 9,95% of students (74/744) completed the survey. Most considered that exercises were useful for their learning, weren't a waste of time, enhanced their learning in their histology skills, showed enough diversity and that feedbacks were useful for learning.

Conclusions: Our results indicate that initial blended learning can be improved with self-evaluation based on ICT techniques, and new active learning resources can be produced. Research and production of ICT modules to stepwise help students to acquire complex competences should be carried out.

Take-home Messages: These results appear as an encouraging signal to further integration of new technologies and 'Socrates pedagogy' in the learning of medicine.

OTT-PA5-02
A continuous assessment tool for teaching and assessing clinical pattern recognition

Sally Krasne (*David Geffen School of Medicine, UCLA, Physiology, Los Angeles, USA)*, Carl Stevens (*David Geffen School of Medicine, UCLA, Emergency Medicine, Los Angeles*), Philip Kellman (*University of California, Los Angeles, Psychology, Los Angeles*)

Background: Clinical pattern recognition is mastered via non-systematic experience in clinical practice. We have developed and implemented a new tool, the Perceptual and Adaptive Learning Module (PALM), for efficiently training clinical pattern recognition to pre-determined criterion levels.

Summary of Work: In PALMs, we categorize images according to the common features (e.g., diagnoses) to be discriminated, with each category containing a large number (25-40) of novel exemplars presented in flash card-like fashion. Unlike flash cards, the sequence in which categories are presented is continuously updated based on the accuracy and response times of the previous exemplars from each category. Categories are retired as the pre-determined criteria for successive, accurate responses within a target response time are met. This continuous sequencing and retirement focuses learning on clinical patterns that are least well recognized. The PALM is complete once all categories have been retired.

Summary of Results: We use this tool for teaching 12-lead ECG interpretation, basic thoracic radiologic anatomy (x-ray and CT), basic histopathology, and dermatology lesion feature recognition. Multiple choice pre- and post-tests are used to evaluate the effectiveness of the PALM. In all cases, we find highly significant ($p<0.0001$) improvement with large effect sizes (typically >1) in the ability of the users to recognize and interpret the clinical images. We also observe long-lasting retention of the learning.

Conclusions: We find PALMs to be an efficient and effective means of teaching interpretation of clinical

patterns to undergraduate medical students and residents.

Take-home Messages: PALMs can provide effective, criterion-based tools for learning, assessing and maintaining skills for a wide range of disciplines.

OTT-PA5-04
DISECKT – cutting edge data-informed self-evaluation

Alaa Aboulhoda (*University of Calgary, Physician Learning Program, Calgary, Canada)*, David Topps (*University of Calgary, Family Medicine, Calgary*), Heather Armson (*University of Calgary, Family Medicine, Calgary*), Elaine Chow-Baker (University of Calgary*, Physician Learning Program, Calgary*), Lara Cooke (*University of Calgary, Office of Continuing Medical Education, Calgary*)

Background: Informed self-assessment has shown promising potential in guiding educational activities in continuing medical education (CME). The Alberta Physician Learning Program (PLP) has been effective in collating data sources from health information systems to inform clinicians about their practices. But simply making data available is not sufficient to produce behaviour change.

Summary of Work: Our program has developed a series of data-informed virtual patients that provide clinicians with challenges that assess their clinical decision making pathways, in an online, easily accessible and reproducible format, using open-source software. We explore the boundaries of clinical practice between the dogma of practice guidelines and the pragma of real life. Clinical areas include low back pain, warfarin management, abdominal pain, headaches and schizophrenia management.

Summary of Results: Linking aspects of each case into local and regional data repositories provides more authentic context and feedback. Small but measurable changes, in how clinicians tackle these cases as they progress in their thinking, show potential for significant savings in health resource utilization when broadly applied.

Conclusions: Basing key clinical elements of virtual patient cases on data derived from regional health information repositories enhances clinical relevance and highlights key system challenges.

Take-home Messages: Grounding such data in clinical scenarios with realistic diagnostic and therapeutic dilemmas is more engaging than simply providing data related to guidelines.

OTT-PA5-05
Medical students' self-assessments and reflections on their learning from a new Virtual Patient model for primary care

Helena Salminen (*Karolinska Institutet, Centre for Family Medicine, Huddinge, Sweden)*, Björklund Karin (*Karolinska Institutet, Centre for Family Medicine, Huddinge*), Eva Toth-Pal (*Karolinska Institutet, Centre for Family Medicine, Huddinge*), Nabil Zary (Karolinska Institutet*, Centre for Learning and Informatics, Solna*), Leanderson Charlotte (*Karolinska Institutet, Centre for Family Medicine, Huddinge*)

Background: A virtual patient (VP) model was created with embedded process skills for communication that facilitates medical students' reflection and clinical reasoning. The VP model consists of iterated learning cycles with a didactic inventory, a concrete experience, pre-formulated feedback and a reflection in each cycle. The students formulate their learning goals and self-assess their learning at the end of the embedded VP case. The aim was to investigate the students' learning goals, iterative reflections and self-assessments of learning from learning activities generated from the VP model.

Summary of Work: The self-formulated learning goals, iterated reflections and self-assessments of 14 last year medical students' VP cases were collected. The students' free text answers while performing the VP case were analysed using content analysis.
The students' self-formulated learning goals related to knowledge, clinical skills, clinical reasoning, communication, strategies to handle patients and to integrate theory with practice.

Summary of Results: Five themes were identified in the students' reflections and self-assessments of their learning.
• The students' level of knowledge and competence was surfaced.
• They found strategies to handle the complex patients in primary care.
• They identified their own needs for learning and professional development and gaps in knowledge.
• They learned about a holistic perspective with the patient as a discussion partner and a resource.
• They regarded the VP model as a facilitator of their learning process.

Conclusions: The VP model for primary care seems a valuable supplement in the students' self-directed learning and assessment.

Take-home Messages: The new VP model may facilitate the students' learning in primary care.

OTT-PA5-06
Examining systematic training and assessment strategies for robotic surgery skill acquisition

Alex Jiang (*Western University, Schulich School of Medicine and Dentistry, London, Canada)*, Karen Siroen (*Western University, Canadian Surgical Technologies and Advanced Robotics, London*), Ana Luisa Trejos (*Western University, Canadian Surgical Technologies and Advanced Robotics, London*), Rajni Patel (Western University*, Department of Electrical and Computer Engineering, London*), Christopher Schlachta (*Western University, Department of Surgery and Oncology, London*), Sayra Cristancho (*Western University, Centre for Education Research and Innovation, London*)

Background: Simulation-based Mastery Learning seeks to promote long-term skill retention. In mastery

learning, educational outcomes are uniform, while educational time varies among trainees. In this poster, we will present a study design that will explore the effectiveness of two conditions of practice in the transfer of robotic surgical skills from the simulation platform (dV-Trainer TM) to the actual robotic system (da Vinci®). A pilot study has been implemented to collect preliminary data for this project.

Summary of Work: Participants are randomized into two groups ('blocked' and 'random' practice schedules) and follow a 4-phase procedure: warm up, pre-test/post-test, retention and transfer. Five tasks are used for assessment using the MIMIC dv-Trainer simulator and the da Vinci system. Performance is recorded using the MIMIC parameters and a customized OSATS scale.

Summary of Results: Pilot data from two subjects showed that reaching proficiency following random practice may take twice the number of attempts than blocked practice. However, once proficiency is achieved, the random-trained participant was able to complete the tasks faster. The OSATS scores were consistently higher for the subject who underwent blocked training, while the MIMIC scores were variable depending on the task.

Conclusions: These preliminary results are in contrast to other studies comparing blocked vs. random practice, in which random practice translates into better performance. However, the full study should be conducted in order to confirm these findings. Moreover, the suitability of expert-based vs. computer-based assessments should be analyzed with respect to the nature and difficulty of the tasks.

Take-home Messages:

OTT-PA5-07
Learning of simulated cardiac arrests in hospital environments to support performance of the multidisciplinary team

Kristina Ziegert (*Halmstad University, School of Social and Health Sciences, Halmstad, Sweden)*, Anne-Marie Schjetlein (*Halmstad County Hospital, Research & Development Unit of Health Care, Halland, Halmstad*)

Background: Survival after cardiac arrest in hospital environments has in various studies reported to be between 15-39%. In Swedish hospitals 30% of patients who suffer cardiac arrest survive to discharge. There are significant opportunities for improvement in cardiac arrests in hospitals, especially those on general wards. There are few studies of the multidisciplinary team performance regarding both technical and non-technical skills in simulated cardiac arrest on general wards.

Summary of Work: The focus of this study was to observe the multidisciplinary team performance in simulated cardiac arrest on general wards to gain knowledge about existing methods and thus be able to identify areas for improvement. The study design was an exploratory observational study. There were 33 participants in a total of three wards. The simulated cardiac arrest was videotaped, and a patient simulator was used.

Summary of Results: The results suggest that guidelines not always were adhered to cooperation. In the order for the guidelines to work in the multidisciplinary team, background knowledge is required. However, the team leader did not coordinated to the multidisciplinary team work, and did not use effective communication.

Conclusions: Finally, the result suggested suboptimal performance regarding leadership and communication, which in turn affected resources and working environment.

The opportunity with simulation is to train together in a situation without any risk for the patients.

Take-home Messages: Simulated cardiac arrests will give the opportunity to train the multidisciplinary team technical and non-technical skills.

OTT-PA5-08
A randomized controlled trial: 'In situ simulation' versus 'off site simulation' in obstetric emergencies and their effect on knowledge, safety attitudes, team performance, stress, and motivation

Jette Led Sorensen (*Rigshospitalet, Copenhagen University Hospital, Juliane Marie Centre for Children, Women and Reproduction, Copenhagen, Denmark)*, Cees Van der Vleuten (*Maastricht University, Netherlands, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht*), Doris Østergaard (*Herlev Hospital, Capital Region of Denmark and Copenhagen University, Danish Institute for Medical Simulation, Copenhagen*), Vicki Vicki LeBlanc (University of Toronto*, The Wilson Centre, Toronto*), Kim Ekelund (*Rigshospitalet, Copenhagen University Hospital, Department of Anaesthesia, Copenhagen*), Bent Ottesen (*Rigshospitalet, Copenhagen University Hospital, Juliane Marie Centre for Children, Women and Reproduction, Copenhagen*)

Background: Obstetric emergencies threaten the safety of pregnant women. As emergencies are rare, they are difficult to learn, therefore, simulation-based medical education (SBME) seems relevant. However, many questions on how SBME can be optimized remain unanswered. One unresolved issue is how 'in situ simulation' (ISS) versus 'off site simulation' (OSS) impact learning. ISS means simulation-based training in the actual patient care unit. OSS means training in facilities away from the actual patient care unit.

Summary of Work: The objective was to study the effect of ISS versus OSS on individual learning outcome, safety attitude, motivation, stress, and team performance. The trial was a randomized trial including 100 participants (Midwifes, anesthesia- and obstetric doctors, anesthesia-, auxiliary- and scrub-nurses) employed at Rigshospitalet, Copenhagen.

The same multi-professional simulation was conducted in the two simulation settings, and the experimental group was exposed to training in the ISS setting, and the control group in the OSS setting.

The outcome was the individual score on a knowledge test, on a safety attitudes questionnaire, a stress

inventory, salivary cortisol levels, an intrinsic motivation inventory, results from a questionnaire evaluating perceptions of the simulation and a team-based score on video-assessed team performance

Summary of Results: The randomized trial was successfully fulfilled April to June 2013 with 97 participating. Response rates to tests and questionnaires were 90-100%. Data processing and statistical analysis are ongoing.

Conclusions: The perspective is to provide new knowledge on contextual effects and fidelity of different simulation settings. Results will be presented at the conference.

ClincialTrials.gov NCT01792674

OTT-PA5-09
A study to investigate the role of SimMan in teaching clinical skills to preclinical medical students

Meenakshi Swamy (*Durham University, School of Medicine, Pharmacy and Health, Stockton-on-Tees, UK)*, Marina Sawdon (*Durham University, School of Medicine, Pharmacy and Health, Stockton-on-Tees*), Andrew Chaytor (*Durham University, School of Medicine, Pharmacy and Health, Stockton-on-Tees*), David Cox (Durham University*, School of Medicine, Pharmacy and Health, Stockton-on-Tees*), Judith Barbaro-Brown (*Durham University, School of Medicine, Pharmacy and Health, Stockton-on-Tees*), John McLachlan (*Durham University, School of Medicine, Pharmacy and Health, Stockton-on-Tees*)

Background: Simulation encourages preclinical medical students to develop clinical skills in a safe environment, but there is little evidence on its effectiveness.

Summary of Work: 79 first year preclinical medical students participated in the crossover study. They were divided into two groups, one performed chest examination on their peer and the other used SimMan. A pretest, midtest and post-test was conducted using a questionnaire (ten questions on knowledge, and confidence levels rated using a Likert scale). They were assessed formatively similar to their OSCE. Students completed a feedback questionnaire. Data was analyzed using one-way ANOVA and independent t-test.

Summary of Results: When the two groups were compared, there was no significant difference seen between the groups in their pretest and post-test scores on knowledge questions whereas midtest scores increased significantly (P< 0.001) with the group using SimMan initially scoring higher. A significant increase in the test scores was seen in both groups after using SimMan (P<0.001). Students' confidence increased significantly in differentiating between normal and abnormal signs (P<0.001).

Conclusions: The study demonstrated a significant improvement in the students' knowledge after simulation with an increase in their confidence. Students' feedback was very positive.

Take-home Messages: SimMan can be a useful adjunct to teach clinical skills to preclinical medical students.

OTT-PA5-10
Hurdles in Simulation Training

Thomas Kersey (*Frimley Park Hospital, Ophthalmology, Esher, UK)*, Geeta Menon (*Frimley Park Hospital, Ophthalmology)*

Background: Surgical simulation is mandatory in training. Encouraging the use of this technology and integrating it into the training curriculum, has had mixed success. We present a structured approach to improving standards utilising 'EyeSi' cataract surgery simulator and a hurdle based score system.

Summary of Work: 'EyeSi' is a high end virtual reality simulator for intraocular surgery. Validation of 'EyeSi' cataract surgery simulator was achieved by measuring the scores in novice, intermediate and expert surgeons. Work published on group marks has been used to create benchmark hurdles.

We planned a phased introduction of mandatory simulator training over three years. Log-books of timings and performance scores will be recorded.

Year 1 - no hurdle set - records achieved will help form the hurdles for the first year students in future years.

Year 2 - soft hurdle based on the previous years experience and the validated scores for each level group (novice, intermediate and expert).

Year 3 – fixed-hurdle benchmarks, made available to the students early in the year to allow time to practice to achieve the scores in their own time.

Summary of Results:
➢ Take up rates and evidence of integration into training
➢ Indicative outcomes of year one
➢ Faculty development requirements
➢ Business planning for the exercise

Conclusions: Encouraging the use of surgical simulation is essential to its successful integration in a program. A hurdle based approach encourages practice on the simulator but tasks must be a validated to be productive.

Take-home Messages: A hurdle based approach may prove a successful way to encourage simulator usage in trainees.

OTT-PA5-11
Validation of a new bronchoscopy assessment tool in a clinical setting

Nha Voduc (*University of Ottawa, Medicine-Division of Respirology, Ottaw, Canada)*, Nancy Dudek (*University of Ottawa, Medicine, Ottawa*), Timothy Wood (*AIME-University of Ottawa, Medicine, Ottawa*)

Background: Flexible bronchoscopy is an important procedure performed by multiple medical specialties, in a variety of settings. The purpose of this study is to create a bronchoscopy assessment tool (BAT) that could be applied in a variety of clinical settings, and to assess the validity of the ratings.

Summary of Work: Focus groups of expert physicians from two academic centers, representing specialties

which routinely perform bronchoscopy (respirology, critical care, thoracic surgery and anesthesia) identified themes and items that are relevant to the assessment of competency in bronchoscopy. From these focus groups, a bronchoscopy assessment tool was developed. This study represents the first attempt at validation of the tool.

Summary of Results: The BAT consists of 19 items, organized into the following areas: pre-procedure planning, sedation and monitoring, technical skill, diagnostic skill, post-procedure planning. The ability of the trainee to complete each item was assessed with a five-point rating scale (1=supervisor needed to take over this part of the procedure, 5 = competent for independent practice). For this study, 23 residents were enrolled. After each bronchoscopy, the BAT was completed by the supervising physician. An average of 6 BATs completed for each resident. With 6 BATs, the reliability was 0.77. The reliability was increased to 0.81 with 8 BATs. The tool was able to distinguish between junior and senior year trainees (M=3.80 vs 4.81). Scores on most of the items were highly correlated (r>.80).

Conclusions: The analysis provides evidence for the validity of the BAT for the assessment of competence to perform bronchoscopy in a variety of clinical settings. Future work will refine the tool to address redundant items.

Take-home Messages: The BAT demonstrates promise as a reliable tool to assess trainee competence for bronchoscopy.

OTT-PA5-12
eField Notes: A collaborative online tool to document teaching, learning, and performance

Gary Viner (*University of Ottawa, Family Medicine, Ottawa, Canada)*, Eric Wooltorton (*University of Ottawa, Family Medicine, Ottawa*), Alison Eyre (*University of Ottawa, Family Medicine, Ottawa*), Doug Archibald (University of Ottawa*, Family Medicine, Ottawa*), Emma Stodel (*University of Ottawa, Family Medicine, Ottawa*)

Background: Family medicine postgraduate training in Canada is shifting towards a competency-based approach, requiring new methods of teaching and evaluation. The Department of Family Medicine at the University of Ottawa (DFM) has responded by developing a web-based field note (eFN) that allows residents and their supervisors to collaborate on a digital record of teaching, learning, and performance. The eFN supports tagging of data (e.g., CanMEDS role, curriculum domain) that will permit a range of resident-specific reporting to help document the pathway to competency, as well as more global reporting to inform curriculum and faculty development.

Summary of Work: Using the structure of our previously described and well-accepted paper field notes, we defined a clear set of requirements for the digital version. We considered the complexities of work flow and business rules necessary to ensure security and usability. Finally, we used a collaborative approach to develop the electronic tool and test it with residents and faculty.

Summary of Results: 1) Feedback was solicited on early prototypes of the faculty and resident portals in August 2013 with a small group and the tool was refined. 2) A beta-version was demonstrated at a department-wide curriculum retreat in October and 15 preceptors and their residents were solicited for a 3-month pilot. 3) Implementation throughout DFM occurred in February 2014.

Conclusions: This poster describes the eFN and its development, including lessons learned in the development process and the nature of the planned reports that will emanate from aggregating large numbers of eFNs.

Take-home Messages: We have designed an effective tool that will collect and strategically aggregate the volume of documented observations required for competency-based education.

OTT-PA5-13
Reliability of an assessment checklist for medical education websites

Diane Poirier (*Médecins francophones du Canada, Montreal, Quebec, Canada)*, Marie-Francoise Mégie (*Médecins francophones du Canada, Montreal, Quebec*), Cristine Lamoureux (*Médecins francophones du Canada, Montreal, Quebec*), Johanne Blais *(Médecins francophones du Canada, Montreal, Quebec*), Presenter: Céline Monette *(Médecins francophones du Canada)*

Background: Developing and then assessing the accuracy of an assessment checklist of medical education websites in view of providing to the international French-speaking medical community a useful, reliable and innovative tool.

Summary of Work: An assessment checklist of 54 indicators grouped into 5 categories was developed. The accuracy of the checklist's overall score was compared using the results of 7 observers who evaluated the same 16 websites twice. The second evaluation of the sites was carried out one year later. The data collection plan helped measure the following accuracy elements: intra-observer variability, inter-observer variability and site-observer interaction. The accuracy of the checklist score was measured by the intra-class correlation coefficient.

Summary of Results: The resulting intra-class correlation coefficient was 0.684, consistent with a good degree of accuracy. Intra-observer variability (repeatability) was the largest source of variation attributable to the measuring method with 21% of total variation. Another performance measure was also used: the discrimination ratio that considers the number of different categories that can be recognized by a measurement system. The resulting ratio was 2.3, indicating that the checklist helps divide the sites identified in two categories: weak and average.

Conclusions: Due to its high degree of accuracy, the assessment checklist of medical education websites is an interesting tool to assess the quality of these sites. The analysis clearly demonstrates that the best avenue in improving the reliability of the checklist would be to improve repeatability (intra-observer variability).

Take-home Messages:

OTT-PA5-14
Development of a Peer-Teaching Review Instrument for use with Self-Guided e-Learning Modules

Christopher R Pierson (*Nationwide Children's Hospital, The Ohio State University, Pathology, Columbus, OH, USA)*, Eric Fox (*The Ohio State University, Evaluation, Curriculum Research & Development, Columbus, OH*), Doug Danforth (*The Ohio State University, Evaluation, Curriculum Research & Development, Columbus, OH*), Susan Mauck (The Ohio State University*, Evaluation, Curriculum Research & Development, Columbus, OH*), David Way (*The Ohio State University, Evaluation, Curriculum Research & Development, Columbus, OH*)

Background: We launched a new UGME curriculum in which we made abundant use of self-guided web-based learning modules (e-learning). Modules were intended to either replace or supplement traditional lectures. Because teaching in this manner is novel to our faculty, we found a need for an evaluation instrument for use with these educational products. The objective of this project is to evaluate the psychometric properties (validity and reliability) of a new peer-teaching review instrument for use with self-guided e-learning modules.

Summary of Work: We developed the peer-teaching review instrument based on literature of best practices for designing e-learning modules. We drew a stratified-random sample of 48 from the total of 229 (20%). A panel of 5 judges will use the new instrument to evaluate the sampled modules. To assess concurrent validity, scores generated on the new instrument will be compared to ratings made by students who completed the modules. We will also evaluate the new instrument for content validity and inter-judge reliability.

Summary of Results: Review of the instrument is currently in progress.

Conclusions: We expect that ratings from the peer-teacher review instrument will be consistent with student ratings of self-guided e-learning modules, and that the peer-review instrument will complement the student ratings with more practical formative feedback to faculty authors.

Take-home Messages: A peer-review instrument and student ratings provide complementary contributions to the process of a continuous quality improvement cycle designed to incorporate e-learning into a modern UGME curriculum.

OTT-PA5-15
Improving the usability of e-learning resources: a randomized trial

M Razeen Davids (*Stellenbosch University and Tygerberg Hospital, Division of Nephrology and Department of Medicine, Cape Town, South Africa)*, Usuf ME Chikte (*Stellenbosch University, Department of Interdisciplinary Health Sciences, Cape Town*), Mitchell L Halperin (*University of Toronto, Li Ka Shing Knowledge Institute of St Michael's Hospital and Division of Nephrology, Toronto*)

Background: Optimizing the usability of e-learning materials is necessary to reduce extraneous cognitive load and maximize their potential educational impact. This is often neglected, especially when time and other resources are limited. We investigated whether a usability evaluation of our e-learning resource, followed by fixing of usability problems identified, would translate into measurable improvements in usability parameters and into improved learning by medical residents.

Summary of Work: Two iterations of our e-learning resource (V1 and V2) were compared in a randomized trial. V1 was the original version and V2 the revised version with usability problems addressed. Residents in medicine and anesthesiology were randomly assigned to one of the versions. Usability was evaluated by user satisfaction questionnaire and by recording and analyzing the interaction of participants with the application. Learning was assessed by questions designed to test retention and transfer of knowledge.

Summary of Results: Participants reported high levels of satisfaction with both versions with good ratings on the System Usability Scale and adjective rating scale. In contrast, there was a large difference in the occurrence of serious usability problems between the two versions, in particular in the interactive HandsOn case where there was a median of 5 (range 0-50) serious problem instances recorded per participant for V1 and 0 instances (range 0-1) for V2 (P<.001). There was no difference in learning measures between the two versions.

Conclusions: Usability evaluation followed by a redesign of our e-learning resource resulted in measurable improvements in usability. While this was not accompanied by increases in learning in our study population of high-knowledge learners it may translate into improved motivation and willingness to engage with the learning material.

Take-home Messages: An iterative design-test-redesign approach can improve the usability of e-learning resources; this increases the chances of achieving desired learning outcomes, especially when the material is difficult, with a high intrinsic cognitive load.

Monday April 28
1:30PM – 3:00PM

SYMPOSIA

OTT-SE-1
Bridging the Gap: How Medical Education and Measurement Science can Better Collaborate to Meet Growing and Broadening Assessment Needs

Presenters: Andre De Champlain (*Medical Council of Canada, Ottawa)*, Kevin Eva (*University of British Columbia*), Brownell Anderson (*National Board of Medical Examiners, Philadelphia*), Lesley Southgate (*St George's Hospital Medical School, London*), Discussant: Ian Bowmer (*Medical Council of Canada, Ottawa*)

Summary: A widened perspective on assessment has been advocated to better meet the systemic nature of medical education. The aim of this symposium is to outline how measurement scientists and medical educators can better collaborate to meet this desire. The symposium will start with a focus on past successful collaborative models between measurement science and medical education that might serve as a platform for moving forward. Issues that will then be addressed are: (1) Overcoming the Unintended Consequences of Competency-based Assessment – A Medical School Perspective; (2) Integrating Assessment Data and Educational Experiences Across the Continuum; (3) Workplace Assessment: Has Measurement Killed Judgement?

OTT-SE-2
Medical Schools Council (MSC) assessment initiatives

Presenters: Val Wass (*Keele University, Keele*), Katie Petty-Saphon (*Medical Schools Council, London*), Simon Maxwell (*University of Edinburgh, Edinburgh*), Siobhan Fitzpatrick, Fiona Patterson (Work Psychology Group, Cambridge)

Summary: All 33 UK Medical Schools have formed an Assessment Alliance, working together to share good practice and resources and address issues of clinical competency standards. Individual medical school examinations are maintained, monitored by their regulator the General Medical Council. This symposium opens for discussion the challenges of shared test formats and compatibility of standards across this National initiative. The symposium includes: (1) Scene setting for those not familiar with UK processes; Role of the MSC/MSC-AA and its relation to the regulator; Academic freedom of medical schools vs. external accountability to stakeholders; Pros and cons of a national examination/need for comparability between schools. (2) The development and utility of a shared question bank: History and buy- in by schools;

Development of good practice in assessment; IT issues; Practical issues
Comparison of passing standards using Rasch modeling: Conceptual issues; Application and initial results. (3) The development and utility of a national prescribing skills assessment: The problem of prescribing errors; Prescribing in relation to pharmacology and therapeutics; Experience with pilot online national assessments. (4) The place of Situational Judgements Tests for entry into residency (UK Foundation Programme): Lessons from industry and selection into general practice; Identification of the key roles of F1 doctors; Applicability to F1 selection; Experience with the SJT in selection.

WORKSHOPS

OTT-WE-1
The Cohen method of standard setting - WHY and HOW

Presenter(s): Scarpa Schoeman(*University of the Free State, Internal Medicine - Medical Education division (G73), Bloemfontein, South Africa*), Janke Cohen-Schotanus(*University Medical Centre Groningen, Centre for Research and Innovation in Medical Education, Groningen, Netherlands*)

Background: The Cohen method is novel and relatively new standard setting method which uses the very top performing students in the assessment to provide a benchmark of what score was possible to achieve in the particular assessment (difficulty marker). An absolute cut-score is then derived from this benchmark score to set the pass standard for the assessment. The method hinges on the stability of the very top performing students between cohorts within a stable selection process, curriculum and assessment regime.
Intended Outcomes: In this workshop, the method's conception, methodology and application will be explained. The latest research on this method's use and application will the presented on undergraduate and postgraduate data in medicine.
Participants will then be encouraged to apply the method in the workshop to some of their own students' performance data to see how it works.
A discussion with questions and answers will follow.
Structure: 1) Opening talk on the background, methodology and application of the Cohen method. 2) Presentation of its application and research findings on large data sets. 3) Participants apply method to their own retrospective data sets to experience the method's results. 4) Discussion with Q&A
Who Should Attend: This workshop is intended for all educators who need a defensible and sustainable method which produces an absolute cut-score for a test.
****PLEASE bring a charged laptop and at least 1 data set in MS Excel of performance data of a written test.
Level of Workshop: Intermediate

OTT-WE-2
Moving from point-in-time assessment of learning to programmatic assessment for learning – implementing an assessment system for competency-based post-graduate medical education

Presenter(s): Farhan Bhanji (*McGill University, Montreal, QC, Canada*), Linda Snell(*McGill University, Montreal, QC, Canada*), Jonathan Sherbino (*McMaster University, Ancaster, ON, Canada),* Jason Frank (*The Royal College of Physicians and Surgeons of Canada, Ottawa, Canada*)

Background: Competency-based medical education (CBME) requires a robust and systematic assessment program that emphasizes performance observation in the clinical environment. Assessments must occur with increased frequency, be criterion-referenced, and assess the developmental progression of competence providing meaningful feedback to post-graduate trainees along the way. The traditional focus on the psychometric properties of tests become somewhat less important as we place greater emphasis on the assessment's educational impact. The assessment program needs to combine quantitative with qualitative data, integrating narrative and allowing for the use of 'subjectivity' by expert physicians to inform high-stakes decision-making. Reductionist approaches to individual competencies may sometimes be appropriate but assessment should not loose sight of the holistic, global picture of the learner. Faculty development will be of paramount importance and overworked clinician-teachers will require resources and support. In this less standardized assessment environment, the 'collective' of assessment groups to judge performance and progression will be critical.
Intended Outcomes: "Upon completion of this workshop, the participants will be able to:
1. Describe an approach to develop an assessment program for CBME in post-graduate medical education.
2. Outline when simulation might be an appropriate assessment strategy in CBME
3. Recognize challenges in faculty development and how inadequate faculty support may undermine the CBME process."
Structure: "Structure:
• Interactive mini-plenary
• Interactive case studies designed to outline challenges and potential solutions in a CBME program,
• Small group exercises
• Wrap-up"
Level of Workshop: All

OTT-WE-3
An annual leadership assessment – raising awareness of the clinical leadership agenda

Presenter(s): Lindsay Hadley (*Health Education Kent Surrey and Sussex, School of Leadership, London, United Kingdom*), David Black (*Health Education Kent Surrey and Sussex, Postgraduate Dean HEKSS, London, United Kingdom*), Patrick Marshall (*Health Education Kent Surrey and Sussex, Education, London, United Kingdom*)

Background: Clinical leadership is considered essential for maintaining and improving patient care. In the UK, leadership is incorporated into the curriculum for all doctors in training and is part of the GMC's duties of a doctor. As an organisation with responsibility for ensuring the delivery of the curriculum, Health Education Kent Surrey and Sussex has used a multifaceted approach to teaching leadership with a progression of different initiatives supported by an infrastructure based in the local education providers. Trainees and their supervisors have had a target of one formative assessment of their leadership skills each year and trainees need to submit written evidence for their annual review of competency progression.To support the target, workshops, training days and on-line resources have been developed and a formative assessment tool LEADER is used widely. The submitted assessments provide evidence of an increase in understanding of what might afford a good leadership learning opportunity and how every day clinical activities can be used to develop good clinical leadership practice.
Intended Outcomes: Participants will develop an understanding of how they might implement a leadership learning strategy relevant for use in their organisation.
Participants will be able to identify potential leadership learning opportunities within the daily work of postgraduate doctors.
Participants will understand how to use the formative LEADER assessment tool.
Structure: Interactive workshop. Following a didactic session, small groups will explore different leadership learning opportunities from examples provided and their own experience. They will look at how they might implement a leadership learning strategy for their organisation. They will have the opportunity to use the LEADER tool and feed back on its usefulness
Who Should Attend: Educators and trainees interested in leadership teaching and learning
Level of Workshop: Intermediate

OTT-WE-4
Assessing Students in Interprofessional Education (IPE) Activities: Developing Assessment Skills in Tutors

Presenter(s): Linlea Armstrong (*UBC, Medical Genetics, Vancouver, Canada*), Lesley Bainbridge (*UBC, College of Health Disciplines)*

Background: We have developed simulation activities which bring students from various health professions together to learn about, with, and from each other. Typically, both common and profession-specific learning objectives are addressed. Inherent in this type of learning is the need to assess how well students perform

the skills required for effective interprofessional collaboration. Most tutors come to teaching without a fluency in core IPE competencies, nor a clear sense of the expected levels of functioning for senior students from professions other than their own. In doing assessments across professions within multidisciplinary settings, attention to sensitivity and explicit demonstrations of respect for all is especially important. We have identified orientation and training strategies which prepare tutors to provide effective assessment and feedback in this type of interprofessional sessions.
Intended Outcomes: 1. To improve the ability of tutors to assess students in interprofessional learning sessions. 2. To develop tutor skills in facilitating interprofessional observation and feedback.
Structure: 1. A think-pair-share session to enable participants to reflect upon and learn about assessment strategies in interprofessional education.
2. A short didactic presentation on the competencies for interprofessional collaboration and the status quo of assessment in interprofessional education. We will discuss how student evaluations of our sessions over the years shaped our tutor training regarding assessment in such sessions.
3. Small group work designed to explore the challenges associated with assessment of collaboration, and develop strategies for interprofessional assessment that can be tested by participants in their home setting that.
Who Should Attend: Teachers; faculty developers; course directors; assessment and evaluation specialist
Level of Workshop: Intermediate

OTT-WE-5
Reasonable adjustments in assessment – when and how should they be used?

Presenter(s): Drew Gilliland (*Queen's University Belfast, Centre for Medical Education, Belfast, United Kingdom*), Gerry Gormley (*Queen's University Belfast, Centre for Medical Education, Belfast, United Kingdom*), Mairead Boohan (*Queen's University Belfast, Centre for Medical Education, Belfast, United Kingdom*)

Background: Medical schools around the world are required to comply with disability legislation when implementing assessments. In the UK the GMC have provided examples of reasonable adjustments that may be employed to accommodate students taking assessments. These guidelines can relate to the documentation provided, adjustments through allowances, for example, extra time to complete the assessment. Other adjustments include the provision of specialist equipment e.g. computers or special lighting. Alternative assessment venues may also be used.
Intended Outcomes: Gather examples from an international audience on current practice.
Explore the adjustments that should be made for different disabilities and levels of disability.
Consider if adjustments can impact on ability to practice in future career.

Structure: Discussion of examples from Queen's University Belfast of adjustments made during the past five years.
Group work using scenarios to highlight differences in approach between medical schools.
Who Should Attend: Those involved in undergraduate assessment.
Level of Workshop: Intermediate

OTT-WE-6
Standardized Patient Cases: All Scenarios Do Not Equally Assess Communication Skills

Presenter(s): Colette Scott (*National Board of Medical Examiners, Clinical Skills Evaluation Collaboration, Philadelphia, United States*), Ann King (*National Board of Medical Examiners, Measurement Consulting Services, Philadelphia, United States*), Gail Furman (*National Board of Medical Examiners, Clinical Skills Evaluation Collaboration, Philadelphia, United States),* Ruth Hoppe (*Michigan State University, College of Human Medicine, East Lansing, United States*)

Background: Case development is an essential component of standardized patient-based assessment. Cases are the foundation for scenario development, scoring and consequently, examinee behavior. This workshop focuses on developing or transforming case material to allow for a better assessment of communication skills.
Intended Outcomes: Following participation in this workshop, learners will be able to:
1. Understand the impact of scenario development on assessment instruments.
2. Demonstrate the steps required for developing communication cases for an SP-based examination;
3.Critique developed cases.
Structure: 1. Introduction to case development for communication skills
2. Small group work to develop cases
3. Large group critique of the cases
4. Question and answer session
Who Should Attend: All involved in the assessment of communication skills using standardized patients.
Level of Workshop: Intermediate

ORALS

OTT-OE-1: PORTFOLIO ASSESSMENT

OTT-OE-1-1
The utility of an integrated portfolio to assess a primary and community care clinical placement

Chris Roberts (*University of Sydney, Academic GP Unit, Sydney, Australia)*, Shadbolt Narelle (*University of Sydney, Academic GP Unit, Sydney*), Tyler Clark

(*University of Sydney, Office of Medical Education, Sydney*)

Background: We explored the influencing factors on the precision of scoring within a portfolio based assessment of students' achievement of primary and community care learning outcomes within an eight week integrated clinical placement.

Summary of Work: A generalisability analysis on portfolio task scores estimated variance components. We calculated a reliability co-efficient and the effect of varying the number of assessment tasks and raters, and a standard error provided a confidence interval around the pass/fail standard.

Summary of Results: Aggregated portfolio marks for the 257 students, across 372 raters, on six tasks, were normally distributed with a mean of 70.56 and a standard deviation of 8.68. For a single student with six assessment tasks, 11% of the variance in scores was due to true differences in the capability of the student. The most significant interaction was context specificity (49%), the tendency for one student to engage with one task and not engage with another task. The next significant source of measurement error rater subjectivity was 26%. Given the purpose of the portfolio was to determine acheivement of a standard, a modified standard error of 3.7%, gave a 95% CI of +/- 7.2% around a pass/fail score of 57%

Conclusions: Portfolio assessment has modest precision in assessing students' achievement of a standard in required learning outcomes. Whilst a traditional reliability co-efficient was unhelpful, a modified standard error was acceptable. There were identifiable areas where measurement error could be reduced to provide more certainty around decision making. Reducing the noise in the measurement would require engaging with student body on the value of the tasks, more focused academic and supervisor training, and revisiting the rubric of the assessment in the light of rater feedback.

Take-home Messages: We encourage an international consensus in how to demonstrate technical adequacy in reporting competency-based portfolio assessments.

OTT-OE-1-2

Portfolio as summative assessment in an undergraduate orthopedic course

Yosef Tyson (*Institution for Surgical Science, Uppsala University, Uppsala, Sweden)*, Martin Wohlin (*Institution for Medical Science, Uppsala University, Uppsala*)

Background: Most research on portfolios is retrospective and focus on the development of professionalism. Few, if any studies regarding portfolio assessment are randomized and quantitative. This project aims at creating the first randomized controlled study of portfolio in an orthopedic undergraduate setting.

At Uppsala University medical school the orthopedic course is given as a 4 week course in the 4th year. The assessments used today do not measure all orthopedic outcomes and room for improvement is needed. This project will introduce and evaluate a portfolio to fill some of these gaps and research the portfolios effects on knowledge, reasoning, patient relations, skills, professionalism among other areas.

Summary of Work: A pilot study has been conducted to evaluate the format and content (e.g. work place based assessment tools, reflective instructions and criteria for assessment) of the portfolio to be used in the final study.

Summary of Results: Most work place based assessment tools, reflective instructions and criteria for assessment had to be revised and adjusted to suit the orthopedic setting.

Conclusions: It is crucial to develop a shared ownership regarding criteria for assessment in a project like this and to develop criteria and content in a multiprofessional collaboration with all stakeholders including students.

Take-home Messages: This might be the first randomized controlled study on portfolio for summativ assessment in an undergraduate orthopedic setting. For this reason a strong emphasis must be on developing tools, instructions and criteria.

OTT-OE-1-3
Portfolio as reflection instrument for undergraduate students in community medicine

Erike Anggraini Suwarsono (*Faculty of Medicine and Health Sciences Islamic State University Syarif Hidayatullah Jakarta, Medical Education, Jakarta, INdonesia)*, Fika Ekayanti (*Faculty of Medicine and Health Sciences Islamic State University Syarif Hidayatullah Jakarta, Community Medicine, Jakarta*), Widyandana (*Faculty of Medicine, Gadjah Mada University, Medical Education, Yogyakarta*)

Background: Self reflection is a skill that is required to be achieved by medical students. It is needed to establish the life long learner. A proficient competent doctor should conduct self reflection to enhance their professional performance.

Summary of Work: This study was done in the clinical phase of community medicine module for 6th year medical students. The total was 55 students. They were asked to make their portfolios during their session in the primary health services and interaction with the community. The portfolio is also used as one of the assessment tools for the module. At the end of the module, they had to write a narrative reflection for their portfolio task. The students' self reflection was analyzed qualitatively by two inter-raters to distinguish the process development of their professional performance in using the portfolio.

Summary of Results: The first time they completed their self reflection, most of the students said that it was quite difficult to routinely prepare the portfolio in the community setting. The reasons were hard to memorize the correct drugs and condition of the health services. The knowledge gap between theory and reality of using portfolio were recognized and understood by all of the students. It was getting better over time. By reflecting on their experiences during every single task, they can

analyze their performance and become aware of their problems at an earlier stage. By reflection, the development of students' professional performance can be monitored whether increasing or not.

Conclusions: Portfolio as an instrument for self reflection was a suitable choice for learning in the clinical setting.

Take-home Messages: The skill of self reflection is important for doctors to achieve in order to develop and manage their characters on becoming a professionally competent doctor.

OTT-OE-1-4
A tablet PC-web based e-portfolio system for clinical rotation to support reflective learning and alliance among teachers

Kazunobu Ishikawa (*Fukushima Medical University, Center for Medical Education and Career Development, Fukushima, Japan)*, Gen Kobayashi (*Fukushima Medical University, Center for Medical Education and Career Development, Fukushima*), Akiko Sugawara (*Fukushima Medical University, Center for Medical Education and Career Development, Fukushima*), Yoko Moroi (Fukushima Medical University*, Center for Medical Education and Career Development, Fukushima*), Tatsuo Suzutani (*Fukushima Medical University, Center for Medical Education and Career Development, Fukushima*), Testuhito Fukushima (*Fukushima Medical University, Center for Medical Education and Career Development, Fukushima*)

Background: Understanding of the entire goal of clerkship, sharing students' performance, assessment of students' competency and organization of feedback process from learners and teachers are common problems in clinical rotation of medical students. To improve these, we developed a tablet PC based-learning portfolio system.

Summary of Work: During clinical rotation in 28 clinical units, 81 medical students self-assessed their performance in 5 levels. Teachers in each department assessed students' performance using identical specific goals for feedback. Teachers shared students' performance on a tablet PC based-learning portfolio system. Students' also self-assessed their competencies during and after rotation using 68 checklists, which remind entire goal. This system also provided self-produced learning contents (movies & pictures) and mini-test.

Summary of Results: 85% students sufficiently fed their assessment during rotation. Although 36% students perfectly self-assessed, 15% students entered their assessment less than half. Almost all teachers have completely fed feedback. Post-questionnaire and interviews to students and teachers revealed that (1) usefulness for prior-confirmation and reflection of learning points, (2) convenience of tablet PC-web based system for browsing contents and (3) benefits for supporting students by visualization of their learning process in other clinical units. However, there remain the problems about grade calculation and intervention for passive students.

Conclusions: A tablet PC-web based e-portfolio system for clinical rotation has successfully placed in service.

Take-home Messages: Development of a tablet PC-web based system for clinical rotation have possibilities to promote reflective learning of medical students and alliance among teachers in multiple units.

OTT-OE-2: ASSESSING THE PRACTISING DOCTOR 1

OTT-OE-2-1
Implementation of revalidation for trainees in the UK

David Black (*Health Education Kent, Surrey and Sussex, Dean Director, London, UK)*

Background: The revalidation of a doctor's licence to practise finally started in December 2012 for all UK doctors, including the 54,000 doctors in postgraduate training in the UK. Revalidation is an affirmation every five years that a doctor remains up-to-date and fit to practise. A recommendation is made by every doctor's Responsible Officer (RO).

Summary of Work: The RO for trainees is the Postgraduate Dean, who on average is responsible for around 3,000 trainees. Systems needed to be developed to ensure that at the point of revalidation the RO has not only access to all information about a doctor's training record, but also data on employment, health, conduct, together with outcomes of complaints and serious clinical incidents. All factors that might influence a revalidation decision. Systems also needed to be put in place to ensure accurate information transfer for movement between employers.

Summary of Results: Operational UK implementation was achieved by summer 2013. There have been particular challenges in ensuring accurate information to be able to make professional judgements as to whether complaints and incidents have been successfully resolved

Conclusions: The impact of the revalidation process on UK trainees as opposed to the current annual assessment of educational progress has yet to be determined.

Take-home Messages: Decisions on fitness to practice require systems to ensure a global view of a trainee's professional practice, not just an assessment of educational progress.

OTT-OE-2-2
Assessment of senior physicians (70+) in Quebec

Marc Billard (*College des medecins du Quebec, Practice enhancement division, Montréal, Canada)*, Johanne Thiffault (*College des medecins du Quebec, Practice enhancement division, Montreal*), Francois Goulet (*College des medecins du Quebec, Practice enhancement division, Montreal*), Monique Robert (College des medecins du Quebec*, Practice enhancement division, Montreal*)

Background: In 1996, a study conducted by the College des medecins du Quebec showed that 97% of family physicians had a safe practice. In 2013, 6% of active physicians in Quebec were 70+ years old. The literature and our experience indicate that performance declines significantly with age.

Summary of Work: We created a questionnaire, based on known risk factors, that we sent to all 1200 active physicians aged 70 and over. We wanted to know the type and scope of practice of those physicians. We also wanted to raise their awareness of the possible decline in the quality of their work.

Summary of Results: Preliminary results show that 13% of them were judged to have a practice at risk and were selected for a formal peer review. Of those visited, 39% had to do a retraining session. On the positive side, 87% had a practice that was judged safe based on the questionnaire and the questionnaire was well accepted by a vast majority of practitioners. We will present the specificity and sensitivity of our questions. Narrowing the scope of practice is a good strategy.

Conclusions: Short questionnaires fulfill the twofold objective of enticing senior physicians to think about their future and helping us understand their practice better. We intend to pursue that kind of contact with our physicians and might expand it to other age categories.

Take-home Messages: Older physicians need to be assessed. Short questionnaires are well accepted. Narrowing of practice is the key for senior physicians.

OTT-OE-2-3
The Value of a General Medical Science Examination in a Comprehensive Competence Assessment for Practicing Physicians

Elizabeth Wenghofer (*Laurentian University, Sudbury, Ontario, Canada)*, Tom Henzel (*National Board of Medical Examiners, Philadelphia, PA*), Steve Miller (*University of California, San Diego Physician Assessment and Clin, San Diego, CA*), William Norcross (University of California, San Diego Physician Assessment and Clin*, San Diego, CA*), Peter Boal (*University of California, San Diego Physician Assessment and Clin, San Diego, CA*)

Background: The University of California, San Diego Physician Assessment and Clinical Education (PACE) Program conducts a comprehensive assessment program occurring in 2-Phases over seven days. It is designed to evaluate physician competence. The outcome of the PACE Assessment is reported on a 4 point scale: three levels of Pass (Pass, Minor Recommendations, Major Recommendations), and one Fail level. It uses a variety of assessment tools including a test, Mechanisms of Disease (MoD), which evaluates basic medical science knowledge that is expected to be common to all physicians. This study examines the association between the MoD examination and the four levels of outcome of the assessment.

Summary of Work: 234 physicians assessed between 2008 and 2012 were included in the study. ANOVA with post-hoc analyses were used to determine if significant differences in exam scores existed among the four assessment outcomes.

Summary of Results: The majority of physicians had a satisfactory outcome on the PACE assessment with 89.7% passing and 10.3% failing. The overall mean MoD score was 66.9% correct with a median of 68.0%. ANOVA was significant (F=14.2; P<0.001) and post-hoc analysis indicated physicians with the best outcomes scored significantly higher on MoD examination (P<0.05). Mean MoD scores for Pass, Minor recommendations, Major recommendations and Fail outcomes were 71.0%, 64.9%, 63.8% and 59.7%, respectively.

Conclusions: Higher MoD examination scores are associated with better PACE assessment outcomes. The MoD examination is a useful tool in the assessment of medical science knowledge in practicing physicians and can assist in differentiating in outcome levels of comprehensive competence assessment.

Take-home Messages:

OTT-OE-2-4
PROGRESS: assessing the quality of the appraisal discussion

Susi Caesar (*Wessex School of General Practice, Wessex Deanery, UK, Wessex Deanery Appraisal & Revalidation Service, Winchester, UK)*, Karen Bentley (*Wessex School of General Practice, Wessex Deanery, UK, Wessex Deanery Appraisal & Revalidation Service, Winchester*), Sue Bowen (*Wessex School of General Practice, Wessex Deanery, UK, Wessex Deanery Appraisal & Revalidation Service, Winchester*), Steve Scott (Wessex School of General Practice, Wessex Deanery, UK*, Wessex Deanery Appraisal & Revalidation Service, Winchester*)

Background: Revalidation for all doctors was introduced in the UK in December 2012 to ensure doctors remain up to date and fit to practise. The process is based on an annual medical appraisal. The documentary outputs of the appraisal serve as a written summary of the appraisal discussion and a Personal Development Plan (PDP), and provide sign off statements from the appraiser to the Responsible Officer who will make the revalidation recommendation. To ensure a robust process, these outputs need to meet certain standards and provide equivalent levels of information across appraisals.

Summary of Work: The PROGRESS QA tool was developed by the Wessex Appraisal and Revalidation Team for the internal quality assurance of the appraisal outputs. Assessors score the outputs along 8 criteria and can provide free text feedback. The tool was designed to promote an approach to writing the appraisal outputs that would be developmental and encourage a focus on quality improvements in patient care, rather than just serving the outcome of revalidation. It has been used by 8 appraisal leads since 2012 to QA the outputs of around 200 appraisers, who have appraised over 2000 doctors.

Summary of Results: Scores were initially mediocre and demonstrated that the summaries were being written mainly to serve revalidation. Appraiser awareness of this was raised through training with the

tool, after which scores improved. It is argued that outputs scoring highly indicate that the appraisal discussion is supportive and challenging, and that excellence and reflection are encouraged.

Conclusions: A carefully designed quality assurance assessment tool can help appraisers to promote reflection and quality improvements in patient care.

OTT-OE-3: EVALUATION OF THE TEACHER/TRAINER

OTT-OE-3-1
Can student assessments of faculty's professional behaviours support positive behavioural change? A case study in catalytic validity

Meredith Young (*McGill University, Medicine, Centre for Medical Education, Montreal, Canada*), Sylvia Cruess (*McGill University, Centre for Medical Education, Montreal*), Richard Cruess (*McGill University, Centre for Medical Education, Montreal*), Yvonne Steinert (McGill University*, Centre for Medical Education, Montreal*)

Background: One challenge in teaching professionalism is the unprofessional behaviour of some clinical teachers. This study reports the impact of a large-scale assessment system where students assess the professional behaviours of clinical faculty.

Summary of Work: Between 2009 and 2013 clinical clerks anonymously evaluated 2 clinical teachers following each rotation using the Professionalism Assessment of Clinical Teachers (PACT; a locally developed tool for student assessment of professional behaviours of faculty). Faculty performance was analyzed for improvement in ratings across time (teachers evaluated, and provided performance feedback, in two academic years; 8010 forms for 377 faculty). Written comments were coded for presence of educational critical incidents (e.g. inappropriate behaviour), and frequency was recorded across academic year. Using data from internal quality assurance tools, mean proportion of reported mistreatment incidents was used to determine consequential validity of the assessment program

Summary of Results: Mean performance improved across time (mean=4.25 in year 1; 4.32 in year 2 (5-point scale) F(1,6392)=5.4, p<.05). Reported critical incidents fell from 4.3% to 2.1% in year 2, and a significant decrease in the proportion of mistreatment was found following the implementation of the PACT and other interventions (F(3,15)=3.5, p<.05).

Conclusions: Increases in faculty peformance profiles, decreases in free-text reports of educational critical incidents and decreases in institutionally maintained reports of mistreatment appear to support the validity of this assessment program.

Take-home Messages: It appears that the implementation of an assessment system focused on professional behaviours of clinical teachers, paired with systematic feedback, can support positive behavioural change.

OTT-OE-3-2
Aligning Clinical Teaching Evaluation Forms with Physician Competencies

Deborah Simpson (*Aurora Health Care, Academic Affairs, Milwaukee, USA)*, Danish Siddiqui (*Aurora Health Care, Ob/Gyn, Milwaukee*), Jacob Bidwell (*Aurora Health Care, Family Medicine, Milwaukee*), Richard Battiola (Aurora Health Care*, Internal Medicine, Milwaukee*), Danielle Greer (*Aurora Health Care, Center for Urban Population Health/Aurora UW Medical Group, Milwaukee*)

Background: Clinical teaching evaluation forms (CTEFs) typically focus on literature-based attributes of effective teachers including enthusiasm, clarity, time to teach, promotion of critical reasoning, respect for learner/patients, and feedback. As ACGME, CanMeds and others emphasize physician competencies in quality, safety and teamwork, our trainees learn what their clinical teachers' role model (e.g., patient safety hand-washing). However, our CTEFs typically omit these crucial competency-related behaviors.

Summary of Work: To align CTEFs with physician competency priorities, CTEF's must include items related to physicians as role models for quality, safety, and teamwork. To achieve this alignment, we developed a 20-item CTEF: 13 items from established clinical teaching effectiveness literature; 6 items associated with the ability to promote and champion patient safety (including infection prevention), clinical quality (including quality improvement) and teamwork; and 1 overall teaching effectiveness item.

Summary of Results: The new CTEF has been incorporated into 3 residency programs to date, with residency management software supporting delivery, analysis and reporting. Medical students evaluate residents and faculty at the end of each rotation, while residents evaluate faculty at the end of each rotation or quarterly depending on specialty. Residents and faculty report CTEF is quick/easy to use and preliminary analysis suggests strong reliability with potential for deleting select items.

Conclusions: Explicit inclusion of competency-specific teacher behaviors on CTEFs more effectively aligns the teaching evaluation process with desired physician competencies, emphasizing their importance to teachers and learners.

Take-home Messages: As with learners, what we measure is what we improve. CTEFs can be used to align physician competencies with teaching expectations and evaluations.

OTT-OE-3-3
An Instrument to Assess the Competencies of Residency Program Directors in a Multi-Source Feedback System

Jodi Herold (*University of Toronto, Postgraduate Medical Education, Toronto, Canada)*, Susan Lieff (*University of Toronto, Toronto*), Ari Zaretsky (*University of Toronto, Toronto*), Glen Bandiera (University of Toronto*,*

Emergency Medcine, Toronto), Sal Spadafora (*University of Toronto, Anaesthesia, Toronto*), Melissa Hynes (*University of Toronto, Postgraduate Medical Education, Toronto*)

Background: Currently residency program directors (RPDs) get little formal feedback on their performance. This pilot project evaluates the measurement properties of a competency-based multisource feedback (MSF) instrument developed to provide Postgraduate RPDs with feedback on their ability to fulfill the role. The 28-item instrument is based on a review of competency inventories and an iterative process involving national and local input.

Summary of Work: 17 (16%) of volunteer University of Toronto RPD's identified a total of 389 potential raters (mean 22 per RPD). Raters included residents, faculty (including deans, chairs and heads), and administrative staff. 166 (response rate 43%) raters completed the MSF, with a mean of 9 per RPD. RPDs completed self-evaluations using the same instrument.
The 28 competencies were rated on a 5 point scale with 1 being "unsatisfactory" and 5 being "superior". Generalizability analysis was performed where raters are nested within subjects who are crossed with items (r:s x i).

Summary of Results: RPD self-ratings were consistently lower than the mean of IP ratings on all competencies. Generalizability analysis of the pilot data revealed a generalizability coefficient of 0.57. Decision studies conducted indicate that 15 raters per PD are required to attain acceptable dependability (i.e., 0.7 or greater) of scores.

Conclusions: Preliminary results based on a self-selected sample of 17 suggests good measurement characteristics. Likely, reliability coefficients will increase with a more heterogeneous sample of all 100+ RPDs. This instrument must be complemented with other MSF system components including written feedback report and debriefing / goal-setting meeting.

Take-home Messages: MSF is a feasible, valid and reliable approach to providing feedback to residency program directors

OTT-OE-3-4
Location Of Clinicians And Trainee Education at Dalhousie (LOCATED) Project: a health human resource tracking initiative

Peggy Alexiadis Brown (*Dalhousie Medicine New Brunswick, Faculty of Medicine, Halifax, Canada)*, John Steeves (*Dalhousie Medicine New Brunswick, Faculty of Medicine, Saint John*), Tim Fedak (*Dalhousie University, Faculty of Medicine, Halifax*), Silvane Paixao (Dalhousie Medicine New Brunswick*, Faculty of Medicine, Saint John*), Preston Smith (*Dalhousie, Faculty of Medicine, Halifax*), Kathleen MacPherson (*Dalhousie University, Faculty of Medicine, Halifax*)

Background: Previous rural background experience, early exposure in rural communities during medical training and distributed models of undergraduate education can increase recruitment/retention of

physicians in rural settings (Rabinowitz et al. 2001; Pathman, 2004; Snadden, 2011). However it is unclear how successful these initiatives are when increasing recruitment and retention of physicians within rural communities.

Summary of Work: Dalhousie Medicine New Brunswick (DMNB) is one of a number of distributed models of education in Canada. Like other Canadian medical schools, Dalhousie's program uses a variety of approaches to facilitate recruitment of graduates within specific rural communities. Although too early to tell, Dalhousie is developing a process for capturing data to evaluate the success of these initiatives in recruiting and retaining graduates in rural locations and generalist specialties.

Summary of Results: Stakeholders (policy makers, researchers, senior administrators and information technology specialists) were invited to provide input on a process for measuring student/resident/physician pathways into practice. Medical schools already engaged in this activity were consulted as well.

Conclusions: Development of a process that measures how distributed models of medical education influence health human resources ensures medical schools are accountable to communities.

Take-home Messages: Evaluating long term outcomes of distributed medical education is complex and requires a longitudinal approach that includes buy-in from multiple stakeholders.

OTT-OE-3-5
Moneyball: The art of statistical analysis in health-related professions

Brad Marcum (*University of Pikeville/Kentucky College of Osteopathic Medicine, Academic Affairs, Pikeville, USA)*, Margaret Sidle (*University of Pikeville/Kentucky College of Osteopathic Medicine, Insitutional Effectiveness and Research)*

Background: Moneyball was a look at using statistical analysis to make important decisions to help a baseball team perform better This presentation looks at how one medical school applied a similar approach to teaching medical students using computer-based testing using an analytics platform to be more effective.

Summary of Work: We devised a plan for ensuring the success of students. There were multiple activities that occurred to demonstrate fulfillment of this objective. First-year students began taking their block exams as computer-based using ExamSoft. The software makes tracking student performance much easier and provides substantially more data to both faculty and students for use in improving performance. We use the data to make decisions and to effectively assess the achievement of student learning outcomes.This change has required faculty to make lecture materials available well ahead of lecture time students can optimize the learning experience. Our students receive iPads. The use of this technology assists in addressing our learning outcome of providing better options for students with various learning styles.

Summary of Results: These changes collectively address our learning objectives of making a better learning environment for our students. These changes also addressed our learning objectives to better prepare students for their board exams.

Conclusions: We have created a number of reports that include indicators or early warning signs for students who need additional support. This addresses the learning outcome of improving instruction through the use of data.

Take-home Messages: Participants will have a better understanding of what steps to follow to engage faculty in using data to improve their teaching.
address the learning outcomes.

OTT-OE-4: PROFESSIONALISM 2

OTT-OE-4-1
Building a GME-wide Initiative on Assessing Professionalism

P. Preston Reynolds (*University of Virginia, Medicine and Center for Biomedical Ethics and Humanities, Charlottesville, USA)*, Jim Martindale (*University of Virginia, Medical Education, Charlottesville*), Linda Waggoner-Fountain (*University of Virginia, Pediatrics, Charlottesville*), John Gazewood (University of Virginia*, Family Medicine, Charlottesville*), Mike Moxley (*University of Virginia, Obstetrics and Gynecology, Charlottesville*), Casey White (*University of Virginia, Medical Education, Charlottesville*)

Background: Assessing professionalism in graduate medical education remains a challenge. Members of the Project Team (PT) began in 2009 to lay the foundation for a GME-wide initiative to integrate professionalism assessment using the Medical Professionalism Behavior Assessment Tool (MPB).

Summary of Work: The first step was research to determine if residency program directors (PDs) provided seminars or lectures on professionalism, if they assessed their residents' professionalism using a unique assessment tool, and if not, would they use the MPB. The second step was statistical analysis of the MPB.

Summary of Results: Nearly all PDs stated the need for a curriculum on professionalism for residents and faculty, and all thought the MPB included behaviors relevant to professionalism in clinical medicine. The PT presented the statistical analysis of the MPB which has a Cronbach alpha of 0.85 at a GME Council meeting. To date, 7 of 13 PDs have agreed to participate in an initiative to use the MPB in resident assessment of profession.

Conclusions: Building a GME-wide initiative on assessing professionalism required data on the need for assistance in developing curricula and implementing faculty development and resident assessment programs. The statistical characteristics of the MPB as a valid and reliable assessment tool enabled the PT to suggest its use to PDs.

Take-home Messages: Creating a GME-wide initiative to teach and assess professionalism requires engaging PDs at each step of the process. Success in adoption of a professionalism assessment tool depends on its validity and reliability, and whether such an initiative aligns with ACGME Milestones requirements.

OTT-OE-4-2
Measuring professionalism in paramedics

Madeline Carter (*Durham University, Centre for Medical Education Research, Durham, UK)*, Bryan Burford (*Newcastle University, Newcastle*), Gill Morrow (*Durham University, Centre for Medical Education Research, Durham*), Charlotte Rothwell (Durham University*, Centre for Medical Education Research, Durham*), John McLachlan (*Durham University, Centre for Medical Education Research, Durham*), Jan Illing (*Durham University, Centre for Medical Education Research, Durham*)

Background: Professionalism is critical for effective performance in healthcare professionals. This study describes the development of a questionnaire to assess professionalism as a multi-dimensional construct in paramedics.

Summary of Work: Following a literature review, focus groups and workshops, a questionnaire was developed to measure professionalism in paramedics. Items referred to professional status, professional identity, attitudes, behaviours, organisational context, and situational awareness. Attitudes and behaviours were further sub-divided with reference to the five 'clusters' of professionalism identified by Wilkinson et al. (2009) and four additional dimensions identified from the literature review and qualitative work (pride in profession, appearance, flexibility, and behaviour outside work). Two global measures of professionalism were also included. Student and qualified paramedics completed the questionnaire. Global ratings of student professionalism were also obtained from educators.

Summary of Results: Initial factor analysis identified potential dimensions of professionalism in paramedics, including: pride in professional identity, organisational support, focus on time, comparative professional status, focus on professional development, flexible communication, appropriate behaviour, confidence in action, appearance, communication with patients, and adherence to rules. The correlation between student and educator global ratings of professionalism was low, although this relationship was stronger for students who were rated by educators as high on global professionalism. Regression analyses highlighted important predictors of educator ratings of global professionalism.

Conclusions: A questionnaire was developed to measure professionalism as a multi-dimensional construct in paramedics. Initial analyses indicate that the tool is valid and reliable.

Take-home Messages: Professionalism is a complex, multi-dimensional construct. However, it is possible to develop tools to assess elements of professionalism in paramedics.

OTT-OE-4-3
Assessing adherence to principles of patient confidentiality in a clinical skills laboratory

Maria-Louise Barilla-LaBarca (*Hofstra North Shore LIJ School of Medicine, Hempstead, New York, USA)*, Alice Fornari (*Hofstra North Shore LIJ School of Medicine, Hempstead, New York*), Judy Brenner (*Hofstra North Shore LIJ School of Medicine, Hempstead, New York*), Lauren Bloch (Hofstra North Shore LIJ School of Medicine*, Hempstead, New York*), Ellen Pearlman (*Hofstra North Shore LIJ School of Medicine, Hempstead, New York*), Joseph Weiner (*Hofstra North Shore LIJ School of Medicine, Hempstead, New York*)

Background: One area of professionalism, patient confidentiality, can be evaluated by traditional models of assessment, but professional adherence can be challenging in certain situations.
Summary of Work: Two clinical skills stations were developed for first year medical students to address adherence to patient confidentiality. The first case assessed both the medical knowledge related to the ethical principles underlying patient confidentiality through a written exercise as well as the ability to apply these principles in a clinical situation with a standardized patient (SP). In contrast, the simulated "relative" of an established SP asked for personal health information (PHI) of that patient. We compared student performance in these two cases using a 2-sample probability test.
Summary of Results: In the oral portion of the first case, all (59/59) students recognized PHI and none violated HIPAA regulations, as documented by SP checklist and videotape review. In the analysis of written work, only 1 student out of 59 failed to identify the underlying ethical principles applying to the case. In contrast, in the second case several students (4/59) divulged PHI within the context of disguised acceptable social banter (59/59 vs 55/59, p-0.04).
Conclusions: The majority of students is familiar with the ethical principles of confidentiality and can apply them to challenging clinical encounters. Knowledge of HIPAA regulation is not enough to ensure compliance with regulations. Challenging patient scenarios can help students better recognize and apply ethical principles.
Take-home Messages: A clinical skills lab is a useful way of highlighting challenges to patient confidentiality.

OTT-OE-4-4
How professionalism themes in the medical school curriculum are perceived and assessed among faculty members: a qualitative research

Rulliana Agustin (*School of Medicine University of Western Sydney, Higher Degree Research Student/Medical Education Unit, Sydney, Australia),* Ian Wilson (*University of Wollongong, Graduate School of Medicine, Sydney*), David Mahns (*University of Western Sydney, School of Medicine, Sydney*)

Background: Through years, society has asked for better health service. Better health service comes from better health professionals. Professionalism in a medical school curriculum should be well-integrated across continuum of the curriculum. It needs involvement of all stakeholders including faculty members who have same understanding about the importance of professionalism themes in the curriculum. Barriers and challenges sometimes come from faculty members.
Summary of Work: As part of a mixed-methods research, twelve in-depth interview involving faculty members were performed. Informed consent from participant were given after participant were given explanation.
Summary of Results: All in-depth interviews were analysed manually. Vary of themes were raised and then coding were being made. Codes are being created for quantifying qualitative data. Some themes are raised according to professionalism; assessment, role models, curriculum, and the definition itself.
Conclusions: Professionalism is an important part of a medical school curriculum. An examination of the curriculum regarding professionalism is needed including the involvement of faculty members. An in-depth interview can benefit to study this involvement.
Take-home Messages: 1. The importance of professionalism in the medical school curriculum should be realised by all stakeholders.
2. Study shows barriers to application of professionalism in the curriculum.
3. There are different perception among faculty members about professionalism in the curriculum.
4. Further study is needed to address the barriers and challenges.

OTT-OE-4-5
Development of professionalism in clerkship with an on-line concordance-of-judgment educational activity: a pilot study

Amelie Foucault (*University of Montreal, Montreal, Quebec, Canada)*, Serge Dube (*University of Montreal, Montreal, Quebec*), Nicholas Fernandez (*University of Montreal, Montreal, Quebec*), Robert Gagnon (University of Montreal*, Montreal, Quebec*), Bernard Charlin (*University of Montreal, Monteal, Quebec*)

Background: Learning professionalism requires making judgments in ambiguous situations, where multiple behaviours and answers are possible. A Concordance of Judgment Test (CJT), made up of twenty authentic vignettes involving professionalism issues, was conceived and tested in the pilot study. A panel of experts provides detailed explanations for their judgment in each situation. Students obtain a quantitative measure of how concordant their answers are with experts and learn from explanations. While tool is formative, measurement of professionalism can be obtained by aggregate scoring.
Summary of Work: Professionalism vignettes were written by experienced medical educators and validated in focus groups with clerks to ensure validity and reliability. To select experts, all clerks were asked to

nominate attending physicians that represented best professionalism role models. Most frequently named were asked to undertake the online CJT and to provide explanations. Seventy-nine clerks undertook the CJT. Score ranged from 53.0 to 77.0 (mean: 64.6).

Summary of Results: Satisfaction survey results indicated high satisfaction and relevance of tool despite some pitfalls. Post-test focus group data revealed strong desirability bias while answering and showed that students appreciated the exercise and thought that it helped them in professionalism training.

Conclusions: The purpose was to collect students' scores and perceptions of this tool in view of determining pedagogic usefulness and relevance of our instrument to foster the development of professionalism during clerkship. This user-friendly learning activity appears promising both as educational activity and as a way to assess development of professionalism during clerkship.

OTT-OE-5: MULTI SOURCE FEEDBACK

OTT-OE-5-1
Co-workers and Physician Multisource Feedback

Gregg Trueman (*University of Calgary, Calgary, Canada)*, Jocelyn Lockyer (*University of Calgary, Calgary*), Kent Hecker (*University of Calgary, -Calgary*), Tanya Beran (University of Calgary*, Calgary*), Pam Nordstrom (*Mount Royal University, Calgary*)

Background: Multisource feedback (MSF) is increasingly being used as one of the components in revalidation and recertification processes to guide physician development. Data provided by co-workers (e.g., nurses, pharmacists, technicians) are recognized as integral for assessing a physician's teamwork and interprofessional abilities. The purpose of this study was to examine both the reliability of co-worker scores and the association between co-worker familiarity and physician ratings as both affect perceptions of quality.

Summary of Work: MSF data from 9674 co-workers of 1341 Alberta physicians across 9 specialty groups were analyzed. Analyses for internal consistency and generalizability theory (G and D-studies) were used to assess reliability. The association between co-worker familiarity and the MSF scores they provided to physicians was assessed using ANOVA.

Summary of Results: Cronbach's alpha for all co-worker tools was > 0.90. Generalizability coefficients (EP2) ranged from 0.56 to 0.72 depending upon medical specialty. D studies revealed that a minimum of 11 co-workers are necessary to achieve stability (i.e., EP2 > 0.70). Co-worker familiarity exerted a significant (p < .001) positive main effect on physician performance scores, across all specialty groupings [F scores range from 35.50 (i.e., pediatrics) to 61.30 (i.e., psychiatry)]

Conclusions: This study confirms the reliability of co-worker scores and provides evidence that co-worker MSF data is stable and consistent for the purposes of physicians' continuing professional development.

Attention however needs to be paid to co-worker/physician familiarity as this relationship may favourably bias physician performance scores.

Take-home Messages:

OTT-OE-5-2
Perspectives on the outcome from MSF in specialist training – influence of feedback facilitator experience

Bente Malling (*Center of Medical Education, Aarhus University, Aarhus University Hospital, Department of Human Resources, Aarhus, Denmark)*, Christina Hjøllund (*PeoplePartner, PeoplePartner, Aarhus*), Hans Ehlert (*Business Psychology, Knebel*), Gitte Eriksen (Center of Medical Education, Aarhus University*, Aarhus University Hospital, Department of Human Resources, Aarhus*)

Background: Personal feedback in multi-source feedback (MSF) seems to enhance trainees' outcome. Some recommend that personal feedback is provided by specially trained feedback facilitators (FF). It is tempting to assume that outcome of MSF is associated with the experience of the FF. No previous studies have described MSF outcome provided by novices and experienced FFs, respectively.

Summary of Work: A questionnaire exploring FF's performance in the MSF dialogue was developed(1). Outcome was measured through FFs' self-evaluation and trainees' evaluation. Data was analysed descriptively and comments analyzed by inductive methods. The difference between novice (N=14) and experienced FF´s (N=18) was explored.

Summary of Results: In average novice FFs were less critical about own performance. FFs´ and trainees reported that experienced FFs gave less advice and included own experience to a minor degree in the dialogue. Experienced FFs´ reports higher focus on identification of developmental areas. Trainees' confirmed a shift in learning outcome from selfreflection to formulation of plans for future personal development if experienced FFs were involved. A positive correlation was found between follow-up after MSF and experience.

Conclusions: With experience the FF performance changed from just presenting data to facilitation of trainees' own interpretation and usability of MSF for future personal development. By time the FF use MSF as formative assessment tool which most properly will increase outcome.

Take-home Messages: From a trainee perspective the outcome of MSF is positively associated with the experiences of the FF. It can be speculated if supervision can affect the development from novice to skilled FF. reference1; Malling B, Eriksen G, Bjørg G & Ehlert H. Criteria for feedback facilitator performance in multi-source feedback in specialist training

OTT-OE-5-3
Measuring CanMEDs Roles in Internal Medicine Residents Using Multi-Source Feedback (MSF): The Development and Pilot of an Online Instrument

Aliya Kassam (*University of Calgary, Community Health Sciences, Calgary, Canada*), Aleem M. Bharwani (*University of Calgary, Medicine, Calgary*)

Background: MSF is a questionnaire-based assessment which involves evaluating physician performance from multiple sources. Current MSF tools assess clinical competencies but do not reflect the CanMEDs roles entirely; few have been applied to residents. We sought to develop and pilot MSF instruments for residents, peers, medical clerks and nurses using CanMEDS roles.

Summary of Work: Existing MSF questionnaires were reviewed and gaps were identified. Training objectives were also reviewed and items were generated. Focus groups were held with residents, nurses and allied health care staff to review the items. An attending physician and medical clerk also reviewed the items. Separate MSF questionnaires were developed for each group and items were revised accordingly. Each MSF tool was made available online for piloting with residents, peers, medical clerks and nurses. Participants were e-mailed the link to the questionnaire along with a picture of the resident they were to assess.

Summary of Results: Most responses came from nurses (n=20 questionnaires) who were asked to assess residents on three CanMEDs roles. The internal consistency for the questionnaire for nurses was 0.92 (0.87, 0.85 and 0.32 for the collaborator, manager and health advocate subscales respectively).

Conclusions: MSF could provide reliable feedback to residents regarding their performance on CanMEDs roles. Not all roles are relevant or can be measured by participants during an encounter. Encounters between residents and participant groups also vary significantly and participants may be unable to assess all residents.

Take-home Messages: The idea of MSF may be beneficial for residents however its utility and feasibility needs further investigation.

OTT-OE-5-4
Multisource feedback in post-graduate medical education: Residency and beyond

April J Kam (*McMaster University, Department of Pediatrics, Hamilton, Canada*), Kelly Dore (*McMaster University, Department of Pediatrics, Hamilton*), Christina Grant (*McMaster University, Department of Pediatrics, Hamilton*), Moyez Ladhani (McMaster University*, Department of Pediatrics, Hamilton*)

Background: Multisource feedback (MSF) is an evaluation system being utilized in many clinical settings, which is comprised of questionnaire-based evaluation tools completed by faculty, nurses, and ancillary staff. MSF provides specific feedback about observable behaviours from multiple perspectives. Studies show that MSF is particularly effective in assessing observable, non-quantifiable skills, including professionalism, interpersonal and communication competencies. MSF was found to be reliable and valid for resident evaluations, but literature on MSF for sub-specialty residency/fellowship programs is scarce.

Summary of Work: A literature review on MSF was completed highlighting benefits and challenges to its implementation. A questionnaire with 30 questions was distributed to an inter-professional sample (nurses, physicians, social workers, etc.). The top 10 questions selected were included in the final questionnaire. The Pediatrics residency program and four sub-specialty programs (Academic General Pediatrics, Pediatric Gastroenterology, Pediatric Endocrinology, Pediatric Allergy/Immunology) at McMaster University introduced MSF for their incoming residents/fellows. This is the first study looking at the implementation and outcome of MSF in both residency and sub-specialty programs simultaneously.

Conclusions: MSF is a reliable and valid assessment of residents. It is superior to traditional feedback in improving performance after feedback administration. It has also been shown to improve the ease with which faculty deliver feedback, facilitating a generally more constructive evaluation process. Limitations of MSF include the logistics of requiring a large number of assessors and a potential negative response from the residents. These factors must be considered in the implementation of MSF in a PGME program and will be evaluated in the outcome of our program implementation.

OTT-OE-5-5
Ward round assessment of performance (WRAP) – A 360 degree assessment tool for ward round leadership skill

Indumathy Lakshminarayana (*Birmingham Community Trust, Paediatrics, Cannock, UK)*, David Wall (*Retired Deputy Post Graduate Dean, West Midlands Deanery, Medical Education, Birmingham*), Helen Goodyear (*Consultant Paediatrician and Associate Postgraduate Dean for less than fulltime training, careers and professional support and Head of the School of Paediatrics , Health Education West Midlands, Paediatrics, Birmingham*), Taruna Bindal (Consultant Paediatrician at the Alexandra Hospital, Redditch, West Midlands, United Kingdom.*, Paediatriatrics, Redditch*)

Background: Leading ward rounds represents a complex task requiring good managerial and interpersonal skills, an essential pre requisite for hospital consultants. Ward round leadership skills is rarely assessed during training and there is often a lack of supervision and feedback resulting in inadequate preparation of senior trainees for this role.

Summary of Work: The WRAP assessment form was derived from opinions of consultants, senior nurses and trainees using Delphi technique. Paediatric trainees based in the West Midlands UK were asked to pilot tool during hospital ward rounds. Additionally, trainees completed self-WRAP so that scores could be compared to that of assessors which included nursing and medical staff.

Summary of Results: 100 trainees returned 324 assessments. Reliability of WRAP using Cronbach's alpha is 0.843. Data for all 5 domains was non – parametric shifted to the right with scores of 3 (good) and 4

(excellent). Consultants gave significantly lower scores and the same was true with self – WRAP scores. Generalisability theory showed good reliability with a G coefficient of 0.8 for 3 raters. Factor analysis returned 3 factors which accounted for 85% of variance. The factors are 1) Managerial skills - preparation and organisation, team working and communication 2) Teaching and enthusiasm 3) Punctuality

Conclusions: WRAP is a reliable, practical and feasible tool that can be used in both large and small units.

Take-home Messages: WRAP is an effective tool which assesses ward round leadership skills thereby easing the transition from trainee to consultant.

OTT-OE-6: OSCE 5

OTT-OE-6-2
Relationship of Individual Competency and Overall Global Ratings in Practice Readiness OSCEs

Saad Chahine (*Mount Saint Vincent University, Halifax, Canada)*, Bruce Holmes (*Dalhousie University, Halifax*)

Background: The Clinician Assessment for Practice Program uses an OSCE to assess IMGs readiness for practice. Eight competencies rated are History Taking, Physical Exam, Professional Behaviour, Quality of Spoken English, Communication, Counseling, Problem Definition & Diagnosis, and Investigation & Management. The Physician Examiner (PE) also rates the candidate's overall global performance. Our research question is, "Which competencies are the most prevalent in determining a satisfactory rating for the overall global?"

Summary of Work: The study contained 36 IMG examinees and in total 432 observations of their performance.

Summary of Results: Our results show that the odds of satisfactory performance are 3:7. Investigation & Management (p=0.00), History Taking (p=0.00), and Counseling (p=0.04) competencies significantly predict the overall global rating at p<0.05. We conducted a forward stepwise regression, with Investigation & Management as the only predictor. We can predict unsatisfactory at 90%, however only predict satisfactory at 60%. Including the other competencies only marginally improves the predictive capacity of the model.

Conclusions:

Take-home Messages: In our model, we can highly predict unsatisfactory performance in the overall OSCE, however, we can predict satisfactory an OSCE as approximately 60% . Thus there are elements that may predict satisfactory, that we will need to investigate qualitatively. Intuitively, PEs deem some competencies as more important. If we eliminate the overall global rating, and average the competencies, we should consider weighting competencies such as Investigation & Management. Also, by determining which

competency is most complex, we can incorporate this into OSCE case development.

OTT-OE-6-3
Two-Point OSCE Checklist Items: Do they better capture examiner judgments?

Sydney Smee (*Medical Council of Canada, Evaluation Bureau, Ottawa, Canada)*, Ilona Bartman (*Medical Council of Canada, Research and Development, Ottawa*), Stefanie Sebok (*Queen's University, Education, Kingston*)

Background: Scoring for OSCEs has evolved from itemized checklists to multiple scoring formats. The Medical Council of Canada uses conventional case-specific checklists and behavioural rating scales to score its Qualifying Examination OSCE. Physicians observe and score dichotomous items as "Done Satisfactorily". Anecdotally, some examiners feel compelled to give credit even when performance is unsatisfactory.

Summary of Work: To explore this issue, two-point checklist items, scored as "Attempted" or "Done Satisfactorily", were piloted. Examiners scored the same station across two cohorts. For one cohort they scored dichotomously, for the other they scored two-point items. Based on the encouraging results, two-point scoring is being used for one third of the October 2013 Qualifying Examination OSCE.

Summary of Results: Seven stations have been piloted. Within-station item analyses found higher item-total score correlations for two-point versus dichotomous items (e.g., 0.136 vs. 0.292; 0.203 vs. 0.473). Frequencies for "Done Satisfactorily" items were lower for two-point checklists. Frequencies for "Attempted" plus "Done Satisfactorily" were consistent with the frequencies for the same items when scored dichotomously. However, pilot stations were completed by small cohorts (n=91-616). Analyses of the October 2013 data will replicate earlier work with a larger cohort (n >1000) and assess the contribution of two-point scores to the internal consistency of the total score.

Conclusions: The two-point checklists assisted examiners in making finer grained scoring decisions and increased score discrimination. The pilot results support adopting the two-point format.

Take-home Messages: Further recommendations regarding the use of two-point checklists will be made based on results of the overall analyses.

OTT-OE-6-4
Do individual roleplayers systematically bias the outcome of a high-stakes postgraduate OSCE, in terms of candidate ethnicity?

MeiLing Denney (*Royal College of General Practitioners, London, UK)*, Richard Wakeford (*University of Cambridge, Hughes Hall, Cambridge*)

Background: Many high stakes OSCE examinations use trained roleplayers (RPs) to depict patients. Just as individual examiners can potentially affect results through biases expressed in discriminatory marking, so

could individual roleplayers make their simulations easier or more difficult for different candidate groups, for example by candidate ethnicity. We have not found any work investigating such possible systematic impact by individual non-assessing RPs on examination outcome.

Summary of Work: We studied 92,989 case scores from the MRCGP CSA (clinical skills assessment) with known candidate ethnicity, collected from January 2012 to May 2013. We utilised univariate analysis to inspect the data and then multifaceted analysis (using the statistical package FACETS) to quantify the extent of case variance attributable to interaction between candidate ethnicity and RP.

Summary of Results: Univariate analyses showed that RPs' case score difference by candidate ethnicity (or apparent bias) decreased markedly with increased candidate exposure. Multifaceted differential function analysis estimated the candidate ethnicity by RP interaction to account for a non-significant 2% of overall variance.

Conclusions: If RPs' data is limited to only a few assessment days, their equality statistics may appear extreme (in either direction), due to non-representativeness of candidates encountered; RP score differences by candidate ethnicity reduce markedly as their experience increases. Powerful multifaceted analysis of data covering all RPs showed that variance in candidate scores by ethnicity is not significantly affected by RP.

Take-home Messages: Roleplayers do not constitute a significant source of error variance or unfair bias by candidate ethnicity. Their potential bias should only be estimated when they have seen a representative sample of candidates.

OTT-OE-6-5
Validity evidence for the OSATS (objective structured assessment of technical skill): a systematic review

David A Cook (*Mayo Clinic College of Medicine, Div. of General Internal Medicine, Rochester, MN, USA*), Benjamin Zendejas (*Mayo Clinic College of Medicine, Dept. of Surgery, Rochester, MN*), Ryan Brydges (*University of Toronto, Toronto, ON*), Rose Hatala (University of British Columbia*, Vancouver, BC*)

Background: The OSATS was described in 1995 as an approach to assess resident physicians' basic surgical skills. The original activity involved multiple short stations with performance assessed using a global rating scale and a task-specific checklist. The OSATS has since been adapted to assess a variety of procedural skills in diverse contexts. We sought to summarize the evidence supporting the validity of OSATS scores.

Summary of Work: We systematically searched for studies evaluating the OSATS as an assessment tool (last search update March 2013). We used Messick's framework to classify sources of validity evidence.

Summary of Results: We identified 37 studies reporting validity evidence for OSATS scores. 35 studies involved residents, 14 involved practicing physicians, and 5

involved medical students. Specialties included surgery (N=28 studies), gynecology (N=8), and internal medicine (N=2). Twenty studies reported content evidence, 33 reported internal structure (most often inter-rater reliability, N=25), all reported relations with other variables, one reported consequences, and none reported response process. Evidence nearly always supported score validity. Some adaptations altered potentially important elements (e.g., rating scales or number of stations).

Conclusions: The OSATS is commonly used for assessment of procedural skills. Robust validity evidence for OSATS scores demonstrates favorable relationships with other variables and acceptable reliability. Evidence of consequences and response process are lacking.

Take-home Messages: Good evidence supports the use of OSATS scores (as originally described with multiple stations, and both global ratings and checklist) for assessing procedural skills. Evidence should be re-evaluated for adaptations and implementations in new contexts.

OTT-OE-7: SELECTION FOR POSTGRADUATE TRAINING

OTT-OE-7-1
The pros and cons of combining the MMI and the SJT in postgraduate selection

Roberts Chris (*University of Sydney, Academic GP Unit, Sydney)*, Tyler Clark (*University of Sydney, Office of Medical Education, Sydney, Australia*), Annette Burgess (*University of Sydney, Central Clinical School, Sydney*), Fiona Patterson (Work Psychology Group*, Cambridge*), Marcia Grant (*General Practice Education and Training Ltd, AGPT selection, Canberra*), Karyn Mossman (*University of Sydney, Academic GP Unit, Sydney*)

Background: Within Australia, entry into vocational general practice training is determined by a National Assessment Centre (NAC) process which currently uses a combination of a Situational Judgement Test (SJT) and a Multiple-Mini-Interview (MMI).

Summary of Work: We wished to explore the hypothetical and statistical relationships between a written situational test (SJT) and an observed behavioural MMI and understand the unique ways that these formats contribute to determining the non-cognitive characteristics of candidates within the context of selection into postgraduate specialist training.

Summary of Results: Data were available from 1382 candidates, a six station MMI, and a 50-item SJT. A dissattenuated raw correlation with the SJT (r= 0.34) demonstrated evidence for divergent validity, construct, and concurrent validity.

Conclusions: The formats are testing differing non-cognitive aspects and should be retained on the argument of divergent validity. On the other hand where feasibility and cost constrains the number of assessment formats that can be used, our findings raises

the question as to how non-cognitive factors are most efficiently assessed in postgraduate selection settings

Take-home Messages: In assessing non-cognitive skills for postgraduate selection, more research is needed on determining the best combination of written and observed formats.

OTT-OE-7-2
What information is provided in transcripts and medical student performance records (MSPR) from Canadian medical schools?

Jason Robins (*University of Ottawa - The Ottawa Hospital, Diagnostic Radiology, Ottawa, Canada)*, Matthew McInnes (*University of Ottawa - The Ottawa Hospital, Diagnostic Radiology, Ottawa*), Kaisra Esmail (*University of Ottawa - The Ottawa Hospital, Diagnostic Radiology, Ottawa*)

Background: Resident selection committees must rely on information provided by medical schools in order to evaluate candidates. However, this information varies between institutions, limiting its value in comparing individuals and fairly assessing their quality. This study investigates what is included in candidates' documentation, the heterogeneity therein, as well as its objective data (or lack thereof).

Summary of Work: Samples of recent transcripts and MSPRs were anonymized prior to evaluation. Data was then extracted by two independent reviewers blinded to the submitting university, assessing for the presence of pre-selected criteria; disagreement was resolved through consensus. The data was subsequently analyzed in multiple subgroups.

Summary of Results: Inter-rater agreement equaled 92%. Inclusion of important criteria varied by school, ranging from 22.2% inclusion to 70.4%; the mean equaled 48.8%. The frequency of specific criteria was highly variable as well. Only 12.5% of schools provided any basis for comparison of academic performance; the majority only detailed status regarding pass or fail, without any further qualification.

Conclusions: Significant heterogeneity exists in the information provided by official medical school documentation, as well as markedly little objective data. Standardization is necessary in order to ensure appropriate evaluation of graduates from different institutions, as well as those from within the same school.

Take-home Messages: Standardization of medical school documentation is necessary in order to fairly assess and compare residency candidates.

OTT-OE-7-3
Evaluation of a new selection process for selection into for Trauma & Orthopaedics in UK

Alison Carr (*Health Education England, London, UK)*, Máire Kerrin (*Work Psychology Group*), Emma Rowett (*Work Psychology Group*), Mark Goodwin (Chair SAC Trauma & Orthopaedics UK and Ireland), David Wilkinson (*Health Education Yorkshire and the Humber*),

Julie Honsberger (*Health Education Yorkshire and the Humber*)

Background: Selection for entry into specialty training in the UK have recently moved away from local processes towards a co-ordinated national system. A review of selection practices in 2012 in one specialty highlighted a lack of consistency in how the same selection methods were implemented locally. This paper reports on work conducted to improve the standardisation, reliability and validity of selection methods and practices for entry into Trauma and Orthopaedics training in the UK.

Summary of Work: Selection exercises were designed according to a selection blueprint and included a portfolio interview, a clinical scenario and interactive exercises with an actor. N=469 candidates completed the selection exercises as part of the selection into Trauma and Orthopaedics training. A full psychometric evaluation of candidate performance and feedback was conducted.

Summary of Results: Analysis shows that the SC total score, and scores for each exercise and competency domain were capable of differentiating between candidates. Good levels of internal reliability were found for each exercise and for the SC overall. Each SC exercise significantly positively correlated with the SC total. Multiple regression revealed that the Interactive exercise explained the most unique variance in the SC score, followed by the Clinical Scenario, and the Portfolio Interview.

Conclusions: Overall, the psychometric analysis of the SC data available suggests it to be both effective (i.e. stations differentiate) and reliable (i.e. internal reliability). Further work is recommended on the calibration of assessors.

Take-home Messages: Nationally co-ordinated recruitment can help to improve standardisation of selection methods and practices. Having a 'one centre' approach to selection provides opportunities to further improve and enhance validity, reliability and standardisation.

OTT-OE-7-4
Does UK Foundation training prepare a trainee for a surgical career?

Dhanya Mullassery (*Alder Hey Childrens NHS Foundation Trust, Paediatric Surgery, Liverpool, UK)*, Graham Lamont (*HENW(Mersey Deanery), Liverpool*), David Bowen-Jones (*HENW(Mersey Deanery), Liverpool*)

Background: Future career choice is determined early, especially by doctors seeking to train in surgery. With the change in the U.K to a more general early years training (Foundation Programme), completion of which depends on assessment of a generic portfolio, this may make it more difficult for aspirant surgical trainees to demonstrate the aptitude for surgery that is sought in specialty selection processes.
 The study aim was to compare Foundation Programme portfolios maintained by surgical and non-surgical

career trainees seeking to identify significant differences in the outcomes achieved.

Summary of Work: - Anonymised data was extracted from portfolios of 2 randomly selected groups,27 trainees who entered surgical training and 27 who followed other training pathways in the Mersey Deanery. We analysed the Supervised Learning Events (a selection of work-place based assessments), and other aspects that are known to be scored in selection processes such as academic performance, audits and presentations. Continuous data were compared statistically using a Mann Whitney U test.

Summary of Results: The only significant difference between the 2 groups was in the number of Directly Observed Procedures (DOPS) (median number 5.5 vs 3.75-- $p<0.05$) undertaken by those pursuing a surgical career

Conclusions: This study has confirmed that the Foundation Programme does allow the demonstration of procedural competences by surgical trainees at a very early stage of training. In addition the surgical trainees' portfolios were not different from their non-surgical peers in terms of audit, research and publications.

Take-home Messages: The generic nature of the Foundation Programme does not preclude entry to a surgical career

OTT-OE-7-5
Selection for Family Medicine Residency training: is it time to sharpen our tools? A study of the consistency of selection processes across Programs in Canada

Keith Wycliffe-Jones (*University of Calgary, Family Medicine, Calgary, Canada)*, Shirley Schipper (*University of Alberta, Family Medicine, Edmonton*), Maureen Topps (*University of Calgary, Post-Graduate Medical Education, Calgary*), Jeanine Robinson (*University of Calgary, Family Medicine, Calgary*), Kent Hecker *(University of Calgary, Faculties of Veterinary Medicine and Medicine, Calgary),* Alianu Kingsly Akawung *(University of Calgary, Family Medicine, Calgary)*

Background: Selection for Family Medicine (FM) Residency positions in Canada is a high-stakes assessment activity. Students expect the process to be fair and reliable, with every student having an equal opportunity to gain entry to any Program applied for through whatever selection process is employed. Research in the UK has shown a low level of consistency in how Deaneries ranked the same applicants to General Practice Residency. No similar study has been conducted in Canada. Although Canadian FM Residency Programs are more homogeneous than in the UK, this study challenges the hypothesis that there is a high level of agreement between similar Family Medicine Programs in how they rank Canadian graduates who apply to more than one Program.

Summary of Work: Anonymised data from the Canadian Resident Marching Service (CaRMS) will be used to identify applicants to FM Residency Programs between 2002 and 2013 and how each Program ranked

the same applicants. Cronbach's coefficient alpha will be used to measure the correlation between student rankings by Programs as a whole and by Programs identified as being similar.

Summary of Results: An analysis of Cronbach's coefficient values will be presented indicating the level of correlation between how similar FM Residency Programs ranked the same applicants.

Conclusions: This study will demonstrate the level of consistency between similar FM Residency Programs in Canada in how they rank and select applicants to their Programs. Further analysis of the data may stimulate discussion and review of selection processes.

Take-home Messages: How consistent are Canadian Schools in ranking the same applicants to their FM Residency Programs?

OTT-OE-7-6
Implementation of a national programme for assessing and selecting trainee clinical scientists across multiple specialisms

C J Gibson (*National School of Healthcare Science, Birmingham, UK)*, L Lindsay (*National School of Healthcare Science, Birmingham*), S Chamberlain (*National School of Healthcare Science, Birmingham*), E V Davison (National School of Healthcare Science*, Birmingham*), S Hill (*NHS England, London*)

Background: In the UK a modernised training framework has been introduced for all healthcare science staff, at four educational levels for over 50 scientific specialisms. The National School of Healthcare Science (NSHCS) has a range of responsibilities for implementing this framework, including quality management, assessment, and selection for entry to a post-graduate training scheme leading to registered clinical scientist status.

Summary of Work: A national selection and assessment process has been developed and applied across all scientific specialisms. Candidates are assessed for their scientific skills, clinical problem solving, and situational judgement in three stages: aptitude testing, written application, and structured interviews. The outcome of this process has been reviewed.

Summary of Results: The full process has operated for two recruitment cycles, filling 513 posts from over 16,000 applicants. Approximately 64% of candidates passed numerical and logical aptitude tests; of those 12% were shortlisted; and of those 35% were successful. Independent information was gathered by the written and interview assessments ($R<0.5$), but communication scores at interview correlated strongly with tests of scientific ability ($R>0.85$). Equality and diversity analysis showed no disparities for gender, disability or age. Detailed analysis for ethnicity demonstrated confounding factors, especially level of qualifications and English as first language.

Conclusions: Consistent and transparent national assessment has been successfully implemented for a post-graduate clinical scientist training programme covering multiple specialisms in laboratory sciences, physiological sciences, and physical sciences.

Take-home Messages: Large scale selection and assessment of trainee clinical scientists has been achieved using a single application process, providing efficiencies for NHS employers, and fairness for candidates.

OTT-OE-8: WORK BASED ASSESSMENT 3

OTT-OE-8-1
Re-thinking workplace-based assessment: Trainees' and supervisors' perceptions of a revised system

Winnie Wade (*Royal College of Physicians, Education, London, UK)*, David Parry (*Royal College of Physicians, Education, London*)

Background: Workplace-based assessment for postgraduate trainees is undergoing a radical re-think and a new system has been proposed to address some of the concerns and issues which have arisen since its introduction for UK trainees.
A large pilot study in three UK regions was undertaken in 2012/13 to collect data from both postgraduate trainees and supervisors across ten physician specialities on their perceptions of the viability and usefulness of the new system.
Summary of Work: We aimed to explore trainees' and supervisors' experiences of using a formative and summative set of workplace-based assessments. Qualitative and quantitative data were collected through online questionnaires and focus groups on four separate occasions across three UK sites.
Summary of Results: Several key themes emerged: engagement of trainees and supervisors is key to maximising the educational impact and effectiveness of a new assessment system. Acceptability of both trainees and supervisors of the value of assessment for learning remains a challenge.
Conclusions: Training of both supervisors and trainees is key to the successful implementation of a new system. Active engagement of both participants can only be achieved through constructive feedback.
Take-home Messages: Both parties must have aligned views on the value of workplace-based assessment if a radical revision of assessments is to be widely accepted.

OTT-OE-8-2
In Training Evaluations: Developing a quick and easy automated screening tool for measuring completed report quality

Ramprasad Bismil (*University of Ottawa, Psychiatry, Ottawa, Canada)*, Nancy Dudek (*University of Ottawa, Physiatry, Ottawa*), Timothy Wood (*University of Ottawa, AIME, Ottawa*)

Background: In-Training Evaluation (ITE) is used to assess resident competencies in clinical settings. This assessment is documented on an evaluation report (In-Training Evaluation Report or ITER). Unfortunately, the quality of these reports can be questionable and therefore, training programs to improve report quality are common. The Completed Clinical Evaluation Report Rating (CCERR) was developed to assess completed report quality and has been shown to do so in a reliable manner enabling the evaluation of these programs. The CCERR is resource intensive instrument which could limit its use. The purpose of this study was to create a screening measure that could predict the CCERR outcome but in a less resource intensive manner.
Summary of Work: Using multiple regression, the authors analysed a data set of 269 ITERs to create a model that could predict the associated CCERR scores. The resulting predictive model was tested on the CCERR scores for an additional sample of 300 ITERs.
Summary of Results: The quality of an ITER, as measured by the CCERR, could be predicted with a model involving only three variables (R-square = .64). The predictive variables included the total number of words in the comments, the variability of the ratings and the proportion of completed comment boxes on the form.
Conclusions: It is possible to model the CCERR scores in a highly predictive manner. The predictive variables are ones that can be easily extracted by an automated process. Because it is less resource intensive than applying the CCERR, it opens up the possibility of providing feedback from ITER training programs to large groups of supervisors, institutions and even the possibility of creating automated feedback systems.
Take-home Messages: It is possible to measure quality of In-Training Evaluations using a computerised model.

OTT-OE-8-3
Understanding trainees' and trainers' experiences of supervised learning events in the UK Foundation Programme

Charlotte E. Rees (*University of Dundee, Centre for Medical Education, Dundee, UK)*, Ashley Dennis (*University of Dundee, Centre for Medical Education, Dundee*), Narcie Kelly (*University of Exeter, University of Exeter Medical School, Exeter*), Jennifer Cleland (University of Aberdeen*, School of Medicine, Aberdeen*), Karen Mattick (*University of Exeter, University of Exeter Medical School, Exeter*), Lynn V. Monrouxe (*Cardiff University, Institute of Medical Education, Cardiff*)

Background: A new Foundation Programme curriculum emphasising trainees' formative development through supervised learning events (SLEs) was launched in July 2012. Our external evaluation compared SLEs with workplace-based assessments (WPBAs) to answer the following questions: What are participants' understandings and experiences of SLEs and WPBAs? How do these understandings/experiences differ between trainees and trainers?
Summary of Work: We conducted 55 individual and 19 group interviews with 110 trainees and trainers in England, Scotland and Wales. We asked participants about their understandings of SLEs and WPBAs. Using narrative interviewing techniques, we encouraged participants to share their personal incident narratives

about SLEs and WPBAs. Data were analysed using Framework analysis.

Summary of Results: SLEs were conceptualised in more ways than WPBAs. Trainers frequently considered SLEs as assessment and a 'safety net' for catching struggling trainees. SLEs were mostly evaluated positively and WPBAs mostly negatively. Problem areas for SLEs and WPBAs included their initiation, the learning event itself, and their finalisation. Numerous factors were thought to facilitate/hinder the learning processes within SLEs/WPBAs at the level of the individual, interpersonal, culture, and technology. Trainees narrated more positive evaluations of their experiences but evaluated more WPBA experiences as negative compared with trainers.

Conclusions: Although some of our findings support existing literature, most are original given the novelty of SLEs. Further research is needed to explore the complexities surrounding SLEs within workplace learning. We will present five key educational recommendations from this research.

Take-home Messages: Although the shift from assessment to learning is largely welcomed by trainees and trainers, both require further education to better understand SLEs.

OTT-OE-8-4
Rising to the challenge: stress appraisals and performance in a high stakes assessment

Martin Roberts (*Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK)*, Tom Gale (*Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth*), John McGrath (*University of Exeter Medical School, Exeter*), Mark Wilson (University of Exeter*, College of Life and Environmental Sciences, Exeter*)

Background: The ability to work under acute pressure, so crucial in many clinical specialties, is also important in high-stakes assessments faced by junior doctors applying for specialty training posts - yet little is known about the reactions of doctors under this pressure. Responses to pressure may be analysed through the lens of the biopsychosocial model of challenge and threat. Previous research in psychology indicates that individuals who evaluate stress as a challenge rather than a threat may perform better under pressure. We aimed to investigate this model in the real-life setting of selection to anaesthesia training posts.

Summary of Work: Using the cognitive appraisal ratio (CAR), we measured the challenge / threat state of candidates prior to undertaking each selection centre station. Using regression methods we investigated whether the nature of the station and candidates' demographic and professional profiles predicted the CAR and whether the CAR predicted station performance.

Summary of Results: The CAR was affected by the nature of the stations and by candidate gender, but not ethnicity, country of training, age or clinical experience. After controlling for candidate and station factors the CAR significantly predicted station performance:

'challenge' evaluations were associated with better performance.

Conclusions: The CAR was unrelated to age or experience, suggesting that individuals may be predisposed to appraise situations as challenging or threatening. From a training perspective, interventions aimed at training doctors to evaluate stressful events as challenging rather than threatening could improve performance in clinical emergencies.

Take-home Messages: The biopsychosocial model of challenge and threat helps explain the performance of doctors under acute pressure.

OTT-OE-8-5
Work-place based assessment of Internal Medicine resident diagnostic accuracy

Michael D. Jain (*University of Toronto, Division of Hematology, Toronto, Canada)*, George A. Tomlinson (*University Health Network, Division of Clinical Decision-Making & Health Care, Toronto*), Allan S. Detsky (*University of Toronto, Divsion of General Internal Medicine, Toronto*), Luke A. Devine (University of Toronto*, Division of General Internal Medicine, Toronto*)

Background: Making an accurate diagnosis is a core skill for doctors. In residency, assessments of this skill are often based on global impressions formed by supervising physicians. A workplace-based assessment of diagnostic accuracy could be a valuable tool in the assessment of competence and could inform decisions regarding the degree of independence to grant a resident.

Summary of Work: The aim of this study was to develop a method for determining the diagnostic accuracy of senior internal medicine residents "on-call". To our knowledge, the diagnostic accuracy of Internal Medicine residents has not been previously studied. The diagnostic accuracy of 4 senior Internal Medicine residents was compared to that of their attending physicians for 240 consecutive patients referred to the Internal Medicine service. Residents documented their leading diagnosis for each patient admitted. After reviewing the cases with the resident, the attending physician documented their own diagnosis for each patient. The discharge diagnosis was determined by retrospective chart review.

Summary of Results: Compared to the discharge diagnosis, residents made an accurate diagnosis on the first day of admission on 66% of the cases. Attending physicians were accurate in 79% of the cases (p<0.0001 vs. resident). The accuracy of the attending physician was influenced by the accuracy of the resident.

Conclusions: Measurement of resident diagnostic accuracy can be incorporated into the usual clinical practice. Attending physicians have greater diagnostic accuracy than residents.

Take-home Messages: Measuring diagnostic accuracy can provide valuable information that can be used as part of a competency-based program of assessment.

OTT-OE-9: FEEDBACK POSTGRADUATE 2

OTT-OE-9-1
Numbers or Narratives: Do scores affect residents' recall of examiner feedback in a formative OSCE?

Lynfa Stroud (*University of Toronto, Medicine, Toronto, Canada)*, Matthew Sibbald (*University of Toronto*), Ryan Brydges (*University of Toronto*), Heather McDonald-Blumer (University of Toronto), Kevin Eva (*University of British Columbia, Vancouver*), Rodrigo Cavalcanti (*University of Toronto, Toronto*)

Background: Formative assessment requires effective feedback. In medical education, feedback commonly consists of numerical scores combined with narrative comments. In the broader education literature numerical marks have been reported to reduce the likelihood that learners take advantage of suggestions for improving performance. In this study, we examined whether scores affect residents' recall of narrative feedback.
Summary of Work: During a formative five-station internal medicine OSCE, residents were randomized to receive feedback after each station as comments alone or scores plus comments. Following the OSCE, residents were asked to recall specific feedback points from their last station and from a station of their choice. The outcome measure was concordance between resident-recalled and examiner-documented feedback.
Summary of Results: Among 152 residents, 866 individual feedback points were analyzed (mean 2.8 feedback points/station). Residents correctly recalled more feedback from the last station compared to their station of choice (42%vs.32%,p=.005). However, providing scores did not affect recall (37%vs.37%,p=0.9). Recall of feedback was similar for three distinct station types (clinical scenario=38%, physical exam=35%, communication skills=38%).
Conclusions: Residents recalled only a third of the feedback provided, but providing scores did not affect the rate at which feedback was remembered. This result contrasts with literature that suggests numerical marks may be distracting. Recency had a significant impact on feedback memorability, suggesting that a delayed recall test may be needed to clarify long-term feedback effectiveness.
Take-home Messages: In a formative OSCE, provision of scores with comments did not impair immediate recall of narrative feedback.

OTT-OE-9-2
Evaluation of written feedback given to doctors in the first post graduate year

Sandra Carr (*The University of Western Australia, Faculty of Medicine, Dentistry and Health Sciences, Perth, Australia)*

Background: The performance of Australian junior doctors' clinical, communication and professional skills is assessed up to five times in their first postgraduate year. However, little is documented on the comments assessors write and feedback given to junior doctors.
Summary of Work: This qualitative descriptive study explores assessors' behaviours around documenting written feedback on junior doctor performance and documents a content analysis of these comments.
Summary of Results: Of the 303 junior doctors in the study, 237 (78%) participated, resulting in 780 individual assessments. Only 6.3% of the completed forms did not have any comments recorded. 91% of the forms included comments on strengths and 56% on areas for improvement with junior doctor performance. The majority of assessments (88%) were based on close personal observation, while 37% were based on the assessor's general impression and 46% were based on observations made by others with 21% based on all three. On 80% of occasions the assessor consulted with others, including other consultants (52%), registrars (54%), nursing staff (29%) and other members of staff in 5% of cases. In 19% of occasions three or more sources of feedback were sought by the assessor completing the form. Feedback about the assessment was given to the junior doctor 75% of the time, leaving 25% of junior doctors not given this feedback.
Conclusions: Analysis of assessor comments suggests concern about performance in communication and teamwork skills as indicators to further evaluate the junior doctors' performance. However, the written feedback is often generalised and non- specific.
Take-home Messages: Together these findings represent a lost opportunity of assessors to provide meaningful feedback to these junior doctors.

OTT-OE-9-3
Improving written feedback to resident physicians: are Field Notes facilitating change?

Eric Wooltorton (*University of Ottawa, Family Medicine, Ottawa, Canada)*, Gary Viner (*University of Ottawa, Family Medicine, Ottawa*), Douglas Archibald (*University of Ottawa, Family Medicine, Ottawa*), Alison Eyre (University of Ottawa*, Family Medicine, Ottawa*), Madeleine Montpetit (*University of Ottawa, Family Medicine, Ottawa*), Julie Sutton (*University of Ottawa, Family Medicine, Ottawa*)

Background: Family Medicine residents are supervised by preceptors who document their assessments and feedback to the residents. In the past, supervisors used a 'check box' style direct observation form, but many supervisors often submitted poor quality documentation. In 2012 our Department shifted to using a new form called a "Field Note" different from that used at other universities. We believe the Field Note better documents regular ongoing assessment of a resident's proficiencies, problems, and improvement plans.
Summary of Work: We formally evaluated the implementation of the Field Note form to ensure it improved the quantity and quality of documented feedback. Specifically, we created a rubric that was used to score the quantity and quality of written feedback on

observation forms used prior to and since Field Notes were introduced.

Summary of Results: Prior using the rubric to score feedback forms we assessed a sample of forms, and there was good inter-rater agreement. Subsequently over 4300 feedback forms were scored from dozens of assessors covering hundreds residents Department-wide. Results will be reported.

Conclusions: The purpose of this evaluation was to assess the worth or value of these new forms. We assessed whether individual preceptors demonstrated increased frequency and quality of written feedback, using a novel yet simple rubric. Results will be used to provide individual faculty development on providing written feedback to residents

Take-home Messages: This innovative approach allowed us to assess whether the Field Note changed the behaviour of preceptors.

OTT-OE-9-4
Quality of Written Comments on Clinical Performance in a Four-Week Introductory Anesthesiology Program: Directed Content Analysis

Michael G. Richardson (*Vanderbilt University School of Medicine, Anesthesiology, Nashville, USA)*, Nancy K. Richardson (*Nashville*)

Background: Quality written feedback documents & promotes effective learning, yet is hard to come by. The quality of written performance assessments submitted by clinical preceptors during an intensive, highly structured 20-day introductory anesthesiology program was evaluated.

Summary of Work: Senor resident preceptors evaluated new trainees after 2-4 day clinical continuity pairings, using a REDCap on-line survey tool (24 performance milestones, 3 free-text questions, recommendations for subsequent supervision). Free-text responses (strengths/notable improvements; suggested targeted attention/improvement; specific feedback provided) were coded using directed content analysis.[Canavan C, et al. Acad Med 2010]

Summary of Results: Trainee-preceptor clinical contact averaged 2.7 days (median 3; 25-75% interquartile range 2-3). 25 preceptors completed 93 of 93 solicited evaluations of 15 trainees (mean 6/trainee; range 4-7). 83 evaluations (89.2%) contained ≥1 behavioral assessment phrase, including 74 (79.6%) with general and 41 (44.1%) with specific behavioral phrases. Also, 71 (76.3%) contained ≥1 non-behavioral/global assessment phrase. Although 10 evaluations (10.8%) contained only non-behavioral phrases, all 93 evaluations contained ≥1 targeted development/improvement recommendation, including general (38, 40.9%) and specific strategies (79, 84.9%). 20 evaluations (21.6%) referenced ≥1 specific incidence of behavior.

Conclusions: Despite little training (a single pre-program preceptor meeting; clearly communicated guidelines), resident preceptors submitted high quality written behavioral feedback, as well as substantive

recommendations for targeted development. Although untested, we suspect several responsible factors: direct clinical observation during 2-4 days of preceptor-trainee continuity; feedback instrument with clear completion instructions, and which combines free-text questions and behavior-oriented performance milestone assessments; supportive reinforcement to preceptors via e-mail.

Take-home Messages: Non-expert resident assessors are capable of providing high quality written clinical performance feedback.

OTT-OE-9-5
Exploring documented formative feedback within a competency-based residency program

Rosslyn Zulla (*University of Alberta, Family Medicine, Edmonto, Canada)*, Shelley Ross (*University of Alberta, Family Medicine, Edmonton*), Michel Donoff (*University of Alberta, Family Medicine, Edmonton*), Paul Humphries (University of Alberta*, Family Medicine, Edmonton*), Shirley Schipper (*University of Alberta, Family Medicine, Edmonton*)

Background: As the importance of competency-based education grows within health sciences education, receipt of feedback throughout residency remains key as a way for learners to monitor progress. For residents who learn within different hospital and other clinical teaching sites, feedback is essential to help them nurture as many skills as possible within as many clinical domains as they can.

Summary of Work: In 2009, the Competency-Based Achievement System (CBAS) was developed to provide ongoing and continuous feedback to family medicine residents throughout their residency training. In this presentation, we examine the frequency, distribution, and characteristics (including quality) of the written feedback received by a first-year cohort of residents.

Summary of Results: Feedback for 60 residents (28 female) was included in this study, across 14 teaching locations. Medical knowledge, patient-centered care, and procedural skills were most commented on. Female residents were more likely (but not significantly) to receive negative comments. There were significant differences by site in the characteristics of feedback received. Differences were also found between feedback given by experienced preceptors and that given by newer preceptors.

Conclusions: Elaborative narrative feedback may be instrumental to residents, particularly if they are completing their rotations across sites. If the amount of feedback received by a resident is a small, whether or not feedback will be beneficial to the resident will be contingent on its quality. As such, it is necessary to train preceptors to learn how to provide effective feedback.

Take-home Messages: Providing constructive feedback that enables residents to improve their knowledge and skills remains instrumental to nurturing their competencies within family medicine.

OTT-OE-9-6

Implicit and explicit communication of feedback affects students' retention of the feedback message

J.M.Monica van de Ridder (*Albert Schweitzer hospital, Department of Education, Dordrecht, Netherlands*), Janneke van Meteren (*Academic Medical Center, Surgery, Amsterdam*), Ernst-Jan van Unen (*Hogeschool NCOI, Utrecht*)

Background: Humans' behavior is shaped by their perceptions, this is known as the perception-behavior link (Ferguson & Bargh, 2004). The central research question in this study is: How is a feedback message perceived by students when feedback is implicity or explicitly communicated?

Summary of Work: An experimental study with a crossover design was carried out among clerkship students (n=44). One group received implicit feedback in the second week and explicit feedback in the third week of their clerkships, and this was reversed in the other group. When supervisors gave explicit feedback they used the word 'feedback', while this was avoided in the implicit feedback condition. At the end of a day we asked students to write down all the feedback points they had received. We asked the supervisors which feedback they had given. The given and the reported feedback points were compared using a Wilcoxon rank sum test.

Summary of Results: The response was 56% (n=25). In the implicit feedback dialogues, student reported less feedback points (Mdn=2.00) than given by supervisors. (Mdn=3.00), T=6, p<.05, r=-0.77. In the explicit feedback dialogues there was no difference between the reported feedback by students (Mdn=2.00) and the given feedback by supervisors (Mdn=2.00). Only 8% of the feedback content about professionalism is reported back in the implicit condition, while this is 100% in the explicit feedback condition.

Conclusions: Using the word 'feedback' influences retention of feedback points by students. Future research should explore if it also influences the extent to which students apply the received feedback.

Take-home Messages: Use the word 'feedback' explicitly!

POSTERS

OTT-PA4: CURRICULUM EVALUATION 1

OTT-PA4-01
The two-year study of a teaching model in family medicine for undergraduate medical students at Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin, Slovakia

Renata Pecova (*Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin, Dpt. of Pathophysiology, Martin, Slovakia*), Oto Osina (*Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin, Clinic of Occupational Medicine and Toxicology, Martin*), Jurina Sadlonova (*Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin, Clinic of Internal Medicine I., Martin*), Lubica Jakusova (Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin, *Clinic of Children and Adolescents, Martin*), Jan Danko (*Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin, Clinic of Gynecology and Obstetrics, Martin*)

Background: Family medicine inclusion in medical school curriculum is essential for producing competent general practitioners. The aim is comparative two year study of a teaching model in family medicine for undergraduate medical students at Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin, Slovakia.

Summary of Work: A teaching model in family medicine was developed in our conditions. Participants were medical students enrolled in the fifth year of the study (n=176). The study took place at Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin, Slovakia during the academic years 2011–2012 and 2012-2013.

Summary of Results: The primary endpoints are to launch practical part of family medicine subject, improve knowledge of family medicine and develop essential performance skills. Assessment of this two year study model by participating undergraduate students will help us with the future improvement of planning, organization and implementation of family medicine subject study.

Conclusions: A new subject family medicine is essential for all undergraduate medical students at Comenius University in Bratislava, Jessenius Faculty of Medicine in Martin, Slovakia. The practical part of the subject family medicine is implemented by various relevant departments and local general practitioners.

Take-home Messages: Importance of co-operation in family medicine teaching model.

OTT-PA4-02
Qualitative evaluation of theory-practice integration in morphologic courses at the Faculty of Medicine, National Autonomous University of Mexico

Margarita Varela-Ruiz (*National Autonomous University of Mexico, Medical Education, Faculty of Medicine, Mexico City, Mexico)*, Alicia Hamui-Sutton (*National Autonomous University of Mexico, Educational Research, Faculty of Medicine, Mexico City*), Tania Vives-Varela (*National Autonomous University of Mexico, Biomedical Informatics Department, Faculty of Medicine, México City*)

Background: The 2010-Curriculum at the Faculty of Medicine has a courses structure, including outcome competencies. These competencies integrate knowledge, abilities and attitudes and therefore an

appropriate link between theory and practice is required. The aim of this study is to analyze how theory and practice are integrated in the three morphologic courses in the Faculty first academic year.

Summary of Work: It is a descriptive, qualitative study, using the Focus Groups (FGs) approach. Sixteen FGs were conducted, interviewing 156 subjects in total: 2 professor-FGs and 3 students-FGs on Anatomy; 2 professor-FGs and 3 students-FGs on Embryology; and 2 professor-FGs, 1 assistant-FG, and 2 students-FGs on Cell and Tissue Biology.

Summary of Results: On Cell and Tissue Biology, the current Manual included clinical cases deemed as too complex for general practice. Theory was rich in content, leaving much less time for practice, and less weight was given to practical evaluation. On Anatomy, practice was burdened by overcrowded classes, the lack of corpses, and a gap with theoretical contents. On Embryology, practice relies on clinical cases, but not enough time is given for case analysis.

Conclusions: Clearly, theory outweighs practice. The latter supports the theoretical content, but fails to develop suitable abilities.

Take-home Messages: In a competence-based curriculum, the tradition of giving a higher value to theory in medical education at the expense of practice must be put to end.

OTT-PA4-03

Fourth year medical students' perceptions of their learning in the Health Practice Day (HPD) in the Clinical skills programme (CSP) at Wits

Amina Cassim (*University of the Witwatersrand (Wits), Centre for Health Science Education, Johannesbur, South Africa)*, Detlef Prozesky (*University of the Witwatersrand (Wits), Centre for Health Science Education, Johannesburg*)

Background: The early clinical exposure characteristic of modern undergraduate medical curricula presents organisational challenges since it is of necessity part-time and fragmented. At Wits third and fourth year medical students spend one 'clinical' day per week with three activities: shadowing, formal bedside teaching and systematic skills training.

Summary of Work: To investigate whether the three 'hospital day' activities in the fourth year programme comply with standards set for them, and to assess the effect of improvements instituted following a 2006 evaluation.

A cross-sectional descriptive study was carried out. During an eleven week period data were collected from a random stratified sample of fourth year students using a previously developed self-reporting questionnaire. Descriptive and inferential analysis of quantitative data was carried out supplemented by thematic analysis of qualitative data.

Summary of Results: In 'shadowing' sessions students were exposed to appropriately varied patients and procedures but mostly as spectators. The formal bedside teaching sessions had improved in some respects since 2006 but not in others, tutor availability being

particularly problematic. Skills training sessions were generally highly rated but did not fulfil all objectives. In all three activities problem areas were clearly identified.

Conclusions: A clinical skills programme requires the provision of adequate resources and opportunity to maximise positive student experience.

Take-home Messages: Early clinical exposure activities have particular problems and need constant careful monitoring.

OTT-PA4-04

First Patient Program: Program Renewal through Program Evaluation

Leslie Flynn (*Queen's University, Kingston, Canada)*, Sheila Pinchin (*Queen's University, Kingston*), Anthony Sanfilippo (*Queen's University, Kingston*), Kathy Bowes (Queen's University*, Kingston*)

Background: The First Patient Program (FPP) is an innovative longitudinal educational program for medical students at Queen's University. It was developed to provide relevant, experiential learning in the "pre-clinical" years. Students meet patients with chronic conditions to learn about the illness experience. Medical students are paired with a patient for 18 months, meeting the patients in their homes and accompanying them to clinical appointments. Students learn firsthand about: the impact of chronic illness on patients and families, the roles of physicians and other health care professionals, and barriers to health care.

Summary of Work: A program evaluation has been an intrinsic aspect of the FPP. The original evaluation design served formative, monitoring and participatory purposes. Increasingly, it serves a developmental purpose. Questions have been established to determine whether the program activities are achieving their desired outcomes. Both qualitative and quantitative processes are employed to gather information from stakeholders.

Summary of Results: Results are gathered coincidently with program delivery. These results inform ongoing program improvement. Important findings include the scheduling challenges students face to attend appointments, the uncertainty they find regarding supervising physician expectations, an unexpected barrier to receiving healthcare, and most significantly dealing with the experience of the loss of a patient.

Conclusions: Innovation in medical education and experiential learning is essential. Designing and conducting corresponding program evaluation is equally important. Ongoing program evaluation in partnership with educational leads has facilitated program renewal.

Take-home Messages: "The most important purpose of program evaluation is not to prove but to improve." Stufflebeam (1983)

OTT-PA4-05

Assessment item mapping to inform curricular evaluation of a neurosciences course

B Ostrowerka (*University of Alberta, Edmonton, Canada*), H Lai (*University of Alberta*), T Hillier (*University of Alberta*)

Background: Curriculum mapping and program evaluation are essential requirements for an accredited undergraduate medical education program. The mapping of assessment items can be used to investigate alignment with session, course, and program level objectives. This information serves as a source of validation on the organization and delivery of an individual course, or a longitudinal curriculum. However, as curriculum maps become further specified and complex, more refined mapping of assessment tasks is needed to validate achievement of educational outcomes.

Summary of Work: A third year medical student mapped assessment items in an 11 week Neurosciences course from the Year 2 curriculum of the MD Program at the University of Alberta. Each assessment task (i.e., test item) was mapped to six sets of objectives and outcomes, defined by the program, professional and licensing bodies, and additional external stakeholders.

Summary of Results: A total of 492 tags were amassed during the mapping of the 80 assessment items from the course's cumulative final examination. Data collected was used to: a) determine the coverage of course and session objectives by assessment items, b) identify gaps and redundancies in the assessment strategy, and c) evaluate alignment between curriculum, standards, and student performance.

Conclusions: Examination of the objectives guiding the course, and its assessment, will enable the identification of congruity and incongruity between the two. We will present data from this investigation, and on the actions of curricular change that such a process will encourage.

Take-home Messages: The alignment of the instruction and assessment of a course will facilitate the evaluation of educational outcomes.

OTT-PA4-06
A Biomedical Informatics course at UNAM Faculty of Medicine in Mexico: Program evaluation using a mixed methods approach

Tania Vives-Varela (*UNAM, Biomedical Informatics, Mexico City, Mexico)*, Israel Martínez-Franco (*UNAM, Biomedical Informatics, Mexico City*), Marlette Lobato-Valverde (*UNAM, Biomedical Informatics, Mexico City*), Melchor Sánchez-Mendiola (UNAM*, Medical Education Secretariat, Mexico City*)

Background: A curricular reform started at UNAM Faculty of Medicine in Mexico City in 2010. The program includes two new courses: Biomedical Informatics I (BI-I) and II (BI-II). The latter was first implemented in 2011. BI-II goal is to provide students with knowledge, abilities and attitudes related to clinical reasoning and decision-making. This study evaluated the BI-II course with a mixed methods approach.

Summary of Work: The objective was to identify how well BI-II is being incorporated in the 2010-Curriculum, as gauged by students' and professors' opinions. An explicative design was set, where quantitative data were qualitatively assessed to complement and clarify findings. A 25-item survey questionnaire was applied to 586 students at the end of BI-II course, to inquire about four domains: infrastructure, didactics, academic program, and organization. Analysis revealed some weak elements in the course, and they were discussed with qualitative methods. Five focus groups were conducted, three with students and two with teachers (17 students and 8 teachers).

Summary of Results: The following items were identified as problematic by the quantitative survey: motivation for studying contents, difficulty level of the module, expectations of students, knowledge and abilities acquired, and assessment. The qualitative data showed that examinations give more weight to theoretical knowledge.

Conclusions: A mixed approach allows us to identify weak points in the course, go deeper and explain them. This will help us to better understand educational phenomena and to influence them for quality improvement.

Take-home Messages: Quantitative data are enriched and better understood when qualitatively complemented.

OTT-PA4-07
Synopsis of evaluation criteria as a tool for curricular development in the Aachen medicine curriculum

Johann Arias (*RWTH Aachen Medical Faculty, Deanery, Aachen, Germany)*, Alexander Schiffel (*RWTH Aachen Medical Faculty, Deanery, Aachen*), Stephan Erdtmann (*RWTH Aachen Medical Faculty, Deanery, Aachen*), Sonja Finsterer (RWTH Aachen Medical Faculty, *Deanery, Aachen*), Melanie Simon (*RWTH Aachen Medical Faculty, Deanery, Aachen*)

Background: Knowledge on and thus acceptance of evaluation criteria such as the Progress Test is low among lecturers according to an internal survey in Aachen. Therefore, the faculty's evaluation results are not entirely well accepted and applied for the increase of teaching quality.

Summary of Work: Besides the development of an IT-based system that combines data from different sources, the work consists of an approach for direct access to all data which are relevant for evaluation of courses. These well-structured overviews allow for a better understanding of the connection of course and progress test results of cohorts, and students' survey results. Furthermore, summaries according to individual lecturers' interests have to be considered. To increase the acceptance of evaluation results it was essential to work closely together with lecturers and to give explanation on how to apply evaluation on development of courses.

Summary of Results: Consistent synopsis of different evaluation criteria gives lecturers a broader view on their courses' feedback. In the regarded examples, changes in courses can be assessed longitudinally, hence allowing to track curricular changes. Lecturers can

examine not only survey results but also the increase of knowledge to verify evidence in each of the evaluation criteria. Which further information could be interesting for a lecturer to evaluate his course?

Conclusions: Synopsis of different data is helpful for differentiated feedback on certain aspects like increase of knowledge, results of students' surveys, and effectiveness of a course.

Take-home Messages: More transparency in methods and consequences of evaluation criteria lead to a better acceptance of evaluation results among the persons involved.

OTT-PA4-08
PG ethics and professionalism: Practical issues based on how they interact

Roger P Worthington (*Yale University, General Internal Medicine, New Haven, USA)*, Richard B Hays (*Bond University, Faculty of Health Sciences and Medicine, Gold Coast*)

Background: Most medical schools in the developed world teach medical ethics and law, at least at an introductory level. At postgraduate level, the picture changes and the emphasis is more on practical professionalism, including workplace-based assessment. Ethical values inform professional practice, and while ethics and professionalism are not inseparable, the point where ethics ends and professionalism begins is not clear. While the former could be a subset of the latter, the relationship is less obvious than a model of concentric (as opposed to overlapping) circles implies. We argue that there is merit to be had from analyzing this relationship and considering possible implications for postgraduate training and curricula.

Summary of Work: The authors recently collaborated on a book on practical professionalism in medicine. While background research was mostly analytical, this was informed by practical experience of curriculum planning and development, and leading an extended series of workshops. [Ref. Practical Professionalism in Medicine: A global case-based workbook. Worthington & Hays. Radcliffe UK; 2013]

Summary of Results: Evidence upon which conclusions are based comes from analytic reasoning and group discourse, evaluating different responses to common practical scenarios.

Conclusions: The relationship between ethics and professionalism is not just a subject for academic inquiry. It can influence how curricula are shaped, including how and where such topics are taught and assessed in postgraduate medicine.

Take-home Messages: Analyzing interactions between ethics and professionalism can inform curriculum planning and development, including decisions about assessment.

OTT-PA4-09
The impact of the clinical practical year at the Medical University of Vienna: Are there

differences in clinical reasoning skills between students instructed either locally or remotely?

Philipp Pavelka (*Medical University of Vienna, Department for Medical Education, Vienna, Austria)*, Philip Anner (*Medical University of Vienna, Department for Medical Education, Vienna*), Andrea Praschinger (*Medical University of Vienna, Department for Medical Education, Vienna*), Franz Kainberger (Medical University of Vienna*, Department of Radiology, Vienna*)

Background: At the Medical University of Vienna a clinical practical year will be introduced in 2014. As a consequence, there will be a transition year during the academic year 2013/14 and additional clinical clerkship positions need to be provided. Since the University can't provide enough positions at the domestic hospital, students will be able to complete their clerkships in associate peripheral teaching hospitals. As opposed to local students, remote students don't attend any classes, however their clinical practice comprises more hours. Teaching hospitals received extensive information on how to properly instruct our students.

Summary of Work: The aim of this study is to evaluate the clinical reasoning skills of students instructed either remotely or locally. To measure clinical reasoning the Diagnostic Thinking Inventory (DTI) [1] is being applied. An online-questionnaire is being filled out by students at the beginning and the end of winter term 2013.

Summary of Results: The DTI scores gathered from the questionnaires will hopefully allow us to detect improvements in clinical reasoning and compare the results between both student groups.

Preliminary results from a DTI survey, conducted in 2006 [2], showed that the ability of clinical reasoning is independent of actual medical knowledge as taught in theoretical classes and goals achieved in clinical clerkships.

Conclusions: We will see if the recent study is comparable to the past survey and if the variations in clinical practice and the lack of theoretical classes will have an effect on students' ability of clinical reasoning, allowing us making valuable statements about the quality of clinical instruction received in peripheral hospitals.

Take-home Messages: References
(1) Bourdage G, Grant J, Marsden P. Quantitative assessment of diagnostic ability. Med Educ. 1990;24:413-25.
(2) Stieger S, Praschinger A, Kletter K, Kainberger F. Diagnostic grand rounds: a new teaching concept to train diagnostic reasoning. Eur J Radiol. 2011;78(3):349–52.

OTT-PA4-10
The Quality Accreditation of Medical Careers in Argentina

Mario A. Secchi (*Instituto Universitario Italiano de Rosario - Argentina, Rector, Rosario, Argentina)*, Nicolas Rodriguez Leon (*Instituto Universitario Italiano de Rosario - Argentina, Posgraduate, Rosario*), Walter

Bordino (*Instituto Universitario Italiano de Rosario - Argentina, Academic Secretary, Rosario*), Roberto Garcia Turiella (Instituto Universitario Italiano de Rosario - Argentina*, Medical School, Rosario*)

Background: The Quality Accreditation (QA) of Medical Careers Argentina was made during 2009-2010, by the "National Commission of University Accreditation in Argentina" (CONEAU) created by National Law 24521 in 1995.

Summary of Work: The QA was based on Resolution 1314/07 of National Ministry of Education, who resolved to approve the revision of the quality standard, by the "Association of Medical Schools in Argentina": AFACIMERA, which includes more than 40 medical schools both public and private of Argentina. This Resolution approves the basic curriculum, the minimum workload, intensity criteria of practical training and standards for the accreditation of courses of Medicine and the list of professional activities reserved for those who have obtained a medical degree.

Summary of Results: The process was a highly constructive work product
collaboration between the Faculties of AFACIMERA and the Agency CONEAU (2010). 39 of 40 Medical School were accredited for 3 or 6 years with suggestions for improvements.
The main recommendations were: training and more teaching dedication, expand libraries, creation of "Skill Centers" or "Virtual Hospital" to ensure learning and assessment of clinical skills, and more research.

Conclusions: 1) The National Agency CONEAU ensures the QA of Careers with national valid titles in Argentina.
2) The quality criteria from AFACIMERA were approved by the Ministry of Education.
3) We sought to build a national curriculum medical career based on Professional Skills.
4) Was given two years to fit these quality criteria.
5) Internal self-evaluation was performed (the University), an external evaluation (CONEAU), the recommendations and the presentation of the master plan for each faculty were made.
6) The process was a highly constructive product of a collaborative effort between the Faculties from AFACIMERA and the National QA Agency CONEAU.

Take-home Messages: The quality of curriculum evaluation of medical schools in Argentina is insured by revising standards and compliance by AFACIMERA (then by CRUP and CIN: University Council) evaluated by CONEAU from Education Ministry.

OTT-PA4-11
Academic medicine elective at Laval University: program description and curriculum evaluation preliminary results

Miriam Lacasse (*Université Laval, Département de médecine familiale et de médecine d'urgence, Quebec City, Canada*), Marie-Hélène Dufour (*Université Laval, Département de médecine familiale et de médecine d'urgence, Quebec City*), Josette Castel (*Université Laval, Département de médecine familiale et de médecine d'urgence, Quebec City*)

Background: Residents often do not have opportunities to be guided in their exploration of the various spheres of academic medicine. This project aimed to develop, implement and evaluate a scholarship-based elective in academic medicine.

Summary of Work: A 4-week academic medicine elective based on socio-constructivist and humanist learning theories was designed to allow competency development under the scholar (research and teaching) and manager CanMEDs roles. Curriculum included workshops and individual work for a scholarship project development. The elective was implemented with 7 residents in July 2013. Curriculum evaluation based on the Kirkpatrick levels 1 (appreciation questionnaire), 2 (pre-post learning assessment) and 3 (pre-post intention for academic career) was performed. Analyses used descriptive but not inferential statistics because of the small sample size.

Summary of Results: The elective was highly appreciated (mean score: 97.6%) and improved academic knowledge and skills (pre-test: 49.9%; post-test: 68.6%). Academic career intention slightly decreased from 4.71/5 to 4.56/5 after the elective. It was rated less realistic, more demanding but more motivating than initially perceived.

Conclusions: This academic medicine elective was highly appreciated and was associated with considerable learning outcomes, but a slightly lower interest for academic career after the elective. The 2013 elective cohort of residents will be assessed for academic productivity and career choice at 6 and 12 months, and compared with a control group of residents. This will allow better assessment and understanding of this elective's impact.

Take-home Messages: This academic medicine elective was highly appreciated and was associated with considerable learning outcomes, but a slightly lower interest for academic career after the elective.

OTT-PA4-12
A survey of evaluation systems for educational programs at Tehran University of Medical Sciences (TUMS)

Roghayeh Gandomkar (*Tehran University of Medical Sciences, Medical Education, Tehra, Iran*), Leyla Sadighpour (*Tehran University of Medical Sciences*), Azim Mirzazadeh, Batool Amini, Hamid Khanjani, Mojgan Safari

Background: Despite the emphasis of accreditation standards on program evaluation, establishing an effective evaluation system is challenging. The aim of this survey was to identify the program evaluation procedures in action in TUMS affiliated schools.

Summary of Work: A questionnaire was designed, including open and closed questions and distributed to the affiliated schools in 2013. The questionnaire was also introduced personally to the Dean of Educational Affairs in each of 10 Schools. Descriptive data was reported.

Summary of Results: In minety percent of schools, programs were evaluated using at least one source including students, faculty or course director feedback. Alumni source was used in 20% of programs. Eighty percent of schools were evaluating their faculty. Quality of exams was evaluated at 40% of the programs. Coding the data from open questions demonstrated that program evaluation was not an established process in the majority of surveyed schools and/or the evaluation data was not used appropriately.

Conclusions: The results of the survey showed that there was no systematic program evaluation functioning in TUMS which is a top ranked university of medical sciences nationwide; a conflict that put the development of an efficient evaluation system in the order.

Take-home Messages: The information gathered by the survey could be for development of program evaluation systems tailored to the educational needs and conditions of each school.

OTT-PB2: CLINICAL ASSESSMENT

OTT-PB2-01
Incorporating a test-enhanced learning component to clinical rotations

David Sanford (*University of Western Ontario, London, Ontari, cANADA*), Elaine Zibrowski (*University of Western Ontario, London, Ontario*), Michelle Zeller (*University of Western Ontario, London, Ontario*), Joy Mangel (University of Western Ontario*, London, Ontario*), Mark Goldszmidt (*University of Western Ontario, London, Ontario*)

Background: Testing during learning has been shown to improve retention. The impact of testing on learning has not been well studied in real-world contexts. We introduced and assessed the impact of a test-enhanced learning component to our hematology rotation for internal medicine residents.

Summary of Work: Two unique, content-matched multiple-choice tests are now written at the start and end of the hematology rotation. The content covered is based on the blueprint for the American Board of Internal Medicine (ABIM) exam. Residents were provided with feedback and reference materials after each mandatory test. We evaluated our intervention by: (1) correlating test scores with rotation evaluations (2) performing a needs-assessment based on exam performance and (3) measuring retention by comparing ABIM hematology section performance for our current residents and a historical cohort matched for training level.

Summary of Results: Sixty-seven residents participated. End of Rotation Evaluation scores did not correlate with test performance (Test 1 Spearman -0.02, p=0.9, Test 2 Spearman 0.2, p=0.1). Our needs-assessment suggested that: hematologic neoplasms, thalassemia and transfusion reactions are poorly understood content areas. There was no significant difference between

mean ABIM hematology-section scores for the testing and historical cohorts.

Conclusions: Our intervention appears to have value in needs assessment for residents and may also provide more accurate evaluation of medical knowledge compared with evaluator assessment. While we did not demonstrate long-term retention differences, this may relate to the small number of questions on the ABIM hematology-section rather than a lack of benefit to the intervention.

OTT-PB2-02
A new type of oral assessment at the end of a rotation in intensive care medicine: The "fictitious patient" set up by the resident

António Pais de Lacerda (*Hospital de Santa Maria, Faculty of Medicine of Lisbon, Intensive care Department (Prof. Carlos França), Lisboa, Portugal*)

Background: Residents' rotations in an ICU, is required in the postgraduate curriculum of several other areas of medicine. In addition to continuous assessment, their summative assessment at the end of the rotation validates a basic training with critically ill patients, capital in the growth of the future "hospitalist". During the orals, it was asked to present the clinical history of an ICU patient and discuss his/her diagnosis, treatment and evolution/prognosis.

Summary of Work: Currently a new format for orals was tried, in which each resident structures the clinical history of a fictitious patient, having necessarily to follow two or three clinical elements given at an initial statement ("seizures at hospital admission", ""5th day: fever 39ºC", "continuous renal replacement therapy introduced day 8th"...). The resident should explain the admission to the ICU reason(s), the cause(s) for an extended length of stay, discuss laboratory values and imaging, mechanical ventilation modes (chosen parameters), haemodynamics monitoring/treatment, AKI evolution, feeding problems, any surgical interventions...

Summary of Results: Some questions at the end were helpful to clarify the resident's integrated knowledge and to evaluate other areas not highlighted in the "fictitious case". Residents reported greater difficulties during the initial period, "thinking" what kind of condition would have triggered the whole patient's process.

Conclusions: In the oral assessment of a rotation in critical care medicine, structuring "fictional clinical cases" by the resident can be a useful tool for assessing aspects of knowledge integration.

Take-home Messages: Structured "fictional clinical cases" are helpful in postgraduate ICU rotations assessment.

OTT-PB2-04
Patient evaluation of clinical interaction with medical students

James Simpson (*Queen's University, Kingston, Canada)*, Andrea Winthrop (*Queen's University, Department of Surgery, Kingston*), Cherie Jones (*Queen's University, Department of Surgery, Kingston*), Nina Ramic (Queen's University, *Kingston*), Josie Xu (*Queen's University, Kingston*), John Presvelos (*Queen's University, Kingston*)

Background: In many clinical education environments, there is limited direct assessment of student-patient interactions. This is problematic given the important influence of these interactions on the students' future careers as physicians.

Summary of Work: We performed a pilot study using the Consultation and Relational Empathy (CARE) Measure to assess medical students' interactions with patients. The CARE Measure is a well validated and reliable tool used to evaluate the quality of patient-centered care delivered by a health care provider. In ambulatory clinics, 414 patients completed the CARE Measure for 14 students on their surgical clerkship.

Summary of Results: There was a significant difference between the medical students on all CARE Measure questions (p<.01). There was no significant effect of patient age or patient gender, nor any significant difference noted between evaluations generated before vs after the patient saw the preceptor. Cronbach's alpha for the CARE Measure was 0.96 indicating high reliability. A survey administered to patients that completed the CARE Measure found that 95% agreed/strongly agreed that they would be willing to provide feedback to medical students in the future. Medical students were given their evaluation results at the end of the pilot study. When surveyed, the students thought the feedback they received was beneficial.

Conclusions: These results suggest that patients can provide an important source of medical student evaluation that is useful formative and summative feedback for both students and medical educators.

Take-home Messages: The results of this pilot study support an expanded role for patients in the evaluation of medical students.

OTT-PB2-05
Medical students' perceptions upon student-based care followed by supervision: a potential formative student assessment with patients´ eyes

Ernesto A. Figueiró-Filho (*Faculty of Medicine - Federal University of Mato Grosso do Sul, Obstetrics and Gynecology, Campo Grande – MS, Brazil)*, Eliana Amaral (*UNICAMP, Obstetrics and Gynecology, Campinas - SP*)

Background: As part of clinical training in ambulatory settings at a public Brazilian medical school, undergraduate medical students frequently see patients alone, then discuss their findings with supervisors outside the consulting room, before returning to the patient with suggested management strategies (parallel consultation).

Summary of Work: The aim of this study was to assess students´ perceptions on student-based care supported by preceptor supervision outside the consultation room. We performed a cross-sectional study with convenience sample of students who performed outpatient care, in a pre-natal setting. Fifth and sixth year medical students answered a Likert Scale Questionnaire focusing on their self-perception on consulting performance. An open question asked students to write five words that represented what they felt after the student-based patient care.

Summary of Results: Of the 52 students, 56% were in fifth year and 44% in the sixth year. Students agreed that the proposed format encourages participation (82%). The case discussions with the preceptor were useful (90%) and contributed to their learning (88%). The words "trust", "secure", "responsibility" and "autonomy" were the most representative to express the students´ feelings in meeting the patients by themselves, supported by outside supervision of the preceptor.

Conclusions: Students expressed interest in knowing the opinion of the pregnant women about their perfomance in supporting medical care (92%), confirming that they would feel comfortable in being assessed by patients (84%).

Take-home Messages: We observed positive perceptions of the students regarding this outpatient model, opening up future possibilities to formative assessment of students by patients´opinions.

OTT-PB2-06
A survey on ability of 4th year medical students in EBM skills

Mahvash Alizadeh Naini (*Shiraz University of Medical Sciences, Clinical Skill Lab Center, Shiraz, Iran)*, Mohammad Hossien Dabaghmanesh (*Shiraz University of Medical Sciences, Clinical Skill Lab Center, Shiraz*), Narges Vaseghi (*Shiraz University of Medical Sciences, Clinical Skill Lab Center, Shiraz*)

Background: Evidence-based medicine is the integration of research evidence with clinical expertise to resolve patient problems. This process requires several constituent skills. Our goal was to evaluate these skills in 4th year medical students in Shiraz University of Medical Sciences.

Summary of Work: All 4th year medical students (n = 114) were assessed in 3 steps. In the first step, we developed a self assessment questionnaire and ask them to score their competency in EBM skills by themselves. In the second part of research, we developed another questionnaire for assessing their knowledge about Medline and the other medical search engines. In the following we create and implement a computer–based examination that measured medical students' EBM skills.

Summary of Results: In the first step, we collected 109 questionnaires in which 57 students (52/3%) were failed and 52 students (47.7%) were passed. In the second questionnaire which was collected from 52 students (who had passed the first questionnaire), 30 of them (57.7%) were familiar with Medline, who took part in our final examination. In the final computer based examination 18 students passed the minimal score and 12 were failed.

Conclusions: Self assessment questionnaires are not the efficient tools for evaluation of EBM skills.

Take-home Messages: Opening a part for training medical students to use EBM skills seems to be necessary.

OTT-PB2-07

Survey of teaching methods and assessment tools for Patient Safety competencies in Thai medical education centers: a nationwide survey

Usa Siriboonrit (*Chonburi Medical Education center, Chonburi, Thailand*)

Background: The concern for patient safety in the health care system is increasing tremendously worldwide. "Patient centered and patient safety" is one of the core competencies declared in Professional Standards for Thai Medical Practitioners 2011. Patient safety education preparing medical students to deliver safer care is needed and emphasized. The World Health Organization has developed a Patient Safety curriculum guide for medical students worldwide since 2010. What are the current statuses of patient safety curriculum in Thai Medical Education Centers (TMEC) under the Ministry of Public Health?

Summary of Work: A nationwide cross-sectional survey of educators responsible for patient safety education in TMECs was conducted in academic year 2012. The questionnaires asked about teaching methods and assessment tools being used for WHO 12 patient safety education topics, problems encountered and suggestion for implementing of patient safety education.

Summary of Results: The response rate was 74% (23 of 31). Ninety-five percent of TMECs have integrated patient safety in their curriculum. The most frequently used teaching methods were ward round: lecture: small group discussion (60:39:35%). Topics related to improved infection control, safe surgery were taught in 60% of TMECs. Managing risk: Root cause analysis and Being an effective team player were the less frequently covered topics. Only 11 in 23 TMECs assessed students' patient safety competencies. The assessment tools used were On the Job Evaluation, OSCE and MCQ respectively (50:33:8%).

Conclusions: Most TMECs have integrated patient safety in their curriculum with a diversity of methods but only half have assessment for this competency.

Take-home Messages: Teaching and assessment for patient safety competency are an important and needed improvement.

OTT-PB2-08

Safety Audit Assessment in Medical Students for prevention of accidental blood and bodily fluid contact

Jutatip Nirutterug (*Queen Savang Vadhana Memorial Hospital, Thai Red Cross Society, Medical Education Center, Faculty of Medicine, Burapha University, Chonburi, Thailand*)

Background: Accidental blood and bodily contact incident increased in medical students despite good orientation before working in patient wards. Medical Education Center created instrument to evaluate perception of medical students by using "Safety Audit Assessment Form", searching for the defect items and performed activities to prevent accidents during working in patient wards.

Summary of Work: A survey research was conducted in 84 medical students in clinical years. Using "Safety Audit Assessment Form" to evaluate perception about performance in prevention in 20 items. After finding out the defects of safety, the center performed knowledge management process, demonstration using personal protective equipment (PPE), and close monitoring during working. After 6 months, re-evaluation of incident in accidental contact was done.

Summary of Results: The main defect in medical student perception was Intravenous set disposal. The center launched practice guidelines for medical students to report the accident situation; there is a safety team to help them at the same time. On re-evaluation, the tendency of accidental contact with blood and bodily fluid in medical students gradually decreased.

Conclusions: Safety Audit Assessment Form is effective measurement to evaluate perception of medical students in preventing themselves from accidental blood and bodily contact. The results by this assessment form made the center know the defect items of safety, to be the direction of performing activities and practice guidelines including team to help them. Re-evaluation showed gradually decrease of accidents in medical students.

Take-home Messages: Medication Education Center must play effective role in orientation, assessment, fulfilling the defects and forming a team to help medical students at the same time.

OTT-PB2-09

Assessment of Emergency Medicine junior doctors using a RIME-Dreyfus hybrid model

GG Sim (*Changi General Hospital, Accident and Emergency Department, Singapore*), BYG Leong (*Changi General Hospital, Accident and Emergency Department, Singapore*)

Background: Junior doctors who work in our Emergency Department (ED) come with varying levels of medical experience and competence. In the past, the senior faculty made use of a single standard grading system to assess the junior doctors. We decided to create a hybrid system using elements of both the RIME model and the Dreyfus model to determine whether this was a more equitable method of assessment.

Summary of Work: Instead of using the same numerical scale grading system to assess the whole junior doctor cohort, each junior doctor was scored using a RIME-Dreyfus hybrid model which took into account whether they had any previous Emergency Medicine (EM) experience. We hypothesised that junior doctors with previous EM experience would obtain significantly

different results with both grading systems whereas first posting junior doctors would not.

Summary of Results: Junior doctors who were doing their first posting scored similarly on both grading systems. The mean difference was 0.230 (p= 0.270, 95% CI -0.190 to 0.649).

Regarding the junior doctors who had previous EM experience, there was a larger difference of 0.425 between the mean scores from both scoring systems. However, this did not reach statistical significance (p= 0.4974, 95% CI -0.979 to 1.829).

Conclusions: There was no significant difference in the assessment scores obtained by the junior doctors using either the traditional scoring system or the RIME-Dreyfus hybrid Model.

Take-home Messages: We will continue our efforts to refine the assessment methods for junior doctors in our Emergency Department.

OTT-PB2-10
Assessing social determinants of health in undergraduate medical curricula: pragmatic options

Yuko Takeda (*King's College London School of Medicine, Department of Primary Care and Public Health Sciences, London, UK)*, Ann Wylie (*King's College London School of Medicine, Department of Primary Care and Public Health Sciences, London*)

Background: It is imperative to prepare medical students to be aware of and competently responsive to non-communicable diseases (NCD) and risk factors in the clinical context and at wider community and population levels. A complex multi-source evidence to guide best practice is now available. General Practitioner Teachers (GPTs) have been identified as well placed to facilitate and assess student learning with regard to social determinants of health (SDH). We describe how one London Medical School, an innovative model of learning about and assessing SDH has been instituted and evaluated.

Summary of Work: Students prepare three to five 500-word community case studies in their final year that can be used during GPTs tutorials and are all submitted to the School. A convenient sample of case studies (10% of one rotation N= 14 students; 54 case studies) as well as supplementary student focus group data, were analysed.

Summary of Results: Findings suggest that this is an effective way of assessing student learning related to SDH, based on practical experiences and reflective writing during community placements. Students demonstrated appreciation of social complexity in health care and the "patient's journey".

Conclusions: Many areas of teaching cannot easily be assessed by traditional examination processes, which depend on demonstration of skill and knowledge in the absence of social complexity. This research suggests that five focused community case studies have been effective in assessing how student learned about SDH.

Take-home Messages: Assessment options for SDH can be pragmatic, low cost and an integral part of the learning.

OTT-PB2-11
Logbook: Mapping the Clerkship Experience through the EPITOMISE Framework

Shazia Sadaf (*Aga Khan University, Karachi, Pakistan)*, Naveed Yousuf (*Aga Khan University, Educational Development, Karachi*), Rukhsana Zuberi (*Aga Khan University, Family Medicine & Educational Development, Karachi*)

Background: The new logbooks piloted in 2010 have been specially designed adapting the EPITOMISE framework to help maintain a record of problems seen by the students during the clerkship; identify gaps in learning opportunities and facilitate planning of corrective measures to fulfill the learning needs identified by the students relating to the clinical presentations in the curriculum. Submission of completed logbooks is mandatory for students to be eligible for the Year 3 grand OSCE before being promoted to Year 4 of the MBBS program at AKU.

Summary of Work: By the end of the year 3 clerkships a total of 66 logbooks were reviewed to identify the clinical presentations documented; to define the "Core" presentations to be seen based on the frequently documented clinical presentation/s and EPITOMISE criteria; and clinical presentations or EPITOMISE criteria NOT documented to highlight the gaps in the curriculum / rare presentations or aspects of the curriculum that needs to be improved.

Quantitative analysis of data was done using SPSS 17.

Summary of Results: Factorization yielded the core clinical presentations to be seen, which helped determine the minimum number of cases to be documented by each student during the rotation.

Conclusions: The EPITOMISE Logbook is an effective tool for mapping the clerkship experience; monitoring student's performance and planning corrective measures to ensure completeness and uniformity of clinical experiences.

Take-home Messages: Every "EPITOMISE" entry in the Logbook helps illustrate the particular patient problem observed/clerked maximizing student learning.

OTT-PB2-12
The utility and reliability of the mini-CEX as a summative tool in undergraduate medical students

Steve Capey (*Swansea University, College of Medicine, Swansea, UK)*, Wyn Harris (*Swansea University, College of Medicine, Swansea*)

Background: Work place based assessment has been implemented in postgraduate medical education as formative assessment tools for many years. However the use of these assessments as a summative assessment tool in undergraduate education is limited.

Summary of Work: We evaluated the utility of using the mini-CEX assessment as a summative tool in undergraduate assessment. We evaluated the scores for each mini-CEX assessment encounter and investigated the agreement between central and placement assessors to explore the reliability and the potential of using mini-CEX as a summative assessment tool.

Summary of Results: The use of the mini-CEX as a summative tool is a feasible and an acceptable possibility in undergraduate assessment. We had good agreement between central and placement assessors and the results of the mini-CEX assessments have value as a summative assessment.

Conclusions: This exploratory study has demonstrated that it is possible to deliver a valid, summative mini-CEX assessment in an undergraduate setting.

Take-home Messages: It is possible to deliver a reliable and valid assessment of undergraduate medical students using the mini-CEX.

OTT-PB2-13
"I'm no teacher": Understanding radiation therapists' perception and approach to clinical competence assessment of medical radiation sciences students within an academic clinical setting

Kieng Tan (*Princess Margaret Cancer Centre, Radiation Medicine Program, Toronto, Canada)*, Krista Dawdy (*Odette Cancer Centre, Department of Radiation Therapy, Toronto*), Lisa Di Prospero (*Odette Cancer Centre, Department of Radiation Therapy, Toronto*)

Background: Within the academic clinical setting, radiation therapists are required to assess students' performance and competence as part of their clinical practice portfolio. Anecdotal feedback from the clinical setting identified challenges in assessment that emerged as a result of variation in who, how and the context in which the assessment was being completed. A survey was distributed to gain insight into the current state of clinical competency assessment.

Summary of Work: Radiation Therapists (n=75) from two affiliated clinical sites were surveyed. The survey was divided into three principle themes on the assessment of clinical competence: 1) who completes it; 2) how is it assessed; and 3) what measures are utilized in its' evaluation. The survey also captured the clinical experience and formal training in education and/or assessment of respondents.

Summary of Results: All respondents were aware that clinical teaching was part of their portfolio, but only 60% enjoyed teaching students. 55% rated themselves as very comfortable assessing competence, 93% were aware of documentation and support on how to evaluate with only 11% utilization. All collaborate with team members and slightly less with the clinical educators. 87% defined successful competence as sustainability over a defined period of time; measuring both clinical knowledge and application (97%).

Conclusions: Processes surrounding the assessment are congruent among clinical teachers supporting its

reliability. Enhanced collaboration with clinical educators in both support and training will increase their validity.

Take-home Messages: Focus should be placed on training the assessor not controlling the context.

OTT-PB2-14
The Fundamental Basic Clinical Skills of the Intern. How can we improve them?

Onanong Noomcharoen (*Queen Savang Vadhana Memorial Hospital, Thai Red Cross. Faculty of Medicine, Burapha University, Medical Education Center, Sriracha, Thailand*)

Background: At the medical education center level hospital intern roles consist of both clinical skill practices and medical student teaching. In this study we have conducted various exams in order to evaluate clinical skills and problem solving skills. We expect to use the results to improve the training program.

Summary of Work: In academic year 2011-2012, statistical analysis and data were collected from 32 medical students (Extern) and 32 interns during OB-GYN (Obstetrics and Gynecology) rotation for a period of 8 weeks. We have conducted an OSCE (Objective Structured Clinical Examination) and MEQ (Modified Essay Questions) on both groups in the last week of posting. The exam topic was on the basis of medical competency assessment criteria for national license.

Summary of Results: OSCE scores were significantly higher in medical students compared with interns (124.9 +/- 16.3 vs.107.6 +/- 17.4; $p< 0.001$). The average minimum score of interns is in procedural skills. MEQ scores were significantly higher in interns compared with medical students (189.5 +/- 18.4 vs.176.4 +/- 20.5; $p< 0.05$).

Conclusions: Interns have lower clinical performance when compared with externs, which may reflect on less adequate training classes and depends on personal persistence skills.

Take-home Messages: To improve fundamental basic clinical skills of interns. Pretest scores can be used for planning the management of teaching and learning. Persistence skills should be considered.

OTT-PB2-15
Assessing readiness for re-entry of residents returning from prolonged leaves of absences

Jose Francois (*University of Manitoba, Family Medicine, Winnipeg, Canada)*, Cliff Yaffe (*University of Manitoba, Surgery, Winnipeg*), Jillian Horton (*University of Manitoba, Internal Medicine, Winnipeg*), Bruce Martin (University of Manitoba*, Family Medicine, Winnipeg*)

Background: Accommodating residents returning from prolonged leaves of absence during residency presents significant challenges to residency program directors. Little in the published literature provides academic administrators guidance on how to approach these situations. Most of the present published literature on

re-entry relates to the return of previously licensed non-practicing physicians to clinical activity.

Summary of Work: Building on its in-house expertise in physician assessment for re-entry, the University of Manitoba Faculty of Medicine developed a comprehensive approach for the assessment of residents wishing to return to training following a prolonged leave of absence.

The goal of the reentry program is to assure that resident physicians have retained their clinical competence or to enhance, broaden, and/or develop clinical medical skills. Successful completion of a resident reentry program allows the physicians to return to their original or another residency program after an extended leave.

Summary of Results: The assessment process for re-entry includes a number of components starting with an intake process and pre-clinical setting assessment consisting of an assessment of a written assessment of knowledge, an oral exam, simulated patient encounters, documentation exercises and, if applicable, a neuro-cognitive assessment. This is followed by a period of clinical assessment, which includes observation in the clinical settings relevant to the anticipated re-entry residency program. The resident re-entry specialist compiles a summative assessment which identifies readiness for return to residency and restrictions or accommodations which would be required

Conclusions: The University of Manitoba Faculty of Medicine developed a novel comprehensive approach for the assessment of residents wishing to return to training following a prolonged leave of absence.

Take-home Messages: Assessing readiness for re-entry of residents returning from prolonged leaves of absences is an important but complex process

Monday April 28
3:30PM – 5:00PM

SYMPOSIA

OTT-SF-1
Multisource Feedback: Its controversies and challenges in providing feedback to practicing physicians

Presenters: Jocelyn Lockyer (*University of Calgary, Calgary)*, Joan Sargeant (*Dalhousie University, Halifax, Nova Scotia*), John Campbell (*University of Exeter Medical School, Exeter*), Marianne Xhignesse (*University of Sherbrooke, Sherbrooke*), Karen Mann (*Dalhousie University, Halifax, Nova Scotia*)

Summary: Multisource feedback (MSF) is increasingly being used as part of revalidation to assess physician performance across a range of competencies with particular emphasis on collaboration, communication, and professionalism. Both quantitative and qualitative data maybe collected. Feedback from medical colleagues, co-workers (e.g., nurses, pharmacists, technicians), and patients are aggregated and form the basis of the data. MSF approaches have been extensively examined for evidence of validity, reliability, feasibility, acceptability, equivalence, and educational/catalytic effect. This research has identified areas of concern and opportunities to enhance the potential of MSF to support physician learning and change. Four questions emerging from the literature will direct this symposia: (1) Rater selection. What is the optimal approach to selecting raters? Should the physician select the professionals who assess his/her competence? (2) Data presentation. How are MSF data optimally presented to participating physicians? What is the value in collecting qualitative data? (3) Feedback delivery and action plan development. What are the optimal approaches to feedback delivery? (4) Coaching and mentoring. What potential benefits would 'coaching' with a certified coach offer to the MSF process?

OTT-SF-2
Some Promise and Pitfalls of Clinical Reasoning Assessment: A Critical Examination of the Script Concordance Test

Presenters: Matthew Lineberry (*University of Illinois at Chicago, Chicago)*, Clare Kreiter (*University of Iowa*), Georges Bordage (*University of Illinois at Chicago, Chicago*), Discussant: Jack Boulet (*Foundation for Advancement of International Medical Education and Research (FAIMER, Philadelphia*)

Summary: Sound diagnostic reasoning during clinical encounters is a key competency of the effective clinician. However, the inherent complexity of such reasoning makes it challenging to assess, either for formative or summative purposes. In this session, we discuss one type of clinical reasoning assessment, the Script Concordance Test (SCTs). Two recently-published reviews on SCTs' psychometric properties have claimed that the methodology generally produces reliable and valid assessments of clinical reasoning, and that such tests may soon be suitable for high-stakes testing. Through a review of published SCT reports and a re-analysis of a previously-published SCT report, we have identified three critical threats to the valid interpretation and use of SCT scores which were not identified in previous reviews. These threats consist of logical inconsistencies in the scoring procedures, unexamined sources of measurement error, and construct confounding with examinees' response tendencies on Likert-type assessment items. This third issue risks bias against racial or ethnic groups with certain response tendencies; it also makes the test susceptible to score inflation due to coached test-taking strategies. Our research shows that examinees could drastically inflate their scores by never endorsing the two extreme scale points on the tests' 5-point scale. Even examinees that simply endorse "0" for every item could still outperform most examinees that responded as the test intended. In this symposium, we present our research on these validity threats and seek to stimulate discussion of alternative methodologies for assessment of clinical reasoning moving forward.

WORKSHOPS

OTT-WF-1
International cooperation for professional medical education and assessment: How can it work? Who benefits?

Presenter(s): Reinhard Westkaemper (*University of Bern, Bern, Switzerland*), John Norcini (*FAIMER, Philadelphia, United States*)

Background: Certifying exams are established in most medical specialties. There is a wide range of methods and quality. The validity and reliability are often hard to quantify or do not conform to the standards. There are goals for a good certifying process, but there are often very limited resources (financial, lack of assessment knowledge, educated faculty etc.) Professional education, training and assessment becomes more and more important, specially considering the growing migration of medical trainees, as can be seen in Europe. Several European Societies reinforce the discussion about common standards and are running examinations. There is increasing interest for such 'testing', and it seems there is a lack of self-assessment based on accepted standards.
Intended Outcomes: After attending this workshop, the learner will be able to:

• Describe the various methods used in certifying exams.
• Decide on an acceptable standard for quality
• Decide on how the criteria can be achieved
• Identify potential necessary measures.
Structure: The session will briefly describe the mean types of methods used in certifying exams (written, oral, practical) and the criteria for quality. It will be discussed how and where the criteria can be applied, which compromises can be made, and whether there are minimal standards. Are there successful models, how can international collaboration be organized, what is needed, how can it work, what are benefits? Does it improve international education and quality? How can acceptable levels be achieved across national borders?
Who Should Attend: Educators interested in learning about the assessment of professionalism in an international context, trainees interested in international training and (self-)assessment based on high standards.
Level of Workshop: Intermediate

OTT-WF-2
Use of Generalizability Theory in Designing and Analyzing Performance-Based Tests

Presenter(s): David Swanson (*National Board of Medical Examiners, Assessment Programs, Philadelphia, United States*)

Background: Simulation and performance-based testing methods (eg, OSCEs, oral exams, workplace-based assessments) are commonly used in the health professions. Because these methods involve multiple sources of measurement error (eg, examiner stringency, case/task difficulty, content specificity), classical reliability theory does not furnish the conceptual and statistical tools that are needed to investigate their psychometric characteristics. Generalizability theory (g-theory) does provide the necessary tools for estimation of the reproducibility (reliability, precision) of scores and for evaluating alternate approaches to test design and use of testing resources.
Intended Outcomes: At the conclusion of this workshop, participants will be able to
1. Describe the advantages of g-theory over classical test theory
2. View assessment situations from a g-theory perspective
3. Interpret commonly used indices of reproducibility that g-theory provides
4. Decide if they want to learn more about g-theory
Structure: The workshop uses an interactive, seminar-style format to provide a general introduction to g-theory, review commonly used assessment methods and problems from a g-theory perspective, describe statistical procedures and software for conducting generalizability analysis, and identify additional readings for learning more about g-theory.
Who Should Attend: Medical faculty and others involved in designing and implementing assessment procedures in the health professions. The workshop

doesnot assume any familiarity with generalizability theory; however, participants should be comfortable with analysis of varia
Level of Workshop: Intermediate

OTT-WF-3
Competency-based assessment on a shoestring

Presenter(s): Tom Laughlin (*Dalhousie Medical School, Family Medicine, Moncton, Canada*), Jennifer Hall (*Dalhousie Medical School, Family Medicine, Halifax, Canada*), Karlyne Dufour (*Dalhousie Medical School, Family Medicine, Moncton, Canada),* Katherine Quackenbush (*Dalhousie Medical School, Family Medicine, Halifax*)

Background: The College of Family Physicians of Canada have developed the Triple C Competency-Based Curriculum, and mandated that all family medicine departments develop a competency-based teaching and assessment program. Dalhousie Family Medicine has been working on this project for the past nine years, and would like to share some of their experiences, successes, challenges and the current tools in use. The focus of the assessment is, by necessity, primarily gathered by busy family medicine and specialty preceptors in the clinical practice settings of multiple distributed communities.
Intended Outcomes: 1. Participants will become familiar with an effective tool (field note) and its use in gathering behaviour-based information in the workplace.
2. Participants will learn methods to use and organize this information in summative and formative assessment.
3. Participants will be provided with a set of tools and a process that may be used or modified for their own programs.
4. Participants will have the opportunity to share the successes and challenges associated with this change in teaching and assessment
Structure: A short didactic presentation will be followed by role plays, small group work integrating provided tools and small and large group discussion.
Who Should Attend: Program Directors, faculty interested in competency-based assessment in the workplace
Level of Workshop: Intermediate

OTT-WF-4
Leadership as a competency: Entrustable professional activities for learning and assessment

Presenter(s): Larry Gruppen (*University of Michigan Medical School, Medical Education, Ann Arbor, United States*), J. Thomas Fitzgerald (*University of Michigan Medical School, Medical Education, Ann Arbor, United States*), Patricia Mullan (*University of Michigan Medical School, Medical Education, Ann Arbor, United States*)

Background: Some competency-based educational systems have adopted Entrustable Professional Activities (EPAs) both as tools for learning and vehicles for assessment (ten Cate, 2007). EPAs have appeal because they reflect relevant professional tasks and functions, but they are also complex and difficult to define and assess. This is particularly true for leadership outcomes, which are influenced by multiple environmental and social factors. Thus, describing leadership EPAs and using them for assessment is a challenge that remains unresolved.

Intended Outcomes: Generate assessment and standard setting processes appropriate for complex and multi-dimensional leadership competencies

Structure: This workshop will present three leadership EPAs: implementing change in an organizational setting, developing a proposal for organizational change, and mentoring colleagues. The workshop facilitators will describe the origin of these EPAs and how they are used for learning (20 min). The workshop audience will then divide into three small groups to evaluate the EPAs as assessment tools, how they might be quantified or scored, and how to define standards to judge "competence" level (40 min). The small groups will then reassemble and share the highlights of their discussions with the others (30 min). Participants will leave with the three EPAs and the notes from the group discussions and their own observations

Who Should Attend: Medical educators, especially of faculty and residents; curriculum evaluators

Level of Workshop: Intermediate

OTT-WF-5
The Objective Structured Clinical Examination (OSCE) – Identification of stations level flaws through the use of an OSCE item writing error detection tool

Presenter(s): Kathy Brotchie (*Monash University, Gippsland Medical School, Churchill, Australia*), Linda Sweet (*Flinders University, School of Nursing and Midwifery, Adelaide, Australia*)

Background: The Objective Structured Clinical Examination is an internationally accepted format for assessing clinical skills. Despite being in use now for more than three decades, OSCE item writers have been without a best practice "how to" guide. An expensive undertaking in any health professional education program, it is essential that OSCE stations perform as expected providing a valid and reliable assessment. Quality assurance processes such as piloting of stations does not always occur and pre-exam reviewers may fail to identify errors. Post-exam analysis has been used to identify flaws in station documentation and composition. Removal of errors prior to the assessment should form part of a quality improvement cycle and may benefit from a structured approach. A tool for identifying errors has been created and compared against a small database of OSCE stations for evaluation and revision. This workshop will explore the use of the

tool in the recognition and remediation of station level flaws.

Intended Outcomes: Participants will: reflect on the prevalence of OSCE item writing errors in their own settings; review a sample OSCE to identify potential flaws; investigate the use of a tool designed for the systematic identification of errors; reflect on the utility of the tool.

Structure: Interactive exercises including discussions, large and small group activities and opportunities for individual reflection.

Who Should Attend: Academics, clinical educators and health professionals involved with OSCE question writing or reviewing.

Level of Workshop: Intermediate

OTT-WF-6
Developing Situational Judgement Tests (SJTs) to assess professionalism in medical students

Presenter(s): Richard Knox (*University of Nottingham Medical School, Division of Primary Care, School of Medicine, Nottingham, United Kingdom*), Rodger Charlton (*University of Nottingham Medical School, Division of Primary Care, School of Medicine, Nottingham, United Kingdom*), Pirashanthie Vivekananda-Schmidt (*University of Sheffield Medical School, Academic Unit of Medical Education, Sheffield, United Kingdom),* Ryan Prince (*Health Education England, West Midlands, Birmingham, United Kingdom*)

Background: Regulatory bodies across the globe are keen for professionalism to be incorporated into medical school curricula. The complexity of assessing professionalism is much debated but the BEME review (AMEE 2013) indicates that the best premise to develop a curriculum for professionalism from is Situated Learning Theory (SLT).
Realistic clinical and professional scenarios are a key feature of situational judgement tests (SJTs), which may therefore be a useful tool in the assessment portfolio for the professionalism domain. SJTs have been used in non-medical pre-employment testing for more than 50 years. SJTs are increasingly used as part of the selection process for specialty training posts in different branches of medicine; for example, as part of family physician selection in the UK since 2007 and in Australia since 2012. Although validated as a tool to assess 'non-academic' attributes of doctors, there is little experience of using SJTs as assessment tools of professionalism in medical students. Careful development of their validity and reliability is required.

Intended Outcomes: • Shared understanding of the concept of professionalism
• Increased awareness of the possibility of using SJTs to assess attributes of professionalism in medical students.
• Greater awareness of the limitations of the use of SJTs.
• Increased understanding of how to incorporate international cultural interpretations of professionalism in SJTs.

• Collaborative working to ensure robust development of SJTs as an emerging assessment modality.
Structure: Short didactic presentations followed by facilitated small group discussions to achieve a consensus at a plenary
Who Should Attend: •Medical educators with an interest in the assessment of professionalism.
Level of Workshop: Introductory

OTT-WF-7
Assessing patient-centred clinical encounters and providing effective feedback – using video for clinical supervisor training and calibration

Presenter(s): Jenepher Martin (*Monash University, Eastern Health Clinical School, Box Hill, Australia*), Noel Roberts (*Monash University, Eastern Health Clinical School, Box Hill, Australia*), Kathryn Ogden (*University of Tasmania, Launceston Clinical School, Launceston, Australia),* Jennifer Barr (*University of Tasmania, Launceston Clinical School, Launceston, Australia*)

Background: Meaningfully and defensibly deconstructing and assessing the clinical encounter, in formative and high stakes examination, is challenging for all clinicians. The Patient Partner Program (P3: University of Tasmania, and PTA: Monash University) equips undergraduate students for true patient-centred care, focusing on the development and integration of key clinical competencies.
Supervising clinicians are well positioned to provide valuable feedback and foster student skill development. The group has developed, evaluated and refined a Rating Instrument for Consultation Skills (RICS) to facilitate this process. Training clinicians to reliably assess consultations and provide effective feedback is integral to P3/PTA.
Using the experience of P3/PTA, this practical session will bring participants together to discuss and workshop assessment strategies for clinical encounters. The example assessment for the workshop is RICS, and discussion will generalize the content to other clinical assessments relevant to participants.
Intended Outcomes: • Develop skills in assessment of students in a clinical context
• Experience and develop skills in assesor calibration
• Determine relevance of RICS and other instruments to participants' own contexts
• Share experiences of clinical assessment methods
• Opportunity to collaborate on further development of RICS
Structure: • Introduction to P3/PTA, the RICS tool, and use of video for formative assessment.
• Sharing of clinical assessment experiences and challenges
• View and assess two videotaped student consultations using RICS
• Assessor callibration exercise and discussion.
• Discussion about clinical assessment, and applicability of workshop to participants' contexts.

Who Should Attend: Health professionals and educators assessing students in clinical settings.
Level of Workshop: Intermediate

ORALS

OTT-OF-1: STANDARD SETTING

OTT-OF-1-1
Setting a Performance Standard for the National Assessment Collaboration Examination: Application of Generalizability Theory

Andrea Gotzmann (*Medical Council of Canada, Research and Development, Ottawa, Canada)*, André F. De Champlain (*Medical Council of Canada, Research and Development, Ottawa*), Sirius Qin (*Medical Council of Canada, Research and Development, Ottawa*), Fang Tian (Medical Council of Canada*, Research and Development, Ottawa*)

Background: Setting and evaluating the performance standard for high-stake medical examinations is critical in valid score interpretation. This study evaluated the generalizability of the ratings for the standard setting exercise for the National Assessment Collaboration Examination. This examination consists of 10 OSCE stations and 24 multiple-choice questions. The focus of this study was on the OSCE component.
Summary of Work: A panel of 18 physician experts were selected that represented and balanced demographic variables, and were divided into two sub-panels with two facilitators. A stratified random sample of 50 candidates across the score range were used to set the cut-score, and the same 50 candidate station score sheets for each station were presented to the panelists. Each score sheet provided rating scales on up to nine competencies.
Each panelist reviewed and provided per station their overall judgment which category each candidate belonged: "Unacceptable", "Borderline" and "Acceptable". The process was repeated for all 10 stations. Generalizability-Theory analysis was conducted to evaluate the effects of candidates, stations, raters, and sub-panel.
Summary of Results: The results indicated that largest effect was for candidate by station, followed by rater by candidate by station, and candidate. There was a small effect for station, rater and sub-panel.
Conclusions: These results indicate that a very small amount of variance was due to differences across raters or sub-panels which provided additional validation evidence for the recommended cut-score.
Take-home Messages: The panelists' ratings are important to evaluate in the context of standard setting, taking into account the candidate effects and stations specificity.

OTT-OF-1-2

Standard setting: a review of the literature and mathematical models on the impact of the number of re-sits on the diagnostic accuracy of professional assessments

John McLachlan (*Durham University, Centre for Medical Education Research, Durham, UK)*, Jan Illing (*Durham University, Centre for Medical Education Research, Durham*), Charlotte Rothwell (*Durham University, Centre for Medical Education Research, Durham*), Julian Archer (Plymouth University*, Medical School, Plymouth*

Background: Most professional examinations in medicine allow re-sits (re-takes) for those who fail. Treating the assessment as a diagnostic test, some failures will be True Negatives, who should fail. In addition, there will be False Negatives, False Positives, and True Positives. Increasing the number of re-sits permitted, decreases the number of False Negatives, but increases the number of False Positives to a greater extent. If the costs of these are asymmetric, such that False Positives (doctors who are practicing medicine at some level when they should not be) are more expensive than False Negatives, the consequences are exacerbated. Discussions of standard setting ought to include considerations of re-sit policy.

Summary of Work: We reviewed the international literature and mathematical models on the impact of the number of re-sits on the diagnostic accuracy of professional assessments.

Summary of Results: Approaches vary from country to country and discipline to discipline. The ideal strategy in each case depends on the specific nature of the assessment and candidates, but none the less, some general considerations can be drawn.

Conclusions: We recommend that employing more than 3 re-sits should only be permitted when particular evidence justifies it. After the allowed number of re-sits, a 'refractory' period should be required, to allow the candidate a realistic chance of gathering more knowledge and expertise before any further re-sits are possible.

Take-home Messages: The number of re-sits allowed should not be unlimited.

OTT-OF-1-3
Standard Setting: A Comparison of Modified Angoff Method and Borderline Method

Chew Fei Sow (*International Medical University, Clinical Skills and Simulation Centre, Kuala Lumpur, Malaysia)*

Background: We determine our undergraduate medical students' clinical and communication skills competency using the OSCE. A large number of examiners from various specialities were needed for 240 students per cohort with two intakes a year. We used Modified Angoff (Angoff) previously as our standard setting method (SSM). However, at least 10 examiners were needed to obtain acceptable reliability but meeting as a group is difficult due to workload and different timing of convenience. Moreover, new examiners had difficulty in

visualising hypothetical borderline candidates. Recently we switched our SSM from Angoff to the Borderline Group Method (BLM).

Summary of Work: In a recent OSCE exam, both BLM and Angoff were applied simultaneously to compare the two methods. Pre-OSCE, Angoff produced overall pass mark of 56%. On actual OSCE, examiners completed the checklist scoring and allocated a global score based on the student's overall performance, ere using the BLM generated pass mark of 57%. Post-OSCE, the examiners collectively reconsidered the borderline students' performance and reviewed the score individually and the total post-OSCE passing mark was 55%.

Summary of Results: Total candidates were 210 Year Three students. Number of candidates who failed using Angoff pre-OSCE was 10 and post-OSCE SS produced 11 failures. Likewise, there were 12 students who failed using BLM. There was a minimal difference between the pass mark using Angoff and BLM. Only one extra student out of 210 (0.5%) had failed using the BLM.

Conclusions: BLM is as reliable as Angoff. BLM is easily implemented and may be considered a better SSM due to feasibility based on the organization's available resources.

OTT-OF-1-4
Inflationary effect of unmasking numerical rating scale values in a clinical clerkship oral examination

Daniel M Panisko (*University of Toronto, Department of Medicine, Toronto, Canada)*, Lynfa Stroud (*University of Toronto, Department of Medicine, Toronto*), Edmund Lorens (*University of Toronto, Department of Medicine, Toronto*), Sumitra Robertson (University of Toronto*, Department of Medicine, Toronto*), Rajesh Gupta (*University of Toronto, Department of Medicine, Toronto*), Luke Devine (*University of Toronto, Department of Medicine, Toronto*)

Background: Examiners are reluctant to penalize students for poor clinical performance on oral examinations. This tendency may be overcome by masking actual numerical values on rating scales and by relying on verbal anchors alone.

Summary of Work: In a case-based structured clinical oral examination (SCOE) in the University of Toronto Internal Medicine clinical clerkship, examiners rated students on global domains using a 5 point Likert scale with verbal anchors for each point on the scale. Numerical scores were hidden from examiners in the 2010-11 academic year (n= 226 students) but were unmasked in 2011-12 (n=218).

Summary of Results: With the unmasking of numerical values, mean student scores on global domains increased from 79.9% to 82.0% (p < 0.001). Overall mean clinical case scores (a combination of checklist and global scores) also increased from 69.6 to 71.3 (p = .002). There was no difference in student performance on other assessment modalities used in the Internal Medicine clerkship between the two academic years.

Conclusions: There was an inflation of student marks in a clerkship SCOE when examiners were provided explicit

numerical values, in addition to verbal anchors, for items on global rating scales.

Take-home Messages: The inflationary impact of transparent numerical rating scales on overall scores in clinical oral examinations needs to be considered when designing student assessment tools in clerkships.

OTT-OF-2: ASSESSING THE PRACTISING DOCTOR 2

OTT-OF-2-1
Examinations for MOC and Certification: Where are the differences?

Linjun Shen (*American Board of Psychiatry and Neurology, Buffalo Grove, USA)*, Dorthea Juul (*American Board of Psychiatry and Neurology,Buffalo Grove*)

Background: In the US, psychiatrists typically take initial certification examinations after completion of residency training. Thereafter, they are required to take a maintenance of certification (MOC) examination every ten years. Although the two examinations measure the same competencies, empirical data are needed to determine how similar or different the designs of the two examinations should be. To this end, this study was designed to explore how certification and MOC examinees perform on the same group of items.

Summary of Work: Forty questions from the 2013 psychiatry MOC examination that closely fit the exam's blueprint were selected to be embedded in 2013 psychiatry certification examination as non-scorable questions. Similarly, forty psychiatry certification examination questions were embedded in the psychiatry MOC examination. About 200 examinees from each examination were randomly selected to answer these questions.

Summary of Results: MOC examinees performed slightly better on the MOC questions and certification examinees scored slightly higher on the certification items. When the questions were classified as application-oriented or scientifically-oriented, MOC examinees scored better on application-oriented questions, and certification examinees scored better on scientifically-oriented questions. The performance differences were in the range of 4-7%. The differences in item discrimination indexes were negligible.

Conclusions: MOC examinees were more capable of solving clinical problems and certification examinees were more familiar with scientific topics. This is understandable since MOC examinees are more experienced practitioners, and certification examinees are recent residency graduates.

Take-home Messages: In general, test questions for initial certification and MOC can be used interchangeably. However, how to differentiate or equalize the scientific and application aspects between the two examinations needs further deliberation.

OTT-OF-2-2

Assuring fitness to practice and fitness for purpose of UK doctors

Mike Bewick (*NHS England, Leeds, UK)*, Jenny Simpson (*NHS England, London*), Kim Hoyland (*NHS England, Newcastle*)

Background: Work-place based assessment of doctors' performance in their day-to-day practice has presented a major challenge to educators, the service and medical managers in the UK for the last 20 years. Within the country's nationalized system of healthcare, provided the potential for collecting the relevant information, yet the organizational mechanism of achieving a regular, robust and consistent assessment on every doctor practicing in the UK, proved elusive.

Summary of Work: The introduction and implementation of a system of medical revalidation, in December 2012, mandated by law, dependent upon standardised annual appraisal.

Summary of Results: The UK now has a single system in which every doctor, of every discipline, every sector, NHS or non-NHS, is required to engage in a standardised system of annual appraisal. Failure to do so results in loss of the doctor's licence to practice. A new role, the Responsible Officer has been introduced in every healthcare organisation.

Conclusions: The responsible officer's duty, enshrined in law, includes review of the output of each doctor's appraisal over 5 years, triangulating this against standardised patient and colleague feedback and other systems of clinical governance and audit. The responsible officer must also ensure that the appraisal system meets quality assurance standards and that action is taken to address any concerns about a doctor's practice.

Take-home Messages: 20% of UK doctors are going through revalidation this year, 40% in 14/15 and by March 2016, the fitness to practice and fitness for purpose of every UK doctor will have been assessed.

OTT-OF-2-3
A comprehensive approach to workplace based assessment

Marilyn Singer (*University of Manitoba, Division of Continuing Professional Development, Winnipeg, Canada)*, Brenda Stutsky (*University of Manitoba, Division of Continuing Professional Development, Winnipeg*)

Background: The Division of Continuing Professional Development Clinician Assessment Programs, Faculty of Medicine, University of Manitoba, was approached by the College of Physicians and Surgeons of Manitoba (CPSM) to develop a rigorous comprehensive assessment process for physicians practicing on the Conditional Register of the CPSM. As a result, the Manitoba Practice Assessment Program (MPAP), based on the CanMEDS competency framework, was created.

Summary of Work: A comprehensive literature review and a survey of Canadian assessment programs was

completed. Based on the findings, the MPAP was developed to include an in-depth self-assessment/practice profile, multi-source feedback, key interviews, and an onsite assessment that includes a chart review, a chart-stimulated recall session, and direct observation of practice.

Assessors are trained through blended learning (i.e., online and in-person). A pilot study of volunteer physicians was completed to further train assessors and validate assessment tools.

All data are entered into an online computer program which maps back to the CanMEDS competencies. Reports are generated by component and by competency.

Summary of Results: The MPAP started live assessments in January 2013. To date, five specialty assessments and one family medicine assessment have been completed. Five candidates were successful, and remediation was recommended in one case.

Conclusions: Workplace based assessment is the most suitable way to assess physicians in practice. The assessment must gather information from a variety of sources. Assessors must be well-trained.

Take-home Messages: High stakes assessments of practicing physicians must use a variety of tools. Assessments must be individualized to suit each practice.

OTT-OF-2-4
Identifying transitions, supports and feedback to support continued competence

Mary Clark (*College of Occupational Therapists of BC, Victoria, Canada)*, Jodi Herold (*University of Toronto, Faculty of Medicine, Toronto*), Susan Glover Takahashi (*University of Toronto, Faculty of Medicine, Toronto*), Christopher Corbett (CSCW Systems Corporation*, Victoria*), Kathy Corbett (*College of Occupational Therapists of BC, Victoria*)

Background: Throughout their careers, health practitioners experience transitions which may impact their clinical competence. These transitions may or may not trigger learning to maintain or acquire needed competencies. Several factors influence the outcomes of these transitions, including the individual's competence, contexts, access to resources that support competence, and feedback that triggers the realization that competence is in question.

Summary of Work: This case report describes the implementation of an online tool, the Annual Competence Review (ACR), which supports continued competence by helping occupational therapists (OTs) identify transitions and strategies to maintain competence. Completed by almost 2000 OTs in 2013, this practice-based assessment and interactive risk management tool provides feedback on the quality of their reasoning and decision-making, plus an ongoing, accessible record of individual risks and supports to competence.

Summary of Results: Results indicate 77% of OTs reported that currently, or in the next 6 months, they will be in transitions (e.g. in/out of practice/clinical

setting/leave). Interestingly, the frequency of transitions was less for the subgroup classified as lower performers on a clinical decision making test (ANOVA resulted in an $F_{1,1986} = 6.3$, $p < 0.05$). The feedback from participants indicated high satisfaction ratings with the ACR, with most (85.3%) agreeing it's relevant or appropriate to supporting continued competence and many (62.7%) planning to use the ACR feedback to guide their continuing competence over the next year.

Conclusions: Further study will evaluate the effectiveness of this tool to support health practitioners' continued competence throughout their careers.

Take-home Messages: The ACR shows promise in providing individual feedback to support continued professional competence.

OTT-OF-2-5
Accredited Scientific Practice (ASP): an approach to formal continuing professional development (CPD) in healthcare science reflecting patient and service needs

Richard Billings (*Health Education England, Modernising Scientific Careers, London, UK)*, Gillian Manning (*Health Education England, Modernising Scientific Careers, London*), Maxine Foster (*Health Education England, Modernising Scientific Careers, London*), Shelley Heard (Health Education England*, Modernising Scientific Careers, London*), Sue Hill (*Health Education England and NHS England, Modernising Scientific Careers, London*)

Background: Healthcare science (HCS) is essential for safe/effective patient care. Developing the current and future workforce through a structured programme of CPD (Accredited Scientific Practice [ASP]) reflecting changing patient/service requirements is part of the Modernising Scientific Careers (MSC) programme, a UK wide educational/training framework for HCS which defines four levels of training through to Consultant Clinical Scientists.

Summary of Work: The ASP framework provides formal recognition of CPD in the form of assessed academic and work-based learning across the career framework, allowing employers to develop the skills they require from their HCS workforce and scientists to progress in their careers. ASP is offered at three levels with the academic component being set at one level higher than the qualification base for each of the healthcare science staff groups. Workplace assessment is employer based but standardised/recorded through the National School for HCS (NSHCS).

Summary of Results: An online assessment/learning tool (OLAT) developed by the NSHCS is used to assess clinical competencies acquired through structured ASP, while also providing workforce information about the number of staff undertaking development in specific scientific areas of HCS. The NSHCS provides certification of satisfactory assessment of the work-based skills; the academic component is assessed by the relevant university.

Conclusions: Seventy-nine areas at 3 different levels of practice have been identified by senior scientists/service

for ASP development. Work is on-going with the NHS to develop further a coherent, quality-assured framework of workplace and academic training.

Take-home Messages: Developing coherent training/ assessment for HCS CPD reflecting patient/service needs will ensure safer, high quality scientific/diagnostic services for patients.

OTT-OF-3: THE STUDENT

OTT-OF-3-1
Effects of a test-taking strategies course in reducing test anxiety related to medical students' licensure exams

John Encandela (*Yale University, School of Medicine, New Haven, CT, USA)*, Nancy Angoff (*Yale University, School of Medicine, New Haven, CT*), Gary Leydon (*Yale University, School of Medicine, New Haven, CT*), Michael Green (Yale University, *School of Medicine, New Haven*)

Background: Test anxiety interferes with preparing for and taking the U.S. Medical Licensing Exam (USMLE) for many students. A course on test-taking strategies was piloted and studied for effectiveness in reducing test anxiety.

Summary of Work: Twenty-four students were randomly chosen to take the course from 43 who expressed interest among 101 second-year students. Test anxiety was measured with the validated Westside Test Anxiety Scale (score 1-5) at baseline (T1) and after the USMLE (T3) for all students and immediately after the course (T2) for course-takers. We compared anxiety scores over the 3 periods for cases (80% power to detect 0.5 point change) and compared T1 to T3 change in anxiety scores between cases and controls. Correlation between anxiety and USMLE scores was determined, adjusting for MCAT scores, which provided evidence of past test performance.

Summary of Results: Ninety-three students had a mean baseline anxiety score of 2.48 --"high normal" test anxiety (2.36 for controls; 2.41 for "not chosen" controls; and 2.79 for cases, p =0.02). Anxiety scores in the cases decreased to 2.61 after the course (p = 0.09), and to 2.53 after the USMLE, (p < 0.005). From baseline to post USMLE, case scores decreased by 0.20 points; controls increased by 0.15, p = 0.02. Test anxiety was inversely correlated with USMLE scores (β = -0.24, p = 0.02).

Conclusions: Test anxiety is associated with lower scores on USMLE Step 1. A test-taking strategies course can modestly reduce test anxiety. Further study with larger numbers will determine if such interventions can improve performance.

OTT-OF-3-3
Clerkship-failing students: what indicators are predictive?

M.A.H.E. Latijnhouwers (*Radboud University Nijmegen Medical Centre, IWOO, Nijmegen, Netherlands)*, R. Goorden (*Radboud University Nijmegen Medical Centre, IWOO, Nijmegen*), A.M. Hettinga (*Radboud University Nijmegen Medical Centre, IWOO, Nijmegen*), M.F.P. ter Horst (Radboud University Nijmegen Medical Centre, IWOO, Nijmegen), R.F.J.M. Laan (*Radboud University Nijmegen Medical Centre, IWOO, Nijmegen*), C.T. Postma (*Radboud University Nijmegen Medical Centre, Internal Medicine, Nijmegen*)

Background: The three final years of the medical curriculum in our institution comprise of successive clerkships. Unsatisfactory clerkship-markings are infrequent. However, their impact is high as it leads to discontinuation of clerkships until successful remediation. As average clerkship grades are high and thus discriminate rather poorly, other indicators might be valuable in early identification of students 'at risk' for clerkship-failure. This led us to inquire how clerkship grades relate to other assessment results.

Summary of Work: The study was done with the results of 85 students. To obtain a standardized comparator for clinical performance they were assessed on a 10-station OSCE after finishing their clerkships. Further comparators were results on the national medical knowledge progress test and their bachelor's course grades.

Summary of Results: Clerkship grades showed a significant correlation with both the OSCE and progress test results (R= 0,26 and 0,28). Interestingly, some of the students with weakest scores in the OSCE had good to excellent grades for their previous clerkships: lack of clinical performance was not evident from their clerkship grades. Comparison with bachelor's course grades also showed mixed results.

Conclusions: Singular assessments results do not show much predictive value on the individual achievements of learners. Assessment information obtained throughout medical curriculum should be longitudinally integrated to obtain a more accurate view of strengths and weaknesses of individual students.

Take-home Messages: Assessment information should be integrated and applied at the level of individual students to optimize their studying and provide feedback and timely remedial guidance.

OTT-OF-3-4
Innovative Approach to Remediation: Give Students Responsibility

Elaine Dannefer (*Cleveland Clinic Lerner College of Medicine of CWRU, Education Institute, Cleveland, USA)*, Beth Bierer (*Cleveland Clinic Lerner College of Medicine of CWRU, Education Institute, Cleveland*), John Tetzlaff (*Cleveland Clinic Lerner College of Medicine of CWRU, Anesthesiology, Cleveland*)

Background: Learners typically consider remediation as a punishment. We present a process that frames remediation as a learning opportunity to self-regulate performance. Students who make insufficient progress in a competency-based curriculum are identified for remediation through portfolio reviews or formal referrals. Students placed in remediation are given

responsibility to develop learning plans and submit progress reports to a promotions committee until goals are met. This process requires students to engage in the self-regulation tasks of setting goals, monitoring performance, and self-assessing progress. Key components involve continuous formative feedback and advisor support to help students monitor and self-assess performance.

Summary of Work: With IRB approval, we examined the student identification process, success rate and duration, and remediation plans of eight medical student classes.

Summary of Results: Of 224 eligible students, 29 were identified for remediation, with 93% successfully completing remediation. Remediation duration ranged from 1 to 19 months (Mean = 14.4, SD = 10.2). About 45% of students were identified through their portfolios and the remainder through faculty referrals. Of the available nine competencies, more students needed to address professionalism (59%) rather than medical knowledge (14%). Student-constructed remediation plans became more structured over time.

Conclusions: Giving students ownership to generate a remediation plan and provide progress reports for targeted competencies resulted in successful outcomes for most students and provided students with a facilitated experience in self-regulation.

Take-home Messages: Giving students responsibility to design and implement remediation plans with faculty support is feasible and supports skills in self-regulation.

OTT-OF-3-5
Psychometric and Validity Study of Emotional Intelligence in Undergraduate Medical Students Employing Confirmatory Factor Analysis

Naghma Naeem (*Batterjee Medical College, Department of Medical Education, Jeddah, Saudi Arabia*)

Background: A substantial body of scientific knowledge supports the claim that Emotional intelligence (EI) improves psychological well-being in individuals and contributes, enhances performance in the work place and thus leads to happier, healthier lives and successful careers.

Schutte Self Report Emotional Intelligence (SSREI) Scale is a popular EI scale consisting of 33-items. Developers of the Scale (Schutte, 1998) recommended using a total score to reflect a single factor or composite EI score.

Summary of Work: The current study examined the reliability, validity and factor structure of the bilingual English-Arabic version of the SSREI Scale in an undergraduate medical student population.

Summary of Results: Confirmatory factor analysis demonstrated a three factor structure comprising of 16 items. The reliability (α) of the three sub-scales was Optimism=0.76 (9 items), Awareness of Emotions=0.72 (2 items) and Use of emotions=0.55 (5 items).

Conclusions: The current study is the first to investigate the psychometric properties of a bilingual English-Arabic version of SSREI in an undergraduate medical student population. The bilingual version demonstrated a different factor structure (three factors) than the

original English language scale (uni-dimensional). Based on the findings, it is recommended that a separate score should be reported for each sub-scale to provide more meaningful information and interpretation.

Take-home Messages: EI is a multidimensional construct. SSREI (bilingual) scale is a useful, valid and reliable tool to measure emotional intelligence, which can be used for measurement, development and research on emotional intelligence.

OTT-OF-3-6
Impact of Changes in Mindfulness on Perceived Stress and Empathic Concern in Medical Students

Aviad Haramati (*Georgetown University, Washington, DC, USA*), Neha Harwani, Kevin Motz, Kristi Graves, Hakima Amri, Nancy Harazduk

Background: Recent reports indicate that mindfulness training decreases burnout and improves physician empathy and well-being. Medical students also suffer from burnout and exhibit declines in empathy, perhaps related to their stress level. Yet most medical curricula lack interventions that foster mindfulness, which may improve student stress management. The aim of this study was to determine whether a mind-body medicine course would alter mindfulness in medical students and affect their stress and empathy.

Summary of Work: Georgetown University SOM offers an 11-week experiential course in mind-body medicine (MBM) to expose first-year medical students to various mind-body approaches (e.g. meditation and guided imagery). The group sessions, led by 2 trained facilitators, involve sharing openly and listening without judgment. Data were obtained in 118 first-year medical students who completed the surveys before and after participating in the MBM course. Survey instruments included: Perceived Stress Scale (PSS), Freiberg Mindfulness Inventory (FMI), Positive and Negative Affect Scale (PANAS), and the Interpersonal Reactivity Index (IRI).

Summary of Results: Significant increases (P<0.001) were observed in mindfulness (FMI), positive affect (PANAS) and empathic concern (IRI), while significant declines were seen in perceived stress (PSS), in negative affect (PANAS), and in personal distress in response to distress in others (IRI). Furthermore, the changes in perceived stress and affect were significantly correlated (P<0.001) with improvements in mindfulness.

Conclusions: A one-semester course in mind-body medicine skills during the first year of medical school is effective in enhancing traits such as mindfulness, positive affect and empathic concern, while reducing students' perceived stress and negative affect. Further, the mindfulness level was an important predictor for the changes in perceived stress.

Take-home Messages: Fostering mindfulness through an experiential mind-body medicine skills course may help decrease medical student stress and enhance emotional intelligence. Such curricular interventions may promote better physician-patient communication and improve the quality of health care.

OTT-OF-4: ASSESSMENT AND INTERPROFESSIONAL EDUCATION 1

OTT-OF-4-1
Evaluation of an Interprofessional Training Ward Experience for Final Year Medical Students

Jean McKendree (*Hull York Medical School, York, UK*), Patricia McGettigan (*Barts and the London School of Medicine and Dentistry, William Harvey Research Institute, London*)

Background: Multiple care failings in hospitals in England have highlighted the lack of interprofessional work experience in health professional education. This study evaluated the impact on ward staff and medical students of a two-week interprofessional training placement undertaken by three cohorts of final year medical students.
Summary of Work: The HYMS interprofessional training placement was undertaken on a rehabilitation unit. Students were teamed with the qualified staff, each team responsible for 4-5 patients during each shift. Students and staff shared their skills and demonstrated to each other what their individual professional roles involved. Impact on staff was evaluated with focus groups and QPSNordic questionnaire before and after the first cohort of students. Impact on medical students was evaluated via a placement survey and the Readiness for Interprofessional Learning Survey (RIPLS) before and after the ward experience.
Summary of Results: Staff found that apprehension of a deterioration of patient care quality or an unmanageable workload were not realised, and mentioned multiple positive benefits of the placements. Students evaluated the placement positively and showed significant improvements in the three areas measured by the RIPLS (Teamwork, Professional Identity, Patient-Centred Care).
Conclusions: The ward experience provided opportunities for learning that had a positive impact on both staff and students. Students gained understanding that they felt would have a positive effect on working with other healthcare colleagues in multidisciplinary teams in the future.
Take-home Messages: Interprofessional working by students in a real ward environment, while challenging to fund and run, are becoming increasingly important for ensuring the quality of care delivered in teams and should be a requirement of healthcare education.

OTT-OF-4-2
Toward a common framework for IPE: lessons regarding assessment, evaluation, and implementation, from a multi-method study of IPE programs

Amy V Blue (*University of Florida, College of Public Health and Health Professions, Gainesville, FL, USA*), Benjamin J Chesluk (*American Board of Internal Medicine, Philadelphia, PA*), Lisa N Conforti (*American Board of Internal Medicine, , Philadelphia, PA*), Eric S Holmboe (American Board of Internal Medicine, *Philadelphia, PA*)

Background: The practice of interprofessional education (IPE) has recently grown in the United States. The publication of competencies from the Interprofessional Education Collaborative (IPEC) was a step forward to recognize the importance of collaboration among health professions. However, there remains substantial difficulty in implementation and considerable variability in evaluation of IPE.
Summary of Work: A multi-methods project involving literature review, interviews, and an expert meeting to 1) provide recommendations to organizations wishing to implement or augment IPE programs, 2) uncover gaps in current assessment/evaluation practices, and 3) describe next steps for the field.
Summary of Results: A diverse collection of methods and tools are used to assess/evaluate IPE learners/programs. Unlike some of our Canadian counterparts, most US institutions have not developed an explicit program-evaluation framework. There is a small and growing literature indicating evidence of the effectiveness of IPE, but it is based on studies in different settings, with different interventions and measured outcomes.
Conclusions: For the field to move ahead there is need for robust assessment/evaluation. Standardized use of common tools can help new programs avoid "re-inventing the wheel", and allow for local adaptations. Longitudinal assessment from diverse data streams is crucial.
Take-home Messages: To align with the demands of the clinical care system, there must be an increased focus on learner assessment and program evaluation of IPE. Assessment/evaluation can be built into evolving, or added to existing, IPE programs. Faculty can benefit from training in assessment of interprofessional skills and enhancement of their own collaborative skills.

OTT-OF-4-3
An interprofessional theme assessment instrument (KidSIM) translated and evaluated in HiFi Simulations in medical and nursing undergraduate programs

Martin Wohlin (*Uppsala University, Department of Medical Science, Uppsala, Sweden*), Jan Larsson (*Uppsala University, Uppsala*), Maria Magnusson (*Uppsala University, Uppsala*)

Background: Although the implementation of interprofessional education (IPE) is straightforward, there is a lack of interprofessional team assessment instruments on the undergraduate level. The purpose of this present study is to translate and pilot a newly developed instrument (the Kid SIM team performance scale) for interprofessional team performance in two undergraduate medical and nursing programs at Uppsala University in Sweden.

Summary of Work: Translation and validation of the instrument is planned for autumn 2013. A pilot study examining the psychometric and contextual characteristics of the Swedish KidSIM is scheduled for the winter 2013/2014. Analysis and interpretation will be finished by end of March 2014.

Conclusions: Interprofessional undergraduate team initiatives lack reliable and valid assessment instruments. KIDSim is the first translated to Swedish and used in high fidelity interprofessional team simulations on an undergraduate level in Sweden. This study contribute to the body of knowledge supporting the development of reliable and valid instruments for interprofessional team assessment on the undergraduate level.

Take-home Messages: Translation and validation of assessment instruments might be well worth the effort.

OTT-OF-4-4
Contextual activity sampling system impacts on clinical interprofessional learning

Hanna Lachmann (*Karolinska Institutet, Department of Clinical Science and Education, Södersjukhuset, Stockholm, Sweden)*, Bjöörn Fossum (*Karolinska Institutet, Sophiahemmet University*), Unn-Britt Johansson (*Karolinska Institutet, Sophiahemmet University*), Klas Karlgren (Karolinska Institutet, *Department of Learning, Informatics, Management and Ethics, Stockholm*), Sari Ponzer (*Karolinska Institutet, Department of Clinical Science and Education, Södersjukhuset, Stockholm*)

Background: The contextual activity sampling system (CASS) is a methodology designed for collecting data about experiences of on-going activities. Frequent distribution of questionnaires by using mobile phones provides detailed data and to repeatedly raise key issues about, e.g., clinical interprofessional learning. Students' engagement and their reflections about on-going activities are requested for encouragement during clinical practice.

Summary of Work: The intention with this work was to investigate if using CASS impacts on the students' academic emotions, readiness for interprofessional learning and their experiences of team collaboration during a clinical interprofessional training ward (IPTW) course. Three different data collection methods generating both quantitative and qualitative data were used. Thirty-three students from four different healthcare programs (medical, nurse, occupational therapy and physiotherapy) participated. The teams of students were randomised to an intervention group (exposed to CASS) or to a control group (not exposed to CASS).

Summary of Results: The result showed that the intervention group rated teamwork and collaboration significantly higher (p = 0.02) after conducted course than before the course. This difference was not seen among the control group students who experienced stress significantly higher (p = 0.03) compared to the intervention group.

Conclusions: CASS methodology seems to support the students' reflection, impact on their experiences of on-going activities, interprofessional collaboration and also to decrease their experience of stress during clinical practice.

Take-home Messages: CASS develops students' ability to reflect on ongoing clinical practice, on why activities are performed and needed as well as on the importance of interprofessional collaboration within health care.

OTT-OF-4-5
A functional framework for development and assessment of inter-professional collaboration as a professional competency

John Pugsley (*The Pharmacy Examining Board of Canada, Toronto, Canada)*, Carol O'Byrne (*The Pharmacy Examining Board of Canada, , Toronto*)

Background: Health care professionals are increasingly embracing collaborative practice to enhance patient outcomes and for effective use of resources.

There is a plethora of research and practical experience in inter-professional collaboration as a competency. Few models include well-defined, measurable indicators that can be used to guide observations, assessments and professional development in practice and educational settings, licensing examinations or other contexts.

Summary of Work: The Pharmacy Examining Board of Canada hosted an inter-professional workshop to clearly articulate what collaboration in health care involves, to guide and support learning, assessment and practice in all health care settings.

Based on a literature search on collaboration in the health professions a 'Functional Framework' was drafted using a functional analysis approach. Workshop participants then modified and/or added to each Framework component and identified indicators of good performance. The Knowledge and Skills required to perform each of the functional Units and Activities were identified.

Summary of Results: Workshop participants agreed that the Framework developed articulates the functions, activities and indicators of effective collaborative practice, from multiple perspectives.

Conclusions: The Framework is a relevant, useful tool to improve professional education and collaborative practice, particularly for patients who have complex, multi-dimensional health needs. The Framework can also be used in professional certification, ongoing professional development, professional regulation, workplace and public policy development.

Take-home Messages: Inter-professional collaboration is a complex competency that can be learned and assessed using the Functional Framework functions, activities and performance indicators

OTT-OF-5: MINI-CEX

OTT-OF-5-1

A comprehensive evaluation of a workplace-based assessment task

Catherine Haigh (*Monash University, School of Rural Health, Gippsland Regional Clinical School, Traralgon, Australia*), David Birks (*Monash University, School of Rural Health, Gippsland Regional Clinical School, Traralgon*)

Background: Teaching, learning and assessment in the clinical workplace has been a tradition in medical education. An adaptation of Norcini's mini CEX termed the mini case record (MCR) is used for this purpose during the first clinical year of the undergraduate MBBS at Monash University. The purpose of this study was to examine the utility of this task according to Van der Vleuten's pentagram.

Summary of Work: Classifications of patient presentations encountered by students completing this assessment task across one academic year were mapped to the curriculum content map to measure validity. MCR scores recorded for the cohorts of students from 2005 to 2013 were analysed to explore trends in performance and assessment for history taking, physical examination, clinical reasoning and professional behaviours as an index of reliability. Numbers of qualified clinicians willing and available to act as assessors were charted against student enrolments to evaluate feasibility. Written feedback provided to students was thematically analysed in terms of potential educational impact of the MCR experience. Finally, focus groups of students and assessors discussed the educational impact and acceptability of this assessment task.

Summary of Results: The task satisfies the five criteria outlined by Van Der Vleuten. The task was acceptable to both groups of stakeholders. Students' concerns related to the reliability of this high stakes assessment, and assessors acknowledged differences in perspectives according to clinical expertise.

Conclusions: Assessors and students should be supported appropriately to undertake workplace-based assessment.

Take-home Messages: Alignment between curriculum, clinical learning, and assessment tasks to progress competence is facilitated by regular feedback to stakeholders on performance.

OTT-OF-5-2

Development of a set of milestone-anchored direct observation tools (MADOT)

Marielena DiBartolo (*University of Calgary, Department of Pediatrics, Calgary, Canada*), Glenda Bendiak (*University of Calgary, Department of Pediatrics, Calgary*)

Background: Direct observation is essential to the assessment of clinical skills performance. Available assessment tools generally employ performance scales, with or without adjective anchors. Few utilize behavioural anchors, though these may reduce leniency/severity bias, minimize range restriction, and curtail halo/millstone effects. The University of Calgary Pediatric Respirology training program uses the mini-Clinical Evaluation Exercise (mini-CEX) for formative assessment during ambulatory clinic rotations. Concerns exist regarding the utility of the numerical ratings and the specificity of feedback generated. We sought to develop tools to optimize the effectiveness of assessment by direct observation.

Summary of Work: Clinical tasks integral to ambulatory clinic encounters were identified. Using the Pediatric Milestones (Accreditation Council for Graduate Medical Education, American Board of Pediatrics), key subcompetencies were mapped to each task, and their component milestones used as anchors for performance assessment. These were cross-referenced to the Royal College of Physicians and Surgeons of Canada CanMEDS roles.

Summary of Results: Four key clinical tasks were identified: history-taking, physical examination, case presentation, and counselling. Subcompetencies were chosen with attention to balancing comprehensiveness of assessment with clarity and ease of completion. Space was provided to record specific feedback, including examples of observed behaviour and suggestions for improvement.

Conclusions: A novel set of direct observation tools has been developed which use narrative, behavioural anchors . Uniquely, these tools draw from both Canadian and American standards, which may increase their generalizability. The next phase entails establishing the validity and reliability of these tools.

Take-home Messages: As competency-based assessment evolves, milestone-anchored tools will assist in accurate assessment and facilitation of formative feedback.

OTT-OF-5-3

Finding consensus within the noise: Shared idiosyncrasy helps explain rating variance

Andrea Gingerich (*University of Northern British Columbia (UBC Medicine), Northern Medical Program, Prince George, Canada*), Glenn Regehr (*University of British Columbia, Centre for Health Education Scholarship, Vancouver*), Kevin Eva (*University of British Columbia, Centre for Health Education Scholarship, Vancouver*)

Background: Rater-based assessments are often disparaged for their lack of inter-rater reliability. This study explored the extent to which rater variability can be accounted for by subsets of raters generating different inferences while assessing a trainee's performance.

Summary of Work: We asked volunteer physicians to assess then reflect on video-recorded patient encounters for 7 residents. For each video, raters completed a mini-CEX and provided open-text responses to questions regarding what they observed in the video as well as their subjective inferences about each resident. Subsequently, we gave naïve participants the set of raters' written inferences for a given video and asked them to sort the responses into piles based on

similarity. We assessed the consistency of the piles generated by sorters and the amount of variance that these piles explained in mini-CEX ratings.

Summary of Results: 48 physicians assessed at least 1 video (34 completed all 7). Variability in ratings within each video was high (range=4-8 on 9-point scale). Preliminary analysis of the piles generated by 14 sorters suggests that subsets of raters make inferences similar to each other but different from those made by other subsets, and that the majority of variance in ratings can be explained by these groupings of inferences.

Conclusions: Physicians who drew similar inferences about a given resident gave similar ratings. However, multiple, often incompatible inferences about the same resident performance were consistently seen across subsets of raters.

Take-home Messages: Inter-rater "error" variance may contain agreement amongst rater subsets, suggesting multiple "signals" within the "noise".

OTT-OF-5-4
Enhancing Mini-CEX reliability in anaesthesia training using an intuitive scoring system based on trainee independence

Jennifer Weller (*University of Auckland, Centre for Medical and Health Sciences Education, Auckland, New Zealand)*, Brian Jolly (*Newcastle University, Faculty of Medical and Health Sciences, Newcastle*), James Crossley (*University of Sheffield, AUME, Sheffield*)

Background: The perceived value of the Mini-CEX has been undermined by low reliability and limited ability to identify underperformance. We proposed that anaesthesia supervisors intuitively know the extent to which a trainee can be left to manage a case independently, and a scoring system based on trainee independence would improve the reliability of supervisors' scores, and the ability to identify underperforming trainees.

Summary of Work: We analysed data from 338 Mini-CEX assessments where supervisors scored trainees on: overall performance against expected level for stage of training; and level of independence with the case. In addition, the expected independence level for trainees at different stages for each of the 338 cases was calculated from the mean of three experienced supervisors' scores, who judged each case on the basis of surgical complexity and patient factors. This allowed calculation of the independence score corrected for case complexity for all 338 cases.

Summary of Results: With the independence scale, reliability co-efficients of 0.7 were attained with seven Mini-CEX assessments compared to over 50 assessments using the conventional scoring system. Furthermore, the corrected independence scoring system identified eight trainees performing well below expectations compared to none in the conventional system.

Conclusions: Using a scoring system anchored in trainee independence, Mini-CEX is a feasible and reliable option in anaesthesia specialist training programmes to make judgements on trainee progression, with potential to

identify underperforming trainees to facilitate timely remediation.

Take-home Messages: Mini-CEX is reliable if supervisors use scoring systems aligned with the way they make judgements on trainees, and enhanced by a rubric of expected independence with varying case complexity.

OTT-OF-5-5
Implementation of mini-CEX assessments in an outpatient clinic-oriented specialty

Joanne Alfieri (*McGill University, Radiation Oncology, Montreal, Canada)*, Khalil Sultanem (*McGill University, Radiation Oncology, Montreal*)

Background: The Mini Clinical Evaluation Exercise (mini-CEX) is an assessment of an observed clinical encounter with immediate developmental feedback. It is designed to provide constructive feedback on skills essential to the provision of good clinical care.

Summary of Work: A mini-CEX assessment tailored specifically to the radiation oncology resident was developed as a complimentary assessment to the already existing Rotation In-Training Assessment report (ITER). Attending staff were trained in this form of assessment using verbal and written instruction.
How and when is the mini-CEX used?
The resident and supervisor jointly agree on a clinic day, preferably toward the end of the rotation, when the mini-CEX will take place ensuring that there is adequate time for the encounter and subsequent feedback.

Summary of Results: A formal evaluation of the program is planned, however, initial feedback reveals that both residents and staff appreciate the constructive feedback that results from such an intervention.
It has also proven to be complementary to the ITER in that it provides very different information as to how the resident is performing clinically. Its formal and structured nature lends itself well to a busy outpatient clinic-oriented specialty such as radiation oncology.

Conclusions: A mini-CEX assessment program was developed and implemented for radiation oncology residents. Training attending staff to use this evaluation method is straightforward.

Take-home Messages: Mini-CEX assessments are feasible in radiation oncology and have been found to be valuable additions to the existing ITER evaluations.

OTT-OF-5-6
Development of a Patient Safety Mini-CEX Tool: Rating Interns' and Residents' Patient Safety Competence in Clinical and Simulated Patient Encounters

Tangerine Holt (*Australian-American Fulbright Commission, Executive Director, Canberra, Australia)*, Beverley Bird (*Monash University, Faculty of Medicine, Nursing & Health Sciences, Clayton*), Brian Jolly (*University of Newcastle, School of Medicine & Public Health, Newcastle*)

Background: Patient Safety has been addressed internationally and at national levels through regulated patient safety policies and healthcare facility standards. There is, however, little practical guidance for assessing individual clinician's awareness of patient safety in clinical encounters.

A Patient Safety (PSM) Tool, funded by the Health Department of Victoria, was initially developed to raise awareness of patient safety issues for International Medical Graduates newly employed in acute care settings in Australia. Its relevance as a formative assessment package for other trainees was recognized in the tool development phase.

Summary of Work: A review of patient safety and medical education literature guided the PSM development process through the identification, testing, refining and validation of its dimensions; a seven stage process over a two-year period. Four dimensions, 18 Knowledge Items and 128 Competency Indicators were ranked for Importance and Relevance by key stakeholders including trainees and clinicians. Testing of the resulting PSM on 150 participants utilised OSCEs, High/Low Simulation and Clinical (General Hospital) settings.

Summary of Results: Data analysis further refined the PSM's four Competency Areas to 13 Knowledge Items and 41 Competency Standards.

Managing Information, Clinical Judgement and Decision Making, Communication, and Infection Control were identified as areas of consistently poor or borderline performance.

Conclusions: The PSM is applicable to all levels of clinicians in a variety of settings and is adaptable to web-based applications, It identifies patient safety issues in practice and raises clinicians' awareness of patient safety in their individual practice.

Take-home Messages: The PSM is a valuable patient safety focussed formative assessment tool.

OTT-OF-6: ONLINE ASSESSMENT

OTT-OF-6-1
What a great piece of e-learning...Or is it?

Henry Fuller (*Leeds Teaching Hospitals NHS Trust, Medical Education, Leeds, UK)*, Rebecca Brown (*Nottingham University Hospitals NHS Trust, Paediatrics, Nottingham*), Claire Burnell-Hornby (*NHS, General Practice, Leeds*), Andrea Fox-Hiley (Leeds Teaching Hospitals NHS Trust*, Medical Education, Leeds*), Clare Donnellan (*Leeds Teaching Hospitals NHS Trust, Gastroenterology, Leeds*), Alison Cracknell (*Leeds Teaching Hospitals NHS Trust, Acute and Elderly Medicine, Leeds*)

Background: 21 deaths and 79 cases of harm have arisen in the UK over a 5-year period from the misplacement of nasogastric tubes. A steering group at Leeds Teaching Hospitals NHS Trust decided that e-learning would be an appropriate medium to deliver key

theoretical knowledge and bring about behavioural change.

Summary of Work: The e-learning uses a scaffolding approach in its instructional design, mixing theory with simulated practice. It concludes with a formal assessment quiz, which requires a minimum score of 90% to pass. It is hosted on the NHS National Learning Management System (NLMS), which – after 3 attempts - records a pass or fail grade.

Summary of Results: 185 of 196 members of clinical staff passed the final assessment suggesting a high level of theoretical knowledge across multiple disciplines. However, 128 members of staff got at least one question wrong which, – if isolated to a particular area of patient safety is potentially significant.

Conclusions: Because reporting at individual question level goes beyond the SCORM standards accepted by the NLMS, we are unable to identify any significant trends in understanding from the assessment data that could later focus design remediation. Therefore we should ask – are we collecting data to support and develop our learners – or are we just collecting data?

Take-home Messages: The NLMS is only suitable to deliver one-size-fits-all approach to training and assessment. NHS Trusts should consider adopting Virtual Learning Environments or Learning Management Systems purpose-built for education.

OTT-OF-6-2
The web-based computer system that enhances medical students' skill and knowledge in progress notes reporting

Wansa Paoin (*Faculty of Medicine, Thammasat University, Department of Surgery, Pathumthani, Thailand*)

Background: Since June 2013, Faculty of Medicine, Thammasat University implemented a web-based computer system that enables the medical students to write their progress notes and send the report via the internet. The teachers can comment and send feedback to the students online.

Summary of Work: Forty-eight medical students used the system to report their progress notes. The teachers can edit, comment in the notes and send feedback to the student.

Summary of Results: Most of the students used personal computers, but some used mobile devices to send the report. The student and teachers found some benefits from the system i.e. the ability to save their notes in digital format, sending report and getting teacher feedback on-line, the elimination of difficulty when reading the notes written by hand etc. After several rounds of feedback, every student showed improvement on the quality of writing their progress notes.

Conclusions: The web-based patient reporting system adds another way for medical student-teacher interaction. The usage of the system should be expanded to cover other type of reports in the future.

Take-home Messages: The web-based system could be used to increase other ways of medical students and

teachers interaction, eliminate the difficulty of reading handwritten notes, report the time of report sent accurately, help the users to access the notes using computers or mobile devices. The system helps the medical students improve their skill and quality of writing the progress notes.

OTT-OF-6-3
Upgrading usability in eLearning – Case ePortfolio for CPD in Finland

Topi Litmanen (*Pro Medico, Helsinki, Finland*), Kristiina Patja (*Pro Medico, Helsinki*)

Background: Poor usability in an eLearning solution can discourage active use and hinder learning. Usability testing is an important part of developing software and websites. Unfortunately, testing seems to be rather rare when developing eLearning tools. We will illustrate methods used in improving usability in designing and building an ePortfolio tool for practicing physicians in Finland.

Summary of Work: Taitoni.fi-platform is a personal web tool designed to manage and support physicians' CPD. Its' core function is to encourage physicians to set learning goals and to document their development in formal and informal learning activities. The platform was developed using User-Centered Design methods. Design process can be divided into three phases, accordingly to the phases of the project: analysis, design and evaluation. Analysis started with exploring similar products in medical and other fields. Prototype was then tested with users and the nearly finished ePortfolio was piloted for six months in real life setting in a large health centre. After the pilot, external specialist in usability evaluated the tool. Modifications were made web tool based on the results of each phase and method.

Summary of Results: While questionnaires are sometimes presented as an economical way to collect user experiences, they represent retrospective presentations of experiences. More accurate and useful data can be gathered with contextual data, such as thinking aloud methods and discussing with users.

Take-home Messages: Most of the important decisions concerning a web-tool are made before the actual coding, e.g. in need assessment and content analyses. Testing with authentic users in an early phase is essential for improving usability.

OTT-OF-6-4
Cloud-based mobile technology for assessing medical student competencies

Gary Ferenchick (*Michigan State University, Internal Medicine, East Lansing, USA*), David Solomon (*Michigan State University, Internal Medicine, East Lansing*)

Background: The assessment of medical students' competence in clinical settings is challenging. Included among the challenges is the lack of effective approaches for assuring that transparent standards are used when judging competence. We present a generalizable

initiative that improves and simplifies observation-based assessments in authentic clinical settings.

Summary of Work: We developed a cloud-based content management and assessment system named Just in Time Medicine (JIT). The primary goals of JIT, included: 1) establishing a self-service interface capable of creating criterion-based assessment tools displayable on mobile devices which are capable of capturing learners' progress related to potentially hundreds of clinical competencies; and 2) the creation of a permanent record of the trainees' observed skill.

Summary of Results: From July 2012 through June 2103 we used JIT to implement clinical evaluation exercises (CEXs) among 189 third-year internal medicine students in our geographically dispersed system.
A total of 1944 CEXs were completed by 322 observers (residents and attendings); 36.8% of the CEXs assessed communication skills, 20.0% assessed history-taking and 43.2% assessed physical exam skills. Forty-eight percent of the assessments took less than 10 minutes, as did 72.6% of feedback sessions. Observers were either satisfied or highly-satisfied with 95.0% of the evaluations. A significant correlation existed between student CEX performance and their end-of-clerkship overall grade, (correlation .105, p = .0001).

Conclusions: Use of mobile cloud-based technology easily captures discrete clinical performance data in work-place settings with high user satisfaction and with evidence of predictive validity.

Take-home Messages: Cloud-based assessment systems delivered via existing mobile technology represents a viable strategy in accomplishing work-place assessments.

OTT-OF-6-5
Developing and evaluating an e-Learning resource as an example of Assessment of learning

David Kandiah (*University of Western Australia, Perth, Australia*), Diana Jonas-Dwyer (*University of Western Australia, Perth*)

Background: The electronic resource "Rheumatology Case Scenarios" was created as a web-based module to maximise the interactive learning opportunities for medical students and residents.

Summary of Work: The web-based module was based on paper-based resource that had been used for 20 years to teach about common and important rheumatic conditions that all doctors will have to manage in their careers. In addition, this allowed the incorporation of assessment and supporting information for the students.

Summary of Results: 131 students completed the pre-test questionnaire. Their mean score for the module (maximum score 26) was 14.3 (SD 2.9, range 6-21). The student scores were normally distributed. Students who had completed all 8 cases were then given a follow-up questionnaire with the same questions but randomised differently. The students who completed this questionnaire had a mean score of 17.4 (SD 3.5, range 11-22) Paired samples t-test of the students who had completed the 2 questionnaires confirmed a statistically

significant improvement of the scores of each student (range 2-4 points, p<0.001). The answers for the questionnaire were only provided to the students after they had completed the second questionnaire.

Conclusions: To assess their retention of knowledge, students were also given the option of completing the same questionnaire 3 months later. The mean score rose to 23. This was statistically significant for these students when compared to their scores at both previous levels suggesting that there was retention of their knowledge.

Take-home Messages: Web-based resources can consolidate learning in the clinical arena. This process can be applied to all medical disciplines.

OTT-OF-7: APPROACHES TO UNDERGRADUATE SELECTION

OTT-OF-7-1
Admission interview and academic entry scores are associated with different aspects of student performance in an Australian Physiotherapy programme

Susan Edgar (*The University of Notre Dame Australia, Physiotherapy, Perth, Australia)*, Annette Mercer (*University of Western Australia, Medicine, Dentistry and Health Sciences, Perth*), Peter Hamer (*The University of Notre Dame Australia, Physiotherapy, Perth*)

Background: In Australia, the admission interview is currently utilised in selection processes for two undergraduate Physiotherapy programmes and is gaining popularity for selection into graduate-entry programmes. To date there has been no evidence of the predictive validity of the admission interview for entry into Australian Physiotherapy programmes. The Consensus statement from the 2010 Ottawa Conference reported a lack of evidence from selection processes for health-care professions outside medicine. The aim of this study was to determine any association between interview and academic scores on entry into Physiotherapy at The University of Notre Dame Australia and subsequent performance through course.

Summary of Work: A retrospective observational study was undertaken of two entry year groups (n= 141). Predictor variables included interview score, admission academic scores and demographic data. Outcome measures included individual student performance in all aspects of the four year programme. Bivariate correlations between each admission score and each outcome variable were calculated followed by linear regression modelling.

Summary of Results: Interview score demonstrated a significant relationship with performance in three of six clinical placements throughout the course; more than any other admission criteria. Further, analysis of attrition data revealed that students who failed to complete the course scored on average 2.65 points lower on the admission interview. No other admission

score was significantly associated with withdrawal from course.

Conclusions: Interview score for entry into Physiotherapy is associated with clinical performance in an Australian undergraduate programme.

Take-home Messages: There is a role for both academic and non-academic selection processes for entry into Physiotherapy.

OTT-OF-7-2
A new approach to selecting medical students – National University of Singapore, School of Medicine Experience

Dujeepa D. Samarasekera (*Yong Loo Lin School of Medicine, National University of Singapore, Medical Education Unit, Singapore*), Sin Chuan Tay (*Yong Loo Lin School of Medicine, National University of Singapore, Dean's Office, Singapore*), Shing Chuan Hooi (*Yong Loo Lin School of Medicine, National University of Singapore, Dean's Office, Singapore*)

Background: Evaluating medical school applicants' suitability is always a formidable task for any medical school.

Summary of Work: School of Medicine, National University of Singapore changed its admission process from having a semi-structured interview to a score based on a Focussed Skills Assessment (FSA) and a Situational Judgement Test (SJT). From 2013, FSA and SJT are used to select the student cohort to the faculty. FSA evaluates 8 domain areas namely empathy, communications,integrity, general knowledge, resilience and their personal portfolio. SJT is a 23-question MCQ type test which all shortlisted candidates sit.

Summary of Results: We will discuss why the FSA and SJT tests were developed, the station structure and the reasons behind replacing the old format of semi-structured interview. The post-hoc analysis results of FSA:

1.Cronbach's alpha data shows that there is little/no consistency among the eight FSA stations. Every station appears to be independent from each other.

2.Weak correlation was found between the FSA and SJT. This concludes that both tools measure different attributes

The new format of FSA has also reduced the number of examiners from 196 to 188 and the time duration by half (from 16 days to 7.5) compared to the previous semi-structured interviews.In addition, the majority of the candidates who underwent FSA were happy with the format (92.4% selected the "agree" and "strongly agree").

The majority of FSA interviewers were also had positive perceptions regarding their experience as 82% of them rated 6 or above in a rating scale of 8.

Conclusions: FSA and SJT analysis shows that they are effective in reducing the school's resources and focus on the individual domain performances. The new FSA was perceived well by both the faculty and candidates.

Take-home Messages: The necessity of CQI in admission process to employ best evidence techniques

in selecting candidates. Ways to contextualize formats for higher returns on investment.

OTT-OF-7-3
Selecting MD/PhD Applicants Using a Modified Personal Interview (MPI)

Lindsey Fechtig (*University of Toronto, Toronto, Ontario, Canada)*, Kulamakanam Kulasegaram (*University of Toronto, Toronto, Ontario*), Nicole Woods (*University of Toronto, Toronto, Ontario*), Norman Rosenblum (University of Toronto*, Toronto, Ontario*), Mark Hanson (*University of Toronto, Toronto, Ontario*)

Background: MD/PhD student selection requires assessment of characteristics related to research and the physician scientist career (psc). Personal interviews are commonly used to select students though single traditional personal interview is limited by low reliability. Recent work suggests that a modified personal interview (MPI) has utility for selection tasks focusing a few specific competencies suggesting possible utility for MD/PhD selection. The MPI shows acceptable reliability when using 4 brief, semi-structured interviews. We investigated the reliability, acceptability, and construct validity of the MPI in the context of MD/PhD selection.

Summary of Work: Applicants to the University of Toronto MD/PhD program were evaluated via the MPI with 4 brief semi-structured personal interviews each with a single rater. Interviewers were given pre-selected questions but were encouraged to be flexible in interviewing. Each interview was mapped to attributes valued for MD/PhD training: research experience, self-reflection, motivation, and the physician scientist perspective. Three attributes (maturity, curiosity, analysis of psc) were mapped across all interviews. All attribute-items were evaluated via 7-point Likert scales. Results were analyzed for reliability using generalizability analyzes and factor analysis for construct validation.

Summary of Results: Forty-two applicants completed the MPI. Inter-interview reliability averaged across 4 interviews was 0.76. Inter-item reliability within interview was 0.94. Factor analysis showed all items loaded on a single factor with maturity items having highest loadings. A majority of applicants and interviewers found the MPI to be acceptable.

Conclusions: The MPI has utility for selection of MD/PhD students as it shows acceptable reliability, acceptability and preliminary validation. Future research will determine the predictive validity of this process.

OTT-OF-7-4
Portfolio in a program of selection for a rurally focussed medical school

Ian Wilson (*University of Wollongong, Graduate School of medicine, Wollongong, Australia)*, Lyndal Parker-Newlyn (*University of Wollongong, Graduate School of Medicine, Wollongong*)

Background: Very little has been written concerning the use of portfolios in the selection of medical students. Those papers that have been written have not demonstrated the psychometrics of the process. This presentation will describe the development of and psychometric analysis of the data generated.

Summary of Work: The University of Wollongong commenced taking medical students in 2007. The focus of the program was to be regional, rural and remote medical practice. Selection was based on GPA, GAMSAT, MMI and portfolio. The portfolio is computer based and was designed to measure personal characteristics and rural background. Students completed the portfolio on line and submitted it with their application. Portfolio assessors were trained and each portfolio scored using a structured template. Two scores result, one relating to personal attributes and the other to rural background.

Summary of Results: The psychometrics of the portfolio section were good. The personal attributes section had a Cronbach's alpha of 0.648 which rose to 0.685 when a poorly performing section was removed. The rural section was even better with an alpha of 0.877 rising to 0.914 with a poorly performing section removed. The portfolio scores correlated negatively with the GAMSAT and GPA, while poorly with the MMI. Preliminary predictive data will also be presented.

Conclusions: A selection portfolio can provide additional information to that obtained by more standard methods.

Take-home Messages: A well-constructed portfolio can provide addition valuable information for selection decisions.

OTT-OF-7-5
The BioMedical Admissions Test (BMAT) for medical student selection: overview of research evidence

Joanne Emery (*Cambridge Assessment, Research and Validation, Cambridge, UK)*, Sarah McElwee (*Cambridge Assessment, Research and Validation, Cambridge*)

Background: The BMAT has been in use since 2003 and forms a part of the selection processes of a number of UK institutions. The test consists of three sections measuring scientific aptitude, scientific understanding and written communication. It is designed to assess readiness for demanding, science-based study and not fitness to practise. Institutions use BMAT scores in conjunction with other selection criteria such as national examination results and interview performance, although some institutions use the test as a hurdle to the interview stage. This presentation gives a brief overview of the validation work carried out for the BMAT.

Summary of Work: Early validation work focussed mainly on the predictive validity of the test. The relationship between BMAT scores and test preparation has also been investigated. Group differences in BMAT scores continues to be an important research focus.

Summary of Results: The BMAT predicts early medicine course performance, with the scientific understanding section generally showing the strongest relationship

with both examination success and examination failure. The effect of candidate preparation on test scores is difficult to establish with correlational designs but no relationship has been found between BMAT scores and the preparation help that candidates receive from their schools. Small gender differences in BMAT performance are evident but these do not appear to be a result of test bias.

Conclusions: Research evidence supports the predictive validity and fairness of the test for student selection.

Take-home Messages: The BMAT appears to be a useful and valid addition to the medical student selection process.

OTT-OF-7-6
Lottery as selection procedure: how (un)fair is it?

Janke Cohen-Schotanus (*University Medical Center Groningen, Groningen, Netherlands)*, Nienke Schripsema (*University Medical Center Groningen, Groningen*), Axel Themmen (*Erasmus Medical Center, Rotterdam*), Suzanne Lucieer (Erasmus Medical Center*, Rotterdam*), Anouk Wouters (*Free University Medical Center, Amsterdam*), Gerda Croiset (*Free University Medical Center, Amsterdam*)

Background: In the debate about the fairness of selection procedures, lottery is hardly mentioned as an alternative. "We want only excellent students", is the most heard argument against lottery. However, does selection of students really improve the cohorts of students in terms of grades and study progress?

Summary of Work: In the Netherlands, due to national policy, students enter medical school through; (1) direct access based on extreme high pre-university GPA, (2) (voluntary) selection, or (3) national lottery. This situation offers the unique opportunity to compare selection to lottery because, after rejection, students can be admitted through lottery. At different medical schools the effects are studied and we compared the outcomes.

Summary of Results: All studies were single-site studies. At one medical school selected students had a lower risk of drop out and higher clerkship grades. Another school found no such differences, however, their selected students scored significantly higher on motivation. A third medical school also found no differences between selected and rejected students, however, lottery students who did not participate in the selection procedure had significantly lower test grades and a higher risk of drop out.

Conclusions: The results of the Dutch studies were inconclusive. However, selection might be a barrier too high for less motivated applicants.

Take-home Messages: From the viewpoint of fairness, one might argue that lottery seems to be a defensible selection procedure after the 'bad apples' are rejected.

OTT-OF-8: ASSESSMENT OF STUDENTS AND PBL

OTT-OF-8-1
Monitoring and evaluation of the problem based learning process assessment scores for minimizing bias and errors

Shital Bhandary (*Patan Academy of Health Sciences, Public Health and Medical Education, Lalitpur, Nepal)*, Satish Raj Ghimire (*Patan Academy of Health Sciences, Anatomy, Lalitpur*), Ira Shrestha (*Patan Academy of Health Sciences, Physiology, Lalitpur*)

Background: School of Medicine, Patan Academy Health Sciences uses a hybrid PBL curriculum to develop and assess the soft skills among the students using a validated Tutor Assessment of Student (TAS) tool. These assessments are summative and thus assessment scores should be monitored and evaluated continuously.

Summary of Work: TAS tool score of 58 students divided into 7 PBL groups in an integrated block was analyzed using Kruskal-Wallis H test and Dunn post-hoc test. A Multi-Facet Rasch Model based on all the responses is then used to further analyze the difference among the tutors, items and students.

Summary of Results: The Kruskall-Wallis test revealed significant different scores among 7 tutors. Dunn post-hoc test revealed the pair of tutors that had significantly different scores than others. Multi-Facet Rasch Model using 1856 responses revealed the generous and strict tutors along with easy and difficult items and performing and non-performing students. Regular tutor trainings were conducted thereafter.

Conclusions: Tutors assessment of students on soft skills is found to be significantly different among different PBL groups. This is worrisome as they were trained extensively a priori on the importance of inter-rater agreement and student could pass or fail based on them. Thus, regular tutor training on the importance of inter-rater agreement should be conducted to minimize the average scores among the tutors.

Take-home Messages: PBL process assessments when used for summative purpose need a continuous monitoring and evaluation, which in turn, reveals difference between tutors, items and student and helps to devise further plans to minimize the bias and errors in the scoring.

OTT-OF-8-2
Assessment of PBL performance using both checklist and global ratings: Comparisons between two blocks across two classes

Ming Lee (*David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, USA)*, Paul Wimmers (*David Geffen School of Medicine at University of California, Los Angeles, Los Angeles*)

Background: Although numerous studies have been conducted to compare the effects of PBL and traditional teaching methods, only a few have reported on the assessment of PBL performance. How tutors assess performance in a PBL setting is still unclear.

Summary of Work: An assessment tool with four domains (Problem Solving, Use of Information, Group Process and Professionalism) and 17 checklist items was used to assess medical students' PBL performance. Block 1 and block 6 PBL ratings of two classes were used in the study. Pearson correlation coefficients were calculated to examine their relationships. Multiple regression analyses were conducted to examine the predictive value of checklist items on associated domains, and of domains on the overall rating.

Summary of Results: Correlations between domain and checklist ratings within the same block were higher than those between blocks. Each domain was significantly predicted by associated checklist items as a whole and individually, with few exceptions. All domains except Professionalism significantly predicted the overall rating. Findings of the two blocks bore significant similarity.

Conclusions: Low correlations between blocks may be explained by students' improved performance over time and the effect of being in a different group and working on different cases. High predictive value of checklist items on domain ratings indicates tutors' rating on one may influence the other. Assessed professional behaviors of students (respect, punctuation) did not affect tutors' overall rating.

Take-home Messages: Problem solving and group interaction are considered more important PBL performance than professional behaviors. A multi-domain, multi-method approach provides a comprehensive assessment of PBL performance.

OTT-OF-8-3
Development And Pilot Of A Novel, Descriptive Student Assessment For The PBL Component Of A Therapeutics Course

Suzanne Donnelly (*University College Dublin, School of Medicine and Medical Sciences, Dublin, Ireland)*, Paula Heaphy (*University College Dublin, School of Medicine and Medical Sciences, Dublin*), Annette Liston (*University College Dublin, School of Medicine and Medical Sciences, Dublin*), Eoghan McCarthy (University College Dublin*, School of Medicine and Medical Sciences, Dublin*), Patrick Murray (*University College Dublin, School of Medicine and Medical Sciences, Dublin*)

Background: Summative assessment of participation in the PBL component of a therapeutics course was initiated at the students' request. The key consideration was that assessment would be valid, feasible and acceptable and promote learning aligned to Barrows' four educational objectives. Initial pilot of a 3 domain numerical instrument derived from the 44-item validated Tutotest showed that while scores correlated strongly with those on Tutotest, its validity and educational impact were poor and it was decided to develop a novel instrument with high educational impact.

Summary of Work: Using Pangaro's descriptive RIME synthesis for in-training clerkship as a model, characteristics of four levels of student participation in PBL were debated by experienced PBL tutors until consensus was reached. Detailed descriptors were provided for each, characterized as: Observer; Participant; Hypothesier and Synthesiser in the discussion phase and Bare Reporter; Reliable Reporter; Interpreter and Educator for the report phase. The OPHS: BRIE instrument was used alongside the original numerical instrument for 800+ PBL assessments and the psychometrics, feasibility, acceptability and educational impact of each were compared.

Summary of Results: We report that psychometrics for the RIME based assessment were similar to those for the numerical score. In terms of stakeholder acceptability, feasibility and educational impact, the descriptive assessment scored better. Furthermore, the model enabled faculty to identify and address specific weaknesses in student participation and to document meaningful progression of student participation in PBL over the semester.

Conclusions: Participation in PBL can be reliably, validly and feasibly assessed using a novel descriptive score derived from Pangaro's RIME synthesis which additionally provides rich feedback to faculty and students.

OTT-OF-9: ASSESSMENT OF MILESTONES IN TRAINING

OTT-OF-9-2
Testing Construct Validity of the McMaster Modular Assessment Program (McMAP)

Meghan McConnell (*McMaster University, Hamilton, Canada)*, Teresa Chan (*McMaster University, Hamilton*), Jonathan Sherbino (*McMaster University, Hamilton*)

Background: From October 2012-June 2013, we conducted the pilot phase of a workplace-based, milestone-anchored assessment system. McMAP contains 52 unique work-based assessment (WBA) items that assess performance using a variety of methods, including direct observation tools (checklists, Mini-CEX), multi-source feedback (MSF), and other observed activities.

Summary of Work: During our pilot year we gathered encounter cards (EC) for 16 learners (7 Postgraduate Year 1 [PGY1], 9 PGY2), resulting in 637 assessments from 64 raters across all CanMEDS roles. Each EC contained a single task mapped to a CanMEDS framework key competency (e.g. Communicator – Charting Checklist/Rating Scale), plus a behaviourally-anchored global rating scale (GRS) for the residents' overall daily functioning.

Summary of Results: Each WBA item was mapped to a specific CanMEDS Role a priori. The relationship between CanMEDS Roles and daily GRS was examined using descriptive statistics and Spearman's rho. The GRS ratings were significantly different from the Task EC ratings($p<0.001$). The GRS was associated with the Medical Expert Role (rho=0.747,p=0.03,n=15) and Communicator Role (rho=0.756,p=0.001,n=15). Medical Expert was associated with 3 Intrinsic Roles: Communicator (rho=0.576,p=0.03,n=14), Manager

(rho=0.761, p=0.02, n=9), and Professional (rho=0.639,p=0.03,n=11). Scholar and Advocate were associated to each other (rho=0.841,p=0.04, =n=6), but this may be due to low performing residents not completing enough Scholar or Advocate ECs (n=0 low performers, n=5 for high performers).

Conclusions: McMAP is a WBA system that allows for the comprehensive assessment of the CanMEDS Roles. Single daily ratings of global functioning may be insufficient. Instead a mix of tools that assess various competencies are necessary to assess the global competency of learners.

Take-home Messages: Milestone-based education requires a mix of tools to effectively assess professional and clinical competencies

OTT-OF-9-3
The McMaster Modular Assessment Program (McMAP) improves quality of in-training evaluation reports via aggregated work-based assessments and guided narrative global assessment

Teresa Chan (*McMaster University, Division of Emergency Medicine, Department of Medicine, Hamilton, Canada)*, Jonathan Sherbino (*McMaster University, Division of Emergency Medicine, Department of Medicine, Hamilton)*

Background: The McMAP system consists of work-based assessment (WBA) tools that are blueprinted to the CanMEDS framework. During clinical training WBAs are completed by various faculty members. The WBAs include checklists, behaviourally-anchored global scores and written comments. At the completion of a block of training the WBAs are compiled in a report for a supervising faculty member, who completes a guided narrative global assessment of the resident's performance (McMAP in-training evaluation record, ITER).

Summary of Work: 41 McMAP ITERs written by 25 faculty members were available. We compared McMAP ITERs to older ITERs generated in the same program. A random cohort of 25 McMAP ITERs and a matched historical cohort of 25 ITERs from PGY1 & 2 residents from the same program two years prior were randomly selected. All ITERs were redacted. All 50 ITERs (both pre- and post-McMAP) were scored by two investigators(TC,JS) using the validated Completed Clinical Evaluation Record Rating (CCERR).

Summary of Results: The level of agreement between the two raters on the CCERR scale was high (Cronbach's alpha=0.916 (df=49, p<0.001). There was a significant increase (p<0.001) in median CCERR scores after McMAP was introduced, resulting in a 15.7 point increase (out of a total 45 point CCERR scale). The median CCERR score of the McMAP ITERs was 27.5 (IQR 20.5-23.5), whereas the historical cohort was 13.8 (IQR 11.3-15.8).

Conclusions: Our McMAP ITER system generated significantly improved quality of assessments in our ITERs, as judged by the CCERR scale.

Take-home Messages: WBAs (mapped to a competency framework) and guided narrative global assessment greatly improve ITER quality.

OTT-OF-9-4
Assessment of level 1 EM milestones in incoming interns

Laura Hopson (*University of Michigan, Emergency Medicine, Ann Arbor, USA)*, Samantha Hauff (*University of Michigan, Emergency Medicine, Ann Arbor*), Marcia Perry (*University of Michigan, Emergency Medicine, Ann Arbor*), Eve Losman (University of Michigan*, Emergency Medicine, Ann Arbor*), Monica Lypson (*University of Michigan, Internal Medicine, Ann Arbor*), Sally Santen (*University of Michigan, Emergency Medicine, Ann Arbor*)

Background: With the ACGME's Next Accreditation System (NAS), residencies are required to report residents' progress through the milestones. The milestones include 5 progressively advancing skill levels, with Level 1 defining the skill level of a medical student graduate, and Level 5, an attending. The ACGME stresses that multiple forms of assessment should be used to ensure capture of the multi-faceted competencies.

Summary of Work: The purpose of this study was to pilot multisource assessments of Level 1 milestones for incoming interns during July. Assessment was done by 3 separate instruments. 1. The Post-graduate Orientation Assessment (POA) administered by the GME Office. This multi-station exam covers 9 of the EM milestones. 2. Direct observation of patient encounters using a form developed from the EM milestones. 3. The global monthly evaluation, updated to reflect EM milestones.

Summary of Results: In the POA, interns were competent in 40-100% of milestones assessed. Overall competency was 70% with low scores noted in Aseptic Technique (PC13) and Hand-off (ICS2). In excess of 80% demonstrated competency in EBM (PLBI); Patient Safety (SBP1); Values Interpretation (PC3); Consent (PC9); Pain Assessment (PC11); Image Interpretation (PC3); and Geriatric Assessment (PC2). On direct observation, all but one intern met all Level 1 milestones. On global monthly evaluations, all interns met Level 1 milestones.

Conclusions: A multisource assessment of EM 1 milestones is feasible and useful to determine Level 1 milestones for incoming interns.

Take-home Messages: A structured assessment such as the POA may be useful in detecting areas of weakness that may be overlooked or difficult to assess on direct observation.

OTT-OF-9-5
Alignment of undergraduate and postgraduate accreditation standards in Canada

Susan Andrew (*University of Alberta, Edmonton, Canada)*, Anna Oswald (*University of Alberta, Edmonton*), Kent Stobart (*University of Alberta, Edmonton*)

Background: The Future of Medical Education in Canada (FMEC) and the Carnegie Foundation's Educating Physicians: A Call for Reform of Medical School and Residency both call for a transformation of medical education. FMEC recommends that "Recognizing that accreditation is a powerful lever, Canadian medical leaders must review and realign existing standards of the Committee on Accreditation of Canadian Medical Schools (CACMS) and the Liaison Committee on Medical Education (LCME) and develop new ones, as necessary, to respond to the recommendations in this report. This may involve the alignment of undergraduate and postgraduate accreditation standards". The Carnegie Foundation's Educating Physicians seven recommendations include a similar call for alignment "that accrediting, certifying and licensing bodies together develop a coherent framework for the continuum of medical education and establish effective mechanisms to coordinate standards and resolve jurisdictional conflicts."

Summary of Work: As a first step in the Canadian context, we assessed how well LCME/CACMS undergraduate accreditation standards align with the Royal College of Physician and Surgeons of Canada (RCPSC) and the College of Family Physicians of Canada postgraduate training standards.

Summary of Results: The 132 LCME/CACMS Medical School accreditation standards were aligned with the 155 post-graduate standards from the RCPSC accreditation General Standards (A and B). This comparative evaluation demonstrates that the standards do not align closely. Gaps, redundancies and key differences are highlighted.

Conclusions: These results are the first step in understanding how accreditation needs to be adapted and re-aligned across the education continuum to provide consistent and coordinated training.

Take-home Messages: Accreditation needs to be coordinated across the continuum.

POSTERS

OTT-PA7: PORTFOLIOS

OTT-PA7-02
Assessment of an Educational Portfolio Program for Medical Students

Jack Kopechek (*The Ohio State University College of Medicine, Pediatrics, Columbus, USA*), Cynthia Ledford (*The Ohio State University College of Medicine, Internal Medicine, Columbus*), Rollin Nagel (*The Ohio State University College of Medicine, Medical Administration, Columbus*), Curt Walker (The Ohio State University College of Medicine*, Internal Medicine, Columbus*), Carmine Alexander Grieco (*The Ohio State University College of Medicine, Radiology, Columbus*)

Background: Our program was designed to help students develop reflective, self-directed learning skills. The program consists of a personal on-line portfolio for each student, a series of reflective assignments related to the core educational objectives (CEOs) and a relationship with a trained faculty coach who facilitates student reflective learning. Detailed performance reports serve as a starting point for student reflection on their progress towards meeting the CEOs.

Summary of Work: To evaluate our program, we implemented several surveys including coach and student evaluations of the program, and coach and student evaluations of each other. Surveys items rated program and coach effectiveness in promoting student reflective learning and in addressing the pertinent CEOs.

Summary of Results: Coaches and students generally agreed on the strengths and weaknesses of the program, although coaches tended to rate the program higher than the students. Over 60 % of students agreed or strongly agreed that program goals were met except for helping them identify the best way to learn and how to maintain on appropriate balance or personal and professional commitments. Coaches and students placed high value on the coach/student relationship and the role of performance reports. Students perceived reflective writing to be of less value

Conclusions: Surveying students and coaches in an educational portfolio program can provide useful information for program improvement.

Take-home Messages: E-Portfolio programs should survey their students and coaches annually to assess progress towards program objectives.

OTT-PA7-03
Portfolio Assessment of a Four Year Skilled Clinician Program

Susan Love (*Dalhousie University, Medicine, Halifax, Canada)*, Marie Matte (*Dalhousie University, Medicine, Halifax*)

Background: The acquisition of knowledge, skills, and attitudes relating to clinical skills performance is cumulative over time and requires repetition for mastery. This concept is consistent with the way student learning occurs, and more specifically for medicine, it is consistent with the maintenance of clinical competence in the context of clinical practice.

Summary of Work: The four-year skilled clinician program is designed to provide students with a variety of formative and summative experiential and explanatory (knowledge) learning assessments across the four year continuum of undergraduate medical education. Students will be required to build their own "learning portfolio" as evidence of their achievement of program learning objectives. Central to the learning portfolio will be the practice of self-reflection as a means for all students to enhance questioning of their performance, thus moving them along the learning continuum from habit to reflection.

Summary of Results: Through the use of a variety of assessment methodologies, students will begin to develop insights into the connections and interactions

between their learning experiences, and more importantly, through the practice of self-reflection, begin to understand how these experiences contribute to their development as physicians.

Conclusions: The Program began in September 2013 for the Med 1-3 classes and will be continually monitored and evaluated for effectiveness.

Take-home Messages: This learning portfolio can serve as an important means of monitoring ongoing professional behaviours, examination of self, examination of gaps in learning, and provision of appropriate and timely remediation exercises.

OTT-PA7-04
The impact of a Portfolio Assessment and Support System (PASS) on faculty views of assessment. Are we changing culture?

Jane Griffiths (*Queen's University, Family Medicine, Kingston, Canada)*, Elaine van Melle (*Queen's University, Family Medicine, Kingston*), Nancy Craig (*Queen's University, Kingston*)

Background: Competency Based Medical Education (CBME), represents a paradigm shift from the traditional time and process based educational training.
To meet the requirements of the CBME approach, the Department of Family Medicine has implemented a robust assessment system, the Portfolio Assessment and Support System (PASS).
The Academic Advisor (AA) role is considered central to the success of the PASS. This new, expanded role of the AA from the previous mentorship model represents a marked change, and may present challenges to existing expertise and attitudes regarding assessment.
The purpose of this study is to determine whether PASS has facilitated a paradigm shift in how AAs view and approach assessment.

Summary of Work: Semi-structured interviews were held with 12 Queen's University Family Medicine AAs. The data was analyzed using a thematic analysis.

Summary of Results: Culture is shifting; AAs do understand their role and the tenets of CBME.
They incorporate multiple measures in assessment and are able to proactively identify the resident in difficulty. However, they are still using "internal yardsticks" to assess and see the need for benchmarks. They see the potential that PASS could give them the "proof" to make competency declarations.

Conclusions: Over a 4 year period we have seen a tangible shift in assessment culture. Faculty think about assessment in a more rigorous way. This study has allowed us to better identify faculty development needs and support the change.

Take-home Messages: This evolving paradigm shift requires careful management of change.

OTT-PA7-05
Evaluation Based on Community Experience Report by Medical Students

Srimuang Palangrit (*Faculty of Medicine, Thammasat University, Community Medicine and Family Medicine, Pathumthani, Thailand)*

Background: Community experience report is the reflection of medical students' opinions and materials learnt by them. Faculty of Medicine at Thammasat University has organised for medical students to study rural communities by staying overnight in the communities for nine and 14 nights during their sophomore and junior years, respectively. Each medical student's performance is evaluated by his/her community experience report.

Summary of Work: Information from community experience reports by 11 medical students was gathered. The students were those in the group facilitated during their sophomore year and followed during their junior year by the researcher. There were 22 reports altogether. The information was analysed in terms of three items, namely overall activities, materials learnt and analytical thinking; all of which were based on Rubric Score on the scale of one to five. Report content analysis was also included.

Summary of Results: Mean scores for the overall activities, materials learnt and analytical thinking were 4.40, 4.35 and 4.05, respectively. The medical students wrote that it had been impressive and inspiring to become good medics; that they had been exposed to the livelihood and got to practice in real situations; that they had had great impression, and it had been good attitude fostering; that communication was a vital aspect to keep in mind; and that patients were to be treated as a whole, not as diseases.

Conclusions: Community experience report is an evaluation tool effective in measuring medical students' direct experience, perception, attitudes, analytical thinking and other aspects.

Take-home Messages: Rubric Score should be standardised further.

OTT-PA7-06
Use of Placement Supervision Groups in formative assessments

Heidi Davina Breed (*Health Education East Midlands, Educator and School Development, Nottingham, UK)*, Bridget Langham (*Health Education East Midlands, Postgraduate Medical Management Team, Nottingham*)

Background: The UK Foundation Programme introduced the concept of multi-professional Placement Supervision Groups (PSG) in 2012 to facilitate the named Clinical Supervisor in the completion of their trainee's End of Placement Report. We were interested in the utilisation of the PSG across our region and whether there were any emergent patterns.

Summary of Work: We carried out an audit using ePortfolio data on a random 10% sample of our 2012-13 trainee cohort. The following data items were extracted from the Clinical Supervisor reports in ePortfolio:
• Local Education Provider (LEP)
• Specialty Grouping
• Utilisation of PSG

Summary of Results: 279 reports were analysed, in 42% of cases a PSG had been utilised. We then analysed the data at LEP and specialty grouping level which revealed that a trainee's clinical placement bore no influence over whether a PSG was used, except in Psychiatry where the PSG was used in 81% of instances.
When the results were analysed by LEP a more significant trend emerged with a PSG usage range of 22.2% to 100%. This directly correlated with LEP performance data from our quality management visits.
Conclusions: The use of PSG in a LEP is one indicator of the culture and quality of that education environment. Data from ePortfolio could be used to develop metrics to further inform our quality management processes, with the ultimate aim of ensuring the education environment is fit for purpose.
Take-home Messages: Application of assessment methodology should be considered when evaluating the culture and quality of a postgraduate medical education environment

OTT-PA7-07
Clinical portfolio (log book) for undergraduate medical students

Muhammad Abid Bashir (*Nawaz Sharif Medical College, University of Gujrat, Surgery, Medical Education, Gujrat, Pakistan*)

Background: Undergraduate medical education in Pakistan is passing through a transitional phase with a predominantly traditional system of education in most of the medical institutions based on a traditional subject based curriculum. In surgery and other clinical subjects, there is end of clinical rotation ward test (in 3rd, 4th & 5th year) and summative assessment as the exit exam after fifth year (one single exam in three years) comprising of short & long cases and OSPE without any emphasis on achievement of specific learning outcomes or continuous monitoring during clinical training.
Summary of Work: The clinical portfolio (structured log book) is developed on the basis of a competency based curriculum with clearly laid down learning outcomes and provision of continuous monitoring. It is developed to clearly communicate to the students what is expected from them at the end of clinical rotation, monitor the progress, identify learning gaps and rectify deficiencies.
Summary of Results: It is the very first effort to introduce structured training and monitoring at undergraduate level within the framework of traditional educational system. Various components of the logbook will be presented along with barriers and difficulties as well as response of the students to the new development.

OTT-PA7-08
The comparison of the depth of reflection: Dental trainees reflect more critically by failure experience

Taiji Obayashi (*Hiroshima University, Graduate School of Biomedical and Health Sciences, Hiroshima, Japan*),

Takayuki Oto (*Hiroshima University, Graduate School of Biomedical and Health Sciences, Hiroshima*), Yukiko Nagatani (*Hiroshima University, Graduate School of Biomedical and Health Sciences, Hiroshima*), Kayo Kajitani (Hiroshima University Hospital*, Dental Hygiene Section, Clinical Support Department, Hiroshima*), Tetsuji Ogawa (*Hiroshima University Hospital, Department of Advanced General Dentistry, Hiroshima*)

Background: Dental trainees undergo various clinical experiences during an internship. They develop by experiencing both success and failure. When they look back on an event, it is not apparent which experience becomes a more critical reflection. We analyzed their portfolio qualitatively.
Summary of Work: For Hiroshima University Hospital dental trainees from 2010 to 2012, we asked them to describe the event that they looked back on during training and remembered most after completion of one-year of training. We coded these texts in the qualitative analysis method of SCAT (Steps for Coding and Theorization, Otani, 2008). Then, we classified them into a positive image and a negative image. In addition, we evaluated the depth of reflection in a pragmatic approach to categorising reflective material (Sandars, 2009) and Reflective ability rubric and user guide (O'Sullivan et al. 2010), and compared these results. We performed the evaluation with two researchers and confirmed a good rate of agreement.
Summary of Results: The mean of the depth of reflection (2010-2012) became 2.8, 2.75, 2.58 in the positive image, 1.63, 2.09, 1.97 in the negative image at each year. As for the dental trainees who described the negative image in each year, reflection was significantly deeper than he/she who described the positive image (P<0.05).
Conclusions: When the dental trainees experience failure in the Hiroshima university hospital, he/she reflects more critically.
Take-home Messages: We think that the dental trainees develop more by experiencing failure than success.

OTT-PA7-09
The Finnish mPortfolio Taitoni.fi for continuous professional development

Kristiina Patja (*Pro Medico, Helsinki, Finland*)

Background: Medical professional requires multiple skills and there is a growing need to document and assess professional's development and to acknowledge work environment as a major learning environment.
Summary of Work: Taitoni.fi is a personal web tool designed to manage physician CPD. Its core function is to encourage physicians to set learning goals and to document their development in formal and informal learning activities. It applies CanMEDS-roles in setting targets and reporting learning activities. The tool is accompanied and linked with a national CME-calendar, certificate databank and Curriculum Vitae-tool. Feedback for physician is provided in numbers and graphics. It has been developed together with

practitioners and tested in Helsinki primary health care centre in 2012. Taitoni.fi is launched for all physicians in Finland (n=26,000) in October 2013. In future Taitoni may form a national CPD assessment tool.
Employers will get reports of groups of physicians in numbers and in graphics.

Summary of Results: Results from national launch will be presented. In the pilot study, a total of 78 physicians tested Taitoni-tool for 6 months. Of them 67 (87%) were women and their mean age was 40.2 years (SD=9.8 years). Formal Educational events formed a majority of recorded learning, whereas informal learning was reported less often. Competencies related to the learning events as measured with CanMEDS-roles distributed rather evenly.

Conclusions:

Take-home Messages: The linkage of CPD documentation to work environment is essential and collaborating with employers will stimulate personal development management in health care.

OTT-PA7-10
An innovative reflective practice model for paramedicine students to direct and self-assess their learning during on-road clinical placements

Adam Rolley (*The University of Queensland, School of Medicine, Ipswich, Australia)*, Emma Bartle (*The University of Queensland, School of Medicine, Brisbane*)

Background: No standardised model exists for paramedic students' on-road clinical placements. The variation between paramedicine courses in Australia means students on placements have different levels of knowledge and clinical skills, and therefore different learning requirements. A new reflective practice model for on-road clinical placements was implemented into The University of Queensland's Paramedicine program, requiring students to set their own learning objectives. This study investigated the ability of students to self-assess their learning using this approach.

Summary of Work: Students were required to develop an individual learning plan for their clinical placement, based on their identified knowledge gaps, capabilities assessment, learning needs and course requirements. The final placement portfolio, after their placement, included a self-assessment of how well they met their learning objectives. The self-assessment data were used to explore the ability of students to develop appropriate and achievable learning objectives, through analysis of the discourse they used and range of learning objectives set.

Summary of Results: This model provides paramedicine students with a more structured learning experience during on-road clinical placements, allowing them to address gaps in their knowledge; however students need to learn more about what makes a good learning objective, as many were broad with no measurable outcome.

Conclusions: Students need support and guidance when setting their own learning plans. Requiring students to take control over their learning during on-road clinical placements can improve learning outcomes, however

students need support and guidance when setting their own learning objectives.

OTT-PA7-11
Assessment of medical students' professional identity formation: follow-up of reflective descriptions during preclinical courses

Masami Tagawa (*Kagoshima University Graduate School of Medical and Dental Sciences, Center for Innovation in Medical and Dental Education, Kagoshima, Japan)*, Kazunori Ganjitsuda (*Kagoshima University Graduate School of Medical and Dental Sciences, Center for Innovation in Medical and Dental Education, Kagoshima*), Saori Kijima (*Kagoshima University Graduate School of Medical and Dental Sciences, Center for Innovation in Medical and Dental Education, Kagoshima*)

Background: Medical education has to focus on developing students' professional identity as well as training of competency. Students' understanding professionalism and identity formation is a dynamic and ongoing process of meaning-making by interactions with teachers, patients, and others. Monrouxe et al. reported that the identity formation of medical students begins at an individual level and continues on to a collective level in clinical training.

Summary of Work: We analyzed medical students' reflections to assess identity formation related to preclinical experiences in a formal curriculum. Students recorded reflections on their learning using an e-portfolio during courses of communication, teamwork, physical examination, and other clinical experiences. Students learned knowledge, skills, and attitudes by lectures, clinical observations, role-playing with simulated patients and peers, and group discussions. A cohort group of students' descriptions from January 2012 to July 2013 was analyzed by quantitatively (frequency and length), as well as qualitatively to determine stages of identity.

Summary of Results: Ninety-four students, who gave consent to this research, wrote 2245 descriptions (mean 24, max 38, min 6 per student) during 19 months. Individual students' behavior on recording their reflections varied. Female students wrote more frequently and longer descriptions than male students. Thirty-four percent of students reported a collective level of identity during the follow-up period.

Conclusions: Medical students' diverse behaviors of reflection were revealed. Different identity development during the formal preclinical curriculum was suggested.

Take-home Messages: Continuous recording of reflective descriptions is useful to detect and follow up students' identity development.

OTT-PA7-12
Lessons learned in the implementation of a competency-based portfolio-supported assessment system for UME

Kimberly Lomis (*Vanderbilt University School of Medicine, Medical Education and Administration, Nashville, TN, USA*)

Background: As part of a larger curricular revision, we have implemented a competency-based assessment system "fit for purpose" (Van der Vleuten) and supported by a portfolio-based process (Dannefer). Our goal is to train students in informed self-assessment, applying performance evidence to direct future learning.
Summary of Work: We generated a standardized set of performance measures aligned with ACGME competency domains. Descriptors are detailed enough to inform future learning, yet can be applied across multiple settings. Performance data from multiple courses is resorted digitally to create "competency domain reports" that supplement traditional course grades. At set intervals, each student reviews performance evidence and completes a self-assessment of milestone progress. The student meets with his/her portfolio coach to further clarify understanding of the data and to generate a personalized learning plan.
Summary of Results: The standardized milestone behavioral anchors are currently being used in multiple settings (classroom, clinical), courses (science blocks, clerkships) and levels of learners (years 1, 3, 4). We will present early experience with collection of nearly 3000 forms and completion of two rounds of competency domain reports, portfolio review and personalized learning plans for first year students.
Conclusions: The development of standardized milestone behavioral anchors to be applied across settings was a challenging endeavor that required ongoing communication with multiple stakeholders. Repeated revisions were required to align with practicalities of implementation. Technological solutions and extensive training for faculty and students are critical.
Take-home Messages: The introduction of a competency-based assessment system is an attainable, though complex, change management process.

OTT-PB4: CURRICULUM EVALUATION 2

OTT-PB4-01
Evaluation of the educational quality of the residency training programmes at the Karolinska University Hospital – a stakeholder discrepancy?

Kristina Sundberg (*Karolinska Institutet/Karolinska University Hospital, Department of Medicine (Huddinge)/Department of Research and Development, Stockholm/Solna, Sweden*), Eva Johansson (*Karolinska University Hospital, Department of Research and Development, Solna*), Lars Kihlström (*Karolinska University Hospital, Department of Research and Development, Solna*)

Background: Sweden has a non-accrediting, speciality-based peer-review system for monitoring the quality development in residency training programmes. To promote further continuous quality development in residency training at the Karolinska University Hospital, different stakeholders are engaged in a newly developed educational opinion survey on educational aspects; residents, program directors and heads of clinic.
Summary of Work: During the winter of 2012 the Karolinska University Hospital and six other university hospitals in Sweden co-designed an educational opinion survey including 19 items. The survey was distributed during spring 2013 to residents (n = 509), program directors (n = 51) and heads of clinics (n = 46).
Summary of Results: The opinion survey resulted in descriptive statistics revealing differences in how the three stakeholder groups view quality aspects of the residency training programmes. The largest discrepancies of opinions on educational quality are found between the residents and the heads of clinic. The two themes that created the largest discrepancies between the two stakeholder groups are work-place based assessment as well as feedback on both generic and clinical competency skills.
Conclusions: Making the stakeholders aware of their different views on the quality of residency training programs is the first step towards continuous quality development. A follow-up study on how the results have been disseminated among the three groups and their next step of action will be conducted during the fall of 2013.
Take-home Messages: To be able to work pro-actively with continuous quality development of the residency training programs at Swedish university hospitals, different stakeholders need to be involved in the evaluation process.

OTT-PB4-02
Evaluation of Educating for Equity (E4E) Intervention: How Health Professional Education Can Reduce Disparities in Diabetes Care and Improve Outcomes for Indigenous Populations

Michael Green (*Queen's University, Department of Family Medicine, Kingston, Canada*), Lynden Crowshoe (*University of Calgary, Department of Family Medicine, Calgary*), Kristen Jacklin (*Laurentian University, Northern Ontario School of Medicine, Sudbury*), Betty Calam (University of British Columbia*, Department of Family Practice, Vancouver*), Han Han (*Queen's University, Department of Family Medicine, Kingston*), Leah Walker (*University of British Columbia, School of Population and Public Health, Vancouver*)

Background: Educating for Equity (E4E) is an international collaborative research project aimed at developing Indigenous health professional education capacity as a vehicle for improving care and health outcomes. The focus of the Canadian team is on improving diabetes care in Indigenous populations through a research and evidence informed Indigenous health continuing medical education intervention for family physicians. In phase 1, focus groups and interviews were conducted with Indigenous patients and healthcare providers to understand needs and gaps in

health services. This knowledge was used to generate a care framework which was translated into an educational intervention. Phase 2 aims to implement the intervention and evaluate its impact on Indigenous patients' outcomes and healthcare experience. This poster presents the evaluation framework of the E4E intervention.

Summary of Work: The E4E intervention consists of a one-day workshop with pre-workshop needs assessment and post workshop reflective components. Family physicians will be randomly assigned to intervention and delayed intervention control groups. A mixed-method, multi-measure, controlled design is used to evaluate the intervention. Data collection includes patient chart reviews, patient experience surveys and interviews.

Summary of Results: Co-primary outcomes are HbAIC and a summary patient experience score. Chart reviews prior to and post intervention will indicate impact of physicians' participation on clinical outcomes. Patient surveys and interviews will reflect changes perceived by patients in physician care delivery.

Conclusions: Primary and secondary outcomes and the theoretical framework underpinning the intervention will be described.

Take-home Messages: Appropriate evaluation of Indigenous medical education requires a multi-modal approach including measuring patient experiences, outcomes and providers' perspectives on the training.

OTT-PB4-04
Assessment of changes in teaching physiology and pharmacology courses

Juraj Mokry (*Jessenius School of Medicine, Comenius University, Department of Pharmacology, Martin, Slovakia),* Daniela Mokry (*Jessenius School of Medicine, Comenius University, Department of Physiology, Martin*)

Background: Traditional way of teaching using lectures, labs and seminars is still very popular at medical schools in Slovakia. Physiology and pharmacology are both among the most important pre-clinical subjects and their mastering is a key to pass successfully in clinical stage of under-graduate medical study.

Summary of Work: In order to change the retention of knowledge among the students and better preparation for clinical stage, several new tools were introduced into the curriculum and their impact on students as well as teachers was analyzed. As the most important, use of simulations (both software and hardware), case-studies, problem-based learning, and interactive formative assessment were implemented into curriculum. The opinions of students and teachers were evaluated by a specially designed questionnaire and their retention knowledge was compared with students from previous years before involvement of simulation.

Summary of Results: The students` and teachers` feedback showed significantly increased interest in learning pre-clinical subjects and increased rate of their theoretical preparation before respective lessons. Furthermore, slight increase in results of retention knowledge tests was demonstrated.

Conclusions: Due to very positive feedback from students, more interactive tools are to be introduced, e.g. e-learning courses on selected topics, electronic lectures and several study materials available on specially designed portal. However, more detailed evaluation of these changes is required, in order to determine the real impact on students` retention of knowledge and their ability to use them in clinical settings.

Take-home Messages: Direct and post-poned assessment of the impact of teaching ways in pre-clinical subjects is essential tool for further progress in education and curriculum development.

OTT-PB4-05
Development and Initial Evaluation of a Bedside Teaching Service: Impact on Resident, Faculty and Patient Outcomes

Danise Schiliro (*New York University School of Medicine, Medicine, New York, USA)*, Douglas Bails (*New York University School of Medicine, Medicine, New York*), Patrick Cocks (*New York University School of Medicine, Medicine, New York*), Michael Janjigian (New York University School of Medicine*, Medicine, New York*), Lisa Altshuler (*New York University School of Medicine, Program for Medical Innovations and Research, New York*)

Background: Studies have documented decreased time for resident in direct patient care, and declines in residents skill in history-taking, physical examination, and communication with patients. Work hour restrictions have increased patient hand-offs, increasing risk for impact on quality of care.

Summary of Work: We developed a 2-week inpatient rotation for medicine residents, focusing on bedside teaching, master clinicians to model/ teach high-level physical exam (PE) and clinical reasoning skills, and training in oral/written communication skills. An evaluation model was developed linking resident skill on performance-based measures of clinical reasoning and communication to quality of care outcomes: length of stay, readmissions, patient satisfaction and disease-specific measures. Resident performance is being tracked post-intervention to assess maintenance of effects. Impact of master clinician teaching on attending performance is being tracked to assess outcome as a faculty development effort, thus broadening its reach.

Summary of Results: Initial results suggest curricula is well received by learners and faculty, is providing in-depth training in PE and clinical reasoning skills significantly adding to residents' learning. We will present data exploring the links to patient care outcomes.

Conclusions: We present a model to evaluate an inpatient teaching curriculum utilizing performance-based-assessments and patient outcome measures. Initial results point to the value to the learning environment. The evaluation model provides a framework for linking educational interventions in inpatient setting to care outcomes.

Take-home Messages: The evaluation of a structured inpatient-teaching ward can include meaningful assessment of impact on learners, faculty and patient care targets, and can contribute to linking curriculum, provision of care and health outcomes.

OTT-PB4-06
The first stage questionnaire analysis of reform test on Medical Ethics course teaching of undergraduates in Sichuan University

Yan Li (*Political Science College, Sichuan University, Chengdu, People's Republic of China*), Youping Li (*The Chinese Cochrane Center, Chengdu*), Liji Lan (*Political Science College, Sichuan University, Chengdu*), Shangwei Zheng (West China Hospital, Sichuan University, *Chengdu*), Yixin He (*Political Science College of Sichuan University, Chengdu*)

Background: According to the improvement of global health care and requirement of Good Practice Transformation(GPT) in medical research ethics, the reform of medical ethics course in Sichuan University is imperative.

Summary of Work: Self-made questionnaires were used including pre-class requirement questionnaire and effect feedback questionnaire after class. 178 sophomores in College of Stomatology took part in this questionnaire research. The information were analyzed comparatively and mapped out by excel software.

Summary of Results: Pre-class requirement questionnaire revealed the students' preference on course contents and teaching methods which led teacher to adjust tuition during this term. The information of effect questionnaires after class demonstrated high learning satisfaction. But there are some complaints concentrating on four areas: too short class length, too low grade to take this course, out-of-date content and didactics still need improvement.

Conclusions: Comparing curriculum provision of medical ethics course with other domestic universities and some foreign universities, much evidence was found to help us to reform our class.

Take-home Messages: Adjusting medical ethics didactics and curriculum according to the requirements of students can improve the satisfaction of class. But our questionnaire should be modified.

OTT-PB4-07
Effective and efficient one-week teaching of integrated endocrine system

Nuanthip Kamolvarin (*Chulalongkorn University, Biochemistry, Bangkok, Thailand*)

Background: Faculty of Medicine, Chulalongkorn Univesity is one of the best medical schools of Thailand, founded in 1947. Medical curriculum is 6 years. Our curriculum has always undergone evaluation and improvement processes: from conventional, to community-based, to problem-based and then outcomes-based curriculum. Our school recruits 300 students each year.

Summary of Work: Endocrine system is taught in the second semester of the 2nd year of curriculum, and teaching period is 1 week (30 hours) with integration of disciplines (Anatomy, Biochemistry, Pharmacology and Physiology), and integration of teaching methods (lecture, gross anatomy, histology laboratory, symposium, micro-PBL, group discussion, Q and A forum, formative and summative evaluation, and facebook consultation). Evaluation is performed on each lecturer and teaching method. Scoring of the course consists of 95% MCQ, 3% output of micro-PBL, and 2% attention. Examination questions are screened for quality and standardization by course committee.

Summary of Results: Evaluation of students for 8 years (2005-2012), average number of students is 278 per year, found learning outcomes of students are: 73.9% of student scores are more than 71.49 score, 21.1% are more than 57.8 scores. Average failure rate is less than 1%. Average student satisfaction and student happiness with the course are > 4.5 from 5 score. Symposium is the most favorite session of the course.

Conclusions: Integration of disciplines and integration of teaching methods are an effective and efficient practice to help large groups of students to learn happily and successfully.

Take-home Messages: Good practice of disciplines and teaching methods integration of course design for 1 week is possible.

OTT-PB4-12
Assessment of a current didactic format for a subspecialty academic half-day curriculum

Christopher Hillis (*McMaster University, Medicine, Hamilton, Canada*), P Wasi (*McMaster, Medicine, Hamilton*), W Lim (*McMaster, Medicine, Hamilton*)

Background: There is scant literature to support the 'ideal' format for a subspecialty academic half day (AHD) curriculum. AHD provides protected time to supplement clinical learning. In developing a new AHD curriculum for the Hematology Program at McMaster we sought to assess the utility of the current format of didactic sessions in achieving residents' learning needs.

Summary of Work: A survey was distributed to current Hematology residents (N=7) and Faculty (N=24).

Summary of Results: We received a 100% and 66.7% response rate from residents and Faculty, respectively. All residents and 92.3% of Faculty felt that the current AHD was somewhat useful/useful for consolidation of clinical rotation knowledge. Whereas 57.1% of both residents and Faculty felt the same for laboratory rotations. Forty-six percent of Faculty and 28.6% of residents felt the curriculum was somewhat useful/useful for gaining competency in the intrinsic roles. For concepts not learned on rotations, 57.1% of residents felt AHD was somewhat useful/useful versus 84.6% of Faculty.

Conclusions: Similar proportions of residents and Faculty agree that a didactic AHD is useful for consolidation of knowledge gained on rotations. There is

disagreement about its utility for knowledge not obtained on rotations. A minority of respondents indicated that a didactic format was useful for learning intrinsic roles.

Take-home Messages: According to our survey, a didactic format of AHD is perceived as useful for consolidating knowledge gained on clinical rotations. When designing a new curriculum attention should be paid to intrinsic role content. More work is needed to determine how best to deliver material not learned on rotations during AHD.

OTT-PB4-13
Evaluation of Procedural Skills Training in Undergraduate Students: A qualitative study

Suchada Anotayanonth (*Chonburi Medical Education Center, Chonburi, Thailand)*, Taweesak Nopkesorn

Background: Outcome-based curriculum has been launched at Chonburi Medical Education center, Thailand in 2002. Procedural skill is one of twelve outcomes. We sought to evaluate whether procedural skills training in the undergraduate curriculum is sufficient.

Summary of Work: A qualitative approach using twenty in-depth interviews (twelve undergraduate students, four graduates, and four medical teachers), five focus groups of stakeholders (three undergraduate students and two consumers), six logbooks and incident reports from the clinical risk database were collected, reviewed, and triangulated in academic year 2010. The verbatim notes, transcribed recordings of interviews or focus groups, field notes, and the researcher's reflective notes made during the research were parsed into fragments, and coded for emergent themes.

Summary of Results: Three major thematic categories emerged from the qualitative analysis. The "curriculum" domain was described by inadequate curriculum mapping through three clinical years and incapable on some skills during an internship in the community hospital. The "Teaching and learning environment" domain mentioned in mentor preparation, encourage the patient safety mind, a procedure guide book for all skills base on the Thai Medical Council Standard, practicing in non-patient cycles before contact real patient,. The "student engagement" domain highlighted in self-directed learning, motivation, and self-assessment of their competence and coverage.

Conclusions: Procedural skills training in undergraduate students in the clerkship experience seems poorly prepared and insufficient.

Take-home Messages: Teaching students basic procedural skills in a formal and safe environment is necessary.

OTT-PB4-14
Creation of an Institutional Assessment Plan that Drives Medical Education Research

Robbyn L. Tolles (*University of Nevada School of Medicine, Office of Medical Education, Reno, NV, USA)*, Gwen S. Shonkwiler (*University of Nevada School of Medicine, Office of Medical Education, Reno, NV*), N. Nicole Jacobs (*University of Nevada School of Medicine, Department of Psychiatry and Behavioral Sciences, Reno, NV*), Ramona Houmanfar (University of Nevada, Reno*, Department of Psychology, Reno, NV*), Daniel Reimer (*University of Nevada, Reno, Department of Psychology, Reno, NV*), Melissa Piasecki (*University of Nevada School of Medicine, Office of Academic Affairs, Reno, NV*)

Background: In 2010 the University of Nevada School of Medicine (UNSOM) began a curriculum reform process from a traditional, discipline-based curriculum to an integrated, systems-based one. The primary concern of UNSOM faculty was how the new curriculum would affect student learning outcomes. In response to this issue, the Office of Medical Education and the Behavior Analysis Program with the Psychology Department at the University of Nevada, Reno formed a collaborative Curriculum Evaluation Group (CEG).

Summary of Work: The CEG revised the institutional assessment plan to determine the impact of the new curriculum on student learning outcomes. Three domains of institutional assessment were identified: a) curricular program, b) student achievement and professionalism, and c) faculty engagement in teaching. Medical education research questions were generated for the overarching plan and each domain of assessment.

Summary of Results: The poster will display our program of assessment for each of the three assessment domains with research questions and hypotheses pertinent to each domain. It will also show the data collection schedule and progress to date.

Conclusions: Curricular change allows all stakeholders in medical education an opportunity for self-assessment and program improvement. Assessment and evaluation are tools that can be used to study and measure the effectiveness of that change.

Take-home Messages: A change in curricular structure creates an excellent opportunity to review anticipated outcomes, generate a fresh institutional assessment plan that will measure the impact of implemented changes, and conduct research in medical education.

OTT-PB4-15
Student's perspective on patient advocacy training in an undergraduate medical curriculum

Michelle Huie (*University of Alberta Faculty of Medicine & Dentistry, Edmonton, Canada)*, Lucy Ma (*University of Alberta Faculty of Medicine & Dentistry, Edmonton*), Tracey Hillier (*University of Alberta Faculty of Medicine & Dentistry, Edmonton*), Hollis Lai (University of Alberta Faculty of Medicine & Dentistry*, Edmonton*)

Background: Health advocate is one of seven key physician roles. In an undergraduate medical curriculum, this competency needs to be built upon through repeated exposure to content, role modelling and practice in multiple clinical contexts. A single lecture or even a series of lectures cannot sufficiently prepare students to apply these skills in practice. Our study

examines how advocacy content is being delivered in our school, and whether such content is delivered on a repeated basis.

Summary of Work: Through a student's lens, we use curriculum mapping to determine how advocacy was taught throughout the first year of medical school at the University of Alberta. All sessions in the first year of the program were systematically reviewed. For each learning event, the taskforce identified whether core competencies of advocacy were included.

Summary of Results: Initial analysis of existing content suggests patient advocacy was emphasized consistently throughout the year in multiple courses under a variety of delivery methods. Full report of our findings will determine coverage across topic, frequency of coverage, and compare the combination of delivery methods.

Conclusions: A student-led taskforce can use curriculum mapping to identify areas of strength and weakness regarding specific core competencies. This information can augment data generated by faculty to demonstrate curricular integration and inform curricular change.

Take-home Messages: Curriculum mapping is a valuable tool to assess whether important patient centered concepts are taught throughout a program, and evaluate the consistency of emphasis across different courses.

Tuesday April 29
8:30AM – 10:00AM

SYMPOSIA

OTT-SG-1
Issues and Controversies in the Use of Portfolios for Assessment in Undergraduate and Graduate Medical Education

Presenters: Kenneth Locke (*University of Toronto, Faculty of Medicine, Toronto*), Anthony Donato (*Reading Health System, Reading, PA*), Pippa Hall (*University of Ottawa, Faculty of Medicine, Ottawa*), Margaret McKenzie (*Cleveland Clinic Lerner College of Medicine, Cleveland, OH*), Hedy Wald (*Warren Alpert Medical School of Brown University, Providence*)

Summary: Portfolios are increasingly used in health professional education to complement, or in some cases replace, other forms of competency assessment. In many cases, they are used solely to promote reflective practice skill development in learners; in others, portfolios play a significant role in progress decisions. In this symposium, we will examine underlying theoretical principles, and explore issues and dilemmas that can arise, when using portfolios for formative and/or summative assessment. In doing so, we will discuss assessment strategies in light of Schuwirth and Van der Vleuten's conceptual framework of assessment tools as instruments of learning, with emphasis on systems of assessment for learning, rather than reliance on individual assessments of learning. We aim to clarify and justify common elements of successful portfolio assessment systems. Portfolio implementation strategies and assessment systems within presenters' undergraduate and graduate medical education programs, with associated practical and institutional issues, will serve as exemplars.

OTT-SG-2
Narrative description as evaluation data in health professional education

Presenters: Janice L Hanson (*University of Colorado School of Medicine, Aurora, Colorado)*, Lindsey Lane (*University of Colorado School of Medicine, Aurora, Colorado*), TJ Jirasevijinda (*Weill Cornell Medical College, New York*), Paul Hemmer (*Uniformed Services University of the Health Sciences, Bethesda, MD*), Marjan Govaerts *(Maastricht University, Netherlands)*

Summary: This symposium will confront the implicit assumption that "measurement" is preferable to "description" when assessing and evaluating learners in medical education. Symposium presenters will discuss why written narrative descriptions of learners' performance provide a more useful and valid foundation for assessment and evaluation than ratings and scores from scales, checklists and examinations. Presentations

will address challenges of changing a culture of evaluation that has relied on numbers for most evaluation data; methods for building shared understanding among faculty and learners; challenges of relying on narrative data when faculty come from different cultural backgrounds. The symposium will close with a summary of the group's conversation about changing a program's culture of evaluation toward narrative description of learner performance.

WORKSHOPS

OTT-WG-1
Receiving feedback: stimulating feedback seeking within the department culture

Presenter(s): Monica van de Ridder (*Albert Schweitzer hospital, Department of Education, Dordrecht, Netherlands*), Jim Blatt (*George Washington University, Washington DC, United States*), Carol Capello (*Weill Cornell Medical College, New York, United States),* Elizabeth Kachur (*Medical Education Development, New York, United States*)

Background: Feedback reception is an important part of the feedback process. When feedback is not well received, it may not be accepted and may not be applied into practice. When students seek feedback themselves this has a positive influence on the acceptance and use of the feedback. But in some departmental cultures feedback seeking can be hard, because it is seen as a sign of weakness. Or gains and cost are not in balance. This interactive workshop will approach the feedback process from a recipient's perspective.
Intended Outcomes: Identify key characteristics of effective feedback and feedback seeking;
Discuss selected findings from feedback and feedback seeking research;
Identify barriers in the departmental culture to feedback seeking;
Explain how to prepare students and colleagues for the feedback seeking process;
Express more confidence in their own ability to use feedback seeking techniques.
Structure: Welcome and Orientation; Feedback and feedback seeking opportunities and challenges (large group brainstorming); Theoretical Frameworks, Feedback Seeking and Literature Review (mini-lecture); Analyzing barriers in the departmental culture (small group task followed by large group discussion); Practice techniques to stimulate feedback seeking in students and colleagues (small group role play); Summary Exercise (sharing of individual take-home-points).
Who Should Attend: It is intended for faculty to teach learners how to seek feedback.
Level of Workshop: Intermediate

OTT-WG-2

Assessment of the educational environment

Presenter(s): Helen M Goodyear (*Health Education West Midlands, Medical Education, Birmingham, United Kingdom*), Taruna Bindal (*Alexandra Hospital, Paediatrics, Redditch, United Kingdom*)

Background: Quality assurance and quality improvement in medical education is of key importance, in particular ensuring that training of students and doctors takes place in a supportive and constructive educational environment with good educational practice. A systematic approach to the collection, analysis and interpretation of information about aspects of educational programmes needs to be adopted.
Intended Outcomes: At the end of this workshop, attendees will be familiar with what constitutes a good educational environment and with tools to assess the educational environment which are relevant to the context in which they work.
Structure: This highly interactive workshop looks at what makes a good and a bad educational climate using small group work. There will be a short presentation on the educational environment including use of well validated tools which are applicable in a variety of environments and can be used for comparison year on year and between different institutions and countries. Participants will try out different methods to assess the educational environment in their own place of work and have the option of working through scenarios.
Who Should Attend: This workshop is designed for those working in undergraduate or postgraduate medical education who are interested in looking at the educational environment in their place of work and/or their training programmes.
Level of Workshop: Introductory

OTT-WG-3
Designing an Assessment System for Learning

Presenter(s): Elaine Dannefer (*Cleveland Clinic Lerner College of Medicine of CWRU, Education Institute, Clevleland, United States*), Andrew Linn (*The University of Adelaide, School of Medicine, Faculty of Health Sciences, South Australia)*

Background: Assessment of learning in health professions education has traditionally taken form as a barrier to progression, or for confirmation that sufficient learning has been achieved to meet required quality standards. Increasingly, we recognize that assessment can also play an important role in promoting and enhancing the learning experience. Through utilization of a systems-thinking approach, these assessment goals may be met concurrently, primarily by placing more emphasis on immediate and continuous feedback. The overall design as well as the various components of an assessment system to support learning will be considered. Bridging theory and practice, the workshop will provide practice guidelines and interactive, application exercises.

Intended Outcomes: Participants will be able to identify key components and processes of an assessment system designed to promote learning, and leave with an approach for planning and implementing an assessment system that supports learning.
Structure: We will provide a brief overview of assessment for learning, examples of assessment systems that support learning, and system design guidelines. Participants will then work through a case study in small groups to develop an assessment system that accomplishes a defined learning goal. Following case discussion, participants will use design guidelines to outline an assessment system to promote learning in their educational program. The workshop will conclude with a review of guidelines for developing assessment for learning.
Who Should Attend: Program directors, assessment coordinators, and education leaders interested in assessment strategies to promote learning.
Level of Workshop: Intermediate

OTT-WG-4
Building an evaluative culture in education in a hospital setting

Presenter(s): Sandra Cunning (*Centre for Addiction and Mental Health, Education, Toronto, Canada*), Kathryn Parker (*Holland Bloorview Kids Rehabilitation Hospital, Toronto, Canada*), Ivan Silver (*Centre for Addiction and Mental Health, Education, Toronto),* David Wiljer (*Centre for Addiction and Mental Health, Education, Toronto, Canada*)

Background: Since much of medical education is delivered in hospital settings, it is timely to examine the education culture that exists in hospitals to support these activities. In an era of increased accountability to payers and patients, hospital administrators seek ways to advance the implementation of best practices and interprofessional care to improve patient care. Yet we know little of the effectiveness of best practices and interprofessional care or how to ensure they are well implemented. Education scholars and program evaluators are being asked to provide their expertise in measuring such outcomes. Developing and maintaining an evaluative culture in a hospital setting can be challenging. What is needed is a systematic approach that supports the link between evaluation and education scholarship to conduct careful and methodologically appropriate evaluations. The purpose of this workshop is to exchange ideas on how to build an evaluative culture. The workshop is based on the presenters' experiences of leading the design and development of a wide range of formative and summative evaluation initiatives including face-to-face and digital educational and clinical programs. The workshop will explore the continuum of evaluations that assess programmatic, organizational and systems level outcomes.
Intended Outcomes: Participants will increase their understanding of:
• The characteristics of an evaluative culture

• The roles and purpose of evaluation (what it is and what it isn't)
• Evaluation tools that could be adapted or applied to their setting
Structure: •Brief didactic presentations outlining evaluation and implementation science frameworks
•Reflection through interactive activities, including a novel ice-breaker, write-pair-share, table talk and the use of bouncing balls, on how these frameworks might be applied to other settings
Who Should Attend: Healthcare professionals, hospital administrators, health policy and program evaluators, medical educators and scholars.
Level of Workshop: Intermediate

OTT-WG-5
Supporting doctors to transform consultations: programme outcomes and evaluation

Presenter(s): D Blaney (*Medical Protection Society, -, Leeds, United Kingdom*), J Kidd (*University of Warwick, Warwick Medical School, Coventry, United Kingdom*), J Purkis (*University of Warwick, Warwick Medical School, Coventry, United Kingdom*)

Background: Poor doctor-patient communication is a major source of patient complaints. About 5% of doctors are responsible for 50% of patient complaints and a significant number of the complaints result from ineffective clinical communication. The best ways of offering remediation and evaluating the outcomes of such a process and the impact of the remediation on the knowledge, attitudes and skills of the participants are not known as the literature is limited and most interventions are with students or doctors in training. We report on the outcome of an intervention, the Clinical Communication Programme (CCP) that has been running since 2005 in UK, Australia, South Africa and SE Asia for the remediation of doctors working both in secondary and primary care and across all disciplines. The workshop will focus on a tool to assess communication behaviour within consultations.
Intended Outcomes: Processes to identify doctors who are poor communicators.
The main communication issues for this group.
Existing processes for remediation.
How these interventions may be evaluated.
How doctors changes in behaviour (outcomes)may be assessed.
Structure: Short presentation on scope of problem, outline of the CCP and evaluation of the intervention. Group discussion on identification, remediation processes, the evaluation of the processes and how the doctor outcomes of the remediation may be assessed. Summarise workshop, drawing together the outcomes of the group discussions.
Who Should Attend: Medical educators, clinical directors, and researchers interested in clinical communication.
Level of Workshop: Intermediate

OTT-WG-6
Collaborative OSCE creation: administrating a quality OSCE using combined resources

Presenter(s): Moyez Ladhani (*McMaster University, Pediatrics, Hamilton, Canada*), Hilary Writer (*University of Ottawa, Pediatrics, Ottawa, Canada*)

Background: The Objective Structured Clinical Examination (OSCE) is a reliable and valid tool that is now a mainstay of formative and or summative assessment of learners at all levels of medical education. As such medical education program directors often find themselves charged with the administration of multiple OSCEs to learners across different levels of the education spectrum. Implementation of the exam is resource intensive, from development of a question bank, to people resources that include faculty, standardized patients and program assistants. Many programs lack the infrastructure and manpower to run a successful and comprehensive OSCE to provide the best assessment and feedback to trainees.
The facilitators of this workshop have implemented a national collaboration, which has successfully administered OSCEs across 17 pediatric residency programs, with substantial workload reduction and improved quality at each program.
Intended Outcomes: Following attendance at this workshop participant will be able to:
1. Understand the elements necessary for a successful collaboration.
2. Participants will be able to effectively implement the steps required for the successful administration of a collaborative OSCE including blueprint creation, station development, and sharing of results.
3. Effectively administer and measure in collaboration with similar medical education programs a standardized and valid OSCE.
Structure: Highly Interactive workshop. Following a short didactic session, small groups will complete interactive activities including 1. Blueprint development. 2. Standardized template for station creation. 3. Scoring and rating systems. 4. Trouble shooting collaboration challenges. A final group session will review progress and share strategies.
Who Should Attend: Program Directors PG and UG
Level of Workshop: Introductory

ORALS

OTT-OG-1: ASSESSMENT OF CLINICAL REASONING

OTT-OG-1-1
Development of an Instrument to Measure Clinical Reasoning in Pediatric Residents: The Pediatric Script Concordance Test (PSCT)

Suzette Cooke (*University of Calgary, Pediatrics, Calgary, Canada*)

Background: The aim of this research was to determine if targeted strategies applied to the design and development of a PSCT could enhance the validity and reliability of this method sufficiently to demonstrate its value and justify its utilization in: 1) Formative assessment and tracking of the development of competency in clinical reasoning over the course of pediatric residency training (PGY1-PGY4), and, 2) Summative assessment in the Royal College of Physicians and Surgeons of Canada (RCPSC) Pediatric qualifying examinations.

Summary of Work: The PSCT was constructed by three RCPSC certified general pediatricians. PSCT content was selected based on acute care topics in the RCPSC Pediatrics Objectives of Training. The ninety-minute PSCT was comprised of 24 clinical cases and 137 items. General pediatricians were recruited to serve on the panel of experts. Pediatric residents from the Universities of British Columbia, Alberta, Calgary and Saskatchewan were recruited to participate in the study.

Summary of Results: Ninety pediatric residents and twenty general pediatricians completed the PSCT. The PSCT demonstrated very good reliability (Cronbach alpha = 0.89). Senior pediatric residents (PGY3-4) scored significantly higher than junior pediatric residents (PGY1-2), supporting the PSCT's construct and discriminant validity.

Conclusions: Targeted strategies applied to the design and development of a PSCT can enhance the validity and reliability of this method sufficiently to demonstrate its value and justify its utilization in formative assessments and begin piloting in summative assessments in pediatric residency education.

Take-home Messages: The PSCT holds tremendous potential as a method to assess clinical reasoning skills in formative and summative assessments in pediatric residency education.

OTT-OG-1-2
Assessing clinical reasoning in junior students with the CBCR-test: reliability and feasibility

Olle ten Cate (*University Medical Center Utrecht, Center for Research and Development of Education, Utrecht, Netherlands*), Angela van Zijl (*University Medical Center Utrecht, Center for Research and Development of Education, Utrecht*), Lisette van Bruggen (*University Medical Center Utrecht, Center for Research and Development of Education, Utrecht*), Eugène Custers (University Medical Center Utrecht*, Center for Research and Development of Education, Utrecht*)

Background: Clinical reasoning requires declarative content knowledge of illnesses scripts, combined with procedural knowledge of what reasoning steps to take. At UMC Utrecht a case-based clinical reasoning course focuses on early building of illness scripts using a structured approach in small groups. For summative testing, initially the Script Concordance Test produced disappointing results. This led to the initiative to build a new online assessment tool for the assessment of clinical reasoning, integrating extended matching, long case and integrative comprehensive puzzle approaches.

Summary of Work: Case-Based Clinical Reasoning-tests consist of a series of clinical cases with nested, varying scenarios. After a short vignette an item could be "Give a differential diagnosis of three". A subsequent item could be: "Scenario1: if diagnosis X would be true, which findings with physical examination would you expect to find?" Each case is provided with long lists of items, from which 1-5 items must be picked, about diagnoses, history findings, physical examination findings, diagnostic procedures findings and therapies or follow-up policy.

Summary of Results: The online CBCR-test has been applied six times since 2010 with groups of about 285 second year medical students, yielding Cronbach's alpha reliabilities of 0.66-0.86 for testing times of 60-75 minutes with 44-57 items. Item construction is intuitive for clinicians and shows face validity.

Conclusions: The CBCR-test has been applied with success and satisfactory reasonable reliability. Further studies to support validity must be undertaken.

Take-home Messages: The new practical tool for the assessment of clinical reasoning is promising.

OTT-OG-1-4
Context Specificity and Cognitive Load in Intermediates

Temple Ratcliffe (*Uniformed Services University at San Antonio Military Medical Center, Medicine, San Antonio, USA*), Amy Flanagan Risdal (*Uniformed Services University, Bethesda*), Steven Durning (*Uniformed Services University, Medicine, Bethesda*)

Background: Context specificity (CS) refers to the variability observed in clinical reasoning performance across different situations. Cognitive load (CL) refers to limitations in working memory that impact clinicians during the clinical reasoning process. Studies suggest multiple factors (patient, physician and/or encounter) may lead to CS for experts (board-certified physicians) during clinical reasoning. Although CL during clinical reasoning would be expected to be higher in intermediates (internal medicine residents), CL's effect on CS in intermediates has not been studied.

Summary of Work: Participants watched a series of three cases portrayed on videos. Following each case, participants filled out a validated post-encounter form (PEF) with sections on history, physical exam, problem list, differential diagnosis, and treatment and included a validated measure of CL.

Summary of Results: Ten intermediates completed all three cases. Consistent with CS, performance varied by case and different sections of the PEF within a case (e.g., history, physical exam, etc.) were not highly correlated. For case 1, higher CL was associated with missing the leading diagnosis ($r = -0.9$, $p < .05$). Trends were also observed with CL and additional history requested ($r = -0.56$). In case 2, CL was negatively associated with the summed diagnostic and therapeutic scores ($r = -0.71$, $p < .05$). Case 3 revealed negative trends between

reported CL and ability to produce a differential diagnosis (r= -0.6) and stating the correct leading diagnosis (r= -0.34).

Conclusions: Consistent with prior studies in experts, CS appears to occur in intermediates and our data suggests that CL is an important mechanism for this phenomenon. In each case, while typically only one or two predetermined aspects of diagnostic or therapeutic performance were significantly associated with CL, most trends revealed the expected negative correlation between reported CL and diagnostic accuracy.

Take-home Messages: Implications of this work include measuring and mitigating CL to potentially improve performance.

OTT-OG-1-5
Comparing Automated and Human Scoring of the USMLE® Step 2 CS® Patient Note

Su Baldwin (*National Board of Medical Examiners, Measurement Consulting Services, Philadelphia, USA*), Polina Harik (*National Board of Medical Examiners, Measurement Consulting Services, Philadelphia*), Brian Clauser (*National Board of Medical Examiners, Measurement Consulting Services, Philadelphia*)

Background: Step 2 CS is a 12-case standardized patient examination that evaluates students' ability to gather medical history information, perform physical examinations, and communicate clinical findings to patients and colleagues. Following each encounter, the examinee is required to document their findings in a patient note (PN) which are rated by physician raters. In order to enhance the assessment of clinical reasoning, a new version of PN was incorporated into the operational exam in June 2012. In addition to recording history and physical examination findings and listing possible diagnoses with follow-up tests, examinees are now required to explicitly support their diagnoses with relevant history and physical exam findings.

Summary of Work: This study examines the extent to which automated scores created using Natural Language Processing (NLP) technology can predict expert ratings on the new PN. In contrast to automated scoring of essays, which typically focuses on formal and stylistic attributes, PN scoring focuses solely on content. Using a NLP algorithm, features are identified using patient notes cross-referenced with scoring rubrics, medical glossary, and dictionaries. A subset of salient features is then manually selected by a panel of content experts for each case. Using multiple linear regression analyses, case-specific prediction equations are built using physician-rater scores as the dependent variable. The resultant regression weights are then applied to a cross-validation sample of double-scored notes to compare the agreement between humans and between the machine and humans.

Summary of Results: Preliminary results show that the machine performance was comparable to humans' on two important criteria: correlation and agreement.

Conclusions: While many opportunities for refining our scoring architecture remain, findings thus far suggest automated scoring has great promise.

Take-home Messages: In addition to addressing issues like low rater reliability, high costs, and time intensiveness, successful implemetation of automated scoring could improve test development, quality control, and score reporting.

OTT-OG-2: ASSESSMENT OF CLINICAL TEACHING

OTT-OG-2-1
The use of peer-assisted learning in teaching clinical examination skills

Anthony Campbell (*Queen's University, Kingston, Canada)*, Aristithes Doumouras (*McMaster University, Surgery, Hamilton*), Raphael Rush (*University of Toronto, Internal Medicine, Kingston*), Taylor David (Queen's University*, Internal Medicine, Kingston*)

Background: Peer Assisted Learning (PAL) is being incorporated into clinical skills programs with increasing frequency. The purpose of this review was to summarize and critically evaluate the evidence of PAL's impact on the teaching of physical examination skills in undergraduate medical education.

Summary of Work: This systematic review addressed the question, "How has peer-assisted learning of clinical skills been incorporated into undergraduate medical education, and how has its impact been assessed?" The Kirkpatrick model of evaluation was used to assess outcomes on 4 levels: reactions, learning, behaviour and results.

Summary of Results: Fourteen articles including 1871 students consistently reported a high level of satisfaction with PAL clinical skills programs, both as learners, and as teachers. Participation in PAL programs was associated with an increase in self-perceived skills and an increase in confidence in clinical examinations. Overall, objective OSCE station marks in students taught by peers were non-inferior to those taught by staff physicians. Additionally, authors reported and intent to incorporate PAL programs for clinical examination into regular medical school curricula.

Conclusions: This review demonstrates that medical students, in learner and teacher roles, benefit from PAL clinical examination programs. There is also evidence that peer led clinical skills tutorials have objective outcomes comparable to standard teaching programs. PAL programs provide a comfortable and effective environment in which students can successfully educate their peers in clinical examination skills.

Take-home Messages: PAL clinical examination programs provide both subjective and objective benefits to student participants, and are comparable in their outcomes to staff-led tutorials.

OTT-OG-2-2
Dental Students' Perception Towards A Communication Skills Course (CSC) Based on Transactional Analysis Theory (TA)

Hanan Omar (*International Medical University (IMU), Bukit Jalil, Kuala Lumpur, Malaysia)*, Liang Lin Seow (*IMU, Kuala Lumpur*), Allan Pau (*IMU, Kuala Lumpur*), Saad Khan (University Malaya*, Kuala Lumpur*), Ammar Jasim

Background: Communication skill training represents an integral part of dental education. The transactional analysis (TA) theory is a psychoanalytical tool based on watching people interact and dividing the ego states into adult, parent and child. The utilization of this theory aims to facilitate awareness of unhealthy communication patterns and move towards functioning adult thinking.

Summary of Work: The course development comprised of several stages 1)Identifying the expected outcomes. 2)Planning the activities and development of the course guide. 3)Developing sequenced teaching units 4)Training the facilitators and simulated patients followed by conducting the course over a period of 4 weeks and 7 sessions, 2 hrs each. Forty six second year dental, students attended the course that involved the introduction of transactional analysis theory and related activities, utilization of simulated patients, video-taped interviews in addition to role playing. Perception of the students towards the course was assessed using an 8 - item questionnaire addressing their identification of their strengths in communication, the usefulness of simulated patients in addition to the usefulness of TA theory in teaching CS.

Summary of Results: 79% of the students agreed that the use of simulated patients made them aware of their strengths in communicating with patients while 69% felt comfortable practicing communication skills in front of class mates. All the students agreed that overall, the course was worthwhile, 65% reported that they would like to participate in such activities and 35 % felt neutral towards that. Regarding the use of transactional analysis theory 68% agreed that the introduction of TA was useful while 32 were neutral.

Conclusions: CS course improved that students understanding of their weaknesses and strengths and that the TA can represent an added value to the CS course conducted.

Take-home Messages: Transactional analysis theory can contribute to improving the CS among dental students.

OTT-OG-2-5
The Clinical Workplace Learning Culture Survey (CWLCS) – a new tool

Jennifer M Newton (*Monash University, School of Nursing & Midwifery, Melbourne, Australia)*, Brian Jolly (*University of Newcastle, School of Medicine & Public Health, New South Wales*), Amanda Henderson (*Princess Alexandra Hospital, Nursing Education, Brisbane)*

Background: The learning environment is fundamental to the individual's learning and expectations of learning. Research has highlighted the disparity between academic and practice learning environments due to the competing values of academia and healthcare organisations. This has a direct impact on novice practitioners' learning. This presentation reports a new survey tool, the Clinical Workplace Learning Culture Survey, funded through an Australian Research Council Discovery grant. The survey will inform an empirically derived model of workplace learning for health professionals that will provide a supportive structure for planning education and evaluating learning in clinical environments.

Summary of Work: Development of a survey of 64 items drawing upon previous published tools was undertaken and piloted with registered nurses and students (n= 39). Following field work observations and interviews with registered nurses and nursing students (n=91) the survey was further refined to 55 items. Registered nurses and nursing students (n= 541) have been surveyed at the time of abstract submission with surveying still in process.

Summary of Results: The pilot survey demonstrated the survey tool yielded approximately 90% appropriate responses. Using SPSS a factor analysis of the 55 item survey will be undertaken to determine the factors in this new survey tool.

Conclusions: This study will make a significant contribution to the theoretical understanding of educational and cultural workplace learning in healthcare that impact on health workforce education.

Take-home Messages: For transference of clinical skills to occur the workplace needs to be receptive, and one that promotes innovation and creative partnerships in learning that foster support for assessment.

OTT-OG-3: ASSESSING COMMUNICATION SKILLS

OTT-OG-3-1
Validation of a Family Meeting Behavioral Skills (FMBS) instrument

Jillian Gustin (*The Ohio State University, Internal Medicine, Division of Palliative Medicine, Columbus, USA)*, David Way (*The Ohio State University, Office of Evaluation, Curricular Research and Development, Columbus*), Jennifer McCallister (*The Ohio State University, Internal Medicine, Division of Pulmonary and Critical Care Medicine, Columbus*), Sharla Wells-DiGregorio (The Ohio State University*, Psychiatry, Columbus*)

Background: The FMBS instrument was developed to provide formative feedback and summative assessment to subspecialty fellows. The objective of this study is to outline the data supporting the validity of this instrument to measure communication skills during family meetings.

Summary of Work: 16 pulmonary/critical care (PCC) fellows were videotaped performing a standardized family meeting. Two communication experts used the FMBS instrument to evaluate each encounter independently by coding for skills performed, level of performance and comparison to another

communication instrument (SECURE). We performed Pearson correlations and Cronbach's Alpha reliabilities to demonstrate evidence of content and construct validity, and Intra-Class Correlations (ICC) to evaluate inter-judge reliability.

Summary of Results: We found internal consistency across the checklist items on the FMBS instrument (α reliability=0.855). The number of skills performed and the level of skill were strongly correlated (r= +0.827, p<0.01). There was a statistically significant correlation between SECURE and the FMBS instrument as a checklist and as a rating scale respectively (0.649 and 0.632, p<0.01). The FMBS checklist showed better inter-judge reliability (ICC(2,2)=0.571) than SECURE (ICC(2,2)=0.322) or the FMBS rating scales (ICC(2,2)=0.231).

Conclusions: Despite the limited sample size, the structure of the FMBS instrument is consistent with the SECURE. These analyses provide evidence for an internally consistent and structurally valid instrument to measure advanced skills in communication during standardized family meetings. Inter-rater reliability was better for the FMBS checklist than the rating scale or SECURE.

Take-home Messages: The FMBS instrument is an effective tool for formative and summative evaluation in the PCC fellowship setting and should be generalizable to other subspecialty fellowships.

OTT-OG-3-2
Longitudinal evaluation of doctor-patient communication, interpersonal skills and professionalism among osteopathic medical trainees

Jeanne Sandella (*National Board of Osteopathic Medical Examiners (NBOME), Clinical Skills Testing, Conshohocken, USA)*, William Roberts (*National Board of Osteopathic Medical Examiners (NBOME), Clinical Skills Testing, Conshohocken*)

Background: The Global Patient Assessment ™ (GPA) was developed by the National Board of Osteopathic Medical Examiners (NBOME) for the Comprehensive Osteopathic Medical Licensing Examination Level 2-Performance Evaluation (COMLEX-USA Level 2-PE). The GPA is a Likert-based tool for the evaluation of proficiency in Eliciting Information, Listening, Giving Information, Respectfulness, Empathy, and Professionalism. Residents' ratings for interpersonal and communication skills have been shown to be positively correlated to the GPA. Reliability indices indicate fair to moderate rater agreement when the GPA was used for evaluation of resident performance. The purpose of this study is to determine if scores increase from medical school to residency.

Summary of Work: Forty-two osteopathic medical residents rotated through stations including triadic encounters, team-based encounters and encounters with partial-task simulators. Standardized patients completed the GPA. Scores were compared to the scores that the individuals received when they took their COMLEX-USA Level 2-PE in their 4th year of medical school.

Summary of Results: The mean score for this cohort on the GPA as osteopathic medical students was 6.01 (SD = 0.44) and increased to 6.24 (SD = 1.16) in the pilot. Analysis of Covariance (ANCOVA) showed the undergraduate GPA scores were a statistically significant predictor of residents' performance (F value 50.02 p <.001)

Conclusions: Residents' scores on this tool improved which suggests that skills evaluated by this tool continue to develop in residency training in spite of greater communication challenges.

Take-home Messages: Post-graduate education assessment to assure residents are competent in humanistic skills could be advised by measurement methods presently used in medical licensure examinations.

OTT-OG-3-4
Assessment of the Variations in Ethical and Communication Skills of Internal Medicine Residents in United States using Observed Structured Clinical Exercises (OSCEs)

Vijay Rajput (*Cooper University Hospital, Cooper Medical School of Rowan University, Internal Medicine, Camden, NJ, USA)*, Emily Seeling (*American College of Physicians, Philadelphia*), Krystal Hunter (*Cooper University Hospital, Statistics and Research, Camden, NJ*), Prianka Bhattacharya (Cooper University Hospital, Cooper Medical School of Rowan University*, Internal Medicine, Camden, NJ*), Krysta Contino (*Cooper University Hospital, Cooper Medical School of Rowan University, Internal Medicine, Camden, NJ*), Anuradha Lele Mookejee (*Cooper University Hospital, Cooper Medical School of Rowan University, Internal Medicine, Camden, NJ*)

Background: Variations in ethical and sensitive communication skills among United States Medical Graduates (USMG), U.S. Born International Medical Graduate (USIMG), International Medical Graduate (IMG) and Osteopathic graduates (DO) need to be assessed.

Summary of Work: Between 2007-2013 we assessed the ethical and communication skills of 126 first year residents. We developed stations using the communication competency framework in five clinical-ethical scenarios: obtaining informed consent, delivering bad news, discussing advance directives, disclosing medical errors, and communicating effectively using an interpreter. Trained faculty observed and assessed each station using a standardized evaluation sheet. We analyzed individual station performance and compared the score of USMG, USIMG, IMG ,and Osteopathic residents in their first year using the one way anova method. We performed the Pearson correlation between the five stations.

Summary of Results: There was significant improvement among residents in discussing advance directives and disclosing medical errors skills from 2007 to 2013 (P< 0.001). USMG, USIMG and DO residents

consistently performed better than IMGs in delivering bad news (P<0.008) and communicating effectively through cultural and linguistic barriers (P <0.011). There was poor performance among all residents in informed consent (P<0.077). There was a moderate positive relationship between residents who performed on disclosing medical errors and obtaining advanced directives(r 0.37, P<0.01).

Conclusions: Assessment of communication skills using OSCEs should incorporate areas of obtaining informed consent, effective use of Interpreter and disclosing medical error .

Take-home Messages: A robust curriculum in simulated training on ethical and communication skills is helpful in undergraduate and early residency training.

OTT-OG-3-5
Holistic versus analytic rating scales for communication skills in an internal medicine residency objective structured clinical examination (OSCE)

Vijay Daniels (*University of Alberta, Medicine, Edmonton, Canada)*, Dwight Harley (*University of Alberta, Dentistry, Edmonton*)

Background: Different types of rating scales are used for measuring communication skills in an OSCE. The purpose of this study was to gather some validity evidence for a holistic compared to an analytic scale.

Summary of Work: Examiners of residents taking either a fourth-year or first and second-year Internal Medicine OSCE scored a published four-point analytic scale measuring empathy, coherence, non-verbal, and verbal skills, and a holistic six-point scale assessing communication. We used generalizability studies to estimate the number of stations required to reach a phi coefficient 0.8 for each scale for each training level. We used independent samples t-tests to compare scores from each scale across the two levels of residents.

Summary of Results: For the fourth-year resident OSCE, 12 stations would be needed to reach 0.8 for both the holistic and analytic scales. For the combined first and second-year resident OSCE, 16 stations would be needed for the holistic scale, and 17 stations would be needed for the analytic scale. A significant difference emerged for level of training such that fourth-year residents had higher non-verbal and coherence subscores than first and second-year residents. There were no differences on the holistic or summated analytic scales. These results contradict previous research in medical students demonstrating differences in verbal skills and coherence.

Conclusions: Holistic and analytic rating scales required similar number of stations to achieve similar reliability. The analytic, unlike the holistic scale, was sensitive to level of training in terms of non-verbal and coherence skills.

Take-home Messages: An analytic communication rating scale had more validity evidence than a holistic scale.

OTT-OG-4: ASSESSMENT AND INTERPROFESSIONAL EDUCATION 2

OTT-OG-4-1
Using OSCE cases to assess resident physicians' competence in inter-professional collaborative practice

Colleen Gillespie (*New York University School of Medicine, New York, USA)*, Barbara Porter (*New York University School of Medicine*), Kathleen Hanley (*New York University School of Medicine*), Jennifer Adams (New York University School of Medicine), Jaclyn Fox (*New York University School of Medicine*), Sondra Zabar (*New York University School of Medicine*)

Background: Inter-professional collaboration (IPC) is essential to effective and safe practice. This study uses OSCE cases to assess resident physicians' current level of IPC competence and its association with other clinical skills.

Summary of Work: Checklist items assessed: role clarification; inter-professional communication; respect (valuing team contributions); and teamwork (developing a plan, delegating). Behaviorally anchored response options were not done, partly and well done. One outpatient case called for the physician to work with an RN to assess a patient and recognize and correct an error (Primary Care IM Residents, n=44) and the other case required the resident to evaluate and respond to acute chest pain (Categorical IM residents, n=35). Checklists were completed by the "Standardized Nurse" and included assessment of communication, history - gathering and physical exam skills. Summary scores were % of items rated well done (Cronbach's alpha >.75). Analyses include frequencies and correlations among summary scores.

Summary of Results: We found deficits in IPC skills. 38% of residents clarified their role. 41% fully explored the RN's knowledge and 37% their assessment of the situation. Two-thirds demonstrated respect for RN contributions. 34% failed to discuss the follow-up plan with the RN. Residents' average IPC score was 54% well done (SD 26%) which was not significantly associated with the other core clinical skills.

Conclusions: Findings document need for education/ training in IPC and suggest that IPC is a distinct competence domain.

Take-home Messages: OSCEs can be used to assess IPC skills and inform education and training goals.

OTT-OG-4-2
Getting Started in Interprofesional Education Scholarship - Fostering New Talents

Ewa Szumacher (*University of Toronto and SHSC, Radiation Oncology, Toronto, Canada)*, Lisa Di Prospero (*University of Toronto and SHSC, Radiation Therapy, Toronto*), Merrylee Mc Guffin (*University of Toronto and SHSC, Radiation Therapy, Toronto*)

Background: Collaboration among radiation oncology educators within UTDRO, Medical Radiation Sciences Program as well as Peters Boyd Academy, Faculty of Medicine, is an important goal of the educational mission of the University Of Toronto Department Of Radiation Oncology. Undergraduate students from MRSP, Peter's Boyd Academy and other programs have been working with educators to develop understanding of scholarship and interprofessional collaboration through active participation in scholarship. The program is evaluated through the formal evaluation process. This presentation will discuss the objectives of the student scholarship in radiation oncology including their outcomes such as publications, presentations at the local, national and international meetings. Suggestions how this program can be expanded will be discussed.

Summary of Work: Retrospective review of the undergraduate students scholarship activities between 2003-2 working with radiation oncology educators within the department of radiation oncology Sunnybrook Odette Cancer Center have been conducted from 2003-2013.

Summary of Results: Twenty-two individuals participated in development of scholarship projects at various stages of implementations.18/22(80%) were undergraduate students, 1/22(4%) graduate student and 3/22 (13%) radiation therapy staff. The projects were related to two major themes 1) understanding the needs and preferences of cancer patients in radiotherapy practice and 2) working collaboratively as an interprofessional team in radiotherapy practice to improve the care of cancer patients. Nineteen peer review publications were published which included at least one student as a co-author, 28 international presentations, 22 national presentations and 16 local presentations were presented either by students or educators within the program.

Conclusions: Interprofessional collaboration allows students to increase their exposure to interprofessional academic practice and research.

Take-home Messages: This successful collaboration allows students to disseminate their scholarly work through presentations, publications and interprofessional collaborations. The list of manuscripts and presentations will be discussed.

OTT-OG-4-3
Building collective impact in interprofessional oral health education and practice

Anita Glicken (*NCCPA Health Foundation/University of Colorado School of Medicine, Johns Creek, USA)*

Background: To address complex issues impacting patient health, new delivery models must engage a diverse primary care workforce in integrated care. This is particularly true in oral health, where a fragmented delivery system has divorced oral health from traditional medicine. Oral health remains a major source of health care disparities and calls for engaging primary care clinicians such as physicians, nurses and physician assistants in screening and preventive oral health services.

Summary of Work: This paper illustrates how a profession neutral online core curriculum, Smiles for Life, can help IPE programs build a culture in oral health that embraces diverse perspectives, builds a shared knowledge base and creates an infrastructure for engagement of resources to support innovative and change. The curriculum includes a built in assessment and tracking system that reports learner outcomes.

Summary of Results: Over the past two years 18 professional organizations representing over six different health professions have endorsed the Smiles for Life Curriculum. Over 180,000 discrete users have completed some portion of the curriculum. In the first six months of the year, over 100,000 individuals visited the website, more than four times the number over the entire prior year.

Conclusions: The National Interprofessional Initiative on Oral Health (NIIOH) and the Smiles for Life Curriculum provide a case study of collective impact demonstrating how interprofessional education has created a movement to build shared understanding of oral health and primary care clinicians who are prepared to deliver oral health preventive services and who are effective at partnering with dental specialists and others to advance overall health.

Take-home Messages: Smiles for Life, an online shared curricular tools, with built-in mechanisms for assessment and reporting of learner outcomes, builds a shared understanding that provides a foundation for embedding inteprofessional oral health education across institutional education programs.

OTT-OG-4-4
Evaluating the outcomes of an interdisciplinary, intercultural undergraduate course on global health

K Hall (*University of Alberta, Pharmacy and Pharmaceutical Sciences, Edmonton, Canada)*, V Nyguen (*University of Alberta, Pharmacy and Pharmaceutical Sciences, Edmonton*), C Cox (*University of Alberta, Pharmacy and Pharmaceutical Sciences, Edmonton*), D Zakus (University of Alberta*, Medicine, Edmonton*), A Kozyrskyj (*University of Alberta, Pediatrics, Edmonton*), <u>K Cor</u> (*University of Alberta, Pharmacy and Pharmaceutical Sciences, Edmonton*)

Background: An innovative undergraduate, interdisciplinary (pharmacy, medicine, public health) and intercultural course, jointly designed by Faculty from the University of Alberta and Fudan University (Shanghai, China), was delivered at Fudan University in July 2013. The course consisted of 3 modules: global health, determinants of health and health care systems. Students worked in interuniversity and interdisciplinary groups to complete assignments and give presentations. On-line discussion forums and site visits to health care facilities/health policy research centres in Shanghai were also utilized.

Summary of Work: A mixed methods approach was used to evaluate the course. A pre-post quasi-experimental design comparing means on a series of self-reported Likert questions, organized into

meaningful scales, was used to estimate effects on learning and attitudes. Student perceptions of the impact of situating the course in China, as well as course design/delivery, were measured using both open-ended qualitative questions and self-reported Likert questions.

Summary of Results: Statistically significant increases in self-reported knowledge and understanding of global health, health determinants, and health care systems and self-reported attitudes toward alternative healthcare practices were revealed. Descriptive summaries revealed that setting the course in China contributed to an enhanced learning experience.

Conclusions: Students who took the course improved their knowledge and appreciation for global health, alternative healthcare practices, and different healthcare systems.

Take-home Messages: Interdisciplinary, cross-cultural courses in situated learning settings appear to be an effective way to teach global health.

OTT-OG-5: ASSESSMENT OF LEARNING OPPORTUNITIES

OTT-OG-5-1
E-induction with assessment - an effective tool for enhancing patient safety?

K Nathavitharana (*Health Education West Midlands, PMDE, Birmingham, UK*)

Background: Induction is an essential component of employment, not least for trainee doctors. The West Midlands region in the United Kingdom serves a population of over 7 million. Trainee doctors in the region work within 23 NHS (National Health Service) hospital and community Trusts (providing secondary care) and over 200 primary care (General Practice) centres. A mandatory, regional e-induction consisting of 18 modules, each with integral assessment, was provided between 2011 – 2013.

Summary of Work: The modules consisted of Statutory (eg fire safety); Patient Safety (eg prescribing); and General (eg clinical audit) topics. Over 6,000 trainee doctors (of all grades) across the West Midlands region successfully completed the generic induction provided online.

Summary of Results: The verifiable e-induction was effective in reducing repetition of the same topics at individual hospital Trusts, enabling local induction to be focused on organisation and department-specific topics. A certificate of completion (valid for three years) was recognised by all NHS organisations throughout the region, eliminating the need for repetition as doctors rotated from one organisation to another during the course of their training.

The effect of e-induction on prescribing errors, healthcare associated infections, blood transfusion errors, adverse effects of anticoagulation and thromboprophylaxis, and critical incidents, is reported. The complex nature of collecting such data, attributing

benefit or otherwise to e-induction, whilst acknowledging confounding variables, is discussed.

Conclusions: E-induction, with integral assessment, enhances patient safety, meets SMART (Specific – Measurable – Achievable – Realistic – Timebound) criteria, and provides an effective and efficient medium for delivering generic induction.

Take-home Messages: Online generic induction with integral assessment:
1. Enhances patient safety 2. Is effective and efficient
3. Meets SMART criteria

OTT-OG-5-2
Evaluation of a surgical decision-making eLearning programme

Sarah C Rennie (*University of Otago, Department of Surgical Sciences, Dunedin, Australia*), Andre van Rij (*University of Otago, Department of Surgical Sciences, Dunedin*), Chrys Jaye (*University of Otago, Department of General Practice & Rural Health, Dunedin*)

Background: Surgical decision-making is an important competency for surgeons, but is often neglected in teaching programmes. An eLearning package was developed to evaluate this method for engaging learners in surgical decision-making and assess its application to both undergraduate and postgraduate participants.

Summary of Work: The eLearning package presented information about surgical decision-making in interactive tutorials and short cases. It compared learners at different levels of training; surgical registrars and senior medical students were recruited face-to-face or by e-mail. Pre and post MCQs were completed to assess knowledge gained; participants also completed an evaluation questionnaire. Paired t-test was used to analyse differences between pre and post MCQ results with Mann-Whitney U and Kruskal Wallis tests to look at differences between different participant groups.

Summary of Results: The majority of participants' MCQ scores increased for all groups (64%), some remained static (16%) and some decreased (20%). The majority of participants rated the site as user-friendly, enjoyable, encouraging reflection and learning. There was no significant difference in the performance and perceptions of the participant groups, other than surgical trainees were more likely to report the interactive short cases as being below their level of training (10.6%) compared to medical students (1.6%)(p=0.0002).

Conclusions: The package achieved its aims of being user friendly, enjoyable, and encouraging reflection. An eLearning package that is user friendly is more likely to be enjoyable and enhance learning. eLearning is a suitable method for teaching surgical decision-making.

Take-home Messages: A well-designed eLearning package can provide a positive learning experience for participants across a wide spectrum of experience.

OTT-OG-5-3
Gaming as a tool to train cognitive skills in Emergency Medicine: how effective is it?

Mary Dankbaar (*Erasmus University Medical Center, Desiderius School, Rotterdam, Netherlands)*, Stephanie Schuit (*Erasmus University Medical Center, Internal Medicine, Rotterdam*), Maartje Bakhuysroozeboom (*TNO, Life Quality, Hoofddorp*), Jan van Saase (Erasmus University Medical Center*, Internal medicine, Rotterdam*), Jeroen van Merrienboer (*Maastricht University, Institute for Education FHML, Maastricht*)

Background: Training in emergency skills is an essential part of medical education. Increasing demands on competences of doctors, lack of time and increasing costs necessitate cost-effective training in hospitals. We developed an educational simulation game using the ABCDE method to better prepare residents for face-to-face (f2f) training. Scenarios are provided in a realistic online emergency department using a high-fidelity physiology model.

Summary of Work: In a quasi-experimental design with family practitioner residents, a control group (n=71) received study materials and did a 2-weeks f2f-training in the ABCDE method; the game group (n=72) received the ABCDE game before the same f2f-training. ABCDE skills were assessed before and at the end of training.

Summary of Results: The control group had more experience with acute patients and less general experience as a trainee compared to the game group. After the game but before f2f-training, the game group performed higher on the ABCDE competency subscale (7-point scale; M=4.25) than the control group (M=3.46, p<.05); their performance on critical decisions was slightly better although not significant (M=63% vs. 48%), and global ratings were similar (M=5.04 vs. 4.92). Evaluations of the game were positive. After the f2f-training, the game group no longer had higher skill levels.

Conclusions: The ABCDE game has a positive impact on the ABCDE competency level among residents. The advantage of the game disappears after two weeks of f2f-training.

Take-home Messages: The ABCDE game is an effective tool to prepare for ABCDE skill training and offers authentic learning experiences.

OTT-OG-5-4
Is content really king? An objective analysis of the public's response to medical videos on YouTube

Tejas Desai (*East Carolina University, Nephrology, Greenville, USA*), Afreen Shariff (*East Carolina University, Internal Medicine, Greenville*), Megan Eure (*NC State University, Physiology, Raleigh*), Vibhu Dhingra (East Carolina University*, Internal Medicine, Greenville*), Deeba Minhas (*East Carolina University, Internal Medicine, Greenville*), Mark Kats (*Northeast Georgia Diagnostic Clinics, Nephrology, Gainseville*)

Background: An increasing number of medical educators and patients have turned to YouTube to teach and learn about medical conditions. These videos are from authors whose credibility cannot be verified & are not peer reviewed. As a result, studies that have analyzed the educational content of YouTube have reported dismal results. We hypothesize that educationally focused YouTube videos, authored by credible sources, are 1) of highly educational, 2) appropriately suited to educate the public, and 3) result in high public engagement.

Summary of Work: Credible videos about cardiovascular diseases were identified from the Mayo Clinic's Center for Social Media Health network. The type of content in each video was assessed by the presence/absence of 7 factors (epidemiology, pathophysiology, screening, diagnosis, complications, treatment, prevention). Each video was also evaluated for understandability using the Suitability Assessment of Materials (SAM). User engagement measurements (views, likes, dislikes, comments, favorites) were obtained for each video.

Summary of Results: A total of 607 videos (35 hours) were analyzed. Half of all videos contained 3 educational factors: treatment, screening, or prevention. There was no difference between the number of educational factors present & all 5 user engagement measurements (ANOVA p NS). SAM scores were higher in videos whose content discussed more educational factors (ANOVA p < 0.0001). However, the 5 user engagement measurements did not correlate with higher SAM scores (R2 0.00025).

Conclusions: Videos with greater educational content are unable to better engage users than lower quality videos. Videos with greater educational breadth are more suitable for patient education, but users do not consider respond to the increased suitability with greater engagement.

Take-home Messages: It is unclear if the notion "content is king" applies to medical videos authored by credible organizations for the purposes of patient education on YouTube.

OTT-OG-5-5
Evaluation of Medical Student Peer Teaching in Simulation

Joseph House (*University of Michigan, Emergency Medicine, Ann Arbor, USA*), Carol Choe (*University of Michigan, Emergency Medicine, Ann Arbor*), Kristin Berg (*University of Michigan, Emergency Medicine, Ann Arbor*), Heather Wourman (University of Michigan*, Emergency Medicine, Ann Arbor*), Sally Santen (*University of Michigan, Emergency Medicine, Ann Arbor*)

Background: The advantage of simulation over standard didactic lectures is fidelity, active learning, and, often, improved retention. However performing simulation is time-consuming and requires significant faculty commitment and time. The purpose of this study was to evaluate a program where medical students taught through case-based simulation. This method would require only one faculty member to oversee the student-run simulation.

Summary of Work: We developed 3 case based scenarios that included stabilization of a cardiac arrhythmia, which could be taught using simulation. We

identified objectives, outcome checklists, and teaching resources for each topic. Medical student peer teaching was compared to faculty led simulation. Evaluation of the program was made through survey of satisfaction, subjective learning, and objective tests of learning.

Summary of Results: Both methods of teaching led to improved knowledge based on the pretest and posttest. Both groups learned equally from the pretest to the posttest (p=0.6). Students in the peer-teaching group agreed with "student directed learning is an effective way to learn new concepts" (4.2 on a strongly disagree to strongly agree 5point scale). However students in the physician led group responded 4.7 on the same scale (p<0.005).

Conclusions: Student-directed simulation is an effective method of learning.

Take-home Messages: Student-directed simulation is seen as an effective method of learning, and does not appear to be extremely time intensive. This type of instruction could be efficiently utilized to enhance education and promote self-directed learning. Additionally, student-directed learning will decrease faculty burden.

OTT-OG-6: PROGRAMME EVALUATION 3

OTT-OG-6-1
A Qualitative study of medical students' experience of the hidden and informal curriculum in obstetrics and gynaecology

Dianne Carmody (*The University of Western Australia, School of Women's and Infants' Health, Perth, Australia)*, Alexandra Tregonning (*The University of Western Australia, School of Women's and Infants' Health, Perth*), Paul McGurgan (*The University of Western Australia, School of Women's and Infants' Health, Perth*)

Background: The hidden and informal curriculum in medical education is now recognised as a powerful teaching forum. Prior studies have highlighted negative and positive experiences in areas such as professionalism, ethical dilemmas and the conflict between the formal and informal curriculum. This study explored student-selected positive and negative experiences in obstetrics and gynaecology and student response to these encounters in the hidden / informal curriculum.

Summary of Work: An on-line questionnaire was developed to collect students' self-nominated positive and negative experiences and their response to these encounters in the hidden / informal curriculum during term. An inductive thematic analysis of student responses was conducted.

Summary of Results: Students' experiences were analysed into four main thematic categories: student experience of health professional(s) delivering clinical care; student observation of health professionals and other team members' interaction; student experience in clinical care delivery; and student's experience of teaching and learning with clinical staff. The students'

responses to experiences were categorised into three themes: student description of response; student noted influence on practice and student rationalization of experience.

Conclusions: Students reported a breadth of notable day-to-day positive and negative learning encounters during their term. The impact of these encounters included a range of negative to extremely positive descriptions of student experience and identified influence on their practice.

Take-home Messages: To maximise learning it may be valuable to include formal teaching to increase student skills in dealing with poor learning encounters in the hidden / informal curriculum.

OTT-OG-6-2
No easy answers: Attempting to improve the student evaluation of teaching process in a Faculty of Pharmaceutical Sciences

Isabeau Iqbal (*University of British Columbia, Faculty of Pharmaceutical Sciences, Vancouver, Canada)*, Marion Pearson (*University of British Columbia, Faculty of Pharmaceutical Sciences, Vancouver*), John Lee (*University of British Columbia, Faculty of Pharmaceutical Sciences, Vancouver*), Simon Albon (University of British Columbia*, Faculty of Pharmaceutical Sciences, Vancouver*)

Background: Student evaluations of teaching (SETs) are used to evaluate quality of teaching and courses, and to inform curricular revisions and decisions on career advancement and teaching awards. Given their importance, SET processes must function well.

Summary of Work: We undertook this study with the purpose of enhancing the SET process in a Faculty where recent student response rates have been low and faculty engagement with results has been variable.

We probed students' (n = 13) and faculty members' (n = 12) perceptions of SETs via semi-structured qualitative interviews to gain an understanding of barriers and drivers to their engagement.

Summary of Results: Students cited inconvenience, lack of incentive, and time restraints as reasons for not completing SET surveys. Those who regularly completed the surveys said they did so out of a sense of responsibility or courtesy. Those who normally did not complete the surveys noted the lack of benefit to their class.

Faculty members stated that low student response rates, students' lack of expertise in teaching and giving constructive feedback, and concerns about reliability presented barriers to their engagement.

Students and faculty provided numerous suggestions for improving the SET process, such as encouraging in-class completion of SETs; having instructors communicate teaching and course changes prompted by SETs; and administering SETs earlier in the term or after the exam period.

Conclusions: The variation in student and faculty perceptions of SETs suggests that no process will satisfy all concerns.

Take-home Messages: Faculties seeking to improve their SET process must implement changes incrementally and based on context-specific evidence.

OTT-OG-6-3
Defining and measuring quality in community based medical education: developing an adaptable audit tool

E Scott (*University of Liverpool Medical School, Liverpool, UK*), O Allam (*University of Liverpool Medical School, Liverpool*), Presenter: S Alexander-White (*University of Liverpool Medical School, Liverpool*)

Background: There has been a shift towards undergraduate education in the community following the GMC recommendations in Tomorrows Doctors. The University of Liverpool's MBChB Community Studies Unit programme has over 200 practices that host medical students however the quality of these placements are not comprehensively audited. The literature review revealed a gap in both research and practice in this area.

Summary of Work: This piece of work defines the standards of a good quality community medicine placement for undergraduates. The data was collected using a combination of focus groups and semistructured on-to-one interviews. Transcripts were analysed to produce a cumulative and collective representation of the stakeholders' opinion. A single analyst approach employed a variant of Thematic Content Analysis. The resulting Quality Toolkit can assess quality of community placements for medical undergraduates.

Summary of Results: The results yielded a simple yet multi-dimension audit tool, which is currently being validated. The feedback produced is useful to reassure and congratulate good practice and suggest improvements in areas of weakness. Although this tool was produced from the stakeholders at the University of Liverpool, the intension is to share good practice; with minimal adaptations this tool could be used to audit community placements hosting students from Universities throughout the UK and internationally.

Conclusions: Ensuring consistent quality of placements for undergraduates requires an evidenced-based, holistic approach. This work proposes an adaptable, practical and useful way to ensure this.

Take-home Messages: An adaptable tool to holistically and consistently assess the quality of community placements for medical undergraduates at the University of Liverpool.

OTT-OG-6-4
Are we producing what Canada needs? Evaluating the new Canadian Family Medicine Triple-C competency-based curriculum

Shelley Ross (*University of Alberta, Family Medicine, Edmonton, Canada*), Ivy Oandasan (*University of Toronto, Family Medicine, Toronto*), Doug Archibald (*University of Ottawa, Family Medicine, Ottawa*), Louise Autier (Universite de Montreal*, Family Medicine, Montreal*), Kathy Lawrence (*University of Saskatchewan, Family Medicine, Saskatoon*), Laura McEwen (*Queen's University, Kingston*)

Background: Many Canadians are unable to access primary care in a way that meets the needs of the community (rural areas, inner city areas, Aboriginal populations, among others). In response, the College of Family Physicians of Canada (CFPC) developed the Triple C Competency-Based Curriculum (Triple C). The goal is to produce family medicine graduates who will be ready to practice comprehensive family medicine in any community with the appropriate competencies.

Summary of Work: A working group of the CFPC has begun a large mixed methods longitudinal program evaluation involving all Canadian Family Medicine residency programs. In this presentation, we will detail the process of the design of the surveys and inventories, and will present preliminary data from the pilot implementation at six programs across Canada.

Summary of Results: Pilot exit data has been collected from 6 pilot programs (n = 273 residents). Just over half intend to practice comprehensive care; about the same number feel confident to do so upon graduating. Graduates reported varying exposure to areas of need: little to no exposure was reported for long-term care facilities (35%), inner-city communities (42%), and aboriginal populations (50%). Intent to practice in those areas paralleled exposure (low exposure = low intent to practice).

Conclusions: It is essential to measure the outcomes of a new curriculum to determine if it is having its intended effect. Our preliminary results suggest there is a relationship between exposure and intent to practice. This can help guide changes to curriculum.

Take-home Messages: The evaluation component of the Triple C project can serve as an example and template to other licensing bodies.

OTT-OG-6-5
Qualitative analysis of collaborative leadership during curricular restructuring at the University of Nevada School of Medicine: A story of faculty engagement and growth

Daniel Reimer (*University of Nevada, Reno, USA*), Ramona Houmanfar (*University of Nevada, Reno*), Gwen Shonkwiler (*University of Nevada School of Medicine, Reno*), Nicole Jacobs (University of Nevada School of Medicine*, Reno*), Robbyn Tolles (*University of Nevada School of Medicine, Reno*), Melissa Piasecki (*University of Nevada School of Medicine, Reno*)

Background: Curricular change in medical education is an inevitable and necessary process that reflects the historical and cultural changes that have occurred in the foundational sciences and the practice of medicine (Cooke, Irby & O'Brien, 2010). The University of Nevada School of Medicine is rapidly making the transition from a discipline-based structure to an integrated, block-based curriculum. External experts in Behavioral

Systems Analysis were enlisted to create a unique, systemic account of the change.

Summary of Work: This study investigated faculty concerns during curricular change and the impact of these types of curricular changes on faculty engagement and leadership development. Faculty identified as emerging leaders in the new curriculum were interviewed at several points during the curricular change process. The interviews were transcribed and coded into thematic categories.

Summary of Results: Key themes were categorized and tracked over time. Money, time and collegial support were identified as major concerns, while peer interaction and improved student outcomes were seen as positive aspects of the process.

Conclusions: Positive and negative perceptions evolve for faculty as curriculum change progresses. Identifying these perceptions can help administrators address concerns of faculty while leveraging the positive areas to reduce the stress of curriculum change.

Take-home Messages: Faculty concerns change over time. Identifying and addressing these concerns can decrease the anxiety of curricular change. Support from external sources of expertise can help enhance the understanding and management of the change process.

OTT-OG-6-6
To assess the knowledge component of a Behavioural Management (BM) module incorporated in the undergraduate dental curriculum of a dental school

Yawar Khan (*Riphah International University, Faculty of Dentistry, Islamabad, Pakistan*)

Background: Dental anxiety has been identified as one of the two most crucial barriers encountered by the general population in accessing dental health care. Dentists have been reported to use a number of BM techniques to manage dental anxiety which may be either learnt during undergraduate or graduate training, or by observing other dentists, or by trial or error. Several of these techniques have been validated as being effective in relieving dental anxiety. It has been suggested that a course in BM should be incorporated into the curriculum of dental schools.
This was a pilot study to assess the knowledge component of a BM module (first phase) incorporated in the undergraduate dental curriculum of a dental school.

Summary of Work: Third and fourth year dental students were divided into two groups during their 'Oral Surgery' rotations. A BM module was incorporated into the rotation schedule of the third year students. The fourth year students were the 'control' group and did not receive any formal BM training. At the end of the six weeks' rotation, the BM knowledge component was assessed using a 16 item validated questionnaire.

Summary of Results: The mean BM knowledge score was 4.3 + 1.6 (third year) and 4.8 + 1.3. There was not much difference in the BM score between the two groups (p > 0.05).

Conclusions: The knowledge component is not increased with the incorporation of a BM component but still gave an important insight into the need of incorporating the module in the undergraduate dental curriculum.

Take-home Messages: The module should be tested for students' competencies in BM skills and attitudes in the second phase to emphasize on the importance of incorporating BM principles and techniques in the undergraduate dental curriculum.

OTT-OG-7: MULTIPLE MINI INTERVIEW

OTT-OG-7-1
The Performance of MMIs in the UK context: 4 Years of Experience in Dundee

Jonathan Dowell (*University of Dundee, Division of Clinical & Population Sciences & Education, Dundee, UK*), Adrian Husbands (*University of Dundee, Division of Clinical & Population Sciences & Education, Dundee*)

Background: The Multiple Mini Interview (MMI) is the primary admissions tool used to assess non-cognitive skills at an increasing number of medical schools. This research seeks to demonstrate the predictive validity of the Dundee MMI relative to other UK pre-admissions measures. Validity coefficients for both raw and Multi-faceted Rasch Model (MFRM) fair scores are compared.

Summary of Work: Examination scores were matched for over 500 students (years 1 to 4) who sat MMIs between 2009 and 2012. Pearson's correlations were used to test relationships between pre-admissions variables namely UK Clinical Aptitude Test (UKCAT), Universities and Colleges Admissions Service (UCAS) form and MMI scores, examination scores and demographic variables. Statistically significant correlations were adjusted for range restriction and were used to select variables for regression analyses to predict examination scores.

Summary of Results: Significant correlations of admissions tools and assessment scores ranged from .18 to .37 and .23 to .61 unrestricted. MMIs scores showed significant positive correlations with exam scores in each cohort and year and remained the most consistent predictor of exam scores. Average validity coefficients for MMI domain-total, and MFRM fair scores were .277 and .284 respectively. UKCAT and UCAS scores showed limited and inconsistent predictive abilities.

Conclusions: The MMI was the most consistent predictor of success in early years at Dundee medical school across four cohorts in pre-clinical and clinical years. Fair score validity coefficients were generally larger than those of domain-totals, suggesting that MFRM adjustments are worthwhile.

Take-home Messages: The MMI demonstrates strong psychometric properties and utility in selection.

OTT-OG-7-2

Do medical students admitted through Multiple-Mini-Interview perform better in Problem-based Learning?

Paul F. Wimmers (*David Geffen School of Medicine at UCLA, -, Los Angeles, USA)*, Rasnik Singh (*David Geffen School of Medicine at UCLA, Los Angeles*), Sara Mottahedan (*David Geffen School of Medicine at UCLA, Los Angeles*), Ming Lee (David Geffen School of Medicine at UCLA*, Los Angeles*)

Background: This study seeks to determine whether students selected through multiple mini-interview (MMI) perform better in problem-based learning (PBL) as assessed in four domains by small-group facilitators.

Summary of Work: Since 2011 all applicants are assessed in a 12-station multiple mini-interview (MMI). PBL performance (of block 1, year 1) in four domains were compared between two classes (class of 2013/2014) traditionally selected and two classes (class of 2015/2016) selected through MMI (total N = 635): (1) Problem Solving; (2) Use of Information; (3) Group Process; and (4) Professionalism (scored on a 7-point Likert scale: unsatisfactory-outstanding). The Likert-scale of each domain is accompanied by milestone descriptors of student behavior (checked 0 or 1). There are 20 PBL groups with 8 students and a facilitator. The facilitator evaluates student PBL performance based on the quality of the weekly learning issue write-up and student observed performance and behavior in the group.

Summary of Results: Analysis revealed significant higher scores for the MMI classes on all domains: (1) Problem Solving, $p<.0001$, η2 (eta-squared) $=.025$; (2) Use of Information score, $p<.0001$, η2 $=.024$; (3) Group Process score, $p<.001$, η2 $=.018$; and (4) Professionalism score, $p<.0001$, η2 $=.023$, and selected milestone descriptors.

Conclusions: Students selected through MMI showed higher scores on all PBL learning domains. Furthermore, these students seem to be more expressive and take on leadership roles more easily, as indicated by the scores on the milestone descriptors.

Take-home Messages: The MMI procedure is capable of selecting cognitive stronger and socially more expressive students who perform better in PBL groups as assessed by PBL facilitators.

OTT-OG-7-3
Predicting performance in medical school with selection measures

Agnes Dodds (*The University of Melbourne, Melbourne Medical School, Parkville, Australia)*, Kate Reid (*The University of Melbourne, Melbourne Medical School, Parkville*)

Background: How to select students into medical school remains a vexed question. Although numerous selection measures are advocated, it is difficult to predict future medical abilities. We were able to establish the usefulness of selection measures as predictors of performance in medical school.

Summary of Work: Melbourne Medical School selects students into a four year graduate MD with equal weightings of Grade Point Average (GPA), Graduate Medical School Admissions Test (GAMSAT) and an eight station Multi-Mini Interview (MMI). We examined relationships between these measures and student performance in the first two years. Students study biomedicine and introductory clinical skills in first year, and general medicine, surgery and ambulatory care in second year. How do their performance in these first two years of medical school relate to selection measures?

Summary of Results: In first year, approximately 40% of variation in the performance of over 300 students in biomedicine could be explained using multiple regression models of selection measures. GPA and GAMSAT were the strongest predictors. Selection measures also predicted 10% of variation in introductory clinical skills, with MMI score the only predictor.
In the second, clinical year, a combination of first year performance and selection measures explained 50% of the variation in the same students' performance. After controlling for first year performance, MMI scores remained a significant predictor.

Conclusions: A combination of selection measures assessing bioscience knowledge and communication skills are related to achievement in the medical course even into the second year.

Take-home Messages: These findings point to the usefulness of the selection procedures and the particular measures used.

OTT-OG-7-4
Does the UBC Multiple Mini Interview (MMI) correlate with the UBC Rural and Remote Suitability Score (RRSS)?

Bruce Fleming, Catherine Macala (*Admissions, Faculty of Medicine, University of British Columbia, Canada*)

Background: The Distributed Undergraduate Medical Program in the Faculty of Medicine at UBC is offered at four distinct geographical sites. A common curriculum with equivalent course objectives and shared common exit competencies has been established at each of the sites. At one of the four sites, located in Prince George, BC, the Northern Medical Program (NMP) was established to provide a learning environment in the type of community where graduates would one day return to live and work.

Summary of Work: The Rural and Remote Suitability Score (RRSS) is a tool that has been developed at UBC to help identify applicants to the medical undergraduate program who are best suited to study, live and eventually work, in a rural or remote setting. The question was asked whether an applicant's RRSS was predictive of performance in UBC's standardized interview (MMI). Both of these quantitative evaluations of applicants are components of the composite formula used in creating an applicant's rank order which in turn may lead to an offer for a seat in the Northern Medical Program.

Summary of Results: Over the 6 year study period, there was no statistically significant relationship between the UBC RRSS and MMI scores. In one of the years there was a small statistically significant positive correlation, in one year a small statistically negative correlation.

Conclusions: Taken together, the results of this study suggest that there is no relationship between the MMI and RRSS at UBC.

OTT-OG-7-5
Generalizability of attributes assessed by the Multiple Mini Interview

Nikki Zaidi (*University of Michigan Medical School, Medical Education Assessment and Evaluation, Ann Arbor, USA)*, Christopher Swoboda (*University of Cincinnati, College of Education, Cincinnati*), Stephen Manuel (*University of Cincinnati, Medical School Admissions, Cincinnati*)

Background: Multiple Mini Interviews (MMI) use scenarios to assess specific attributes of an applicant to medical school. MMI scoring is done for multiple sub-score areas, such as critical thinking, communication skills, ethical decision making and one overall score. Generalizability has been widely studied, yet no studies have examined this at the sub-score level.

Summary of Work: Admissions data for applicants (n=16) to one US medical school were used to create a fully crossed Generalizability study. Using G String IV software, variance components were estimated for sub-score (i) scenarios (s) and applicants (p).

Summary of Results: The majority of estimated variance (37%) is attributable to the scenario (s), 6% to the applicant (p) and 40% to the interaction between applicant and scenario (ps). The sub-scores (i) represent the lowest estimated variance at 0.6% with a high Cronbach's alpha (0.97).

Conclusions: High internal consistency suggests that the MMI is measuring one unidimensional attribute; this is further supported by only 2% of variance attributable to the sub-scores (i). Low estimated variance for the sub-scores (i) supports one of the following assumptions; (1) items are interchangeable; (2) interviewers do not understand how to use the sub-scores or; (3) interviewers are more concerned with the overall score.

Take-home Messages: MMI interviewers may measure attributes with the overall score rather than the sub-scores. Additionally, the scenarios may actually determine the unique attributes assessed and introduce unwanted variance. Consequently, MMI scores are more dependent on scenarios rather than applicant attributes/sub-scores. Special attention should be given to selecting scenarios and training interviewers to understand how to appropriately use sub-scores.

OTT-OG-7-6
Selection to medicine: Aptitude testing may negatively predict student performance. An Irish cohort study

Maureen E Kelly (*National University of Ireland, Galway, Discipline of General Practice, Galway, Ireland)*, Jon Dowell (*University of Dundee, School of Medicine, Dundee*), Adrian Husbands (*University of Dundee, School of Medicine, Dundee*), John Newell (National University of Ireland, Galway*, Clinical Research Facility, Galway*), Siun O Flynn *(School of Medicine, University College Cork)*, Thomas Kropmans (*National University of Ireland Galway, School of Medicine, Galway*), Fidelma Dunne, *(National University of Ireland Galway, School of Medicine, Galway),* Andrew W Murphy (*National University of Ireland Galway, School of Medicine, Galway*)

Background: Aptitude testing and other traditional tools are used for selection in Ireland. Multiple Mini Interview (MMI) is not routinely used. This study compared currently used selection tools with MMI to determine the best predictors of students' subsequent examination performance.

Summary of Work: EU applicants (EU), n= 64, were selected by an aptitude test, Health Professions Admission Test-Ireland (HPAT-Ireland) and academic record measured by the Adjusted Leaving Certificate Score (ALCS). Non-EU applicants, n=45, were selected by interview, Grade Point Average (GPA) and International English Language Testing System (IELTS). A cohort of 109 First Year students completed a 10 station Multiple Mini Interview. Performance was compared with First Year examination results-comprising knowledge based MCQ tests and an OSCE.

Summary of Results: Strong correlations emerged between overall First Year results and ALCS ($r=0.49$; $p<0.001$; n=56) and IELTS ($r=0.44$; $p=0.006$; n=38). HPAT-Ireland was negatively correlated ($r=-0.27$; $p=0.03$; n=63). MMI predicted EU student performance on a communication/ clinical skills OSCE ($r=0.27$; $p=0.03$; n=64).   Multiple regression modelling, using variable selection tools, identified previous academic record as the best predictor of overall First Year Examination results for both EU and Non EU students: ALCS ($p<0.001$); GPA ($p<0.001$).

Conclusions: Academic record and English language proficiency remain strong predictors of performance. It is counter-intuitive and concerning that HPAT-Ireland negatively correlated. MMI appears to have a role in the prediction of communication and clinical skills.

Take-home Messages: Further follow up is warranted to establish the predictive patterns of these selection tools as students become more clinically proficient.

OTT-OG-8: ENTRUSTABLE PROFESSIONAL ACTIVITIES

OTT-OG-8-1
To trust or not trust: Development of Entrustable Professional Activities (EPAs) in Emergency Medicine for Australian medical graduates at the transition to professional practice

James Kwan (*University of Western Sydney, Medical Education Unit, School of Medicine, Sydney, Australia)*, Roslyn Crampton (*Western Sydney Local Health District, Research and Education Network, Sydney*), Roslyn Weaver (*University of Western Sydney, Medical Education Unit, School of Medicine, Sydney*), Lise Mogensen (University of Western Sydney*, Medical Education Unit, School of Medicine, Sydney*), Wendy Hu (*University of Western Sydney, Medical Education Unit, School of Medicine, Sydney*)

Background: EPAs are increasingly being used in clinical postgraduate assessment. To date, no EPAs have been developed for the Emergency Medicine (EM) context, nor has a rigorous methodology for developing EPAs been documented. This study aimed to develop a methodology for designing highly relevant EPAs contextualized for medical graduates at the transition to independent practice in Australian emergency settings.

Summary of Work: An action research methodology, using focus group and interview data collected from 12 panel members experienced in supervising recent medical graduates in specific emergency settings, was used to develop EPAs for two common ED presentations.

Summary of Results: Iterative collections of data from the panel were used to develop the final EPA content. Unexpectedly, data also showed that supervisors have very low expectations of new graduate abilities in EM during their first year of practice. As a result previously published levels of entrustment could not be applied in this context. A new model for entrustment was identified based on three stages of supervision which progress from direct, active supervision to indirect, passive supervision.

Conclusions: A 5 stage process using an action research methodology produced two EPAs and a model of entrustment relevant to medical graduates at the transition to independent practice in Australian emergency settings.

Take-home Messages: Using a rigorous methodology to engage clinical teachers in EPA development is more likely to result in EPAs that will be relevant to specific supervision contexts.

OTT-OG-8-2
When is a Resident "good to go"? Using a modified Delphi technique to define and benchmark entrustable professional activities for Family Medicine Residency training

Shanda Slipp (*University of Calgary, Family Medicine, Calgary, Canada)*, Keith Wycliffe-Jones (*University of Calgary, Family Medicine, Calgary*), Wayne Weston (*Western University, Family Medicine, London*)

Background: A new competency-based assessment program was implemented in the Calgary Family Medicine (FM) Residency Program in 2012.
The detailed sets of competencies developed were helpful for curriculum redesign but impractical for use in the assessment of Residents.

Summary of Work: To provide a more practical aid to assessment, a set of FM Entrustable Professional Activities ("EPA's") was developed utilizing a modified Delphi process.
This set of FM EPA's describes the graduation requirements of the Program and, by defining the expected levels of supervision at various points around these EPA's over 2 years, also acts as a proxy measure of progress of each Resident towards readiness for independent practice.
Initial results were presented at CCME in 2013. Since then additional EPA's have been added and the benchmarks have been updated, based on the results of 2 further Delphi cycles in 2013 and 2014.

Summary of Results: The following results will be presented-
i)   The current set of EPA's for FM Residency training in Calgary.
ii)   Levels of agreement (kappa scores) between Faculty and Residents on benchmark levels of supervision for each EPA over 2 years.

Conclusions: A modified Delphi technique has been used to craft a set of EPA's for Family Medicine and the expected levels of supervision for each allows them to be used to track a Resident's progress over 2 years.

Take-home Messages: This first set of EPA's for FM Residency training in Canada can be used both to define the end points of training and as an assessment decision-aid for Preceptors and Residents.

OTT-OG-8-3
Defining entrustable professional activities: Reaching consensus using a nominal group process

Claire Touchie (*Medical Council of Canada, Ottawa, Canada)*, Georges Bordage (*University of Illinois at Chicago, Department of Medical Education, Chicago*), Steven Downing (*University of Illinois at Chicago, Department of Medical Education, Chicago*), Andre De Champlain (Medical Council of Canada*, Research and Development, Ottawa*), Debra Pugh (*University of Ottawa, Department of Medicine, Ottawa*)

Background: Entrustable professional activities (EPAs) are tasks or responsibilities performed by trainees that require an entrustment decision by clinical supervisors. Defining EPAs for trainees entering first-year residency, regardless of specialty, can require a lengthy consultation process. To date no specific methodology to define these EPAs has been suggested.

Summary of Work: A nominal group process was used to select EPAs that first-year residents should be able to perform independently at the end of their first residency year during inpatient care. Eight faculty members, from five specialties, were purposefully recruited as subject matter experts (SMEs) to define EPAs. Nine other national medical educators reviewed the EPAs for final approval.

Summary of Results: After a brainstorming exercise, 28 EPAs were initially suggested and rated by the SMEs. Following discussions, SMEs agreed that 11 EPAs were appropriate for this level of training. Consultation with

the national educators further reduced the list to 10 EPAs.

Conclusions: The nominal group process offered a structured methodology to select and reach consensus on EPAs for in-patient care for all first-year residents. EPAs provide a framework to define and assess competencies.

Take-home Messages: Defining EPAs can be done systematically using a small-group consensus-building process.

OTT-OG-8-4
A role for heuristic maps in assessing entrustable professional activities

Christine Fessey (*St George's Hospital Medical School, Population Health Science, Londo, UK)*

Background: Critics of competency frameworks (Ten Cate O 2007) despair the omission of cohesive real world work as mediator for achievement. ACME 2013 recently introduced 'milestones' for individual progression, a 'less to more' trajectory where progress varies in terms of time and achievement. This paper proposes a role for heuristic maps in assessing entrustable professional activities.

Summary of Work: Clinicians learn by creating personal case based memories, often tacit arrays guiding interpretation of significance and recognition based decision (Klein 1989). Entrustable professional activities encapsulate competencies into meaningful episodes demonstrated in five stages of decreasing supervision (Ten Cate & Young 2013). Knowledge maps mediate awareness, reflection and co-construction of knowledge use within entrustment episodes (Fessey 2002). This knowledge elicitation study explored UK Fy1 residents' awareness and response to entrustment offered by supervisors at unexpected moments during routine work. Residents' interpretations of these moments are captured/ described within heuristics where know-how and activities interpolate. These are discussed in relation to emotional self control, supervisor support and local social integration.

Summary of Results: Findings show variations in personal trajectories, stamina and disposition to engage in the unanticipated. Residents' who engaged in supervised/ partially supervised entrustable activities accelerated in emotional and performance control.

Conclusions: Individual learning trajectories advance - are shaped by factors within and outside the zone of practice (Sterkenberg 2010); patterns of work deployment, character, attitudes, personal awareness, self control are facilitated through local social endorsement.

Take-home Messages: Residents make sense of competencies embedded in routine practice Heuristic maps make explicit the relationships between skills and roles know-how and entrustable activities.

OTT-OG-8-5
Knowledge and nerve: the importance of self-regulation for demonstrating competence by

junior doctors on entrustable professional activities

Rahil D Mandalia (*Northampton General Hospital, Department of Anaesthesia and Critical Care, Northampton, UK)*, Mark A Fores (*University Hospitals of Leicester NHS Trust, Clinical Skills Centre, Leicester*), Krupa Samani (*University of Leicester, Leicester Medical School, Leicester*), Sophie Mullins (University of Leicester*, Leicester Medical School, Leicester*), Robert Jay (*University of Leicester, Leicester Medical School, Leicester*), Rakesh Patel (*University of Leicester, Department of Medical and Social Care Education, Leicester*)

Background: Although junior doctors gain knowledge and experience rotating through clinical specialties, it may be assumed that self-regulatory skills such self-monitoring or self-assessment, develop appropriately along the way. Given poor performers are often 'unskilled and unaware of it', identifying how well junior doctors calibrate their 'perceived' and 'actual' competence is important for protecting patients from avoidable harm.

Summary of Work: Doctors rotating through renal unit participated in a simulated 'ward-round' with real patients before and after a 4-month placement. Competence on two entrustable professional activities (writing in the notes and prescribing medication) was measured using Royal College of Physicians, England and General Medical Council standards. Self-regulatory behaviours (goal-setting, strategy selection and self-monitoring) as well as perceived self-confidence and satisfaction were evaluated using microanalysis.

Summary of Results: 26 doctors completed 104 clinical encounters. There was a reduction in the quality of note-keeping by the end of the placement (t= 2.740 p<0.05), however there was no difference in the quality of prescriptions (t= 1.024, p>0.05).

Although there was an increase in both confidence and satisfaction by the end of the placement, the differences were not significant (p>0.05).There was no significant difference (p<0.05) in the calibration between self-confidence (p=0.0807), satisfaction (p=0.3059) or sureness (p=0.4983) over correct performance next time and actual competence.

Conclusions: Allowing progression through training using competence on entrustable professional activities (such as note-taking or prescribing medication) alone may overlook other aspects of skill development and 'being a competent doctor' such as self-regulation and calibration.

Take-home Messages: Junior doctors rotating through clinical specialties may gain knowledge or experience, however there may be little increase in their self-regulatory skills, which are also necessary for safe clinical practice.

POSTERS

OTT-PA3: OSCE/STANDARD SETTING

OTT-PA3-01
Eliciting professionalism related feedback from OSCE examiners

Yousef Mousavi (*Medical Council of Canada, Research and Development, Ottawa, Canada)*, Marguerite Roy (*Medical Council of Canada, Research and Development, Ottawa*), Ilona Bartman (*Medical Council of Canada, Research and Development, Ottawa*)

Background: Recently, a change was made to the feedback section of the rating scale form included in a large-scale licensing OSCE. Specifically, the question was changed from "Do you have concerns regarding this candidate's performance?" to, "Do you have any serious concerns about this candidate's professional or ethical behavior?". The motivation for the change came from the observation that the original question elicited many irrelevant comments. The new question was designed to direct examiners to be more selective and discriminating in their comments. This study addresses whether the new question improved the quality of examiners comments.

Summary of Work: Use of the new question in a recent examination involving 445 candidates and 173 examiners was investigated. The quantity and quality of all comments were analyzed for each candidate across 10 OSCE stations. All comments were transcribed and coded according to their valence (positive, negative) and their content (professionalism-related or not). The relationship between the number and type of comments to candidate performance was investigated via statistical analysis.

Summary of Results: Comparison of mean total scores between those with negative professionalism-related vs. negative professionalism unrelated comments (not significant) was conducted.

Conclusions: The presence of both negative professionalism-related and negative professionalism unrelated comments is associated with poorer overall candidate performance. We expected that the specification of serious concerns would have a greater impact on the candidate's total score. The fact that the means did not differ indicates that the examiners did not explicitly focus on the serious professional misconduct of the candidate.

Take-home Messages: The new question for assessing serious professionalism concerns does not currently work as intended and should be redesigned.

OTT-PA3-03
Evaluation of comprehensive OSCE at Thammasat Medical School

Pattharawin Pattharanitima (*Faculty of Medicine, Thammasat University, Internal Medicine, Rangsit, Pathumthani, Thailand)*, Pisit Wattanaruangkowit (*Faculty of Medicine, Thammasat University, Radiology, Rangsit, Pathumthani*)

Background: The comprehensive examinations of three steps are required to graduate medical degree from Thammasat medical school. The third step comprehensive examination of the final year is composed of MEQ, long case and OSCE. The objective is to evaluate failure rate of the whole OSCE and of which clinical tasks.

Summary of Work: The comprehensive OSCE of year 2012 had 27 stations with 5 minutes allocated for each. One of five clinical tasks was assessed in each station. Numbers of stations were: communication=6, interpretation=6, history taking=4, physical examination=4 and procedures=7. Pass mark for each station was determined independently. Standard setting of passing the OSCE was 19 from 27 stations (70%). Failure rate of examinee in the whole OSCE and in each station were analyzed.

Summary of Results: Total examinee was 138. Six students were failed (4.34%). The worst candidate failed maximum 12 stations. Five candidates passed all stations. The percentage of failed candidates in communication, interpretation, history taking, physical examination and procedures were 22.22, 23.38, 6.7, 9.4 and 9.3 respectively. Peripheral blood smear interpretation had the highest rate of failure (47.10%). Diagnosis of skin disease and pediatric advanced life support (PALS) were the second top failure rate (38.41%).

Conclusions: There is very high percentage (95.65%) of passing rate in spite of high standard setting of OSCE (70%). Communication and interpretation are two major clinical tasks that have significant failure rate and candidates do best in history taking.

Take-home Messages: Medical students should be prepared and provided more learning experience with formative monitoring in communication and interpretation.

OTT-PA3-04
Communicating through interpreters. How effective is teaching on practice?

A Rowlands (*Queen Mary University, Barts and the London Medical School, Londo, UK)*

Background: People whose first language is not English are more likely to experience inequalities in health care (Fazil and Kai 2004). Consultations even with interpreters may fall short of identifying patients' concerns and understanding (Seale 2012). Doctors need to be skilled in communicating effectively through a third person. This study evaluates the teaching of such skills to third year medical students.

Summary of Work: Teaching involved video demonstration and practice with interpreters and role players. Of the 328 students in the year, 197 attended the session and 131 failed to attend. The end of year summative OSCE included a station on eliciting a history through an interpreter. An OSCE is an intermediate for practice in clinical settings and constitutes level 3 in Kirkpatrick' taxonomy of educational research (Malick 2010).

Summary of Results: Analysis revealed 87% of attendees and 80% of non attendees passed the station. Further analysis of mean scores and component items identified which aspects of training improves students' skills, in an OSCE setting, of building a relationship and achieving understanding in a triadic consultation using an interpreter.

Conclusions: Training reduced failure rates specifically for the interpreters station of the OSCE.
Intervention was worthwhile for the weaker students but not so much for the competent students.
Attendee cohort more likely to use key skills in interviewing through interpreter - interaction directly with the patient both verbally and non-verbally.

Take-home Messages: Training enhances the key skills in interviewing through interpreter

OTT-PA3-05
Traditional Long case examinations vs Objective Structured Clinical Examination – a comparative critical review

Sankaranarayanan Ramachandran (*The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK),* Bhanumathi Lakshminarayanan (*Oxford University Hospitals NHS Trust, Paediatric Surgery, Oxford*)

Background: To the best of our knowledge, there are only few publications in the peer-reviewed literature comparing Objective Structured Clinical Examination (OSCE) and Traditional Long Case examinations (TLC) as an assessment tool.

Summary of Work: We present a critical review of the principles and utility of two time-honoured methods of assessments in medical education - old style TLC and OSCE. We describe in depth comparison of the advantages and disadvantages of two methods of assessments.

Summary of Results: This analysis is presented in a tabular column format of current evidence-based appraisal of strength and weaknesses of both assessment tools in the present milieu.

Conclusions: Current evidence based review suggests that OSCE is a better assessment tool compared to TLC.

Take-home Messages: Evidence favours Objective Structured Clinical Examination as a better assessment tool.

OTT-PA3-07
The utility of mid-block examination and OSPE scores for formative and summative assessment and its relationship to final performance scores

Pairaya Rujirojindakul (*Prince of Songkla University, Pathology, Songkhla, Thailand*)

Background: With the introduction of problem-based learning (PBL) to medical students since preclinical year, basic science learning has become more integrated with clinical instruction. The components have been organized in sequential blocks. Evaluation mainly based on application of knowledge, a problem related to students fail the final examination existed. Therefore, we aimed to early detect students who may require special educational attention using mid-block assessment scores.

Summary of Work: The mid-block scores from Key Features (KF) and Objective Structured Practical Examination (OSPE) were collected from 186 third-year medical students in a six-week block. The PBL small group and lab attention scores were evaluated. Correlation and linear regression analysis between the final scores assessed by multiple-choice questions (MCQ) and OSPE with mid-block and attention scores were analyzed.

Summary of Results: The correlation coefficient (r) between total mid-block scores and total final scores were 0.75 with 95% confidence interval (CI) of 0.68 and 0.81. There was correlation between MCQ+KF scores and total OSPE scores with r of 0.58 and 95%CI of 0.48, 0.67. The low correlations between PBL attention and MCQ+KF or lab attention and OSPE scores were detected. From multivariate analysis, good predictors for final MCQ scores were KF and OSPE mid-block scores with $P < 0.001$.

Conclusions: There was good correlation between mid-block and final scores. The OPSE showed correlation with MCQ and KF scores.

Take-home Messages: We can use mid-block examination and OSPE scores to predict final scores.

OTT-PA3-08
The standard setting methods to evaluate OSCE pass-scores of sixth year medical students.

Thananit Sangkomkamhang (*Medical Education Center, Khon Kaen Hospital, Orthopaedics, Khon Kaen, Thailand*)

Background: The OSCE is a commonly used performance based assessment tool for evaluation of clinical skills. When OSCEs are used as part of an evaluation process, it is necessary to determine valid pass-scores. In a summative assessment, the decision of cut off point determines pass or fail, but many methods of standard setting have been applied. The purpose of this study was to compare three different standard-setting methods for sixth year medical students OSCE, Borderline Regression, Borderline Group and Angoff's method.

Summary of Work: 137 sixth year medical students were assessed by a 6-station OSCE. The teachers who had trained with standard setting were involved in the 6 stations of comprehensive OSCE station. Standard setting methods were Borderline Regression, Borderline Group, Angoff's method to determine the cut scores.

Summary of Results: The OSCE stations were: X ray interpreted, casting skill, joint aspirate competence, spot diagnosis, approaching of multiple injury and patient education. Using Borderline Regression, Borderline Group, Angoff's method produced cut off points and the different numbers of students who passed or failed.

Conclusions: The methods of standard setting showed different cut off point scores and passing rate. Some

methods produce a lower cut off point than others. Different approaches to standard setting result in differing pass-scores.

Take-home Messages: The considerations to decide the selected standard setting method in assessment should be carefully linked to setting pass-scores for OSCEs.

OTT-PA3-09
A Rasch-based borderline method for setting cut scores in an OSCE station: a preliminary test

Jean-Sebastien Renaud (*Universite Laval, Quebec, Canada*), François Ratté (*Universite Laval, Quebec*), Julie F. Thériault (*Universite Laval, Quebec*)

Background: The borderline regression method (BRM) is currently one of the most popular methods for setting cut scores for OSCE stations. It assumes that the checklist items and the global rating item assess the same competency. However, these assumptions are not always tested and the psychometric properties of the global rating scale are unknown. We present a Rasch-based borderline method (Rasch-BM) for setting cut scores in an OSCE station that addresses the aforementioned shortcomings of the BRM.

Summary of Work: We used scores from an OSCE station (n = 183) scored on an 18-item checklist and a 5-point global rating item (1=Unacceptable, 2=Borderline/Unacceptable, 3=Borderline/Acceptable, 4=Acceptable, 5=Superior). Using the Rasch Partial Credit model (PCM), the cut score was set where a student has a 50% probability of receiving a global rating of 2 and the same probability of receiving a 3. For comparative purposes, we also estimated the BRM cut score.

Summary of Results: Rasch analysis showed that the 18+1 items measured the same underlying competency. Eight students failed the OSCE station using the Rasch-BM (failure rate = 4%), while 24 failed using the BRM (failure rate = 13%).

Conclusions: The Rasch-BM ensures that each item, including the global rating item, measures the same underlying competency. The operationalization of borderline performance in Rasch-BM and BRM is different, leading to different failure rates.

Take-home Messages: The Rasch-BM for setting an OSCE cut score needs further validation but is conceptually appealing: it ensures that all checklist items and the global rating item measure the same competency and have good psychometric properties.

OTT-PA3-12
Implementation of a Pass/Fail Cut-Off for COMAT Exams on Clerkship Rotations

Johnathan Franklin (*Oklahoma State University Center for Health Sciences, Tulsa, USA*), Machelle Linsenmeyer (*Oklahoma State University Center for Health Sciences, , Tulsa*), Matt Vassar (*Oklahoma State University Center for Health Sciences, Tulsa*)

Background: The purpose of this poster and/or presentation is to present the results of Oklahoma State University College of Osteopathic Medicine's transition to adopting a pass/fail cutoff for the National Board of Osteopathic Medical Examiner's (NBOME) COMAT clerkship rotation shelf exams.

Summary of Work: In July 2012, OSU-COM implemented the pass/fail cutoff which required students to achieve a score that is at or above 1 standard deviation below the national mean in order to receive a grade for a clinical rotation. The COMAT exams are an osteopathically distinctive exam. Students must take the respective shelf exam at the end of their rotations in Family Medicine, Pediatrics, Psychiatry, Osteopathic Manipulative Medicine, Surgery, Internal Medicine and OB/Gyn.

Summary of Results: Data will be presented regarding the first-time pass rate for shelf exams in each discipline as well as the pass rates for subsequent attempts. Data will also be presented regarding the changes that occurred in students' COMLEX Level 2 CE scores, first-time pass rate and COMLEX Level 2 CE discipline sub-scores. The presenters will discuss the remediation process that occurs when a student fails a COMAT exam.

Conclusions: Adopting a pass/fail cutoff for the NBOME COMAT clerkship rotation shelf exams has improved student performance on rotations and set the mark high for achievement during rotations.

Take-home Messages: A pass/fail cutoff for NBOME COMAT clerkship rotation shelf exams can improve student performance during rotations.

OTT-PA3-13
Can the rating scale improve the reliability of the data obtained from the Readiness for Clerkship Survey to evaluate the effectiveness of the preclerkship program?

Shayna Rusticus (*University of British Columbia, Centre for Health Education Scholarship, Vancouver, Canada*), Linda Peterson (*University of British Columbia, Evaluation Studies Unit, Vancouver*), Kevin Eva (*University of British Columbia, Centre for Health Education Scholarship, Vancouver*)

Background: Better reliability and discrimination in the assessment of student competencies can be achieved by using rating scales that are aligned with the constructs they are intended to measure.

Summary of Work: We developed a construct-aligned rating scale to maximize the reliability and validity of the data obtained from the Readiness for Clerkship Survey to evaluate the undergraduate MD preclerkship curriculum. Two new response formats were developed and compared to the previous scale, which was based on judgments of competence achieved by students. The independence scale focused on the frequency with which guidance is needed by students and the behavioral/independence scale focused on guidance and the ability of students to correctly complete a task.

Summary of Results: 135 third year medical students and 185 of their faculty responded to the survey using

Ottawa 2014 Abstract Book

one of the three response formats. Correlations between average item scores assigned by students and faculty were highest for the two construct-aligned scales (independence=.92; behavioral/independence=.89; competence=.82). The independence scale showed the greatest spread in item means from highest to lowest rated competencies, and resulted in reliable differentiation (G=.8) between competencies with the fewest raters (independence: student k=8, faculty k=11; behavioral/independence: k=12, faculty k=21; competence: student k=22, faculty k=19).

Conclusions: Both construct-aligned rating scales showed improved reliability over the standard competence scale, with the independence scale demonstrating the greatest improvement.

Take-home Messages: A construct-aligned rating scale improves the reliability of the data obtained from the Readiness for Clerkship Survey to identify relative strengths and weaknesses of the preclerkship curriculum.

OTT-PA3-14
Post Assessment Analysis: Effect on item elimination, test reliability, performance, and standard setting

Cesar Yaghi (*Université Saint Joseph, Beirut, Lebanon*)

Background: Reliability and validity remain a main strength for determining the appropriateness and relevance of tests.

Summary of Work: The main question is whether students' post-examination comments would impact item selection and reliability of tests.

Summary of Results: The questions flagged by students had a significantly higher average difficulty index. These items were simultaneously flagged by reliability analysis in 12/37 cases leading to 3 modifications, but solely identified by students in 25/37 cases. Student intervention led to a very modest grade modification. The changes in test reliability were negligible and increased only to 0.724 on average, with a slight decrease in Cronbach's alpha in two tests, an increase in two other tests, and absence of change in the remaining six tests. Improved reliability standard setting procedure increased class average from 79.1 to 82.2. After intervention, the total number of eliminated items was 81 (10.2%). These eliminations accounted for improving reliability indices from an average of 0.723 to 0.763. The failure events were 30 prior to interventions and dropped to 23 after the optimal reliability intervention.

Conclusions: Reliability analysis with difficulty index, and discrimination indices remain cornerstones in assessing the reliability of tests.

Take-home Messages: Although student intervention did not improve test reliability, it identified validity issues that could be acted upon.

OTT-PA6: COMPETENCY-BASED ASSESSMENT/INTERNATIONAL DIMENSIONS

OTT-PA6-01
Identity formation and competency: A longitudinal conjunctive approach to assessment

Tracey Hillier (*University of Alberta, Undergraduate Medical Education, Edmonton, Canada),* Hollis Lai (*University of Alberta, Undergraduate Medical Education, Edmonton*), Amy Tan (*University of Alberta, , Edmonton*), Andrea Davilas-Cervantes (University of Alberta*, Undergraduate Medical Education, Edmonton*), VJ Daniels (*University of Alberta, Edmonton*)

Background: In addition to developing the interrelated individual identities that form the basis of what it means to be a physician, medical students must develop competencies. Assessment is often limited to the knowledge and behaviors that can be seen and therefore more easily measurable.

Summary of Work: A longitudinal curriculum, informed by the literature, with the overall goal to anchor medical students' learning in generalism was developed. This curriculum includes foundational concepts and skills in Patient-Centred Communication, Physical Examination, Professionalism, Health Ethics, Health Equity, Public Health and Systems, and Evidence-Based Medicine including purposeful integration with curricular elements more focused on the "Medical Expert" topics. Considering the complementary nature of competency and identity a conjunctive approach to assessment over time was developed.

Summary of Results: Physicianship is a longitudinal learning experience that requires the integration of skills, knowledge and attitudes. To facilitate and reinforce student learning as a physician we have developed a longitudinal assessment strategy that: a) captures multiple sources of assessment data, b) is collected across the entire academic year, c) is administered to evaluate the different "threads" involved in becoming a physician. This assessment strategy is further enhanced by both interval formative feedback and remediation interviews for all unsuccessful assessment attempts. By implementing a continuous and interactive assessment process, our purpose is to foster identity formation for our students as a physician and reinforce the multi-faceted, multi-sourced demands that are place upon practicing physicians on a daily basis.

Conclusions: We will discuss the lessons learned from developing and implementing the first year of this program of assessment.

Take-home Messages: By implementing a continuous and interactive assessment process, our purpose is to foster identity formation for our students as a physician and reinforce the multi-faceted, multi-sourced demands that are place upon practicing physicians on a daily basis.

OTT-PA6-02
Identifying tools for assessment of CanMEDS roles in the clinical clerkship: results of a "speed-dating" faculty development workshop

Andrea Winthrop (*Queen's University, Department of Surgery, Kingsto, Canada)*, Ruth Wilson (*Queen's University, Family Medicine, Kingston*), Sheila Pinchin (*Queen's University, Office of Undergraduate Medical Education, Kingston*), Theresa Nowlan Suart (Queen's University*, Office of Undergraduate Medical Education, Kingston*)

Background: In competency-based medical education, assessment of the "intrinsic roles" of CanMEDS in the clinical clerkship is challenging. Our lead faculty for the intrinsic roles and our clinical clerkship rotation leaders needed to identify teaching/learning strategies and valid assessment tools that could be implemented within the context of specific clinical rotations.

Summary of Work: We used a form of "speed dating" to facilitate interaction between clerkship Course Directors and "Competency Leads" (faculty with expertise in the CanMEDS roles). Using 90 minutes of a half-day faculty development retreat, each Lead had a station, while Directors changed stations at 10-minute intervals. Leads completed structured notes during the "date". Timing was strictly enforced to encourage focused dialogue.

Summary of Results: Seven Course Directors and six Competency Leads participated. Each "date" produced seven worksheets with a combined total of 69 suggestions for teaching or assessment strategies, including behavioural anchors.

Conclusions: Six months post-retreat, clerkship courses have integrated two assessment recommendations (one from Advocacy, one from Scholar) from the speed dating. In addition, a pilot project will be initiated on one rotation with use of behavioural anchors as part of the assessment tools. The speed dating method was effective in allowing faculty to review current practices related to the intrinsic roles and to creatively develop additional assessment strategies and tools.

Take-home Messages: Use of "speed-dating" provided focused faculty development to initiate strategies for workplace-based assessment of CanMEDS roles in the clinical clerkship.

OTT-PA6-03
How do students use Intended Learning Outcomes to support their learning?

Robert Norman (*University of Leicester, Medical and Social Care Education, Leicester, UK)*, Sara Brooks (*University of Leicester, Medical and Social Care Education, Leicester*), Kerry Dobbins (*University of Leicester, Medical and Social Care Education, Leicester*), Jon Scott (University of Leicester*, Cell Physiology and Pharmacology, Leicester*), Mark Rawlinson (*University of Leicester, English, Leicester*)

Background: Intended learning outcomes (ILOs) are prescribed within medical training programmes. They define the attainments of students to be achieved at stages in their programme of study against the scope of the curriculum specifications. It is assumed that students use published ILOs to frame their studies and to guide self-assessment of their progress but there is little direct evidence on how medical students engage with ILOs.

Summary of Work: Undergraduate students across the Schools of Biological Sciences, English and Medicine at the University of Leicester were surveyed by questionnaire and focus groups regarding their perception and use of ILOs.

Summary of Results: Students reported ILOs to be useful learning aids that prescribe the subject material for assessment. Varying patterns of use in different disciplines were observed, with many medical students engaging only after teaching sessions or at revision. A proportion of respondents reported difficulty in discerning the depth of learning required for assessment, with over half indicating the possibility of underestimating the level of learning specified and requiring some teaching before ILOs become clear. Prescriptively detailed ILOs were widely requested.

Conclusions: Students do not always perceive the content of ILOs as clearly as the author and often underestimate the depth of learning that is required.

Take-home Messages: More attention is required in the framing of ILOs, particularly so that students can assess more accurately the level of learning expected. Medical students require guidance in using ILOs more evenly throughout their studies and across the span of the curriculum.

OTT-PA6-04
Reviewing Objectives in Musculoskeletal Undergraduate Medical Education: Can we agree on what's important?

Ali Bateman (*University of Western Ontario, London, Ontario, Canada)*, Megan Cashin (*University of Western Ontario, London, Ontario*), Darren Van Dam (*University of Western Ontario, London, Ontario*), Michael Payne (University of Western Ontario*, London, Ontario*)

Background: The scope of medical school objectives should be aimed at the undifferentiated physician (UP). We were concerned that, with multiple iterations by various authors, our objectives had drifted away from this intended target. This study evaluated a large sample of our MSK course objectives in terms of relevance to the goal of preparing the graduating UP; these were compared to related Medical Council of Canada (MCC) objectives.

Summary of Work: 18 physicians from Family Medicine, Orthopedics, Physiatry, and Rheumatology scored a random selection of 150 objectives (135 Course and 15 MCC objectives) for need-to-know, should-know, and nice-to-know for the graduating UP. 10 medical students scored the same objectives for clarity.

Summary of Results: 2655 individual ratings were made; of these, 1255 (47.3%) were need-to-know, 885 (33.3%) were should-know, and 515 (19.4%) were nice-to-know. Based on a predetermined threshold, 57 objectives (38%) were agreed upon by all reviewers; 11 were MCC objectives. 1500 individual ratings were made by students, and from this, 117 (78%) of objectives were clear in. Between specialties, there was a statistically significant difference for need-to-know ratings for the

Course (p=0.0001) but not the MCC objectives (p=0.26). Objectives agreed upon as need-to-know did not correlate with clarity ratings by students compared to objectives without agreement (p=0.83).

Conclusions: The higher concordance between specialties for the MCC objectives indicates that they are universally important for the UP and should guide curriculum development. There was no difference between clarity scores for agreed upon need-to-know and other objectives, suggesting that the content of the non-need-to-know objectives, rather than poor clarity, is responsible for their importance scores.

OTT-PA6-05
Validation of the Laval Developmental Benchmark scale for competency evaluation in family medicine residency

Annie St-Pierre (*Unité de médecine familiale Laurier, Québec, Canada*), Annie Thériault (*Unité de médecine familiale St-François D'Assise, Québec*), Michel Labrecque (*Unité de médecine familiale St-François D'Assise, Québec*), Stéphane Turcotte (*Hôpital Saint-François d'Assise, Québec*), Jean-Sébastien Renaud (*Faculté de médecine de l'Université Laval, Québec*), Miriam Lacasse (*Unité de médecine familiale Laurier, Québec*)

Background: The Laval Developmental Benchmark Scale for Family Medicine (DBS-FM) was recently developed to set competency benchmarks for the two-year family medicine training program. The DBS-FM includes 70 items covering the seven CanMED-FM roles. The goal of the study was to assess the reliability, validity and feasibility of the new DBS-FM.

Summary of Work: 57 preceptors from 6 family practice teaching units of the Laval University network volunteered to use the DBS-FM to assess the progress of 37 family medicine residents. Each resident was independently evaluated by 3 preceptors. Inter-rater reliability was measured using Kendall's coefficient of concordance and by calculating the consensus between the 3 raters for each of the 70 items of the scale (2 out of 3 evaluators gave a similar rating for a specific competency). Convergent validity was assessed by comparing the DBS-FM score to the clinical teacher's global rating of the residents. Feasibility was measured as a function of the average time taken to complete the DBS-FM.

Summary of Results: Inter-rater reliability was moderate (Kendall's statistic: 0.35 – 0.49). The scholar role obtained the lowest score, 0.35. We found a significant positive association between the DBS-FM score and the clinical teacher's global ratings of the residents (Fisher exact probability: 0.009, p=0.04). Mean completion time of the DBS-FM was 17.4 (S.D.=4.1) minutes.

Conclusions: The study demonstrated that the DBS-FM has strong validity and moderate reliability.

Take-home Messages: Further work is needed to fine tune rater instructions and improve inter-reliability, shorten its completion time and replicate the validation study.

OTT-PA6-07
Documental analysis as a first way to assess the competencies development in community-based education

Ana Claudia Germani (*School of Medicine, University of São Paulo, Preventive medicine, São Paulo, Brazil*), Aline Scalisse Silva (*School of Medicine, University of São Paulo, São Paulo*)

Background: Health promotion is an important issue to develop an efficient health system, according to national and international recommendations. In this context, medical graduation should include the development of health promotion competencies in the official curriculum. However, to include it in a traditional curriculum is a huge challenge.

Summary of Work: The Primary Care Discipline (PCD) is an interdepartmental initiative (preventive medicine, pediatrics and internal medicine) to offer community-based education, including health promotion issues. In the first year, medical students´groups have to plan, act and check health promotion solutions focused on the population problems (addressed for the local health service). The aim of this project is to use CompHP (Developing competencies and professional standards for health promotion capacity building in Europe) to analyze the groups´ report and to recognize what and how nine health promotion core competencies had worked in a team in 2012.

Summary of Results: A documental analysis showed that almost all groups developed "Enable changes", "Communication", "Needs Assessment" competencies. "Planning", "Implementation" and "Evaluation" were less recognized in the reports and "advocate for health", "mediate through partnership" and "leadership" were noticed in a few reports.

Conclusions: The health promotion community-intervention contributed to development of health promotion competencies among medical students. Reflection on the strengths and weaknesses of the actual model is essential to improve learning potential of this activity.

Take-home Messages: Documental analysis could be an initial way to assess community-based education, focussed on the health promotion. Identifying what and how the health promotion competencies have been developed could help the educational program.

OTT-PA6-09
Factors Predicting IMG Success in Matching to Residency Programs in Canada

Stephen Rimac (*Canadian Resident Matching Service, Ottawa, Canada*), Presenter: Sandra Banner (*Canadian Resident Matching Service, Ottawa, Canada*)

Background: The purpose of this study to review historical data to look at patterns and trends in how applicants match to residency programs in Canada. A probability model is developed to determine whether

certain applicant characteristics increase the likelihood of matching to a Canadian residency program. Differences between immigrants and CSA applicants will be noted.

Summary of Work: This longitudinal study aggregates data from CaRMS' Post-Match Survey of Canadian and International Medical Graduates, and historical match statistics. Data mining tools were used to predict applicant attributes and application ranking strategies to successfully matching to a residency position.

Summary of Results: In our 2013 match, we have detected that certain attributes, such as applicant demographics, first choice discipline, and applicants' match ranking behaviours predicted match success.

Conclusions: In a review of the historical data on medical graduates applying for residency positions in Canada, several factors have been found to predict whether applicants successfully match to a residency program in Canada or not.

Take-home Messages: Study will increase policy capacity of other institutions through evidence-based research.

OTT-PA6-10
Validating the Impact of Cultural Differences on Residency Education (ICDRE) Questionnaire

Douglas Archibald (*University of Ottawa, Ottaw, Canada)*, Alison Eyre (*University of Ottawa, Ottawa*), Dorota Szczepanik (*University of Ottawa, Ottawa*), Lionel Laroche (Multicultural Business Solutions*, Toronto, Ontario*), Memoona Hasnain (*University of Illinois - Chicago, Chicago*), Simone Dahrouge (*University of Ottawa, Ottawa*)

Background: The assimilation and optimal education of International Medical Graduates (IMGs) continues to be a challenge for residency programs due to areas of cultural discordance between IMGs and the Canadian and American medical systems. The problems related to cultural discordance are often perceived by residency educators and administrators as communication, collaboration, and professionalism issues. To address these issues we will be using modern validity theory as a framework to ensure the validity of scores from The Impact of Cultural Differences on Residency Experiences Questionnaire (ICDRE) which measures self-reported perceptions of a residency experience with regard to the concepts of: sense of hierarchy, individualism/teamwork, and risk tolerance.

Summary of Work: All residents who complete the 50 item questionnaire will receive an individualized report of their results which shows how their responses compared to others. Furthermore, the ICDRE may help learners recognize that these possible differences translate into different behaviours in clinical and educational settings.

Summary of Results: The investigators drew upon their professional experiences as educators and work with foreign trained professionals, and the relevant recent literature on cultural competencies to develop a draft version of the ICDRE. Ten experts from various departments within Faculty of Medicine at the University of Ottawa with experience teaching IMGs formed a panel to review the ICDRE prior to being tested by residents. Other sources of validity evidence will be discussed.

Conclusions: Our findings are intended to enable IMGs as well as residency training program educators and administrators to work collaboratively towards overcoming the obstacles presented by discordance between IMGs and host medical systems.

OTT-PA6-11
Interim assessment of educational program for undergraduate dental students with Japanese-English dual linguistic education system at Hiroshima University (the first and the second year)

H Oka (*Institute of Biomedical & Health Sciences, Hiroshima University, Department of International Collaboration Development for Dentistry, Hiroshim, Japan)*, K Suardita (*Institute of Biomedical & Health Sciences, Hiroshima University/ Faculty of Dentistry, Airlangga University, Department of International Collaboration Development for Dentistry/ Department of Conservative Dentistry, Hiroshima/ Surabaya (Indonesia)*), T Takata (*Faculty of Dentistry, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima*), H Nikawa (Faculty of Dentistry, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima), K Kato (*Faculty of Dentistry, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima*), M Sugai (*Faculty of Dentistry, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima*)

Background: Hiroshima University Faculty of Dentistry (HUD) has positively pursued educational reform in dentistry and oral health sciences from undergraduate level, and made active approaches to develop Asia-based global collaboration in dental education and research. To enhance these initiatives, we began to implement the International Dental Course (IDC) program since 2011. Then, IDC international students and regular dental program students (almost all of them are Japanese) have shared the HUD curriculum for 3 and half years and learnt the dental subjects together with a Japanese-English dual linguistic education system; all the lectures are given in these two languages because students need to understand special Japanese terms for learning about the Japanese system.

Summary of Work: To assess the dual linguistic education system at HUD, we had provided questionnaires to the regular program students and the international students about the affectivity of dual linguistic education and understanding the class contents at the end of each semester.

Summary of Results: The results showed that although the students have a problem in understanding the lecture material, they can cover up the problem with the preparations before the lecture and the reviews after the lecture by themselves. Understanding level of the

class contents and satisfaction level for the class have been increased year by year.

Conclusions: At this time point, it is not established clearly about the effects of the dual linguistic education system on ability for globalization. However, it is confirmed that the dual linguistic education has affected and improved the attitude of the students for learning.

Take-home Messages: HUD has been developing an educational program with Japanese-English dual linguistic education system for global collaborations.

Tuesday April 29
10:30AM – 12:00PM

PLENARY: MIRIAM FRIEDMAN BEN-DAVID LECTURE

Selecting for Personal Suitability

Speaker: Dr. David Powis, Professor, School of Medicine and Public Health, Faculty of Health, University of Newcastle, Australia

Medical school training can give a future doctor the basic knowledge required, and foster their skills for updating that knowledge to ensure continued academic competence. It can also teach or nurture the development of some of the other skills and attitudes in the competency list. But it is unrealistic to expect that medical education can do it all, particularly if the student is attitudinally unsuited or otherwise ill-equipped in their psychological makeup to meet the expectations of the profession and the community outlined above. Acceptance of this line of thought must lead us to acknowledge that we should take particular care in selecting medical students – future medical practitioners – basing our choice on a range of criteria that reflect the picture of the generic good doctor. In my presentation I will describe techniques and methods that have been used to measure some non-academic and non-cognitive qualities and provide empirical data on their reliability, construct validity and, most important, their predictive validity that supports their adoption for the purpose of selecting suitable future health professionals.