UNIVERSIDADE DE LISBOA
FACULDADE DE MEDICINA

A Best Evidence Medical Education (BEME) Systematic Review on the feasibility, reliability and validity of the Objective Structured Clinical Examination (OSCE) in undergraduate medical studies.

Madalena Maria Botelho Folque de Mendoça Patrício

Orientador:
Professor Doutor António Vaz Carneiro

Doutoramento em Ciências e Tecnologias da Saúde
Especialidade em Educação e Comunicação em Ciências da Saúde

Lisboa, 2012

Esta dissertação foi aprovada em reunião do Conselho Científico da Faculdade de Medicina da Universidade de Lisboa a 15 de Maio de 2012.

'The OSCE as a 6 Star Exam'

Acknowledgements

Looking back at the whole process I realize that when we undertake this kind of work we do not progress alone. At the end, when taking a pause to reflect on what was achieved, I have no doubt that by myself it would have been impossible. The work I am presenting today is the result of different contributions, where learning was a shared experience between enthusiastic researchers and supporting staff.

This is why I would like to express my appreciation and gratitude to all those who in different ways contributed to this dissertation.

My first words go to my supervisor, **Professor António Vaz Carneiro**, for his support, comprehension and example as rigorous researcher, but, above all, for his friendship during all the years we have been working together, always very demanding in order to make me work to the best of my ability.

To **Dr. Filipa Fareleira** and **Dr. Miguel Julião**, for their invaluable co-work coding more than one thousand studies, for their generosity and professionalism during a process which they first joined as medical students but ended as MDs. Coding and reaching consensus were the most challenging methodological steps of the whole process, and I will keep forever very good memories of the thought provoking discussions, with their critical and challenging comments at the various stages of the work.

To **Professor Geoffrey Norman** and **Professor Meredith Young**, co-authors of two articles integrated in this dissertation, for their valuable support in terms of statistical analysis and idea sharing for taking the maximum of the study.

To **Professor Alex Haig**, for the wonderful support in terms of literature search, namely for his readiness and efficiency.

To **Sofia Mata** and **Dr. Manuela Castro,** for their high commitment and rigor in obtaining the studies identified in the search process and constructing the OSCE Reference Manager database.

To **Dr. Rui Fonseca** and **Dr. Ana Catarina Silva,** for creating and supporting the electronic dynamic database, which allows coding and establishment of consensus on line. My thanks go also to all team of the **IT Department (FMUL)** for their continuous support solving everyday problems and to the **Editorial (AEFML)** team for their committed collaboration with graphic work namely to **Maria José Silva** for being so graceful with successive reviews of the text.

To **Dr. Isabel Aguiar** and **Dr. Antónia Ferreira** for their support and friendship during all these years.

To the whole team of the Institute of Introduction to Medicine, **Professor António Barbosa, Dr. António Pais de Lacerda, Dr. Miguel Barbosa and Dr. Miguel Andrade** for creating the conditions to make the pursuit of this work possible.

To **Professor Rui Victorino,** for the strong support to this work.

To **Professor João Carlos Gomes-Pedro**, for his great contribution to Medical Education in Portugal and for inviting me to join the Department of Medical Education at the time of its creation in 1985, offering the opportunity for a career on Medical Education.

To **Professor Maria de Lourdes Levy,** a special word of admiration for her personality and her approach to life which greatly influenced my personal and professional development.

To **Professor Henrique Bicha Castelo** for his support and incentive to this work.

To **Professor Ronal Harden,** for his advice and stimulating discussions. I will never forget our inspiring and sustained shared thinking dialogues.

To Professor Tom Haye*s*, an old friend since the time of his collaboration with the FMUL in the context of the first Master Course on Medical Education, for his supportive and challenging feedback which so much enriched the discussion.

To Professor Katherine Boursicot, for her generosity when taking the time to review the OSCE papers, an area of her major interest.

To **Dr. Ana Ramos,** for facilitating my thinking and communication in a second language and for being always available when I needed her support.

To my sister **Assunção**, for her continuous support, stimulus and critical reflection and above all for her sensible and pragmatic approach bringing me back to reality when my dreams were too high.

To my daughter **Madalena** for her creativity, criticism, and companionship, helping me to find practical and theoretical solutions by looking at problems from different perspectives.

To all those friends who during these years kept me motivated by saying '*Go on Madalena … You must finish your thesis*!" Among them let me mention, **Professors Amália Vaz Guedes, António Castanheira Dinis, António Levy, Carlota Saldanha, Jacinto Monteiro, João Lobo Antunes, Joaquim Figueiredo Lima, José Melo Cristino, José Ducla Soares, José Pereira Miguel, Lincoln Justo da Silva, Mário G. Lopes, Paulo Costa and Paulo Leal Filipe** for their constant support and positive expectations about this work.

Finally to my friends, my family and especially my close family - **Luís, Pedro, João, Maria, Madalena** and **Alexandre** for their unconditional support, comprehension and patience when I was unavailable.

All of you in many different ways have been determinant in encouraging and helping the pursuit of this work.

Finally I want to mention two persons who are no longer among us…

Professor José Guilherme Jordão, whose tragic disappearance means an irreplaceable loss for Medicine, for the Faculty, for his friends and family. He supported this systematic review since the very beginning, joining the team as a coder. José Guilherme was a great friend and one of the persons I most admire for his personal and professional qualities. His example remained as a continuous incentive during all these years while I missed his support and guidance.

My father, who was always present during my work. His character, rigor, loyalty, professionalism and exigency were landmarks in the difficult route of this dissertation, where I sometimes felt crossing a desert. I miss him every day and, specially, today because he would love to be among us.

ABSTRACT

Title – A Best Evidence Medical Education (BEME) Systematic Review on the feasibility, reliability and validity of the Objective Structured Clinical Examination (OSCE) in undergraduate medical studies.

Introduction – As the so called 'traditional methods' showed problems namely in terms of psychometric criteria, in 1975 Harden et al. introduced the OSCE, as an attempt to solve the multiple criticisms concerning assessment of clinical competencies. Consisting of multiple objective 'stations' designed to assess a range of clinical or practical skills under similar circumstances (same assessment, same patients, and same examiners), the OSCE has been used exponentially all over the world.

Objectives – To produce scientific evidence about the OSCE suitability to assess learning outcomes in undergraduate medical studies, i.e. to what extent is the OSCE a feasible, reliable, valid, fair and acceptable method of assessment with educational impact, when used to assess learning outcomes in undergraduate medical education?

Methods – The BEME methodology was applied by two independent coders, who scrutinized literature from 1975 until 2009. One thousand and sixty five studies were analysed.

Results – The evidence points to the OSCE as a feasible approach to the assessment of clinical competence for use in different cultural and geographical contexts. It can assess a wide range of learning outcomes in different specialties and disciplines for formative or summative purposes. It may be used to assess students, a curriculum or an educational intervention in the different phases of education and in different health care professions. The study suggests that one reason for the wide-scale adoption of the OSCE is its inherent flexibility in terms of the number of students that can be assessed, the number of examiners included, the type of patients represented (including standardised patients) and the format of the examination itself, including the length of the examination, the number and the duration of stations as well as its capacity in terms of the tasks to be requested from students.
Previous concerns regarding reliability of the exam when using, for example, standardized patients (instead of examiners) and global ratings (instead of checklists) were not confirmed in our study. Results point to the OSCE as a fair, acceptable, relevant and satisfactory exam, well received by students and teachers, as well as examiners and patients, who perceived the OSCE as an exam capable of educational impact with a steering effect on learning and teaching. Evidence was also found on face content validity, with the more exigent design corresponding to stations sampled against blueprinting and course objectives, incorporating contributions from other teachers and other experts, with final decisions on content being reached through consensus meetings.
The OSCE, because of its unique benefits, is recommended in spite of, in some circumstances, being expensive to administer. Alternative forms were found to reduce OSCE costs and transparent categories are needed to report direct *vs.* indirect costs. A higher quality of reports is desirable, namely in terms of economic viability, for this information to support schools when they decide on the OSCE implementation.

Conclusions – Although we must be aware of the possibility of a bias in the results - since the tendency is to publish more the stories of success than negative ones - the evidence that we produced helps to understand why, already a decade ago, Norman (2002) stated '*the objective structured clinical examination, with its multiple samples of performance, has come to dominate performance assessment*'.

RESUMO

Título – Uma Revisão Sistemática da BEME (Melhor Evidência em Educação Médica) sobre a exequibilidade, fiabilidade e validade do Exame Clínico Objectivo e Estruturado (*OSCE, Objective Structured Clinical Examination*) em educação médica pré-graduada.

Introdução – Como os métodos de avaliação 'tradicionais' apresentavam problemas, sobretudo a nível dos critérios psicométricos, em 1975, Harden et al. introduziram o OSCE para tentar resolver as múltiplas críticas relativas à avaliação clínica. Construído com base em múltiplas 'estações' objectivas, concebidas para avaliar uma série de competências clínicas ou práticas em circunstâncias análogas (mesma avaliação, mesmos pacientes, mesmos examinadores, mesma duração), o OSCE tem tido uma utilização crescente em todo o mundo.

Objectivos – Produzir evidência científica quanto à capacidade do OSCE enquanto método adequado para avaliar os resultados da aprendizagem ou, por outras palavras, até que ponto é o OSCE um método de avaliação exequível, fiável, válido, justo, aceitável e com impacto educacional, quando usado para avaliar competências em educação médica pré-graduada?'

Métodos – A metodologia da BEME foi aplicada por dois avaliadores independentes, que analisaram a literatura de 1975 até ao final de 2008. Foram codificados 1065 dos 1085 estudos identificados.

Resultados – A evidência aponta para o OSCE como sendo um método exequível para a avaliação das competências clínicas, que pode ser usado em diferentes contextos culturais e geográficos. O OSCE é capaz de avaliar uma vasta gama de resultados de aprendizagem em diferentes especialidades e disciplinas, com fins formativos ou sumativos. Pode ser usado para avaliar estudantes, um currículo ou uma intervenção educacional, em diferentes fases de ensino, numa enorme variedade de profissões na área da saúde. O estudo sugere que uma das principais razões para a adopção deste método em larga escala é a sua inerente flexibilidade em termos do número de alunos que podem ser avaliados, número de examinadores, tipo dos pacientes (incluindo pacientes normalizados), bem como o seu formato, incluindo a duração do exame, o número e duração das estações bem como as suas potencialidades em termos de tarefas a serem exigidas aos alunos. Preocupações anteriores sobre a fiabilidade do OSCE não foram confirmadas neste estudo quando se utilizam pacientes normalizados (em vez de peritos) e escalas de classificação global (em vez de listas de verificação). Os resultados sugerem o OSCE como um exame justo, relevante e recebido com satisfação por alunos e docentes bem como examinadores e pacientes que reconhecem o impacto do OSCE na aprendizagem e no ensino. A evidência aponta ainda para a 'validade de face e de conteúdo' do OSCE, onde o formato mais exigente pressupõe a selecção das estações com base em *blueprinting* e objectivos do curso, com contributos de outros docentes e outros especialistas e com as decisões sobre o conteúdo final a serem alcançadas através de reuniões de consenso.
A utilização de OSCE é recomendada devido aos seus benefícios únicos apesar, de em algumas circunstâncias, ser um exame dispendioso. Foram identificadas formas alternativas para reduzir os custos do OSCE e são necessárias categorias transparente que permitam descrever quais os custos directos *vs.* indirectos. É desejável uma maior qualidade dos estudos primários, nomeadamente em termos da viabilidade económica do OSCE, para que esta informação possa apoiar as escolas médicas quando estas se encontram no difícil processo de decisão quanto à implementação deste método de avaliação.

Conclusões – Embora tenhamos que ter em conta a possibilidade de algum enviesamento nos resultados, uma vez que a tendência é de publicar estudos de sucesso, a evidência apurada ajuda a perceber porque é que há 10 anos atrás Norman (2002) afirmava que *'o exame clínico objectivo estruturado, com os seus múltiplos formatos de desempenhos, veio para dominar a avaliação de desempenhos'*.

Table of Contents

OUTLINE OF DISSERTATION **21**

CHAPTER 1
Introduction to the systematic review. Setting the scene **25**
- Background 27
- Aim of the study and instrumental objectives 29
- Examining the quality of BEME methodology and of OSCE reports 29
- Methods 29
  - Establishment of a working Systematic Review Group 30
  - Framing the research question(s) 30
  - Defining inclusion/exclusion criteria 30
  - Developing a search strategy 31
  - Retrieving the material under analysis 31
  - Developing an OSCE Reference Manager database 31
  - Developing an OSCE electronic database 32
  - Coders' training and pilot phase 32
  - Analysing and coding of primary studies 32
  - Establishment of consensus 32
  - Analysing data 33
  - Discussion and Synthesis 33
  - Conclusions and application to practice 33
- References 33

CHAPTER 2
Examining BEME as a trustful mean to produce evidence on OSCE. Comparing Cochrane Reviews with BEME reviews **35**
- Abstract 37
- Practice Points 37
- Introduction 37
- The Definition of Causality in Clinical Medicine 38
- Causality in Medical Education 40
- Cochrane Reviews 41
- Best Evidence Medical Education Reviews 41
  - Step 1 – Formulating an educational relevant question 42
  - Step 2 – Search and selection of studies to be included in the review 42
  - Step 3 – Codification and critical appraisal of evidence 42
  - Step 4 – Synthesis of the evidence 42
  - Step 5 – Practical application of the results 43
- Discussion 43
- Notes on Contributors 44
- References 44

CHAPTER 3
Examining the quality of the OSCE primary study reports. A proposal for a comprehensive checklist to improve reporting of OSCE **47**
Abstract 49
Practice Points 49
Introduction 49
Methods 50
Study objectives 50
Working group 50
Search Process 50
Codifying process – OSCE online database 50
Purpose of OSCE in report 50
Results 50
Methodological difficulties in terms of reporting OSCEs 50
Problem I - Data Missing 50
Problem II: Inconsistent Data 52
Heterogeneity in reporting on validity, reliability and feasibility 52
Validity 52
Reliability 52
Feasibility 53
Lack of a standardized vocabulary in reporting 54
Statistical issues 54
Lack of structure within reporting 57
Discussion 57
Conclusions 57
Notes on Contributors 57
References 57
Appendix 1 58
Appendix 2 61

CHAPTER 4
Evidence on technical and economic feasibility. Is the OSCE a feasible method for assessing undergraduate medical students? **63**
Abstract 65
Practice Points 65
Introduction 65
Methods 66
Criteria for inclusion 66
Coding of papers 68
Results 68
Evidence on technical feasibility 68
Evidence based on all retrieved studies 68
OSCE publications 68
Different cultures and contexts 68
Multi-professional Settings 69
Phases of Education 69
Evidence based on all studies with data on the research questions 69
Stage of curriculum 69
OSCE purpose 70
Type of exam 70
Role of exam 70

Feedback 70
Learning Outcomes 70
Subject areas 70
Students 70
OSCE design and time 71
Recording and scoring (ratings) 71
Examiners 72
Staff and patients 72
Evidence on economic feasibility 73
Evidence based on all studies reporting data on OSCE costs and/or time 73
Discussion 76
Acknowledgement 77
Notes on Contributors 78
References 78
Appendices 79
Appendix 1. BEME Methodology 79
Appendix 2. OSCE studies included in the systematic review 79
Appendix 3. OSCE electronic coding sheet 79

CHAPTER 5
Evidence on OSCE assessment criteria. Is the OSCE meeting the requirements for a good assessment tool? 81
Abstract 83
Practice Points 83
Introduction 83
Methods 84
Results 85
Reliability 85
Inter-station reliability 85
Reliability of global ratings compared to station-specific checklists 85
Reliability of expert examiners compared to standardized patients evaluators 85
Inter-rater reliability 86
Predicted overall reliability with varying station number 86
Validity 88
Validity against criterion measure (concurrent validity) 88
Face/content validity 88
OSCE Fairness 89
OSCE Acceptability 91
OSCE Relevance 91
Satisfaction with OSCE 93
OSCE Educational impact 95
Steering effect on learning 95
Steering effect on teaching 97
Discussion 99
Evidence on OSCE reliability 99
Evidence on OSCE validity 99
Evidence on OSCE fairness 99
Evidence on OSCE acceptability 99
Evidence on OSCE educational impact 99
Conclusions 102
Limitations of the Study 102

Acknowledgement 102
Notes on Contributors 103
References 103
Appendices 107
Appendix 1. Studies contributing to reliability 107
Appendix 2. Studies contributing to validity 107
Appendix 3. Major difficulties when reporting reliability and validity 108

CHAPTER 6
Concluding remarks 109
Summary of the BEMER process: from questions to results 111
Evidence on OSCE technical feasibility 111
Evidence on OSCE economic feasibility 111
Evidence on OSCE reliability 112
Evidence on OSCE validity 112
Evidence on OSCE fairness 112
Evidence on OSCE acceptability 112
Evidence on OSCE impact on learning and teaching 112
Critical appraisal of the BEMER 113
How pertinent was the educational research question? 113
How exhaustive was the process of localizing and selecting the evidence? 113
How rigorous was the critical appraisal of the evidence? 113
Examining the quality of BEME as a credible approach to obtain evidence 113
Examining the quality of primary studies in which the evidence is based 114
Examining the quality of data analysis to extract the evidence 114
How trustworthy was the integration of findings with educational judgment? 115
How comprehensive was the transfer of evidence to induce changes in practice? 115
Final considerations on BEMER 115
Limitations of the study 116
OSCE: The way forward 117
OSCE as an 'authentic exam' 117
OSCE as an 'assessment for learning', not just 'assessment of learning' 117
OSCE as an 'exam with wide applicability' 118
OSCE as a 'response to the continuum of education' 118
OSCE as a 'feasibility exam used in a wide range of contexts' 118
OSCE as a 'flexible exam contributing to teachers ownership' 118
OCSE as an 'exam capable of assessing professionalization in a global world' 118
OCSE as an 'exam where new technologies will bring new improvements' 118
OCSE as an 'exam where students will be used as examiners' 118
OSCE as an 'overall assessment' 119
OSCE as an 'exam contributing to research in medical education' 119
References 119

ANNEXES 121
Appendix 1. BEME methodology 123
Appendix 2. References of OSCE papers regarding this BEMER 127
Appendix 3. Structure of OSCE electronic coding sheet 157

List of Tables

CHAPTER 2
Table 1. Evidence of causality between factor and effect 39

CHAPTER 3
Table 1. Purpose of OSCE in report 51
Table 2. Areas where information is missing 52
Table 3. Evidence for validity 53
Table 4. Evidence for realibility 53
Table 5. Evidence for feasibility 54
Table 6. Comprehensive checklist for reporting OSCE studies 55

CHAPTER 4
Table 1. OSCE costs 73
Table 2. Proposals to decrease OSCE costs 75

CHAPTER 5
Table 1. Inter-station reliability 86
Table 2. Effect of context variables on inter-station examiner/checklist reliability 87
Table 3. Inter-rater Reliability 87
Table 4. Validity against criterion measure 88
Table 5. Station design 89
Table 6. OSCE fairness 90
Table 7. OSCE Relevance 92
Table 8. Satisfaction regarding the OSCE 94
Table 9. OSCE steering effect on learning 96
Table 10. OSCE educational value 97
Table 11. OSCE steering effect on teaching 98
Table 12. Comparison the reliability results 100

CHAPTER 6
Table 1. Overall quality control of the BEMER process 116

List of figures

CHAPTER 1
Figure 1. Miller's pyramid (1990) 28

CHAPTER 2
Figure 1. A parallel-group randomized controlled trial 39
Figure 2. Confounding variables 41

CHAPTER 4
Figure 1. Criteria and number of studies per level of analysis 67
Figure 2. Number of OSCE publications per year 68
Figure 3. Publications and date of first publication per continent 69
Figure 4. Distribution of OSCE exams per number of students 71
Figure 5. Distribution of OSCE exams per OSCE total duration in minutes 71
Figure 6. Distribution of OSCE exams per station time in minutes 72
Figure 7. Distribution of OSCE exams per number of examiners 72
Figure 8. Distribution of OSCE exams per number of simulated patients 72

CHAPTER 5
Figure 1. Expected overall reliability *vs.* number of stations 87

List of Appendices

CHAPTER 3

Appendix 1. Studies included in the BEME Systematic Review used for this study 58
Appendix 2. Studies not included in the OSCE Systematic Review 61

CHAPTER 4

Appendix 1. BEME methodology 79
Appendix 2. OSCE papers included in the systematic review 79
Appendix 3. OSCE electronic coding sheet 79

CHAPTER 5

Appendix 1. Studies contributing to reliability 107
Appendix 2. Studies contributing to validity 107
Appendix 3. Major difficulties when analysing OSCE reliability and validity 108

OUTLINE OF DISSERTATION

A Best Evidence Medical Education (BEME) Systematic Review on the feasibility, reliability and validity of the Objective Structured Clinical Examination (OSCE) in undergraduate medical studies.

Outline of Dissertation

The dissertation concerning a Best Evidence Medical Education systematic review (BEMER) on the feasibility, reliability and validity of the OSCE (Objective Structured Clinical Examination) in undergraduate medical studies is structured in six chapters. The contents for each chapter are presented below.

CHAPTER 1 – Introduction to the BEMER systematic review. Setting the scene

CHAPTER 2 – Examining BEME as a trustful mean to produce evidence. Comparing Cochrane Reviews with BEME reviews

The results of this chapter are reported in a paper published in Medical Teacher*:*
Patricio M, Carneiro AV. (2012). *Systematic reviews of evidence in medical education and clinical medicine: is the nature of evidence similar?* Medical Teacher 34: 474-482.

CHAPTER 3 – Examining the quality of the OSCE primary study reports. A proposal for a comprehensive checklist to improve reporting of OSCE

The results of this chapter are reported in a paper published in Medical Teacher*:*
Patricio M, Julião M, Fareleira F, Young M, Norman G, Carneiro AV. (2009). *'A comprehensive checklist for reporting the use of OSCEs',* Medical Teacher 31: 112-124

CHAPTER 4 – Evidence on technical and economic feasibility. Is the OSCE a feasible method for assessing undergraduate medical students?

The results of this chapter are reported in a paper accepted for publication in Medical Teacher*:*
Patricio, M, Julião M, Fareleira F, Carneiro AV. *Is the OSCE a feasible tool for accessing competencies in undergraduate medical education? Evidence from a BEME systematic review.*

CHAPTER 5 – Evidence on OSCE assessment criteria. Is the OSCE meeting the requirements for assessment in undergraduate medical education?

The results of this chapter are reported in a paper submitted to Medical Teacher:
Patricio, M, Julião M, Fareleira F, Young M, Norman G, Carneiro AV. *Is the OSCE a reliable and valid tool for assessing to assess competencies in undergraduate medical education? Evidence from a BEME systematic review.*

CHAPTER 6 – Concluding remarks

ANNEXES

CHAPTER 1
Introduction to the systematic review
Setting the scene

A Best Evidence Medical Education (BEME) Systematic Review on the feasibility, reliability and validity of the Objective Structured Clinical Examination (OSCE) in undergraduate medical studies.

Introduction to the systematic review
Setting the scene

Background

Assessment is a topic highly regarded in the medical education community. This is understandably so because its multifaceted nature: '*it drives and stimulates learning, provides information on educational efficacy to institutions and teachers, and protects patients'* (Norcini et al. 2011). Assessment is a crucial factor in medical education due to its role in the certification of students' competencies, with subsequent short and long-term consequences for students' progress. For Rowtree (1987), the consequences of a rigorous assessment are the most important throughout the educational process, and this is why '*assessment may motivate or destroy a student'*.

Students are aware of this importance and, already in 1971, Snyder stated that '*students do what they know they will be asked to do and do nothing if they know they will not be assessed*'. The recognition of the importance of assessment explains why Newble et al. (1994) stated that '*to invest in assessment is the same as to invest in education*'. Many studies confirm the impact of assessment in the learning-teaching process, namely in the complex interaction between 'curriculum in paper', 'curriculum in action' and 'curriculum experienced by the students' (Grant & Gales 1989).

Due to this central role, more accountability is needed and great pressure is put on medical schools regarding the use of assessment methods. Among them the 'assessment of clinical competences' is probably the most difficult area, because it must guaranty that graduates acquired the necessary knowledge, skills and attitudes needed for future professional practice. During the last decades we witnessed intense debates on clinical assessment, because unfortunately there is not one 'best assessment' and all approaches present problems and limitations.

The major difficulties with 'traditional methods' (long cases, short cases, oral examinations, etc.) relate to psychometric criteria - namely in terms of reliability and validity of the assessment method (examiner variation, unstructured marking, random content, etc.). This is a major problem in the assessment area and new approaches have been developed.

Given the aforementioned problems, in 1975, Harden et al. introduced the Objective Structured Clinical Examination (OSCE) as an attempt to solve them. The OSCE consists of multiple 'objective stations' designed to assess a range of clinical skills under similar circumstances (standardized assessment, standardized patients, standardized examiners and standardized duration).

Throughout 37 years, the OSCE has been widely used to assess multiple competencies, for different purposes, in a range of professional bodies from all phases of education and, to our knowledge, no other method of teaching or assessment has known such a great impact in medical education. Soon it was considered the principal method to assess clinical learning outcomes, recognized as an important contribution for the well known problems associated with assessment of clinical competences.

The two features that made the OSCE an important tool in the assessment toolkit relate to its structure that was designed to increase the reliability and validity of the clinical assessment. To achieve higher reliability (capacity of the exam to present reproducible scores) the format is highly

standardized, with candidates performing the same assignment in each station, marked through the same scoring tool, by the same examiner, or different examiners in different stations (multiple independent assessments), interact and examine the same patient in each station, under the same timing scale. For the achievement of higher validity (capacity of the exam to measure what it is supposed to measure) the OSCE requests demonstration of skills, with stations supposed to be blueprinted and sampled according to course objectives. In an OSCE exam the candidates are assessed at the 'Shows How' level purposed by Miller in 1990 (Figure 1). They must demonstrate they have acquired the skills they are supposed to master, which is also an indicator of OSCE validity.

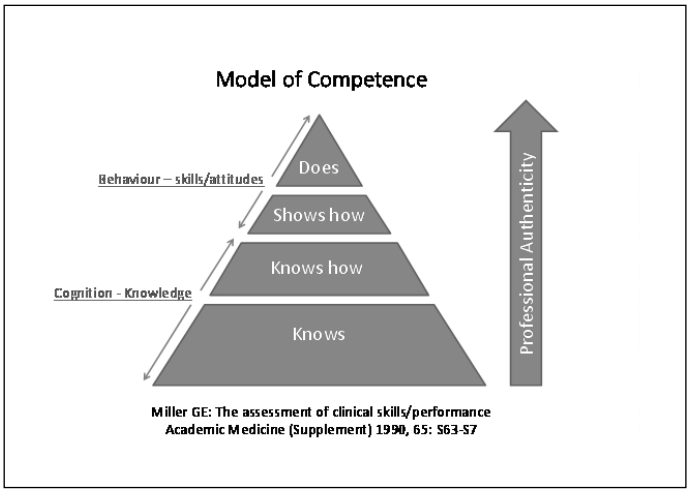

Figure 1. Miller's pyramid (1990).

Another important and interesting characteristic of the OSCE relates to its flexibility i.e. the capacity of being practicable for multiple purposes, for different types of exams, with different designs to accommodate high and low number of students, at different timings, testing a different range of learning outcomes, and using real/simulated patients and examiners with different backgrounds (teachers, experts, patients, students, to name a few).

Due to its structure, the OSCE meets the requirements for 'best practice in assessment', because: 1) it clarifies the purpose of the exam, 2) it defines what is to be tested, 3) it selects appropriate tasks, 4) it addresses practical and technical issues of administration and scoring, and 5) it sets standards for performance, as defined by Newble et al. (1994).

Independently of their use, expansion and success, all assessment methods must be evaluated. Also, medical schools are requested to comply with standardisation criteria. Previously, requirements were focused in the feasibility, reliability and validity of the assessment tools but in recent years the importance of further criteria that relates to the closer relationship between assessment and teaching has been added. Examples of those additional requirements were found in literature for example:

- van der Vleuten in 1996 proposed we should also look at impact on future learning and practice, acceptability to learners and faculty, and costs.
- The 14th Ottawa Conference in 2010 - a meeting fully dedicated to the Assessment of Competence in Medicine and the Healthcare Professions – outlined as additional criteria educational effect, catalytic effect, and acceptability (Norcini et al. 2011).
- The General Medical Council in 2011 set out the specific requirements: educational impact, cost effectiveness, acceptability and defensibility.

At a time when decisions in medical education should not be based only in personal opinions, a new attitude demanding a specific culture to make teachers critically analyse their own practice is needed, as suggested by Davies & Crombie in 2003, who stated that *'the systematic reviews are essential because we never know as much as we think we know'*.

This need for evidence also applies to the use of OSCEs namely on how the exam is meeting the assessment criteria requirements because independently of their use, expansion and success, all assessment methods must be evaluated.

This was why a Best Evidence Medical Education Review (BEMER) (http://www.bemecollaboration.org), was undertaken at the Centre for Evidence Based Medicine (CEMBE) of the Faculty of Medicine of the University of Lisbon, to comprehensively look at the OSCE assessment criteria i.e. at its feasibility, reliability, validity, fairness, acceptability and educational impact. This dissertation reports on the process and results of this BEMER.

Aim of the study and instrumental objectives

The aim of the study was to analyse scientific evidence about the OSCE, by means of a BEMER, on its suitability to assess learning outcomes in undergraduate medical studies (and only these).

Two instrumental objectives were defined:

- To characterize OSCE technical and economic feasibility
- To gather evidence on the reliability, validity, fairness, acceptability and educational impact of the OSCE.

Examining the quality of BEME methodology and of OSCE reports

The BEME systematic review approach was examined to identify if it would be a trustful mean to produce evidence on the OSCE. Additionally, the quality of primary studies was examined after reviewing the initial 400 studies. The objective of these critical analyses was to appraise the quality of the evidence we would obtain in the context of this BEME systematic review.

The results of the appraisal process are reported in two papers already published and included in:

- **Chapter 2 - Examining BEME a trustful mean to produce evidence on OSCE. Comparing Cochrane Reviews with BEME reviews:**
  Patricio M, Carneiro AV. (2012) '*Systematic reviews of evidence in medical education and clinical medicine: is the nature of evidence similar?* Medical Teacher 34: 474-482.
- **Chapter 3 - Examining the quality of OSCE reports reviews:**
  Patricio, M, Julião M, Fareleira F, Young M, Norman G, Carneiro AV. (2009) *'A comprehensive checklist for reporting the use of OSCEs'*. Medical Teacher 31:112-124.

Methods

BEME methodology - as described in the BEME protocol (www.bemecollaboration.org) was undertaken based on the following steps:

(1) Establishment of a working Systematic Review Group
(2) Framing the research question(s)
(3) Defining inclusion/exclusion criteria
(4) Developing a search strategy
(5) Retrieving the material under analysis

(6) Developing an OSCE Reference Manager database
(7) Developing an OSCE electronic database
(8) Coders' training and pilot phase
(9) Analysing and coding of primary studies
(10) Establishment of consensus
(11) Analysing data
(12) Discussion and synthesis
(13) Conclusions and application to practice

1. Establishment of a working Systematic Review Group

A working group in Lisbon was constituted by the coordinator (a MD-PhD and MSc in Medical Education), a research director (Educationalist and MSc in Medical Education - also acting as a coder), two coders (final-year medical students) and two administrative assistants.

Included in the team there was two international consultants: a PhD from the Department of Clinical Epidemiology and Biostatistics at McMaster University, Canada, and a PhD from the Centre for Medical Education at McGill University, Canada.

2. Framing the research question(s)

Whether the OSCE is feasible, reliable and valid as a method of assessment of learning outcomes in undergraduate medical studies were the initial research questions for this BEME Systematic Review. Later, due to current educational developments, other questions were added concerning new assessment criteria requirements, introduced by the GMC in 2011: fairness, acceptability (i.e. OSCE relevance and satisfaction with the OSCE) and educational impact (i.e. OSCE capacity of steering learning and teaching).

3. Defining inclusion/exclusion criteria

Only English studies reporting on 'classical OSCE exams' performed in undergraduate medical education were included in the study. Therefore the following primary studies were excluded:

- Non undergraduate studies
- Non medical studies
- Non English studies
- Non 'classical' OSCEs
- Non primary studies
- OSCE studies for teaching students
- OSCE studies for training teachers.

A study was coded 'non-classical' when it did not conform in general terms with the classical approach of the OSCE, as described by Harden et al. in 1975. Among them we found studies where the candidate was a 'team' or 'group' instead of an individual - for example, TOSCE (Singleton et al. 1999), G-OSCE (Hill et al. 1994), GOSCE (Elliot et al. 1994; Fields et al. 1995; Vooijs et al. 1997), GOSPE (Biran 1991), when assessment was based on video instead of direct observation, for example VIPSCE (Shallaly & Ali 2004), OSVE (Humphris & Kaney 2000), where exams had only written stations (Akici et al. 2004), where exams were only peer-rated (Geddes & Crowe 1998) and 'non-classical' formats, for example OSCEs with only one or two stations (van Dalen et al. 2001; Robins et al. 2001). 'Non-classical' studies were excluded, to avoid a bias when calculating the reliability of the OSCE.

Secondary studies were excluded because systematic reviews must be based on primary studies only. Also excluded were studies where OSCEs' objectives were 'to teach students or train teachers', because the objective of this BEMER concerns the feasibility, reliability and validity of the OSCE when implemented for assessing learning outcomes.

Long case exams and OSLERs which appear as a result of the search were also excluded from analysis.

4. Developing a search strategy

Literature was searched by a BEME information scientist, from 1975 (date of the first publication on the OSCE) until the end of 2008. All identified references were inserted into a Reference Manager database.

Two different phases were considered:

- Phase I - Literature search from 1975 to end of 2001
  We started with the OSCE database material published by Harden et al. in 2003, which covered the OSCE literature from 1975 till the end of 2001. The 712 references were identified through:
  - Electronic search of medical, educational & related databases
  - Hand search on 6 key medical education journals
  - Search of TIMELIT reference database
  - Search of Gray Literature (for example the Proceedings of Ottawa Conference)
  - Search on specialised literature collections, at the Medical Education Centre, University of Dundee.

  The key journals selected for searching were: Academic Medicine, Medical Education, Medical Teacher, Teaching & Learning in Medicine, Advances in Health Sciences Education and Education for Health.

- Phase II - Literature Search from 2002 to the end of 2008
  Previous search was updated until the end of 2008 by the same BEME information scientist who made the initial search. These references were electronically identified (electronic searches have improved considerably since the initial run and there was no need to repeat the intensive hand searches labour) and TIMELIT (used in phase I) was abandoned later.
  The key words used in both phases were base terms which were tested and adapted: 'OSCE', 'OSPE', 'GOSPE', 'objective structured clinical exam$', 'objective structured practical exam$', 'structured clinical exam$', 'structured clinical interview$' (the truncation symbolic $ is fairly generic and is used to pick up all alternative endings).

5. Retrieving the material under analysis

When a reference was identified the process of retrieving the paper started immediately. This was easier in Phase I because the papers were sent by the Medical Education Centre at the University of Dundee. For Phase II the process was more difficult: the majority was obtained through the libraries of the University of Lisbon and the University of Columbia. Finally, the editors of journals of non-retrieved papers were also approached to obtain missing papers.

6. Developing an OSCE Reference Manager database

An OSCE Reference Manager database was created to include the list of identified references. The objective was to facilitate a quick identification of a study through its author(s), date, journal, title, etc., and the insertion of references in publications.

The software Reference Manager is one of the most reliable databases management programs available to the academic world and has the advantage to be compatible with most bibliographic databases. All elements of the working team were trained in its development and use.

7. Developing an OSCE electronic database

A new 'online database' (Lotus software) was developed, since the existing BEME coding sheet was not applicable in our systematic review. Items for the new coding sheet were defined by the whole team according to research questions. Literature was blueprinted and the new database served as a coding sheet supporting coding and establishment of consensus on line. The database was structured upon four main sections: 1) information on publication, 2) background of OSCE exam, 3) results on OSCE feasibility, reliability and validity, and 4) study problems, solutions and conclusions, each of them including several fields. Full description of this software is presented in Chapter 4.

One of the most important characteristics of the new OSCE database was its dynamic structure i.e. a structure, which could be modified during the coding process by adding or reformulating a field. Moreover, the majority of the fields were 'open fields' which could be fed by new items when they show up during the coding process. These features were extremely important since when a systematic review starts its impossible for researchers to have the full picture of what is under investigation.

Fields such as 'existence of a previous pilot', 'number of sub-stations under analysis', 'sub-sample of students under analysis', 'total number of students in the course', 'relevance', 'fairness', are examples of fields inserted later during the coding process. They implied a second review of the papers analysed until that date. In what concerns the number of options within the same field, we found, for instance, 266 different types of OSCE aims, 273 Institutions responsible for OSCE publications and 45 different stations organized in 156 different combinations (depending on the studies).

These are just some examples showing that it would be extremely difficult to code such a complex exam with the traditional paper based coding sheet, as the fields of the coding sheet are defined before the coding process starts.

This new database allowed independent coding by each coder and establishing of consensus online.

8. Coders' training and pilot phase

Two coders were trained by the research director in the BEMER systematic methodology. Background literature was made available to them and several meetings took place to discuss the process of coding and how to reach consensus.

After the initial 'theoretical training period', a 'pilot phase' started concerning the coding of the first 75 papers. During the 'pilot phase' each paper was reviewed and inserted in the database by each independent coder. A discussion followed to justify each decision, before consensus was established leading to the coding of a 'consensus record'.

9. Analysing and coding of primary studies

Each paper was coded by two independent reviewers. As already stated, the new electronic OSCE database was used to support coding.

10. Establishment of consensus

A 'consensus meeting' occurred after independent coding of a certain number of records (usually no more than twenty) was made. Consensus was reached by comparing the classification of the two coders and as a result of the discussion a 'consensus record' was created for each record in the online OSCE database. Consensus could be achieved face to face or electronically.

Disagreements were discussed until consensus was achieved. When disagreements occurred, a discussion took place and, frequently, this implied the coders had to reread the paper before consensus could be established at the next session. During the pilot phase the establishment of consensus for a single paper could easily take more than one hour. Progressively the time allocated to establish consensus diminished and within an hour it was possible to code 4-6 papers depending on its complexity.

Consistency among coders was established throughout the process excluding the pilot phase (see results in Chapters 3, 4 and 5).

11. Analysing data

The procedures for data analysis were determined by each research question. For detailed information on the different levels of analysis see Chapters 3, 4, and 5. The analysis was supported by the Department of Clinical Epidemiology and Biostatistics at McMaster University and the Centre for Medical Education & Department of Medicine, McGill University, both in Canada.

12. Discussion and Synthesis

The discussion and synthesis of results was done in the context of each research question, involving the whole team. Results are presented in Chapters 3, 4, and 5.

13. Conclusions and application to practice

Conclusions were established in order to facilitate the transfer of results into practice when taking into account the limitations of the study.

References

Akici A, Kalaca S, Goren MZ, Akkan AG, Karaalp A, Demir D, Ugurlu U, Oktay S. 2004. Comparison of rational pharmacotherapy decision-making competence of general practitioners with intern doctors European Journal of Clinical Pharmacology 60(2):75-82.

BEME: Best Evidence Medical Education. Retrieved May 1, 2012. Available from: http://www.bemecollaboration.org

Biran LA. 1991. Self-assessment and learning through GOSCE (group objective structured examination). Medical Education 25(6):475-479.

Davies Ht, Crombie IK. 2003. What is a systematic review? Evidence Based Medicine 1(5):1-7.

Elliot DL, Fields SA, Keenan TL, Jaffe AC, Toffler WL. 1994. Use of a group objective structured clinical examination with first-year medical students. Academic Medicine 69(12):990-992.

Fields SA, Toffler WL, Elliot DL. 1995. Principles of clinical medicine: an interdisciplinary integrated 2-year longitudinal course. Medical Education 29(1):53-57.

Geddes EL, Crowe J. 1998. Peer-Related Objective Structured Clinical Examination. Physiotherapy Canada 50(4): 269-275.

General Medical Council. 2011. Assessment in undergraduate Education. Advice supplementary to Tomorrow's Doctors (2009). Retrieved May 1, 2012. Available from: http://www.gmc-uk.org/Assessment_in_undergraduate_web.pdf_38514111.pdf

Grant J, Gales R. 1989. Changing medical education. Medical Education 23: 252-257.

Harden RM, Stevenson M, Downie WW, Wilson GM. 1975. Assessment of clinical competence using objective structured examination. British Medical Journal 1:447-451.

Harden V and Harden RM, AMEE 2003, Dundee, UK: Association for Medical Education in Europe (AMEE).

Hill DA, Guinea AI, McCarthy WH. 1994. Formative assessment: a student perspective. Medical Education 28(5):394-399.

Humphris GM, Kaney S. The objective structured video exam for assessment of communication skills. 2000. Medical Education 34(11):939-945.

Miller GE. 1990. The assessment of Clinical skills/performance. Academic Medicine 65:S63-S67.

Newble D, Dauphinee D, Macdonald M, Mulholland H, Dawson B, Page G, Swanson D, Thomson A, van der Vleuten C. 1994. Guidelines for assessing clinical competence. Teaching and Learning in Medicine 6:3 213-220.

Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V, Roberts T. 2011. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. Medical Education 33:206-214.

Robins LS, White CB, Alexander GL, Gruppen LD, Grum CM. 2001. Assessing medical students' awareness of and sensitivity to diverse health beliefs using a standardized patient station. Academic Medicine 76(1):76-80.

Rowntree D. 1987. Assessing students: how shall we know them? London: Kogan Page.

Singleton A, Smith F, Harris T, Ross-Harper R, Hilton S. 1999. An evaluation of the Team Objective Structured Clinical Examination (TOSCE) Medical Education 33;1:34-41.

van Dalen J, Bartholomeus P, Kerkhofs E, Lulofs R, van Thiel J, Rethans JJ, Scherpbier A, van der Vleuten C. 2001. Teaching and assessing communication skills in Maastricht: the first twenty years Medical Teacher 23(39):245-251.

van der Vleuten CPM. 1996.The assessment of professional competence: developments, research and practical implications. Advances Health Sciences Education 1:41-67.

Vooijs MEEC, Scherpbier A, van Dalen J, Ramsay G, Kootstra G. 1997. Group objective structured clinical examination in an Inter-European accident and emergency course. Education for Health 10(1):69-78.

Shallaly GH, Ali EA. 2004. Use of Video-Projected Structured Clinical Examination (ViPSCE) instead of the traditional oral (VIVA) examination in the assessment of final year medical students. Education for Health 17(1):17-26.

Snyder BR. 1971. The hidden curriculum. New York, Knopft.

CHAPTER 2
Examining BEME as a trustful mean to produce evidence on OSCE. Comparing Cochrane Reviews with BEME reviews

2012; 34: 474–482 MEDICAL TEACHER

Systematic reviews of evidence in medical education and clinical medicine: Is the nature of evidence similar?

MADALENA PATRÍCIO & ANTÓNIO VAZ CARNEIRO
University of Lisbon, Portugal

Abstract

Background: It is accepted worldwide that clinical and educational decisions should be informed by the best available evidence, not individual opinion only.

Aims: This article discusses the epistemological basis of educational evidence, as compared with clinical evidence, looking at the different nature of the science behind each one.

Method: Two models – BEME Reviews in medical education and Cochrane Reviews in clinical medicine – based on our own experience of a soon to be published BEME Review (BEMER) and several systematic reviews our group has published in clinical medicine – were used to identify similarities and differences between the two approaches.

Key findings: The evidence to support clinical as well as educational decision making is different in its nature, as well as in its quality. However, their approach is similar in its fundamental steps (design a question, select evidence, critically appraise it, synthesize and apply), so the differences between BEME and Cochrane are perhaps more a matter of degree, than the existence of fundamental differences.

Conclusions: Two fundamental principles – decision making should be supported by a hierarchy of evidence and evidence alone is never sufficient for sound practice – apply to BEME and Cochrane reviews. The capacity to transfer their results into practice is the most important factor in terms of success of both approaches.

Introduction

The term 'evidence-based medicine' (EBM) is presently a common concept throughout medical systems around the world. Its use has grown exponentially and has taken a central role in clinical practice in several developed health systems.

It was firstly coined in 1991, but fully presented in 1992 when it was defined as a 'paradigm shift' (Evidence-Based Medicine Working Group 1992) because it considers unsystematic clinical experience of the individual doctor and pathophysiologic reasoning (two major tenets of classical clinical practice) to have lower value than scientific evidence from clinical research. Its proponents stress that evidence from research should constitute the basis for clinical practice, that interpreting its results requires a formal set of rules (critical appraisal), and places a lower value on authority than the traditional medical paradigm (Guyatt et al. 2008).

The two fundamental principles of EBM are that clinical decision making should be guided by a hierarchy of evidence and that evidence alone is never sufficient for sound clinical practice.

In the perspective of the EBM practitioner, to achieve the best possible results – that also take into account patient values – the clinician must combine his/hers own clinical experience with evidence from well-designed clinical studies on therapeutics, diagnostics, prognosis, etc. Some would call this the 'art of medicine'... and several studies have been published on the need, problems and practical challenges of EBM teaching (Aiyer & Dorsch 2008; van Dijk et al. 2010; Oude et al. 2011).

Evidence-based education (EBE) can be defined as the methodology by which educationalists and other teaching professionals (policy makers, managers, students, etc.) base

Practice points

- It is well accepted that the teacher's decisions should be informed by the best available evidence, not individual opinion only.
- The differences between BEME and Cochrane reviews are perhaps more a matter of degree of the supporting evidence than the existence of fundamental differences.
- BEME and Cochrane reviews are, and will remain in the foreseeable future, a very demanding task.
- The medical education community is aware of difficulties regarding BEME reviews – common to all human sciences – due to the holistic nature of the object of the study, in addition to lack of resources associated to financial constraints.
- The crux of the question is that BEME evidence should translate scientific knowledge into practice.

Correspondence: António Vaz Carneiro, Faculty of Medicine of the University of Lisbon, Center for Evidence Based Medicine, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal. Tel: +351 217 985 135; Fax: +351 217 940 424; E-mail: avc@fm.ul.pt

ISSN 0142–159X print/ISSN 1466–187X online/12/060474–9

their decisions to support medical education interventions. It replaces 'opinion' with 'evidence'.

The systematic approach to science as a basis for practice has been increasingly present in the area of medical education, and culminated in the Best Evidence in Medical Education (BEME) initiative, launched in 1999. BEME *is the implementation, by teachers in their practice, of methods and approaches to education based on the best evidence available. It involves a professional judgement by the teacher about their teaching taking into account a number of factors of the QUESTS dimensions namely the quality, utility, extent, strength, target and setting of the evidence available* (Harden et al. 1999)

Underpinning EBM and EBE are the scientific data provided by studies furnishing information for clinical practice and medical education. This information has attained astonishing dimensions: presently there exist more than 25,000 journals in science, technology, and medicine, increasing at a rate of 3.5% a year; in 2009, they published 1.5 million articles. PubMed now has more than 20 million papers and EMBASE is not far from this number (Fraser & Dunstan 2010). In addition, there are several more databases available for searching.

Given the fact that for each research question asked there may exist several studies trying to answer it, sometimes with contradictory results, it makes sense to look at all the results and combine them if possible. If, for example, we want to decide on a specific method of assessment the ideal would be to be able to take an 'informed decision' based on results of all studies that have looked at the effectiveness of that approach.

Until the 1990s the experts had the task of combining the results of several studies in what are today called 'narrative reviews': he/she would select (without saying how) the studies that addressing the question, would leave out others (again without a explicit exclusion methodology), would summarize the results of the included ones and, finally, synthesized them, hopefully coming to a specific conclusion (Borenstein et al. 2009).

The problem with this approach is that this is a rather subjective way to get the right conclusions: if we would ask several experts, most likely we would get several answers – one for each analysis – based on different criteria used to select the papers and appraise them (one expert would favour sample size, other follow-up, yet other statistical analysis, etc.). The lack of transparency of this approach also leaves the reader wondering about the accuracy of the findings, especially when two reviews get opposite conclusions (a regular finding in the literature). Finally, as new evidence is accrued to the field, the review becomes less and less useful, not least because some of the research areas these days produce hundreds of studies per year.

To overcome these methodological problems, groups of researchers developed the methodology of so called 'systematic reviews'. This integrative/secondary research made their way into publications, firstly in social sciences, then in medical journals, and finally in medical education journals. This secondary research progressively imposed itself in mainstream publications, and systematic reviews (SR) in medical education are being published more frequently, even though they are far from the number found in clinical medicine. The reason for this is that there is still a somewhat small number of studies that can be included in a medical education SR, so it is not possible to perform this type of analysis as frequently as desired.

One of the main scientific issues concerning SR in general (and in medical education in particular) is the nature of the evidence used to elaborate them. What counts for good evidence? What are the methodologies supporting the elaboration of systematic reviews? How to interpret the results of these reviews? How to apply them in practice? When to update them?

In this paper we will briefly discuss the epistemological basis of educational evidence, as compared with clinical evidence, looking at the different nature of the science behind each one. We will be using for this purpose two models: BEME Reviews in medical education and Cochrane Reviews in clinical medicine, using our own experience of a soon to be published BEME Review on the reliability, validity and feasibility of the Objective Structured Clinical Examination (OSCE) in undergraduate-medical education, and several systematic reviews our group has published in clinical medicine, using also the model of a Cochrane Review.

The definition of causality in clinical medicine: The randomized controlled trial as a model for determining therapeutic efficacy

The causal relationships in medicine and health care in general are of central importance, because answers to several types of clinical questions depend on the assessment of whether a cause and effect relationship truly exists (Haynes et al. 2006).

The definition of causality is that a factor is a cause of an event if its operation increases the frequency of that event.

For example, whenever we ask a diagnostic question (*does this test confirm the presence of disease?*), a therapy question (*does a specific treatment improve patient's condition?*), an aetiology question (*does this risk factor cause that disease?*), or a management question (*do different systems improve services to patients?*) the answers depend on the assessment of the existence of a cause-effect relationship, i.e. the central issue of causation is whether any association seen between an intervention and an outcome indicates a cause–effect relationship, or whether the association is merely spurious (Rothman 2002).

In medical science, there are two types of cause– effect relationship to help define causality: contributory causes and Koch's postulates (Mayer 2004).

When we are faced with complex multifactorial diseases, the best way to prove causation is by applying the so called contributory causes, also called Hill's concepts of causality (after the British epidemiologist A. Bradford Hill that defined them in 1965) (Hill 1965; Gehlbach 2002). These are presented in Table 1.

Other – and different – support for putative cause/effect involves an analysis of the direct relationship between causal

factor and outcome (or effect) – a soft version of the so-called Koch's postulates (Elwood 2007). In this approach:

- a *necessary* causal relationship is one in that the outcome occurs only if the causal effect has operated before
- a causal relationship is *sufficient* if the operation of the causal effect always results in the outcome
- *both* if the causal effect and the outcome have a fixed relationship (neither occurs without the other) and
- *neither* (the most important category) when the operation of the causal factor increases the frequency of the outcome, but this one does not always result, or it can occur without the previous operation of the causal effect.

For example, arterial hypertension and stroke is an example of the latter: hypertension is a major risk factor for stroke, but the great majority of hypertensive patients never develop a stroke, and a very significant percentage of stroke patients are normotensive.

Because most of the time the biological/clinical events are classified in terms of causal relationship as *neither*, one needs a quantitative assessment of causality using observations collected at a group of people/patients, as opposed to a single person.

Hence, the concept of causation in clinical medicine – when talking of therapy or prevention – is best defined and analysed using the design of the randomized clinical trial (RCT), the only design that can answer the question: 'Do the results support a causal relationship between intervention and outcome?' If an association is found, then it may be due mostly to one (or more) of the following mechanisms (Elwood 2007): observation bias, operation of confounding factors, chance and true causality. Hence, the reliable assessment of a beneficial effect of treatment on a pre-determined outcome requires that the studies will control for bias (using the same methods in all the subjects of the study, under the same circumstances, by the same researchers, together with blinding techniques), diminish the influence of potential confounding factors (for example using proper randomization with or without stratification for known factors) and reduce chance (by applying rigorous statistical analysis), so one can believe the results being truly causal.

A typical parallel group RCT (Figure 1) is composed of two groups of patients similar in all the relevant characteristics (factors other than the one under study that affect the frequency of the predefined outcome), constituting a representative sample of the population and treated the same way, except that one of the groups (usually named experimental) suffers an intervention (drug, surgery, etc.). The results are then compared with the other group (control) that was not exposed to the intervention. Everything else being the same, any differences present in the outcome will have to do with the intervention. This effect (benefit or harm) is quantified through calculation of the so called measures of association: classically, relative risk reduction, absolute risk reduction (ARR) and number needed to treat.

Table 1. Evidence of causality between factor and effect (Hill 1965).

Criteria	Comments
Temporality	*Cause precedes effect*
Strength	*Strong cause-effect association (larger relative and absolute risks)*
Dose-response	*The bigger the causal factor, the larger the effect*
Reversibility	*Removal of causal factor diminishes the risk of disease*
Consistency	*Analogous results over different studies*
Biologic plausibility	*Consistent with recent medical knowledge and with a clear scientific conceptual base*
Specificity	*One cause, one effect*
Analogy	*Similar causes for similar diseases*

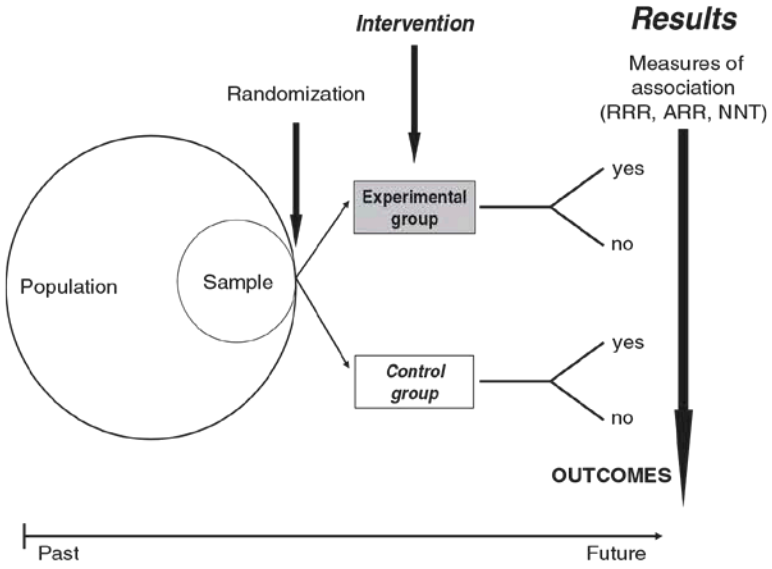

Figure 1. A parallel-group randomized controlled trial.

After determining a beneficial result, the clinician has to apply this research evidence to the patients under his/her care. Defining, in this context, what a treatment is will involve not only common sense, but also the concepts of reliability and relevance (also called internal and external validity, respectively) (Rothwell 2007)

The first is defined by how consistently good the data is only in terms of its intrinsic quality (its internal validity – with its design, can the study answer the scientific question?). The second – relevance – is defined as the degree of appropriateness of the data to the problem of the patient (external validity – can the results be generalized to patients other than the ones included in the study?). Of course this study design constitutes a direct test of causation, but the decision as to whether a relationship is causal must ideally be balanced by professional judgement.

In short, the determination of causality in clinical medicine is best achieved through a study with an experimental design, of which the RCT is the gold standard for determining the benefit of therapeutic or preventive interventions. If controlled for bias, confounding and chance, then one has the basis to believe in the trueness of the results.

Causality in medical education

At a time when educators and policy makers are increasingly aware that decisions in medical education should not be based solely on personal opinions or previous experience, there is a need to invest in scientifically based evidence approaches to support decisions. Also, when society requires that medical schools are accountable, there is a need for medical education researchers to identify the causal relationship between the intervention (curriculum, method, etc.) and students' outcomes.

We believe that all that was said before remains true when it comes to the definition of causality in medical education. Regarding the four premises proposed by Koch, a necessary, sufficient, fixed relationship from causal factor and outcome, and the possibility for outcome to occur without previous operation of the causal factor are valid when applied to medical education, only the identification of the relationship from cause and outcome may be difficult to interpret correctly.

This is not new, as in 2005, Hammick already stated that '*evidence informed practice is increasingly the norm for many professionals, emerging from the seminal discourse of evidence based medicine, through the work of (mainly) the Cochrane Collaboration'*. Also, '*evidence based practice is seen as the way to link knowledge from either primary research or systematic reviews and the logical and reliable application of that knowledge in professional practice. In keeping with this movement, systematic reviews with the aim of providing evidence about the effectiveness of an education intervention are now contributing to knowledge about health care sciences education and providing a valuable resource for education practitioners and policy makers'*.

These were the arguments behind the inception of BEME in 1999, but the complexity of writing systematic reviews in medical education may explain the fact that we did not witness the same exponential increase in the number of BEME published systematic reviews as has the ones occurring in clinical medicine. There are certainly a handful of reasons for this paucity of the number of systematic reviews in medical education, the nature of the definition of what constitutes evidence being one of them.

The answer does not lie in the concept of causation but in the nature of the object of study. As previously mentioned, when we want to identify in clinical medicine the evidence of the effectiveness of a specific intervention, the first thing we have to do is to define both the causal factors and expected outcomes. The identification of such indicators is not simple in medical education which, due to its nature, deals with other kinds of factors.

The evaluation of the impact of a particular teaching method - for instance a lecture - on students' outcomes can be extremely difficult (sometimes impossible). The problem is to differentiate the effect of the lecture (with its specific structure and contents) from the effect of the teacher (his/her communication skills, motivation, humour, empathy, charisma) as well as the characteristics of the students (same lecture, given by same teacher, may have different impact depending on the students). This is why the effectiveness of the lecture may not be reproduced if given by two different teachers, in spite of having exactly the same structure and contents. And when given by same teacher to two different audiences, its impact depends also on the students themselves.

The ultimate question in science is 'what is reality' and, to study human phenomena, one needs to use multiple systems of inquiry. The research associated methods are empirical, psychological/philosophical, and spiritual methods for Filippo (1991) or statistical systems, phenomenological systems, hermeneutic systems or systemic systems for Polkinghorne (1983). In empirical sciences as well as in human sciences, the standard for acceptability of knowledge is that it must be able to withstand the test of experience and experiment. In the human sciences, descriptive and interpretive techniques are accepted and used to validate knowledge, which is fallible, representing *the best explanation and understanding available from which the human being can be confident enough to take action* (Filippo, 1991)

For a full explanation of what a phenomenon is in human sciences we need to consider the confounding factors or confounding covariates (Gelman and Hill 2007). These are defined as external factors that *correlate* with the outcome but are not directly *causal*, i.e. are associated in a non-causal way with both the dependable and the independent variables (Figure 2). The concept is of particular important in observational studies because '*differences between groups are outside the control of the experiment and can affect the outcomes'* (Gelman & Hill 2007). When applied to primary studies, the scientific methodology needs to take into account the confounding variables, in order to avoid false positive errors.

Having said this, we do not mean that confounding variables are not present in other fields, just that education is an area where dissociating them from the causal factor may be extremely difficult, or impossible, due to the holistic nature of the educational phenomenon.

The complexity of the human sciences was highlighted by Filippo (1991) who stated that '*there is not a "single truth" and*

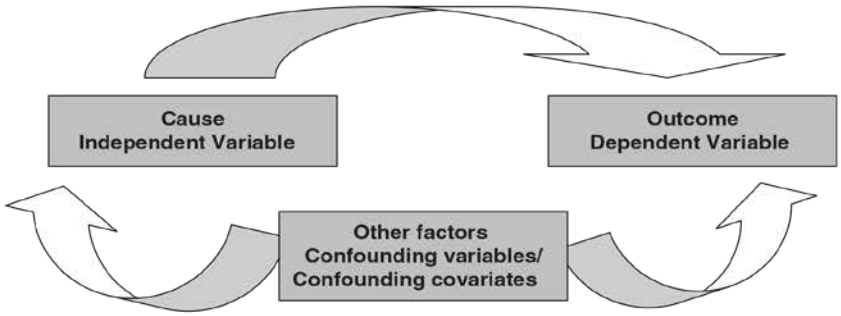

Figure 2. Confounding variables.

that multiple epistemological approaches are needed'. He pointed to different interpretations of the truth depending on the contexts which are indigenous to specific communities and their interpretation of knowledge.

The complexity and the multiple approaches to the different 'truths' explain why the scientific approach in medical education is called BEME, i.e. Best Evidence Medical Education and not EBME (Evidence Based Medical Education) as it happens with Evidence Based Medicine (EBM).

Cochrane Reviews

The Cochrane Collaboration defines a Cochrane Review as a '*scientific investigation in itself, with a pre-planned methods section and an assembly of original studies (predominantly randomized controlled trials and clinical controlled trials, but also sometimes, non-randomized observational studies) as their "subjects". The results of these multiple primary investigations are synthesized by using strategies that limit bias and random error. These strategies include a comprehensive search of all potentially relevant studies and the use of explicit, reproducible criteria in the selection of studies for review. Primary research designs and study characteristics are appraised, data synthesized, and results interpreted.*' (http://www.cochrane.org, retrieved March 2012).

Systematic reviews are analytic integrative instruments of the best scientific evidence available, allowing evidence-based answers to clinical relevant questions. Meta-analyses are systematic reviews with a quantitative analysis of its results.

The process of conducting a systematic review is very rigorous and has a standard procedure that includes: (1) formulating a clinically relevant question, (2) explicit selection of studies to be included in the review, (3) critical appraisal of the relevant evidence, (4) synthesis of the evidence and (5) statistical analysis of the results.

The need for this type of integrative research in therapy arises from the fact that there are usually several RCTs studying a specific therapeutic/preventive intervention and, therefore, if one desires to have a full picture of its efficacy, then a combination of several studies in a systematic review (with or without meta-analysis) will provide it. For example, if all studies show efficacy, than the reader is surer about the use of the intervention in his patients than if the review shows some of the trials being efficacious, some not and other even being harmful.

Best Evidence Medical Education Reviews

Best Evidence Medical Education (BEME) Reviews are high-standard, secondary research papers, which look into specific areas of medical education with the goals of '*disseminating information which allows medical teachers, institutions and all concerned with medical education to make decisions on the basis of the best evidence available; producing appropriate systematic reviews of medical education which reflect the best evidence available and meet the needs of the user; and creating a culture of best evidence medical education amongst individual teachers, institutions and national bodies*' (http:// www.bemecollaboration.org, retrieved March 2012).

Similar to what was said before in terms of Cochrane Collaboration, the methodology of a BEMER implies again a *'comprehensive search of all potentially relevant studies and the use of explicit, reproducible criteria in the selection of studies for review. Primary research designs and study characteristics are appraised, data synthesized, and results interpreted'* (http://www.cochrane.org, retrieved March 2012).

Establishing the parallel with Cochrane the BEME methodology is also very rigorous and follows similar steps: (1) formulation of an educational relevant question; (2) explicit selection of studies to be included in the review; (3) critical appraisal of them, (4) synthesis of this evidence and (5) practical application of the results.

To illustrate a more detailed comparison, it may be of interest to report on authors' experience of running a study of the feasibility of the OSCE undertaken at the Center for Evidence Based Medicine the Faculty of Medicine University of Lisbon. This was carried out as part of a BEMER, which looked more widely at OSCE reliability, validity and feasibility when used to assess learning outcomes in undergraduate medical education.

Let us start with some background information on the OSCE to clarify the object of our review.

Introduced by Harden et al. in 1975, it consists of multiple 'objective stations' designed to assess a range of clinical skills

under similar circumstances. OSCE has emerged as the main method of assessing clinical learning outcomes and has become the method with the greatest impact in medical education. It appeared as an important solution for the identified problems associated with the assessment of clinical competences, namely psychometric ones.

OSCE is probably one of the few assessment tools in medical education where, by definition, the effect of confounding variables is small, because the exam is highly structured – same assessments, same standardized patients, same examiners and same duration – which implies that there is not much scope for the examiner's or patient's confounding interferences, since every task and respective assessment are standardized in advance to the exam.

In this article the intention is not to report on the BEMER findings in terms of OSCE reliability, validity and feasibility, but to highlight the methodological process as well as inform on some lessons taken at each step.

Step 1 – Formulating an educational relevant question

The discussion on clinical assessment continues to be a relevant topic in the medical education agenda and, in spite of OSCE wide and exponential use in a variety of contexts all over the world, we were interested in bringing out evidence on its reliability (are the OSCE results reproducible?), validity (is the OSCE measuring what it is supposed to measure?) and feasibility (is the OSCE financial and physically achievable with the means at hand and circumstances as they are? Why use such a demanding exam (time, money and human resources) if other methods are available for assessing learning outcomes?).

Step 2 – Search and selection of studies to be included in the review

According to Hammerstrøm et al. (2010) '*. . . searches for systematic reviews aim to be as extensive as possible in order to ensure that as many as possible of the necessary and relevant studies are included in the review*'. Both dimensions, exhaustive and relevant data, should coexist in BEME synthesis but, depending on available resources, the scope of the material under analysis can be limited. This was the case in our review, with more than 1000 studies identified through hand and electronic searches (Patricio et al. 2009), where only English reports on undergraduate medical education were accepted, leaving the remaining material for another future systematic review.

Step 3 – Codification and critical appraisal of evidence

All identified studies were reviewed by two independent coders, according to transparent pre-specified inclusion criteria. Let us take the question regarding OSCE reliability as the example to describe the methodology: all accepted studies were scrutinized to identify the reliability of the exam and further analysis was then performed, looking behind the reliability found. In other words, we asked which were the conditions and circumstances associated with OSCE reliability. Could standardized patients be as effective as experts in assessing students? Should global ratings be used instead of checklists? Could feedback be given during the OSCE with no impact in students' performance? Should the number of assessors and stations be increased to obtain higher reliability? How does reliability vary in high vs. low stakes exams, with summative vs. nonsummative exams, with OSCE implemented in hospital vs. community, with high vs. low number of students, with training of patients and examiners, with duration of stations, with number of venues?

These are just some examples of a cascade of questions behind the initial search question 'Is the OSCE Reliable?'

The objective in this phase was 'to go beyond the results' to identify, not only how feasible is the OSCE, but how feasible it is depending on the context (relevant conditions, for example, OSCEs performed in hospital versus community, high versus low stakes exams, etc.) and the circumstances (determining factors, for instance, number of stations, training of assessors and patients. . .). As said above for Cochrane, *when a causal association is found in BEME, it may be due to one or more of the same mechanisms: observation bias, operation of confounding factors, chance and true causality* (Elwood 2007).

This was why each piece of information given by authors was critically appraised, to determine the quality of the individual evidence. After independent codification by two coders, consensus was established and a discussion occurred when agreement was not reached. Studies of 'poor quality' (wrong or inconsistent data when, for instance, the reliability is reported as the average students' grade) or 'evidence free/intuitive papers' (when authors stated the OSCE was reliable without data to support the evidence) were excluded. In our review, a high number of studies were rejected (due to inconsistent or no data) demonstrating how critically the appraisal of evidence is in a BEME review. The major difficulty at this step was to decide where we should stop when codifying data. Our mistake was to have been too inclusive (to avoid missing data) leaving for the next stage the decision on unclear information.

Step 4 – Synthesis of the evidence

This step, probably the most difficult in our BEMER, implies combining the results based on high quality data (quantitative and/or qualitative) to make global sense of them. Frequently, the results point to the same direction, but there are times where contradictory findings are detected, which implies that some questions may well remain unanswered.

In terms of OSCE reliability, it was possible to identify it and how it works when using 'checklists vs. global ratings' and 'experts vs. simulated patients as assessors'. In so doing we tried to produce the best evidence available for analysis since, as proposed by Greenhalgh et al. in 2011 '*in education we have to unpack the context, mechanism and outcomes of each phenomenon*'.

Although we did not get contradictory results, some research questions identified in Step 3 remained without an answer, due to unclear or missing data (requiring future higher

quality in primary reporting). Another difficulty at this stage was the fact that we were too inclusive in Step 3, with the consequence of having to code a lot of useless material.

Step 5 – Practical application of the results

Stating the implications to educational practice is a crucial step. In other words, BEMER findings must be transferred into practice, so that they may support informed decisions in the field. According to Wong (2012). *'the translation of evidence may be more difficult than to generate evidence'*. From our experience of running BEMER and Cochrane SR, differences and similarities between them can be easily identified. The major difference might be the presence of more '*confounding factors*' in BEME reviews as already discussed. This could apply to our OSCE study, in that we know that the success of the OSCE also depends on the training and attitude of the examiners.

Differences of degree were also found in the process:

- Concerning the definition of the research question (Straus et al. 2011) in clinical medicine (CM), one follows the PICO (Patients, Intervention, Comparison, Outcomes) format, while in medical education (ME) the question is much more complex, due to the different nature of the interventions (multiple assessment methods to different students) as well as the different purposes of the research (explanatory objectives)
- The next step – the methodology for selection of studies (inclusion and exclusion criteria) – supports the notion that the studies for ME are never as clear-cut in their methodology and usefulness as the ones found in CM
- Then comes the critical appraisal of the selected studies, using *guides/templates* in CM, adapted to the type of study (therapeutics, diagnostics, prognostic, etc.), while in BEMER the guides/templates – when used – are more variable and complex, depending on the type of study being appraised, again due to the complexity of the interventions (there is no standard of appraisal covering all the types of studies).
- After extracting the data from the papers, one can then synthesize it. In a Cochrane SR this means giving a point estimate (and its relative confidence intervals) of the results of each individual study, as well as a combined global result (around odds ratio or relative risks). In ME, the results should answer the question of what is it about this kind of intervention that works, for whom, in what circumstances, in what respects and why?
- Finally, after all these steps are finished, in a Cochrane SR one can make a statement for dissemination – for example, the role of a drug for a specific disease. In medical education the outcome is different; as previously highlighted, it may be just a general recommendation.

Again, the Cochrane Reviews take stock on hard data from clinical trials (looking at the impact of interventions), selected by pre-defined criteria on validity, importance and applicability. If the degree of heterogeneity between studies is below a certain level (defined by statistical analysis), then a meta-analysis can be performed on top of the results from the SR, allowing precise calculations on the average effect of interventions (through calculations of odds ratio or relative risks).

On the other hand, the BEMER are based on primary studies that are also selected by pre-defined criteria through a keyword search, but allow several types of articles to be included in the analysis (for example, trials and observational data in assessment, as well as descriptive and evaluative studies in learning experiences).

Another example might be that the evidence on the success or failure of the intervention may be less clear in medical education studies because the establishment of causal relationship between the intervention and outcomes may be difficult (sometimes impossible). The practical process in designing BEME reviews appears to be more complex and resource demanding than with CR – this is what we learned from experience when comparing both.

As already stated, in terms of similarities, the two types of systematic reviews discussed before –CR and BEMER – should be based in sound scientific evidence and, although some differences were highlighted, the five methodological steps described above are common to both, namely: identification of a relevant question, exhaustive search, detailed appraisal of the evidence, synthesis of findings, and transfer to practice.

Although we mentioned the degree of confounding variables as the major difference between CR and BEMER, it is important to highlight that confounding factors, as reported above, are also frequently found in clinical medicine (an example is a recent study by Duclos et al. (2012) on the evaluation of thyroidectomy as a treatment in thyrotoxicosis, showed that a confounding factor was the experience of the surgeon).

Also similar in both reviews is the need for rigour behind each methodological step, as well as the importance of experience gained from previous systematic reviews. A need for transparency in justifying any methodological decision is also common in BEME and Cochrane. Transparency is highly important because it allows that '*when using the same search criteria, the same inclusion criteria, the same appraisal criteria, other researchers arrive at the same results'*.

As in CR, a BEMER is a very demanding process, and its value depends not on the generated evidence, but on its capacity to translate results into practice. Similar to Cochrane is also the fact that BEME findings may improve practice *(when showing what works, for who, in what circumstances*), as well as improve research in primary medical education (when highlighting deficiencies or missing areas). Finally there is a need for both systematic reviews (respectively in Clinical medicine and medical education) to support informed decisions at all hierarchical ranks from decision makers to the individual professional.

Discussion

In our day and age, evidence synthesis is a fundamental methodology to support clinical medicine decision making as well as medical education practice.

The nature of the evidence supporting both approaches is naturally different, due to the types of phenomena one is

analysing and the type of studies selected to answer a clinical or educational question.

The differences discussed between BEME and Cochrane reviews are perhaps more a matter of degree than the existence of fundamental differences, and there are examples in literature proving this. Also, some of the more difficult outcomes we struggle with in BEME (namely, looking beyond the results) are increasingly being appreciated as equally important in evidence based medicine. For example, we need to consider not only objective data with regard to blood pressure measurements when studying antihypertensive drugs, but also the importance of patients' well-being, return to employment, etc.

In spite of standardization of the OSCE, our systematic review proved to be a difficult task, confirming the problems found when analysing the initial four hundred studies (Patricio et al. 2009). Several lessons emerged from this BEMER, which eventually may apply to reviews in other fields:

- The difficulties found are essentially methodological, independent of the type of study within a human science;
- The idea that, for a systematic review with large samples, it could be enough to analyse only a small number of studies (for example 50%) to obtain reliable results was not confirmed in our study, due to missing or inconsistent data that preclude discarding the analysis of a single paper
- A professional, highly committed working team is absolutely needed with rigour and persistence, capacity for arguing, listening/accepting other perspectives, strong leadership and group cohesion as essential qualities;
- The resources to perform the task, namely time, are of crucial importance, since the process appeared to be much more exigent than initially foreseen
- Experience may play a crucial role to avoid methodological mistakes, so asking for supervision is essential namely when performing a first BEME SR.
- Attention should be given to increase the quality of primary studies. This is a necessary but not sufficient prerequisite and BEME will only achieve maturity when quality in both reporting and methodology of assessing evidence is of high level.

We have discussed the need for the medical education community to adopt a more rigorous scientific approach, in sync with what the medical community does with clinical science. To identify what seems to works in terms of teachers' interventions – and this is the highest objective of BEME – we need to put in context all variables, due to the holistic nature of the process. Similar to EBM, the two fundamental principles of BEME are that educational decision making should be guided by best evidence available and that evidence alone is never sufficient for sound clinical practice.

The future of BEME systematic reviews is, in our opinion, an optimistic one and we expect to see the publication of more reviews and that they will impact on teaching and therefore in patient care and health care research. No doubt the review methodology will continue to evolve and improve.

As Harden stated already in 1999, *Best evidence medical education has much to offer the teacher, the student, the medical profession and the public.*

Declaration of interest: The author reports no declarations of interest.

Notes on contributors

MADALENA PATRÍCIO is President of the Association for Medical Education in Europe (AMEE), Assistant Professor at the Institute of Introduction to Medicine and Scientific Consultant at the Center for Evidence Based Medicine both at the Faculty of Medicine, University of Lisbon, Portugal.

ANTÓNIO VAZ CARNEIRO is a Professor of Medicine and the Head of the Center for Evidence Based Medicine at the Faculty of Medicine University of Lisbon, and Head of the Cochrane Coordinating Center Portugal.

References

Aiyer MK, Dorsch JL. 2008. The transformation of an EBM curriculum: A 10-year experience. Med Teach 30:377–383.

BEME: Best Evidence Medical Education. Retrieved March 10, 2012. Available from: http:// www.bemecollaboration.org

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. 2009. Introduction to meta-analysis. 1 ed. Chichester: Wiley.

Duclos A, Peix J, Colin C, Kraimps J, Fabrice Menegaux F, Pattou F, Sebag F, Touzet S, Bourdy S, et al. 2012. Influence of experience on performance of individual surgeons in thyroid surgery: Prospective cross sectional multicentre study. Brit Med J 344:d8041.

Elwood M. 2007. Critical appraisal of epidemiological studies and clinical trials. 3d ed. Oxford: Oxford University Press.

Evidence-Based Medicine Working Group. 1992. Evidence-based medicine. Am Med Assoc 268: 2420–2425.

Filippo DS. 1991. What Is Human Science? Retrieved March 10, 2012. Available from: www.lutz-sanfilippo.com/library/general/lsfscience.html

Fraser AG, Dunstan FD. 2010. On the impossibility of being expert. Brit Med J 341:1314–1315.

Gehlbach SH. 2002. Interpreting the medical literature. 1st ed. New York: McGraw-Hill.

Gelman A, Hill J. 2007. Data analysis using regression and multilevel/hierarchical models. 1 ed. Cambridge: Cambridge University Press.

Greenhalgh T, Wong G, Westhorp G, Pawson R. 2011. Protocol-realist and meta-narrative evidence synthesis: Evolving standards (RAMESES). BMC Med Res Methodol 11:115.

Guyatt G, Renni D, Meade MO, Cook DJ. Eds. 2008. User's Guides to the Medical Literature, 2nd ed. New York: McGraw-Hill.

Hammerstrøm K, Wade A, Jørgensen AK 2010. Searching for studies: A guide to information retrieval for Campbell Systematic Reviews, The Campbell Collaboration. Retrieved March 10, 2012. Available from: http://www.google.com/search?q=Hammerstrom+K%2C+Searching+for+studies.+2010&hl=ptPT&source=hp&gbv=2&gs_sm=13&gs_upl=1187l4265l0l7515l3l3l0l0l0l0l406l766l0.2.4l3l0&oq=Hammerstrom+K%2C+Searching+for+studies.+2010&aq=f&aqi=&aql=

Hammick MR. 2005. Evidence-informed education in the health care science professions. J Veterinary Med Edu 32: 399–403.

Haynes RB, Sackett DL, Guyatt GH, Tugwell P. 2006. Clinical epidemiology. 3d ed. Philadelphia: Lippincott Williams & Wilkins.

Harden RM, Grant J, Buckley G, Hart IR. 1999. BEME Guide No. 1: Best Evidence Medical Education. Med Teach 21(6):553–562.

Harden RM, Stevenson M, Downie WW, Wilson GM. 1975. Assessment of clinical competence using objective structured examination. Brit Med J 1:447–451.

Hill AB. 1965. The environment and disease: Association or causation? Roy Soc Med 58:295–300.

Mayer D. 2004. Essential evidence-based medicine. 1st ed. Cambridge: Cambridge University Press.

Oude RK, Thangaratinam S, Barnfield G, Suter K, Horvath AR, Walczak J, Welminska A, Weinbrenner S, Meyerrose B, Arvanitis TN, et al. 2011.

How can we teach EBM in clinical practice? An analysis of barriers to implementation of on-the-job EBM teaching and learning. Med Teach 33:e125–e130.

Patricio M, Juliao M, Fareleira F, Young M, Norman G, Vaz CA. 2009. A comprehensive checklist for reporting the use of OSCEs. Med Teach 31:112–124.

Polkinghorne D. 1983. Methodology for the human sciences: systems of inquiry. 1 ed. New York: University of New York Press.

Rothman KJ. 2002. Epidemiology. 1 ed. Oxford: Oxford University Press.

Rothwell PM. Ed.2007. Treating individuals. From randomised trials to personalised medicine, 1st ed. Edinburgh: Elsevier.

Straus SE, Glasziou P, Richardson WS, Haynes RB. Eds. 2011. Evidence-based Medicine. How to practice and teach it. Churchill Livingstone Elsevier, Edinburgh.

The Cochrane Collaboration. Retrieved March 10, 2012. Available from: http://www.cochrane.org/

van Dijk N, ijk N, Hooft L, Wieringa-de Waard M. 2010. What are the barriers to residents' practicing evidence-based medicine? A systematic review. Acad Med 85(7):1163–1170.

Wong G, Greenhalgh T, Westhorp G, Pawson R. 2012. Realist methods in medical education research: What are they and what can they contribute? Med 46:89–96.

CHAPTER 3
Examining the quality of the OSCE primary study reports A proposal for a comprehensive checklist to improve reporting of OSCE

2009; 31: 112–124 MEDICAL TEACHER

A comprehensive checklist for reporting the use of OSCEs

MADALENA PATRÍCIO[1], MIGUEL JULIÃO[1], FILIPA FARELEIRA[1], MEREDITH YOUNG[2], GEOFFREY NORMAN[2] & ANTÓNIO VAZ CARNEIRO[1]

[1]Faculty of Medicine of the University of Lisbon, Portugal, [2]McMaster University, Canada

Abstract

Background: The Objective Structured Clinical Examination (OSCE) has experienced an explosion of use which has rarely been accompanied by systematic investigations on its validity, reliability and feasibility. A systematic review of OSCE was undertaken as part of Best Evidence Medical Education at the Centre for Evidence Based Medicine of the Faculty of Medicine of the University of Lisbon. Several problems were identified with published papers relating to completeness of information presented, methodological issues or the use of terminology.

Aim: To identify a need for standardization within the reporting of OSCE studies in medical education based in the first 104 papers of the aforementioned review.

Method: Two independent reviewers coded each paper.

Results: The most important problem identified was the lack of information, followed by the degree of inconsistency when reporting on OSCEs (papers with missing data and papers where data was given in a way that interpretation is difficult or impossible in terms of evidence; heterogeneity in reporting, lack of a standardized vocabulary, statistical errors and lack of structure within reporting).

Conclusions: The authors present a 'Comprehensive Checklist for those describing the use of OSCEs in the report of educational literature' as an attempt to encourage better report standards.

Introduction

The Objective Structured Clinical Examination (OSCE) introduced by Harden et al. in 1975 has experienced an explosion of use, crossing most areas of medicine, at all levels of evaluation. The OSCE exam consists of multiple objective 'stations' designed to assess a range of clinical skills under similar circumstances (standardized assessment, patient(s), examiner(s) and duration).

During the last three decades, OSCEs have been used throughout the world with different purposes (assess students, evaluate curricula or an intervention, feedback to teachers and students, etc.), and at all levels of Medical Education (pre-, post-graduation and continuous medical education). Consequently, OSCEs have been well reported in literature with more than 1000 published papers.

Objective structured clinical examination reporting has risen sharply over the last few years, but this increase in publication has rarely been accompanied by systematic investigations of validity, reliability and feasibility. OSCE dissemination as an assessment tool has been witnessed throughout a variety of areas but without systematic evidence supporting its use (Norman 2000; Hart & Harden 2002).

As part of Best Evidence Medical Education (BEME; http://www.bemecollaboration.org/) a systematic review of OSCEs was undertaken at the Centre for Evidence Based Medicine of the Faculty of Medicine of the University of Lisbon.

Practice points

- Attention needs to be paid when OSCEs are reported in the literature namely in terms of
  - Information given
  - Accuracy of statistics
  - Vocabulary used
  - Structure of report
- A checklist should be considered to assist authors in the preparation of OSCE reports.

The objective was to evaluate the validity, reliability and feasibility of the OSCE as an assessment method of the various learning outcomes in undergraduate medical education. These data will be reported on a BEME systematic review for the analysis of OSCE as a method of undergraduate assessment.

In the process of undertaking this review, issues were identified with published papers relating to completeness of information presented, methodological issues and the use of terminology.

This article will identify the problems faced in the analysis of the first 104 papers of the aforementioned systematic review. The focus is not to report the results of the systematic review, but to identify a need for standardization within the reporting of OSCE studies in medical education.

Correspondence: Madalena Patrício, Centre for Evidence Based Medicine, Faculty of Medicine, University of Lisbon, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal. Tel: + 351217985175; fax: +351217985192; email: patricio@fm.ul.pt

ISSN 0142–159X print/ISSN 1466–187X online/09/020112–13

Based on this experience a checklist has been developed and is presented as an attempt to encourage better report standards on OSCE papers.

Methods

Study objectives

The aim is to identify the methodological deficiencies in reporting OSCE studies and suggest a dataset to encourage better standards for OSCE reports.

Working group

A working group in Lisbon was constituted by the coordinator (a MD–PhD and Master in Medical Education), a research director (Educationalist and Master in Medical Education – also acting as a coder), two coders (final-year medical students) and two administrative assistants. The Department of Clinical Epidemiology and Biostatistics at McMaster University, Canada, which served as a consultant for this project, joined this group.

Search process

For the bibliographic search we used the OSCE database material published by Valerie Harden et al. (2002), which covered the OSCE literature from 1975 to end of 2001. The 712 bibliographic references were identified through

- Electronic search of medical, educational and related databases – Hand search on six key journals *(Academic Medicine, Medical Education, Medical Teacher, Teaching & Learning in Medicine, Advances in Health Sciences Education and Education for Health)*;
- Search of TIMELIT reference database;
- Search on the literature collections at Dundee University and on Grey Literature (e.g. the Proceedings of Ottawa Conference).

In 2009, published literature was updated through the end of December 2008. At present time our OSCE Database holds 1094 references for the systematic review. The material under analysis for this article includes the first 104 papers retrieved for the OSCE BEME review.[1] From them only 56 were analysed in depth since they were those with clear evidence of validity, and/or reliability and/or feasibility.

Codifying process – OSCE online database

Two independent reviewers coded each paper based on modified BEME criteria.[2] The coordinator intervened at the start of the project to assist whilst the criteria were being defined. In the subsequent phases of the project, consensus was reached through discussion in cases of non-agreement among coders.

A new 'online database' (*Lotus software*) was developed, since the existing BEME coding sheet was inadequate for our systematic review objectives. This new database allowed independent codification by each researcher. After independent codification of a certain number of records (usually no more than 20) a 'Consensus Meeting' occurs with the full team. Disagreements were resolved by consensus and as a result of this meeting a 'Consensus Record' was created for each record in the online OSCE database.

Consistency among coders was established throughout the process: 100% when codifying OSCE feasibility, 95% for validity and 93% for reliability.

Purpose of OSCE in report

Although in this article we are not reporting details of the design of the OSCE exams, it is important to categorize the OSCE studies in terms of their aims in order to understand the context in which the OSCE was performed. With this objective, coders established a non-exclusive category system through *a posteriori* content analysis technique.

As can be seen in Table 1, in 59 papers the objective was to '*examine the OSCE itself*' (56.7%). Among them we found studies examining: OSCE validity, reliability or feasibility (29.8%), OSCE development including developing formative OSCEs, developing standards and issues related with codification process (27.9%), perceptions regarding the OSCE (14.5%), effects of independent variables on OSCE performance (21.2%), correlation of OSCE with other assessment methods or other variables (13.5%), appraisal of OSCE (5.8%) and OSCE comparison with other assessment methods (13.5%).

In 46 papers the OSCE was used for '*evaluation purposes*' (44.2%). Breaking this down, it was used to assess students' performance (15.4%), to evaluate *curricula* (7.7%) or the effects resulting from an experiment (intervention, method, innovation, teaching at different sites, etc.) (24%).

Finally, in the remaining 10 non-primary studies the OSCE was examined '*from a theoretical perspective*' (9.6%).

Results

Methodological difficulties in terms of reporting OSCEs

The most important issues unearthed in this initial phase of the systematic review were the lack of information followed by the degree of inconsistency when reporting on OSCEs. In other words, major problems identified relate to papers with *missing data* and/or papers with *inconsistent data* that is data given in a way that interpretation is difficult or impossible in terms of evidence.

- *Problem I – Data missing*

The main areas where information is missing are identified in Table 2.

As seen in Table 2, a very high percentage of papers do not present data to support the evidence for feasibility (96.2%), reliability (67.3%) and validity (65.4%).

Moreover, a high proportion of papers (more than 50%) do not report data in essential areas such as '*number and details of faculty involved in the OSCE set-up/development*' (85.7%), *number of years of curriculum* (64.2%), *OSCE duration*

Table 1. Purpose of OSCE in report ($n = 104$).*

Aims	Number of papers	Percentage
1. Studies examining the OSCE itself	**59**	**56.7**
Examining OSCE validity, reliability and feasibility		
Validity, reliability and feasibility	31	29.8
Admission test predictive of OSCE performance	2	1.9
OSCE predictive of residency performance	2	1.9
OSCE predictive of peer rated competence	1	1
OSCE predictive of final examination	1	1
Examining OSCE development		
Developing OSCE, Develop SP	19	18.3
Examining codification processes (checklist/global ratings)	5	4.8
Developing formative OSCE	2	1.9
Developing standards, criterion for pass/fail	2	1.9
OSCE security	1	1
Examining OSCE value		
OSCE versus other assessment methods	14	13.5
OSCE appraisal/OSCE as a mean to provide feedback	6	5.8
Examining the impact of independent variables on OSCE		
Impact/effect of total time/duration/number of stations	5	3.8
Impact/effect of OSCE performed at different sites	3	2.9
Impact/effect of standards	2	1.9
Impact/effect of examiners' background	2	1.9
Impact/effect of examination design/organization of stations	2	1.9
Impact/effect of multiple OSCE examinations	2	1.9
Impact/effect of feedback	2	1.9
Impact/effect of examiners' number	1	1
Impact/effect of examiners' tiredness	1	1
Impact/effect of a written text (on reliability)	1	1
Impact/effect of serial assessments to detect interval learning	1	1
Examining the impact of OSCE on other variables		
Impact/effect of OSCE on 'drive learning'	1	1
Impact/effect of OSCE on 'drive teaching'	1	1
Examining OSCE correlations		
OSCE Correlation with clinical experience/clinical exposure	4	2.9
OSCE Correlation with multiple choice question (MCQ)	3	2.9
OSCE Correlation with traditional clinical examination	3	2.9
OSCE Correlation with long case	2	1.9
OSCE Correlation with written exam	2	1.9
OSCE Correlation with standardized patients' satisfaction	1	1
OSCE Correlation with teaching quality	1	1
OSCE Correlation with teaching expectations about students	1	1
OSCE Correlation with short answer questions	1	1
OSCE Correlation with clinical practical examination	1	1
OSCE Correlation with clinical clerkship evaluation	1	1
OSCE Correlation with Individual knowledge	1	1
OSCE Correlation with extended match questions (EMQ)	1	1
OSCE Correlation with students' confidence	1	1
OSCE Correlation with learning style	1	1
OSCE Correlation with faculty evaluation	1	1
OSCE Correlation with feedback from rotations	1	1
Correlation of admission test with SP satisfaction/OSCE performance	2	1.9
Examining the perceptions of those involved in the exam		
Students' perceptions about OSCE, about motivation to participate	11	10.6
Teachers' perceptions about OSCE = 3	3	2.9
Examiners' perceptions about own health	1	1
2. Studies where OSCE was used to evaluate	**46**	**44.2**
Results from an experiment (Intervention, method, innovation, teaching at different sites)	25	24
Students' performance	16	15.4
Curriculum	8	7.7
3. Studies examining the OSCE from a theoretical perspective	**10**	**9.6**
Non-primary studies	10	9.6

Note: *Non-exclusive categories.

(62.5%), *number and details of standardized patients* (62.5%) *as well as number and details of examiners* (55.4%).

Other important areas are also absent, although at a lower level: *duration of stations* (37.5%), *number of students performing the OSCE* (32.1%), *type of OSCE exam* (32.2%), *information on codification process* (30.4%), *OSCE purpose* (26.8%), *number of stations* (23.2%), *total number of students in the course being assessed* (21.4%), *type of stations* (19.6%), *feedback given to students* (19.6%), *student course year* (16.1%), *outcome/subject under assessment* (5.4%), etc.

Table 2. Areas where information is missing.*

Information missing on validity, reliability or feasibility ($n = 104$)	Number of papers	Percentage
No. data on feasibility	100	96.2
No. data on validity	70	67.3
No. data on reliability	68	65.4
Information missing on other areas ($n = 56$)**		
Number and details of faculty involved in OSCE set up/development	48	85.7
Number of years of the curriculum	36	64.2
OSCE duration (total time and/or station time)	35	62.5
Number and details of simulated/standardized patients	35	62.5
Number and details of examiners	31	55.4
Stations identical in time (yes or no)	21	37.5
Number of students performing the OSCE	18	32.1
Type of exam (high stakes, moderate, no grading, etc)	18	32.1
Codification process (checklist, global ratings, etc)	17	30.4
OSCE purpose (formative, summative, both, etc)	15	26.8
Number of stations	13	23.2
Total number of students in the course being assessed	12	21.4
Type of stations (history taken, physical examination, diagnosis, etc)	11	19.6
Feedback given to students (yes/no)	11	19.6
Student course year	9	16.1
Outcomes under assessment (subjects like paediatrics, surgery, etc)	3	5.4

Note: *Non-exclusive categories.
**Only papers with data on validity and/or reliability and/or feasibility were fully examined.

All the above problems frequently compromised the calculation of the reliability of the studies, since data is not given with the necessary detail for the particular design.

- *Problem II: Inconsistent data*

Difficulties with inconsistent data relate primarily to the following situations:

(1) Heterogeneity in reporting on reliability, validity and feasibility;
(2) Lack of a standardized vocabulary in reporting, that is lack of common vocabulary;
(3) Statistical issues;
(4) Lack of structure within reporting;

Heterogeneity in reporting validity, reliability and feasibility or different approaches on how reliability validity and feasibility are used

Validity. Seventy of the 104 studies did not report on validity (67.3%) and in five studies {18, 58, 76, 78, 664} although authors refer to validity they did not present any data to support its evidence (4.8%).

For the remaining studies (32.7%) *concurrent validity* is mentioned only in one study, *construct validity* is reported in two studies, *face-content* in 17 studies and *correlations with other measures or within OSCE measures* were explored in 13 studies. Please see Table 3 for details.

According to Downing and Haladyna (2004) studies giving information exclusively on 'face-content' should not be considered in terms of validity, since this approach should be seen as a measure of 'fairness and relevance' and not exactly a measure of 'validity construct'. For these authors the term face-validity is rejected as representative of any type of legitimate validity evidence. According to the perspective of this author the evidence for validity is lacking in 87 papers (70 + 17), i.e. on 83.7% of the studies.

We suggest that for studies reporting OSCEs, particular attention should be paid to issues of construct, content, criterion and predictive validity of the exam under question. In order to foster our understanding of OSCEs, there must be a systematic effort to report on issues pertaining to the design and validity of the OSCE.

Reliability. Sixty-eight studies out of 104 papers provided no evidence for reliability (63%). Among them, in nine papers (8.7%), authors stated the OSCE was reliable without giving any evidence for it.

From the remaining 36 papers, only 24 studies (23.1%) present data to support the calculation of the aggregated reliability index (see Table 4).

Due to the nature of the OSCE, there are a number of possible approaches to characterize reliability so it is helpful to think of the OSCE reliability in a framework of Generalizability Theory. There are at least three identifiable sources of error variance in a typical OSCE – the Station (s), the Rater (r) and the Item (i). Rater effects will generally not be computed, since the usual OSCE has one rater per station, but in a study using two or more raters per station, one could separate the effects of raters.

Thus, one could examine Internal Consistency (which amounts to the correlation across items within a station), Inter-rater reliability (the correlation between raters) and the correlation across stations. A second issue is one of aggregation; one could examine the reliability of individual items (the average of the inter-item correlation) or the reliability of the total score across all items within a station (internal consistency); the reliability of a single rater or the reliability of the mean across all raters, the average inter-station correlation or the reliability of the total test. Further, one could look at agreement between raters at the item level (i.e. Do raters agree that the candidate did/did not introduce herself?) or at the

Table 3. Evidence for validity ($n = 104$).*

Papers with evidence	Number/reference of papers	Percentage
Face-content	17 studies {6, 4, 26, 28, 47, 56, 72, 59, 77, 113, 115, 156, 183, 442, 674, 810, 812}	16.3
Correlations with other measures	Nine studies {7, 8, 51, 71, 85, 89, 109, 661, 797}	8.6
Face/content + correlation with other measures	Four studies {1, 104, 310, 317}	3.8
Face/content + construct validity	One study {127}	1
Face/content + convergent/divergent validity	One study {479}	1
Concurrent validity	One study {715}	1
Construct validity	One study {225}	1
Sub-total	34 studies	32.7
Papers without evidence	Sub-total	
No data to support evidence although authors state the study is valid	Five studies {18, 58, 76, 78, 664}	4.8
No data/no reference to validity	65 studies (remaining papers)	62.5
Sub-total	70 studies	67.3

Note: *Exclusive categories.

Table 4. Evidence for reliability ($n = 104$).*

Papers with evidence	Number/reference of papers	Percentage
Data allowing the calculation of aggregated reliability	24 studies {1, 6, 8, 18, 24, 28, 51, 58, 72, 78, 76, 84, 85, 93, 99, 113, 115, 156, 183, 310, 444, 661, 664, 797}	23.1
Data not allowing the calculation of aggregated reliability	12 studies {7, 32, 71, 86, 89, 104, 230, 812, 808, 479, 673, 715}	11.5
Sub-total	36 studies	34.6
Papers without evidence		
No data to support evidence although authors state the study is reliable	Nine studies {127, 33, 49, 126, 225, 307, 317, 810, 442}	8.7
No data/no reference to reliability	59 studies (remaining papers)	54.6
Sub-total	68 studies	63

Note: *Exclusive categories.

score level (Did they agree on the total score of the station?). These differences are all comprehensible in the framework of Generalizability Theory, but are often not made explicit in reporting OSCE reliability.

To further compound this difficulty, there is a confusion in terminology, where some authors report the overall reliability of the test as a Cronbach's alpha {78} or Internal Consistency, where others reserve this to a measure of reliability across items within a station {113}, presumably because the same mathematical form can be used for both. We recommend that authors use the term 'internal consistency' and 'Cronbach's alpha' to apply to the reliability of items across stations, consistent with the original intent of these terms. We suggest that authors describe the overall reliability of the test, where stations are used as items on a test, as 'total reliability'.

Finally, there is some confusion in the use of coefficients. Some authors use Cohen's Kappa to examine agreement between raters at the item level, or weighted kappa to look at rater agreement on rating scales. While this is correct, it is also unnecessary and confusing, as Kappa is mathematically identical to an intraclass correlation, the usual reliability coefficient (Streiner & Norman 2003). Using the intraclass correlation formulation allows the application of Generalizability theory methods. We suggest that all reliability and generalizability coefficients be reported as intraclass correlation coefficients.

It is important to highlight that with reliability we experienced another type of difficulty related to the complexity of the OSCE designs. We found a vast degree of variability in the *types, amounts and depth of information* reported within the papers under review.

Feasibility. In the context of this article, feasibility data was defined as information regarding OSCE costs and OSCE time concerning faculty, examiners, standardized patients or other intervenient staff.

As expected, the number of studies including feasibility data (Table 5) is very low (3.8%) when compared to the studies dealing with validity (32.7%) or reliability (34.6%). Only four studies {56, 60, 76, 442} discussed the feasibility of the OSCEs, with one study giving limited information and sparse concrete data on time {76}, and two studies giving data on costs for faculty and standardized patients per hour {56, 60}. Among them only one study {442} reported on OSCE cost for the total number of students, presenting values in terms of costs of different phases.

The study from Poenaru et al. (1997) reports clearly on feasibility, making a distinction in terms of direct costs (materials, honoraria) indirect costs (hours of salaried work) and hidden costs. This distinction is crucial, for when a total

Table 5. Evidence for feasibility ($n = 104$).*

Papers with evidence	Number/reference of papers	Percentage
Studies with data to support feasibility	Four studies {56, 60, 76, 442}	3.8
Sub-total	Four studies	3.8
Papers without evidence		
No data to support evidence although authors state the study is feasible	Seven studies {6, 58, 156, 479, 797, 673, 808}	6.7
Feasibility mentioned within other contexts (i.e. adequate/non-adequate time of stations)	Three studies {49, 77, 812}	2.9
No data/no reference to feasibility	90 studies (remaining papers)	86.5
Sub-total	100 studies	96.1

Note: *Exclusive categories.

cost is given readers are able to understand what underlies its calculation.

In addition, authors in this study refer to 'confounding variables', 'pocket money' and to the Reznick Model from Reznick et al. (1993). This model can be very useful for reporting on OSCE feasibility since costs are allocated to four different phases: costs concerning *OSCE examination-development*, *OSCE production*, *OSCE administration* and *OSCE analysis-reporting*.

For studies that are systematically examining the feasibility of the OSCE, it is critical that evidence regarding the monetary and time commitments for the OSCE under review be provided. Without this information, it is difficult to examine the overall system costs involved in moving to an OSCE-based testing system.

Lack of a standardized vocabulary in reporting

When searching the literature to select the studies to be included in the BEME Systematic Review we found an impressive number of different terms used to label multi-stationed objective examinations within the sampled literature.[3] Terms included: TOSCE – Team Objective Structured Clinical Examination {79}, OSVE – Objective Structured Video Exam {100}, OSPE – Objective Structured Practical Examination {45, 89, 126}, OSCA – Objective Structured Clinical Assessment {116}, DOSCE – Dental Objective Structured Clinical Examination {670}, G-OSCE – Group Objective Structured Practical Examination {110}, GOSPE – Group Objective Structured Practical Examination {74}. If some of these different designations (for instance OSVE and TOSCE) may be accepted – since they refer to modifications on the structure of the initial OSCE described by Harden et al. (1975) – others (like OSCA) do not have such obvious justification.

In several papers, the manner in which the authors labelled stations was not always clear. Instead of reporting on specific stations, authors report on 'cases and encounters' {15} or just on 'encounters' {237}, just on 'cases' {661}, on 'cases and stations' {235} and on 'encounters and stations' {151}. We suggest that for OSCE studies the term 'station' should be used, leaving other designations for other formats of clinical assessments, for example for Long Case Exam (LCE), Objective Structured Long Examination Record (OSLER), Clinical Skills Exams (CSE), Clinical Skills Assessment (CSA), etc.

The language for the evaluators was equally variable and included terms such as: 'observers' {410}, 'observers and co-observers' {26}, 'examiners' {25, 40, 71, 92, 442, 664}, 'raters' {49, 127, 156, 661}, 'examiners and raters' {32, 33, 230}, sometimes without clear definition of respective functions (i.e. when authors mentioned 'observers', 'raters' and co-observers without clear explanation {58}).

The variability in vocabulary was also ambiguous when referring to the observations collected on the participants, ranging from simple terms such as 'checklists' (reported in 37 studies), 'global ratings' {26, 479} and 'checklist and global scale' {115}, to 'criteria list' {58}, 'structured marking sheets' {127}, 'rating scales' {310, 661}, 'global score' {435}, 'skills ratings' {479}, 'rating forms' {76, 77, 810}, 'checklist/marking schedule/global ratings' {28}, 'detailed evaluation sheets' {673}, 'assessment scale' {86} and the ubiquitous 'check list including two global ratings' {442} or the 'global score/marking sheet/SP narrative data' {237}. We suggest that these can be subsumed under the two descriptors, checklist and global rating.

In addition it is difficult to understand the distinction between OSCEs and SP examinations. Authors refer to different clinical examinations in multiple ways such as Standardized Patient Assessment or Standardized Patient Test or Standardized Patient Examination {151, 154, 161, 235, 237, 239, 240, 241, 675, 676}, Standardized Patient Based Performance Assessment {15}, or Standardized Patient Based Test {244}; occasionally it is not clear if this is equivalent to an OSCE.

Statistical issues

Some fundamental statistical problems were discovered in the review of these 104 papers including misconceptions regarding reliability {32, 237, 435}, errors in calculation {32, 237, 435}, and an absence of reliability calculations on novel OSCEs {442, 479}.

A few fundamental, and surprisingly common errors present in five studies, included using stability of means as a proxy for reliability {32, 33, 435}, mistaking inter- and intra-rater reliability {33}, and assuming that the reliability of the OSCE would 'fall within the normal range' of other published OSCEs {442, 479}.

To explain the above-mentioned misunderstandings: (1) mean differences and stability of means between locations, raters, tests, standardized patients, cases and times will tell us only about the consistency of the average performance, and nothing regarding the individual performance. A *t*-test of

Table 6. Comprehensive checklist for reporting OSCE studies.

1. Information on course background	
School/Institution	School(s) responsible for the implementation/development of the OSCE
Course area	Medicine, Nursing, Dentistry, etc
Curriculum years	Number of curriculum years
Medical education level	Pre-graduate, Post-graduate and CME/PCD
Students' course year	Information on 'course year' students are attending
Total number of students	Total number of students in the discipline(s)/clerkship(s)/residency program(s) etc under assessment
Role of exam	Exams: High Stakes, Moderate, Low stakes (Classroom etc)
2. Information on OSCE design	
OSCE aims	• Developing OSCE, new variations, new uses, new methods • OSCE to evaluate teaching methods, curricula, interventions • OSCE to evaluate individual students • Process of coding: checklist versus global ratings • Validity, reliability & feasibility. Which factors influence it? • OSCE versus other assessment methods • OSCE association with other variables and educational issues (gender, academic background, previous experience, etc) • Other
Role of exam	Formative, summative, both (formative and summative), pilot OSCE, OSCE for training teachers or students, OSCE just for teachers feedback, etc
Total number of students performing the OSCE	Total number of students performing the OSCE. (When students are selected from the total class (students' sub-set) the criteria should be reported.)
Number of students' subset	When OSCE results are reported only for a subset of students, the number and criteria for selection must be explicit.
Maximum number of students per OSCE circuit	Maximum number of students performing in an OSCE circuit.
Briefing the students	Information on how students were briefed on the OSCE structure, process, evaluation criteria etc.
Number of sites	Number of different locations where the same OSCE exam takes place.
Number of exams	Number of different OSCE exams reported in the paper (for instance OSCEs performed in different academic years).
Number of parallel OSCEs (sequential/simultaneous)	Number of identical OSCEs implemented for the same assessment purpose to comply with the need of successive OSCE circuits when the number of students is very large. Parallel OSCEs are different in content but should be as identical as possible in terms of format and degree of difficulty. They may take place at the same time at different places (simultaneously) or sequentially (i.e. same day morning and afternoon or even in different days).
Number of circuits	Within parallel OSCEs there is a possibility of different circuits i.e. the number of times the OSCE is performed to accommodate a large number of student.
Pre-test of stations (Pilot OSCE)	Information if there was a pilot OSCE before the exam.
Subject and specialty areas being assessed	Areas under assessment must be described. Many studies will cover a range of topics like 'end of course skills', 'pre-clinical skills/basic skills' and 'clinical skills' but it is important to report more detailed information for example: • Anaesthetics • A&E – Accident & Emergency Medicine • General Practice/Ambulatory care/family medicine/Primary care • Medicine • Obstetrics & gynaecology • Paediatrics • Pathology • Psychiatry • Public Health or Community Medicine • Surgery • Tropical Medicine • Others It is also important to refer the context of the course under assessment i.e. discipline/block/clerkship/residency program, etc.
Recording and scoring students' performance	Description of the codification criteria (checklists, global ratings, other tools) and on the procedure (paper-based, electronic, etc) should be given. If other tools are used they should be described.
Criteria for pass/fail decision	Clarification of the criteria (Norm or referenced criteria) and on pass/fail cut off should be given (Hofstee method, Angoff method, other).
Number of Stations	Number of stations included in the OSCE.
Number of subset stations	If authors report results just for some stations, criteria for such decision must be explicit. In this case the values for internal consistency and inter coder reliability of the total OSCE must be given.
Station details	Information should be given characterizing the stations in terms of clinical stations, procedural stations, rest stations, written stations and couplets.
Learning outcomes being assessed (Type of stations)	Information on what is assessed in the stations must be given: • History-taking • Physical examination • Diagnosis • Management/Prescription writing/Referral

(*continued*)

OSCE checklist

Table 6. Continued.

	• Practical procedures • Communication skills • Patient education/Health promotion • Patient investigation/Interpretation data/Problem-solving • Ethics/Attitudes • Informatics • Recall of knowledge • Other
Station duration	Information should be given concerning the duration of the stations (time identical or non-identical in all stations and, if not, details must be reported.
Identical station duration	Information should be given clarifying if all stations have the same duration. If not details should be given.
OSCE duration	Information should be given concerning total time spent in each OSCE circuit (this should include the written stations, the rest stations and the couplets). When we are dealing with parallel OSCEs and or different circuits it's important to give the total duration of the OSCE exam.
Number and details of teachers	Teachers involved in the planning/development/implementation of the OSCE (not those involved in marking/marking the students).
Number and details of observers	Teachers, staff or other people present in the stations to supervise OSCE development, functioning just as mere observers (not those involved in marking/marking the students).
Number and details of examiners	Information on the number of people (examiners/raters/coders/markers) in charge of assessing students' performance, with description of respective background (teachers, simulated patients, nurses, other students, peers, medical patients, etc.) level of seniority (senior, junior) and respective school(s) (local or external) should be provided. The process of training should also be described.
Number and details of Simulated Patients	Information should be given namely if they were standardized (yes or no), their background (teachers, other students, nurses, other people) etc. and their nature (people, mannequins/models, computer based, video, etc). Information of selection process must be given.
Number details of real patients	Number and details of real patients. Information of selection process must be given.
Training Process for Raters, SPs, Real Patients	Information on training process must be given: who were the trainers, who are the trainees, how long was the training process, format etc.
Number of stations using video, computer, mannequins, other	Number and details of stations using video, computer or mannequins.
Validity data	• Blueprinting and test preparation/face validity. WHO was involved in planning and developing the OSCE? Just OSCE designers? Other teachers? Other experts? HOW were teachers involved on station design? Just through informal discussions? Through a formal process (questionnaire, focus group, interviews, etc.)? Through a more formal process (Delphi technique)? WHAT was searched in terms of literature to support station design? Looking at examples of other OSCE? Consulting blue prints? Other list of problems? • Content validity • Construct validity or • Criterion validity or • Predictive validity or • Other
Reliability data	• Inter-coder reliability – Kappa or other coefficient. • Internal consistency – Cronbach Alpha or other coefficient. • Correlations among stations • Generalizability • Other
Feasibility data	Information on 'cost' and 'time taken' (planning and development) besides the number of staff and students involved.
Feedback given to students, SPs.	Information should be given in terms on feedback (Who, How, What, When).
3. Information on OSCE results	
OSCE results	OSCE results should be reported according to the aims of the study.
Feedback data on the OCE process	Information regarding opportunity for feedback from students, teachers, examiners, simulated patients, real patients, etc. (when applicable).
Relevance data	Information (when applicable) on how students, teachers, patients, etc evaluate the OSCE relevance (yes or no), depending on what they consider should be under assessment.
Fairness data	Information on how students evaluate OSCE fairness (fair or not fair) according to what was taught in class, opportunities for training etc during the course.
Drive learning	Information regarding drive learning: i.e. data concerning students and teachers report on how the OSCE determined specifically directed learning that occurred specifically focused on the OSCE objectives.
Drive teaching	Information regarding 'Drive Teaching': i.e. data concerning teachers reporting that OSCE highlight students' weaknesses and strengths and from this information curriculum and or teaching methods were modified.
OSCE problems/difficulties vs. solutions	Information on the major difficulties related to the OSCE, and if possible, how to overcome them.

means does not inform your judgments of reliability, and so cannot be used as a proxy for reliability. (2) Since each group of students, standardized patients, cases, evaluators and testing situations differs, one cannot assume that the reliability of the individual OSCE is identical to that found in the literature. We hope that increased attention to the statistics and theory behind reliability will increase the accuracy and validity of the data being published regarding OSCE examinations.

Lack of structure within reporting

Lack of structured abstracts (objectives, methods, results and conclusions) in more than 50% of the papers (precisely in 66.3% corresponding to 69 out of 104 studies) creates difficulty for readers. Well-structured, informative papers would aid in disseminating information regarding the OSCEs, as well as facilitating continuity within an occasionally disparate literature.

Discussion

Lack of information, heterogeneity when reporting data, lack of standardized vocabulary and weak structure within reports suggest a significant concern for readers if they wish to transfer the results to their daily practice, for researchers when interpreting or replicating a study and also for reviewers when conducting a systematic review.

Considering the above-mentioned concerns, the authors suggest a 'Comprehensive checklist for those describing the use of OSCEs in the report of educational literature', structured upon three sections (Table 6):

- Information on Course Background;
- Information on OSCE Design;
- Information on OSCE Results.

Our aim is not only to provide a list of items to be reported, but also to clarify their specific intrinsic meaning, in order to unify OSCE medical terminology. We hope to initiate a discussion amongst editors, publishers, researchers and readers in order to reach consensus on best practice for reporting OSCE studies. We strongly recommend that this type of systematic approach to reporting data be encouraged for individuals who are examining the reliability, validity and feasibility of the OSCE.

We would like to highlight the fact that not all items need be mentioned in every study using an OSCE. Authors must decide what specific analyses to report, and key items will depend on the specific purpose of that particular OSCE. It is evident that the exigency with the information to be reported in a study used to evaluate students, curricula or teaching is not as high as when the study is examining the OSCE itself. Therefore, we suggest the following checklist for investigation in which the analysis and use of the OSCE is the primary thrust of the study reported.

Conclusions

We have provided a checklist for reporting studies on the validity, reliability and feasibility of the OSCE for pre-graduation medical education. Our aim was to identify and describe the problems in the reports we analysed for our BEME Systematic Review and suggest ways to standardize future reporting in this area.

Our suggestion constitutes a first attempt to fill a gap in the research on OSCE and should be regarded as a basis for future work, pending the feedback from the research community in medical education. If implemented, we hope these suggestions can ease improve transferability of results among Medical Educators and, moreover, advance research on this very important assessment method. We also hope that it could simplify future systematic reviews in this area, making them more transparent and easy to perform.

Acknowledgment

The authors wish to thank Ronald Harden and Alex Haig for their invaluable contribution to this report.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

MADALENA FOLQUE PATRÍCIO is president of the Association for Medical Education in Europe (AMEE), project director at the Centre for Evidence Based Medicine and assistant professor at the Institute for Introduction to Medicine, Faculty of Medicine of the University of Lisbon.

MIGUEL JULIÃO is a resident in Family Medicine and Scientific Consultant at Centre for Evidence Based Medicine, Faculty of Medicine of the University of Lisbon.

FILIPA FARELEIRA is a resident in Family Medicine and Scientific Consultant at Centre for Evidence Based Medicine, Faculty of Medicine of the University of Lisbon.

MEREDITH YOUNG is a PhD Candidate in the Departments of Psychology, Neuroscience and Behaviour, McMaster University.

GEOFFREY NORMAN is a professor in the Department of Clinical Epidemiology and Biostatistics, McMaster University.

ANTÓNIO VAZ CARNEIRO is a professor of Medicine and Head of Centre for Evidence Based Medicine, Faculty of Medicine of the University of Lisbon.

Notes

1. A separate list of the 104 references is presented in Appendix 1.
2. http://www.bemecollaboration.org/beme/files/starting%20reviews/Appendix%20IIIA%20BEME%20Coding%20Sheet.pdf
3. A separate list of the 18 studies – not included on the BEME Systematic Review but mentioned in Sections 'Methods' and 'Results' of this article to illustrate the 'lack of standardized vocabulary' and 'statistical issues' – is presented in Appendix 2.

References

Downing SM, Haladyna TM. 2004. Validity threats: Overcoming interference with proposed interpretations of assessment data. Med Educ 38(3):327.

Hart IR, Harden RM. 2002. Best Evidence Medical Education (BEME): A plan for action. Med Educ 22(2):141–144.

Harden V, Harden RM, McManus N, Lilley P. 2002. The objective structured clinical examination (OSCE). 1st ed. http://www.bemecollaboration.org/beme/pages/resources/biblio.html

Harden RM, Stevenson M, Downie WW, Wilson GM. 1975. Assessment of clinical competence using objective structured examination. Br Med J 1:447–451.

Norman GR. 2000. Reflections on BEME. Med Teach 22(2):141–144.

Poenaru D, Morales D, Richards A, Oconnor M. 1997. Running an objective structured clinical examination on a shoestring budget. Acad Med 173(6):538–541.

Reznick RK, Smee S, Baumber JS, Cohen R, Rothman A, Blackmore D, Berard M. 1993. Guidelines for estimating the real cost of an Objective Structured Clinical Examination. Acad Med 68(7):513–517.

Streiner DL, Norman GR. 2003. Health measurement scales: A practical guide to their development and use. Oxford: Oxford University Press.

Appendix

Appendix 1. List of 104 papers included in the BEME Systematic Review used for this study.

1. Remmen R, Scherpbier A, Denekens JO et al. Correlation of a written test of skills and a performance based test: a study in two traditional medical schools. Medical Teacher 2001. 23(1):29–32.
2. Harden RM. Twelve tips for organizing an objective structured clinical examination OSCE. Medical Teacher 1990 12(3/4):259–264.
3. Rezler AG, Woolliscroft JA, Kalishman SG. What is missing from patient histories? Medical Teacher 1991 13(3):245–252.
4. Afroza A. Use of a PMP manual as a teaching tool to accelerate paediatric teaching in Bangladesh. Medical Teacher 2000 22(4):365–369.
5. Sethuraman KR. The use of objective structured clinical examination OSCE for detecting and correcting teaching-learning errors in physical examination. Medical Teacher 1993 15(4):365–368.
6. Wilkinson TJ, Newblw DI, Wilson PD, Carter JM, Helms RM. Development of a three centre simultaneous objective structured clinical examination. Medical Education 2000 34(10):798–807.
7. Townsend AH, McIlvenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination OSCE for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. Medical Education 2001 35(9):841–846.
8. Verhoeven BH, Hamers JGHC, Scherpbier A, Hoogenboom R, Van der Vleuten C. The effect on reliability of adding a separate written assessment component to an objective structured clinical examination. Medical Education 2000 34(9):525–529.
9. Murdoch ED, Cottrell D. Structured teaching methods enhance skill acquisition but not problem-solving abilities: an evaluation of the silent run through. Medical Education 1999 33(1):19–23.
10. Smee SM, Blackmore DE. Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff. Medical Education 2001 35(11):1009–1010.
11. Hodder RV, Rivington RN, Calcut LE, Hart IR. The effectiveness of immediate feedback during the objective structured clinical examination. Medical Education 1989 23(2):184–188.
14. Lewin LO, Papp KK, Hodder SL et al. Performance of third-year primary-care-track students in an integrated curriculum at case Western Reserve University. Academic Medicine 1999 74(1 Suppl):S82–S88.
18. Basco WT, Gilbert GE, Chessman AW, Blue AV. The ability of a medical school admission process to predict clinical performance and patients satisfaction. Academic Medicine 2000 75(7):743–747.
19. Rogers JC, Dains JE. Can first-year students master clinical skills? Academic Medicine 2001 76(10):1065.
24. Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalizability of OSCE measures across content domains. Academic Medicine 1999 74(12):1320–1322.
26. Remmen R, Scherpbier A, Derese A et al. Unsatisfactory basic skills performance by students in traditional medical curricula. Medical Teacher 1998 20(6):579–582.
27. Friedman Ben-David M. The role of assessment in expanding professional horizons. Medical Teacher 2000 22(5):472–477.
28. Dacre JE. The effect of formal instruction in ophthalmoscopy on medical student performance. Medical Teacher 1993 15(4):321–325.
29. Harden RM. Assessment of clinical competence and the OSCE. Medical Teacher 1986 8(3):203–205.
32. Harris I, Ytterberg S, Anderson D, Kofron P, Kvasnicka J, Moller JH. Tailored Response Test: a new approach for teaching in medical education. Medical Teacher 1997 19(3):194–199.
33. Allen SS, Bland CJ, Harris IB et al. Structured clinical teaching strategy. Medical Teacher 1991 13(2):177–184.
34. Macmillan CSA, Crosby JR, Wildsmith JAW. Skilled task teaching and assessment. Medical Teacher 2001 23(6):591–594.
36. Bradley P, Bradley P, Johns V. Can we teach a gentler rectal examination? Medical Teacher 1999 21(2):207–208.
37. OConnor M, McGraw R, Killen L, Reich D. A computer-based self-directed training module for basic suturing [1]. Medical Teacher 1998 20(3):203–206.
38. Cohen R. Assessing professional behaviour and medical error. Medical Teacher 2001 23(2):145–151.
40. Van der Vleuten C, Scherpbier A, Dolmans DHJM, Schuwirth LWT, Verwijnen GM, Wolfhagen HAP. Clerkship assessment assessed. Medical Teacher 2000 22(6):592–600.
44. Whelan GP. Educational commission for foreign medical graduates: lessons learned in a high stakes, high volume medical performance examination. Medical Teacher 2000 22(3):293–296.
45. Dissanayake AS, Ali BA, Nayar U. The influence of the introduction of objective structured practical examinations in physiology on student performance at King Faisal University Medical School. Medical Teacher 1990 12(3–4):297–304.
47. Rahman SA. Promoting learning outcomes in paediatrics through formative assessment. Medical Teacher 2001 23(5):467–470.
48. Cater JI, Forsyth JS, Frost GJ. The use of objective structured clinical examination as an audit of teaching and student performance. Medical Teacher 1991 13(3):253–257.
49. Feickert JAD, Harris IB, Anderson DC et al. Senior medical students as simulated patients in an objective structured clinical examination: motivation and benefits. Medical Teacher 1992 14(2/3):167–177.
50. Harden RM. What is an OSCE? Medical Teacher 1988 10(1):19–22.
51. A-Latif A. An examination of the examinations: the reliability of the objective structured clinical examination and clinical examination. Medical Teacher 1992 14(2/3):179–183.
56. Heard JK, Allen RM, Cason GJ, Cantrell M, Tank PW. Practical issues in developing a program for the objective assessment of clinical skills. Medical Teacher 1998 20(1):15–21.
58. Bouhuijs PAJ, Van der Vleuten CPM, Van Luyk SJ. The OSCE as a part of a systematic skills training approach. Medical Teacher 1987 9(2):183–191.
59. Frost GJ, Cater JI, Forsyth JS. The use of an Objective Structured Clinical Examination OSCE in paediatrics. Medical Teacher 1986 8(3): 261–269.
60. Philip EB, Philip JR. Is it worth it? A look at the costs and benefits of an OSCE for second-year medical students. Medical Teacher 1989 11(3/4):291–293.

(*continued*)

Appendix 1. Continued.

71. Famuyiwa OO, Zachariah MP, Ilechukwu STC. The objective structured clinical examination in undergraduate psychiatry. Medical Education 1991 25(1):45–50.
72. Matsell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured clinical examination in paediatrics. Medical Education 1991 25(4):471–474.
73. Hill DA, Lord RSA. Complementary value of traditional bedside teaching and structured clinical teaching in introductory surgical studies. Medical Education 1991 25(6):471–474.
76. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. Medical Education 1988 22(4): 325–334.
77. Lunenfeld E, Weinreb B, Lavi Y, Amiel GE, Friedman M. Assessment of emergency medicine: a comparison of an experimental objective structured clinical examination with a practical examination. Medical Education 1991 25(1):38–44.
78. Martin IG, Stark P, Jolly B. Benefiting from clinical experience: the influence of learning style and clinical experience on performance in an undergraduate objective structured clinical examination. Medical Education 2000 34(7):530–534.
80. Bradley P, Humphris G. Assessing the ability of medical students to apply evidence in practice the potential of the OSCE. Medical Education 1999 33(11):815–817.
81. Fowell SL, Maudsley G, Maguire P, Leinster SJ, Bligh J. Student assessment in undergraduate medical education in the United Kingdom. Medical Education 2000 34(Suppl 1):1–49.
84. Fox R, Dacre J, Mclure C. The impact of formal instruction in clinical examination skills on medical student performance – the example of peripheral nervous system examination. Medical Education 2001 35(4):371–373.
85. Wass V. Jolly B. Does observation add to the validity of the long case? Medical Education 2001 35(8):729–734.
86. Humphris GM. Kaney S. Examiner fatigue in communication skills objective structured clinical examinations. Medical Education 2001 35(5):444–449.
87. Ribin NJ, Philip EB. Health care perceptions of the standardized patient. Medical Education 1998 32(5):538–542.
89. Nayar U, Malik SL, Bijlani RL. Objective structured practical examination: a new concept in assessment of laboratory exercises in preclinical sciences. Medical Education 1986 20(3):204–209.
90. Black NMI, Harden RM. Providing feedback to students on clinical skills by using the Objective Structured Clinical Examination. Medical Education 1986 20(1):48–52.
92. Humphris GM, Kaney S. Examiner fatigue in communication skills objective structured clinical examinations.. Medical Education 2001 33(5):444–449.
93. WassV, Jones R, Van der Vleuten C. Standardized or real patients to test clinical competence? The long case revisited. Medical Education 2001 35(4):321–325.
94. McGraw RC, OConnor HM. Standardized patients in the early acquisition of clinical skills. Medical Education 1999 33(8):572–578.
96. Johnston BT, Boohan M. Basic clinical skills: don't leave teaching to the teaching hospitals. Medical Education 2000 34(9):692–699.
99. Searle J. Defining competency – the role of standard setting. Medical Education 2000 34(5):363–366.
101. Hudson JN, Vernon-Roberts JM. Assessment – putting it all together. Medical Education 2000 34(11):953–954
102. Schmidts MB. OSCE logistics – handheld computers replace checklists and provide automated feedback. Medical Education 2000 34(11):957–958.
104. Wass V, McGibbon D, Van der Vleuten C. Composite undergraduate clinical examinations: how should components be combined to maximise reliability? Medical Education 2001 35(4):326–330.
105. Kilminster SM, Delmotte A, Frith H, Jolly BC, Stark P, Howdle PD. Teaching in the new NHS: the specialised ward based teacher. Medical Education 2001 35(5):437–443.
108. Kevelighan EH, Duffy S, Walker JJ. Innovations in teaching obstetrics and gynaecology – the Theme Afternoon. Medical Education 1998 32(5):517–521.
109. Collins JP, White GR, Kennedy JA. Entry to medical school: an audit of traditional selection requirements. Medical Education 1995 29(1): 22–28.
111. Jeffery HE, Henderson-Smart DJ, Hill DA. Competency – based learning in neonatology. Medical Education 1996 30(6):440–444.
112. Murray E, Todd C, Modell M. Can general internal medicine be taught in general practice? An evaluation of the University College London model. Medical Education 1997 31(5):369–374.
113. OConnor HM, McGraw R. Clinical skills training: developing objective assessment instruments. Medical Education 1997 31(5):359–363.
114. Finlay IG, Maughan TS, Webster DJT. A randomized controlled study of portfolio learning in undergraduate cancer education. Medical Education 1998 32(2):172–176.
115. Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G. Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. Medical Education 1996 30(1):38–43.
116. Cox K. No Oscar for OSCA. Medical Education 1990 24(6):540–545.
118. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? Medical Education 1999 33(4):276–281.
120. Yelland MJ. Standardized patients in the assessment of general practice consulting skills. Medical Education 1998 32(1):8–13.
126. Malik SL, Manchanda SK, Deepack KK, Sunderam KR. The attitudes of medical students to the objective structured practical examination. Medical Education 1988 22(1):40–46.
127. Newble DI. Eight years experience with a structured clinical examination. Medical Education 1988 22(3):200–204.
153. Kaufman DM, Laidlaw TA, Macleod H. Communication skills in Medical Education. Academic Medicine 2000 75(10 Suppl):S90–S92.
156. Doig CJ, Harasym PH, Fick GH, Baumber JS. The effects of examiner background, station organization and time of exam on OSCE scores assessing undergraduate medical students physical examination skills. Academic Medicine 2000 75(10 Suppl):S96–S98.
158. George J, Taylor C, Conran P. The interdisciplinary generalist curriculum project at the medical college of Ohio. Academic Medicine 2001 76(4 Suppl):S100–S103.
163. Colliver JA, Swartz MH, Robbs RS, Cohen DS. Relationship between clinical competence and interpersonal and communication skills in standardized-patient assessment. Academic Medicine 1999 74(3):271–274.
183. Rosebraugh CJ, Speer AJ, Solomon DJ et al. Setting standards and defining quality of performance in the validation of a standardized-patient examination format. Academic Medicine 1997 72(11):1012–1014.
225. Prislin MD, Giglio M, Lewis EM, Ahearn S, Radecki S. Assessing the acquisition of core clinical skills through the use of serial standardized patient assessments. Academic Medicine 2000 75(5):480–483.
230. Kaufman DM, Mann KV, Muijtjens AMM, Van der Vleuten CPM. A comparison of standard setting procedures for an OSCE in undergraduate medical education. Academic Medicine 2000 75(3):267–271.
303. Gupta P, Bisht HJ. A practical approach to running an objective structured clinical examination in neonatology for formative assessment of medical undergraduates. Indian Pediatrics 2001 38(5):500–513.

(*continued*)

Appendix 1. Continued.

307. Blue AV, Chessman AW, Gilbert GE, Mainous AG. Responding to Patients Emotions: Important for Standardized Patient Satisfaction. Family Medicine 2000 2000(5):326–330.
310. Campos-Outcalt D, Watkins A, Fulginiti J, Kutob R, Gordon P. Correlations of Family Medicine Clerkship Evaluations and Objective Structured Clinical Examination Scores and Residency Directors Ratings. Family Medicine 1999 31(2):90–94.
315. Prislin MD, Fitzpatrick CF, Radecki S. A comparison of Family Medicine Clerkship Student Performance Across Multiple Teaching Sites. Family Medicine 1998 30(4):279–282.
317. Prislin MD, Fitzpatrick CF, Lie D, Giglio M, Radecki S, Lewis E. Use of an Objective Structured Clinical Examination Evaluating Student Performance. Family Medicine 1998 30(5):338–344.
363. Burrows SC, Tylman V. Evaluating medical student searches of MEDLINE for evidence-based information: process and application of results. Bulletin of the Medical Library Association 1999 87(4):471–476.
410. Feather A, Stone SP, Wessier A, Boursicot KA, Pratt C. Now please wash your hands': the handwashing behaviour of final MBS candidates. Journal of Hospital Infection 2000 45(1):62.
442. Poenaru D, Morales D, Richards A, OConnor M. Running an objective structured clinical examination on a shoestring budget. American Journal of Surgery 1997 173(6):538–541.
444. Chateny M, Maguire T, Shakun E, Chang G, Cook D, Warnock GL. Does volume of clinical experience affect performance of clinical clerks on surgery exit examinations? American Journal of Surgery 1996 172(10):366–372.
462. Elnicki DM, Halbritter KA, Antonelli MA, Linger B. Educational and career outcomes of an internal medicine preceptorship for first-year medical students. Journal of General Internal Medicine 1999 14(6):341–346.
464. Smith JL. Cancer education for the generalist physician. Family Medicine 2001 33(5):371–375.
479. Sloan PA, Plymale MA, Johnson M, Vanderveer B, LaFountain P, Sloan DA. Cancer pain management skills among medical students: the development of a cancer pain objective structured clinical examination. Journal of Pain and Symptom Management 2001 21(4):298–306.
659. Othman. The Teaching of Anatomy in the Integrated Medical Curriculum as Practised at the School of Medical Sciences, University Sains Malaysia. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 111–113.
661. Cunnington JPW, Neville AJ, Norman GR. The Risks of Thoroughness: Reliability and Validity of Global Ratings and Checklists in an OSCE. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 143–145.
664. Tann M, Amiel GE, Bitterman A, Ber R, Cohen R. Analysis of the Use of Global Ratings by Standardized Patients and Physicians. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 191–192.
665. Troncon LEA, Rodrigues MLV, Piccinato CE, Figueiredo JFC, Peres LC, Cianflone ARL. Overcoming Difficulties in the Introduction of a Summative Assessment of Clinical Competence in a Brazilian Medical School. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 197–199.
669. Broudo M, White M, Rodenburg D et al. The Effectiveness of Interactive Multimedia as an Instructional Aid for Learning Basic Clinical Skillds and Knowledge. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 321–326.
673. Ban N, Hatao M, Ohtaki J et al. An OSCE Trial in Japan: Feasibility and Correlation with a Written Test. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 404–406.
674. Battles JB, Sprankell SJ, Baker L et al. OSCE Stations for the Longitudinal Assessment of Cancer Screening and Detection. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 407–409.
715. Dijchs R, Prince KJAH, Van der Vleuten CPM, Scherpbier AJJA. Validity of objective tests towards peer-rated competence by students. Medical Teacher 2003 25(3):273–276.
797. Leila N-M, Pirkko H, Eeva P, Eija K, Reino P. Training medical students to manage a chronic pain patient: Both knowledge and communication skills are needed. European Journal of Pain 2006 10(2):167–170.
798. Barman A. Critiques on the objective structured clinical examination. Annals of the Academy of Medicine, Singapore 2005 34(8):478–482.
808. Shriner CJ, Hickey DP. Assessing preclinical medical students' ability to communicate in writing [3]. Family Medicine 2005 37(3):159–160.
810. Mukohara K, Kitamura K, Wakabayashi H, Abe K, Sato J, Ban N. Evaluation of a communication skills seminar for students in a Japanese medical school: A non-randomized controlled study. BMC Medical Education 2004 4(18 NOV 2004).
812. Pierre RB, Wierenga A, Barton M, Branday JM, Christie CDC. Student evaluation of an OSCE in paediatrics at the University of the West Indies, Jamaica. BMC Medical Education 2004 4(16 OCT 2004).

Appendix 2. List of 18 papers not included in the OSCE Systematic Review but referred in sections 2 and 3 of this article.

15. MacRae HM, Vu NV, Graham B, Word-Sims M, Colliver JA, Obbs RS. Comparing checklists and databases with physicians' ratings as measures of students' history and physical-examination skills. Academic Medicine 1995 70(4):313–317.
25. Ram P, Van der Vleuten C, Rethans JJ, Grol R, Aretz K. Assessment of practicing family physicians: comparison of observation in a multiple-station examination using standardized patients with observation of consultations in daily practice. Academic Medicine 1999 74(1):62–69.
74. Biran LA. Self-assessment and learning through GOSCE group objective structured examination. Medical Education 1991 25(6):475–479.
79. Singleton A, Smith F, Harris T, Ross-Harper R, Hilton S. An evaluation of the Team Objective Structured Clinical Examination (TOSCE). Medical Education 1999 33(1):34–41.
100. Humphris GM, Kaney S. The objective structured video exam for assessment of communication skills. Medical Education 2000 34(11):939–945.
110. Hill DA, Guinea AI, McCarthy WH. Formative assessment: a student perspective. Medical Education 1994 28(5):394–399.
151. Macmillan MK, Fletcher EA, Champlain AF, Klass DJ. Assessing post-encounter note documentation by examinees in a field test of a nationally administered standardized patient test. Academic Medicine 2000 75(10 Suppl):S112–S114.
154. Champlain AF, Macmillan MK, Margolis MJ, Klass DJ, Lewis E, Ahearn S. Modelling the effects of a test security breach on a large scale standardized patient examination with a sample of international medical graduates. Academic Medicine 2000 75(10 Suppl):S109–S111.
161. Williams RG, McLaughlin MA, Eulenberg B, Hurm M, Nendaz MR. The patient findings questionnaire: one solution to an important standardized patient examination problem. Academic Medicine 1999 74(10):1118–1124.
235. Pfeiffer CA, Ardolino AJ, Madray H. The impact of a Curriculum Renewal Project on Students Performances on a Fourth-year Clinical Skills Assessment. Academic Medicine 2001 76(2):173–175.
237. Rose M, Wilkerson L. Widening the Lens on Standardized Patient Assessment: what the encounter can reveal about the development of clinical competence. Academic Medicine 2001 76(8):856–859.
239. Swartz MH, Colliver JA, Robbs RS. The interaction of examinees ethnicity and standardized patients ethnicity: an extended analysis. Academic Medicine 2001 76(10 Suppl):S96–S98.
240. Floreck LM, De Champlain AF. Assessing sources of score variability in a multisite medical performance assessment: an application of hierarchial linear modeling. Academic Medicine 2001 76(10 Suppl):S93–S95.
241. Prislin MD, Lie D, Shapiro J, Boker J, Radecki S. Using standardized patients to assess medical students professionalism. Academic Medicine 2001 76(10 Suppl):S90–S92.
435. Fox RA, Dacre JE, Clark CL, Scotland AD. Impact on medical students of incorporating GALS screen teaching into the medical school curriculum. Annals of Rheumatic Diseases 2000 59(9):668–671.
670. Mercer C, Holsgrove G. Developing In-Course and Final Assessments in a New Dental Curriculum. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 374–376.
675. Boulet J, Friedman Ben-David M, Hambleton R, Burdick WP, Ziv A. Assessing the Adequacy of the Post-Encounter Written Scores in Standardized Patient Exams. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 410–412.
676. Burdick WP, Pujol-Farriols R, Martinez-Carretero JM, Descarrega R, Friedman Ben-David M, Boulet J. Reproducibility of Emergency Medicine Standardized Patient Cases from Philadelphia to Barcelona. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997 413–415.

CHAPTER 4
Evidence on technical and economic feasibility. Is the OSCE a feasible method for assessing undergraduate medical students?

Is the OSCE a feasible tool for assessing competencies in undergraduate Medical Education? Evidence from a BEME Systematic Review

MADALENA PATRÍCIO, MIGUEL JULIÃO, FILIPA FARELEIRA & ANTÓNIO VAZ CARNEIRO

Faculty of Medicine of the University of Lisbon

Abstract

Background: The Objective Structured Clinical Examination (OSCE) was introduced by Harden et al. in 1975 trying to answer some of the problems regarding the assessment of clinical competencies. Despite the increasingly widespread use of OSCEs, debate continues with arguments as 'why use such a demanding format if other methods are available?'

Aims: To review and synthesize evidence on technical and economic feasibility of OSCE in undergraduate medical studies.

Methods: BEME methodology was applied by two independent coders to 1083 studies identified by literature search from 1975 until the end of 2008.

Key findings: the OSCE is a feasible approach to the assessment of clinical competence for use in different cultural and geographical contexts; to assess a wide range of learning outcomes; in different specialties and disciplines; for formative and summative purposes; to assess students a curriculum or an educational intervention; in the different phases of education including the early and later years of the undergraduate curriculum; and in different health care professions.

Conclusions: Despite being an expensive test format, evidence suggests that the use of OSCE produces reliable results. The study also suggests that one reason for the wide-scale adoption of the OSCE and the feasibility of its use in different contexts and situations is its inherent flexibility in terms of the number of students that can be assessed, the number of examiners included, the type of patients represented, and the format of the examination itself, including the length of the examination, the number and duration of stations.

Introduction

Increasing attention has focused on assessment in medical education and a number of new approaches have been described. Issues such as validity, reliability, feasibility and impact on learning are at the centre of the educational debate (van der Vleuten & Schuwirth 2005; Norcini & McKinley 2007; Boursicot et al. 2011; Norcini et al. 2011).

The Objective Structured Clinical Examination (OSCE) introduced by Harden et al. in 1975 was described by Norman in 2002 as the '*gold standard for clinical assessment*'. However, teachers may still be concerned as to whether the OSCE is feasible

Practice Points

- OSCE appears to be a feasible way to assess to assess students, a curriculum or an educational intervention, in a variety of learning outcomes and a range of contexts and formats, for formative and summative purposes
- OSCE flexibility is one of its major advantages and a reason for adoption
- The OSCE, because of its unique benefits, is recommended in spite of, in some circumstances, being expensive to administer
- Evidence suggests alternative ways to decrease costs
- Transparent categories for reporting OSCE direct and indirect costs are needed
- Editors should request OSCE reports with information on technical and economic viability to support schools' decisions on OSCE Implementation.

in their own particular settings. Can OSCE be used to assess clinical competencies in specific areas such as psychiatry? While originally described in undergraduate medical education, can it be used to evaluate postgraduate trainees? Is it feasible with a large number of students as opposed to small cohorts? Can it be used when there are financial constraints and available resources are limited? Concerns have also been expressed in relation to the commitment required for faculty, the overall logistics and the costs (Benbow et al. 1988; Frye et al. 1989; Resnick et al. 1993; Cusimano et al. 1994; Heard et al. 1998; Carpenter 1995; Poenaru et al. 1997; Hanson et al. 1998; Feather & Kopelman 1997; Hodges & Lofchy 1997; Kelly & Murphy 2004).

A study of the feasibility of the OSCE was undertaken at the Faculty of Medicine University of Lisbon as part of a Best Evidence Medical Education Systematic Review (BEMER) (http://www.bemecollaboration.org) which looked more widely at OSCE feasibility, reliability and validity. This paper reports on the results and conclusions from the BEMER relating to the feasibility of the OSCE in undergraduate medical education.

In this paper 'feasibility' is defined as '*the capacity of being accomplished or brought about; possible used or dealt with successfully'* and a feasibility study is defined as *'the analysis and evaluation of a proposed project to determine if it is technically feasible, feasible within the estimated cost and profitable'*. Feasibility reflects both 'technical viability' (presence of the necessary elements/conditions to conduct an achievable task) and 'economic viability' (costs necessary for implementation). A feasibility study can help to determine whether a project is financial and practically achievable. Kenkel (2004) related feasibility to the likelihood of success of a project or initiative – basically it (a feasibility study) is answering the question '*Should I do this?*'

The objective of this paper is to describe for papers which reported on the OSCE, the feasibility, both technically and economically. Recognizing the natural bias on submitting and publishing technically feasible OSCEs, we aimed to describe the underlying elements/conditions present in feasible OSCEs and to identify any pattern associated to non-feasible OSCEs.

Methods

BEME methodology as described in the BEME protocol (www.bemecollaboration.org) was undertaken and the specific details are available in Appendix 1. Literature was searched, by a BEME information scientist, from 1975 (date of the first publication on the OSCE), until the end of 2008, resulting in 1083 references that were inserted into a Reference Manager database. Material under analysis was constituted by 1065 papers since, from retrieved studies 10 were unobtainable even after request to authors and editors and 8 were excluded because they report on non-OSCE exams (OSLERs, long case exams, etc). The full list of OSCE studies is available at Appendix 2.

Criteria for inclusion

It was decided to limit the analysis on OSCE reliability, validity and economic feasibility. We looked on studies in English reporting on classical OSCEs performed in undergraduate medical education. The different levels of analysis performed are reported in Figure 1.

A preliminary level analysis was performed in all retrieved studies (1065) which were examined in terms of 'when and where were the studies published' (date and country), 'who published the studies' (name and type of institution) and 'who used the OSCE' (phase of education and professional groups performing the OSCE):

The second level of analysis was performed in all accepted studies (n=366) i.e. classical OSCEs on undergraduate medical education, to collect evidence on the purpose of the OSCE exam. 'non-classical' studies were rejected, to avoid a bias when calculating OSCE reliability. A study was coded 'non-classical' when it did not conform in general terms to the classical approach of the OSCE, as described by Harden et al. in 1975. Among them we found studies: where the candidate is a 'team' or 'group' instead of an individual for example, TOSCE (Singleton et al. 1999), G-OSCE (Hill et al. 1994), GOSCE (Elliot et al. 1994; Fields et al. 1995; Vooijs et al. 1997), GOSPE (Biran 1991), when the assessment is based on video instead of direct observation, for example, VIPSCE (Shallaly & Ali 2004), OSVE

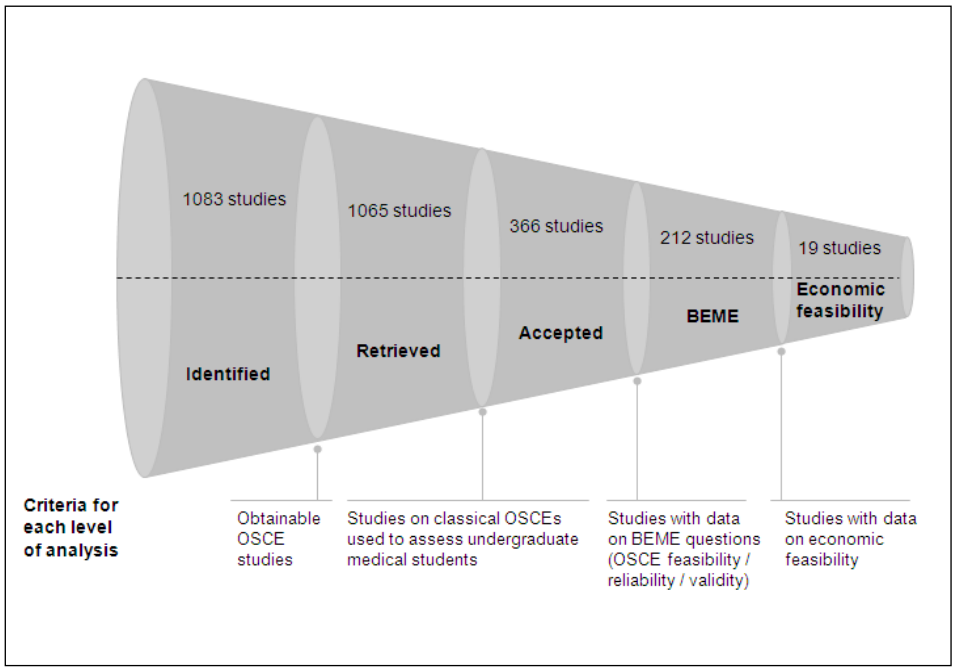

Figure 1. Criteria and number of studies per level of analysis

(Humphris & Kaney 2000), where exams had only written stations (Akici et al. 2004), where exams were only peer-rated (Geddes & Crowe 1998) and 'non-classical' formats for example OSCEs with only one or two stations (van Dalen 2001; Robins 2001).

The third level of analysis was performed in all accepted studies reporting data to support the evidence on the feasibility and/or reliability and/or validity of the OSCE (n=212 studies). They were scrutinized in terms of:

- Learning outcomes assessed by the OSCE (history taking, physical examination, etc.)
- Subject areas under assessment (medicine, dentistry, etc.)
- Underlying elements regarding the design / format of the OSCE exam:
  - Type (high stakes *vs.* non high stakes)
  - Purpose formative *vs.* summative)
  - Feedback provided (by whom, to whom, when and how)
  - Number of students performing the OSCE
  - Number of venues
  - Number of parallel OSCEs
  - Number of circuits
  - Number of stations
  - Number of days
  - Total time
  - Duration of individual stations
  - Scoring tools (checklists *vs.* global ratings)
  - Number and background of examiners
  - Number and background of standardized patients (SPs)
  - Number of real patients
  - Number and background of staff involved
  - Use of mannequins and videos
  - Training process for real patients, SPs and examiners
  - Existence of a pilot study
  - Data on research questions (feasibility, reliability and validity)

The fourth level was based only in 19 studies providing data on OSCE economic feasibility. They were examined in terms of direct and indirect expenses related to 'OSCE development' (design and training), 'production', 'administration' and 'post examination report'. Literature was also scrutinized in terms of alternatives to cope with financial constraints, comparison with costs of other assessment formats and of cost effectiveness.

Coding of papers

The original 'BEME coding sheet' (www.bemecollaboration.org) was modified to accommodate items relevant to the research questions and supporting the above levels of analysis. The coding sheet was replaced by a dynamic Lotus electronic database, allowing coding and establishment of consensus on line. Full description of fields and items under analysis is provided in Appendix 3. Consistency among coders was checked throughout the process and on the results reported in this paper there was 100% agreement.

Results

Results are presented in terms of technical and economic feasibility:

Evidence on technical feasibility

1) Evidence based on all retrieved studies (n=1065)

When analysing all retrieved abstracts (n=1065) no report was found where the OSCE was considered to be not feasible. This does not mean there were no problems identified when the OSCE was implemented only that they did not preclude the realization of the exam.

- *OSCE publications*

Since first described by Harden et al. in 1975 there has been a steady growth in the number of papers reporting the use of OSCE in a wide range of settings. The rise in publications (including gray literature) regarding the OSCE is presented in Figure 2, and shows that from an average of 2.1 studies per year on the first decade (1975-84) a rapid

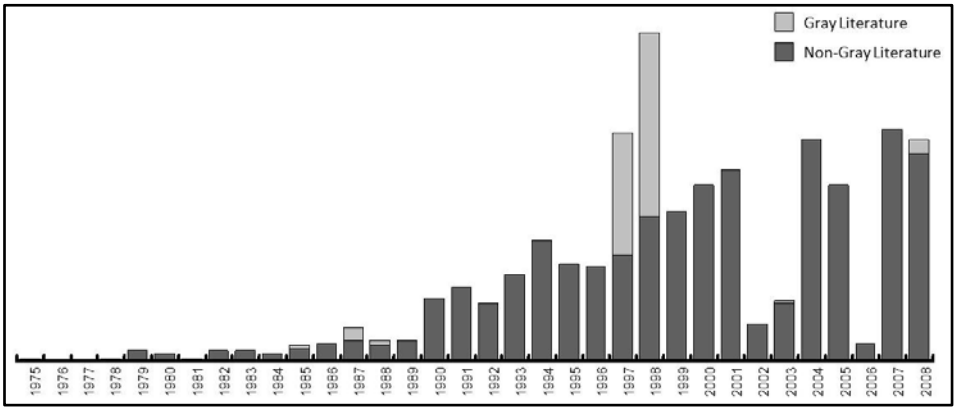

Figure 2. Number of OSCE publications per year (n=1065**)**

increase in numbers followed to an average of 61.5 studies per year in present decade (since 2005). The high peaks in gray literature (1997 & 1998) correspond to the publication of the Proceedings of the 7th and 8th Ottawa Conferences respectively hold in 1996 and 1998 in Maastricht and Philadelphia.

OSCE studies were mainly reported in journals (n=934/88%) with the remaining publications being gray literature (n=131/12%). Two hundred and thirty two journals were identified, with five - *Academic Medicine* (n=176 studies /19%), *Medical Education* (n=134/14%), *Medical Teacher* (n=72/8%), *Teaching and Learning in Medicine* (n=48/5%) and *Advances in Health Sciences Education* (n=32/3%) - publishing 49% of the overall number of OSCE studies.

- *Different cultures and contexts*

The use of OSCE was reported in 5 continents and over 50 countries, confirming the enormous expansion of the OSCE all over the world, and also suggesting there were no strong geographical limitation to OSCE implementation (Figure 3). In some instances (n=337 studies) there was interinstitutional collaboration involving more than one country.

The OSCE was used by 376/35% academic

and 162/15% health institutions some of them linked with academic institutions (n=43/4%), national departments or education/health boards (n=37/3%), commercial firms (n=9/1%) and foundations (n=1%). Usually the institutions publishing the studies were the institution responsible for their implementation.

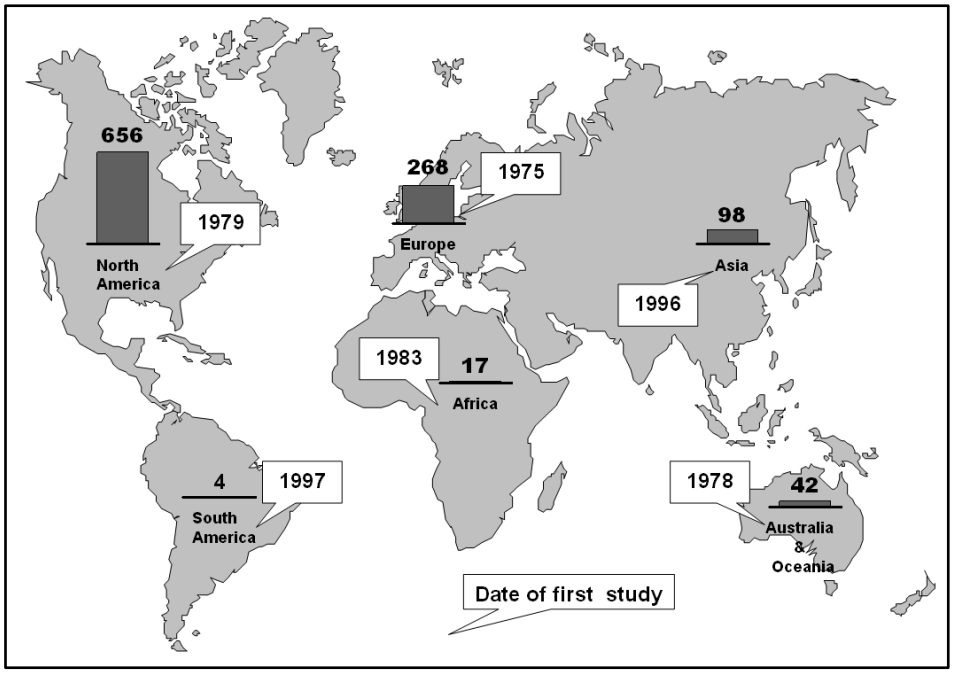

Figure 3. Publications and date of first publication per continent (n= 1065)

- *Multi-professional Settings*

The use of OSCE was documented in 25 health care related professions: Medicine (n=924 studies/87%), Nursing (n=53/5%), Dentistry (n=32/3%) and Pharmacy (n=18/2%) were the fields with more publications. Only a low number of studies report on inter-professional exams probably because, in spite of being encouraged, multi professionalism is not a common feature in medical education. Medicine/dentistry and medicine/nursing, each with 3 studies, were the associated professions with more publications.

- *Phases of Education*

All phases of education – undergraduate, postgraduate and CPD - were covered in OSCE publications predominantly performed by undergraduate (n=672 studies /58%) and postgraduate candidates (n=226/19%). A small number of studies (n=45/4%) relate to continuous professional development, licensure exams for practice (n=60/5%) and overseas registration (n=61/5%). We also found studies applying simultaneously to all medical education levels (n=65/6%) which predominantly correspond to theory papers.

2) Evidence based on all studies with data on the research questions (n=212 studies/263 exams)

As some of the 212 studies report in more than one exam the material under analysis in terms of economic feasibility correspond to 263 OSCE exams. Below we report a summary of main findings since the limitation of space do not allow the report of full results. Detailed description of the above underlying elements and conditions regarding technical feasibility of the OSCE will be reported as a BEME systematic review to be submitted soon.

- *Stage of curriculum*

We found OSCEs performed by students from

every curricular year namely from 1rs year (n=4 studies/2%), 2nd year (n=34/13%), 3th year (n=61/23%), 4th year (n=39/15%), 5th year (n=28/11%) or final year (n=28/10%). We also found studies reporting OSCEs performed simultaneously by students of two different curricular years (n=14/5%). There was not a single reference indicating the OSCE could not be performed within a specific phase of education or curriculum year.

- *OSCE purpose*

The OSCE has been described as a tool to assess competences of students (n=212 studies), to evaluate the curriculum (n=58 studies) and to evaluate a curricular intervention (n=55 studies).

- *Type of exam*

OSCE appears to be feasible with all types of exams, the majority of studies reporting OSCEs performed in the context of high stakes exam (n=115 studies /44%). OSCEs were also performed with volunteer students/no grading (n=44/17%) and data was unclear or missing in 95/36% of the studies.

- *Role of exam*

Concerning the assessment of students' competences, the OSCE has proven to be feasible in formative (n=32 studies /12%), summative (n=74/28%) or both (n=50/19%). In 78/30% of the studies there was unclear or no information on feedback provision.

- *Feedback*

Studies were analysed in terms of

- Who gave the feedback: teachers (n=53 studies /20%), SPs (n=5/2%), examiners (n=1/0,4%), experts (n=1/0,4%), observers (n=1/0,4%) or students (n=2/1% , 2)
- To whom was it given: students (n=72 studies /27%), teachers (n=2/1%), students and teachers (n=2/1%) or students and examiners (n=4/2%)
- When was it given: at the end of the exam (n=22 studies /8%), after each station (n=13%/5%), after both situations (n=2/1%) or after rest stations (n=1/0,4%)
- How was it given: during 1 minute (n=7 studies /3%), 2 minutes (n=2/1%) or 7 minutes (n=3/1%).

- *Learning Outcomes*

Based on the reviewed papers, the OSCE was feasible to assess 27 different types of learning outcomes: physical examination (n=152 studies/58%), history taking (n=142/54%), patient-investigation-data-problem-solving (n=105/40%), communication-skills (n=80/30%), diagnosis (n=71/27%), management-prescription-writing-referral (n=66/25%) and practical-procedures (n=59/22%) were the learning outcomes with more publications.

- *Subject areas*

The OSCE was used to assess learning outcomes in a range of 25 different specialties, for example internal medicine (n=42 studies/16%), ambulatory-family--medicine-primary-care (n=35/13%), surgery (n=34/13%), paediatrics (n=23/9%), obstetrics-gynaecology (n=21/8%) and psychiatry (n=17/ /6%). Information was unclear or not available in 30 studies (11%). The feasibility of using the OSCE in different medical specialties was also highlighted by the fact that many were published in medical specialty journals like Family Medicine (n=17 studies/2%), American Journal of Surgery (n=16/ /2%) and Psychiatric Bulletin (n=12/1%).

- *Students*

OSCE is reported in 75% of the exams as feasible both in a large number of students up to 1237, as well as a low number as 6 candidates (Figure 4).

For figures 4-8, any interval "a" to "b" on the x-scale should be read as]a,b]. The figures represent the distribution of the overall OSCE exam on the x-scale variable, even when the exam was performed in more than one venue, in several parallel OSCES, days, circuits or cycles etc). The column in the right 'NA/NC' correspond to non available (NA) or non clear data (NC).

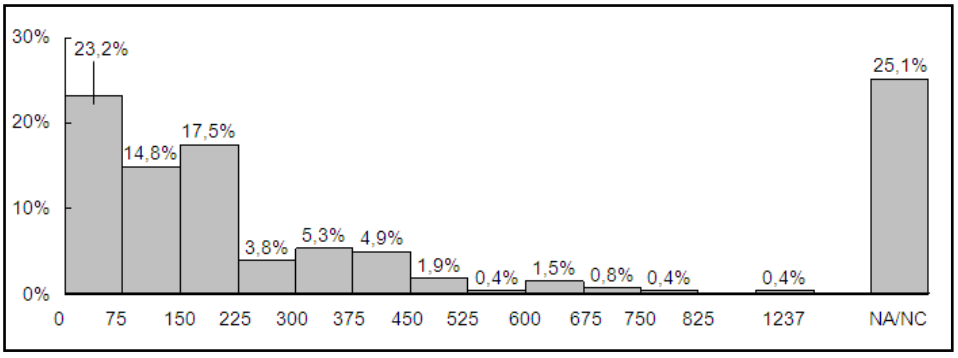

Figure 4. Distribution of OSCE exams per number of students (n= 263 exams)

- *OSCE design and time*

The number of students to be assessed is reflected in the number of stations (varying from 4-40). In order to accommodate a high number of students some studies report on OSCEs performed in: several venues (up to 4), over several days (1-21), in parallel circuits where the examination is repeated sequentially (n=22 studies), simultaneously (n=15 studies) or both (n=1 study). Parallel studies can in turn be repeated in several cycles (2-7).

Information on the duration of the OSCE exam is available in 44% of the exams varying from a minimum of 20 minutes (in a 4 station OSCE) to a maximum of 315 minutes (in an OSCE with 16 stations (and 8 rest stations although this number is not clear and we did not know to what it corresponds: several days, several parallels, several cycles, etc.) (Figure 5).

In what concerns the time per station information is only reported in 30% of the exams from 6 minutes to a maximum of 20 minutes (Figure 6).

- *Recording and scoring (ratings)*

While not all reports documented the type of rating scale used (only available in 30% of the studies), it was clear that checklists (n=138 studies/52%), global ratings (n=12/5%) or both (n=41/16%) were used in the OSCE exams. When some assessors preferred the use of checklists and others the use of global ratings, it was clear that it is possible to use either one.

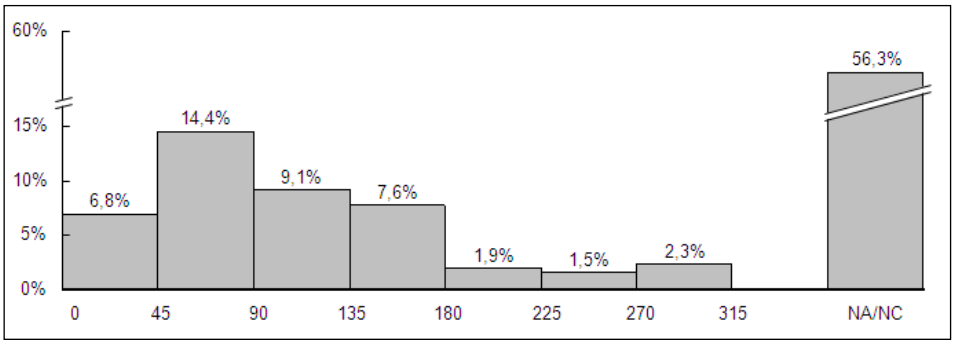

Figure 5. Distribution of OSCE exams per OSCE total duration in minutes (n= 263 exams)

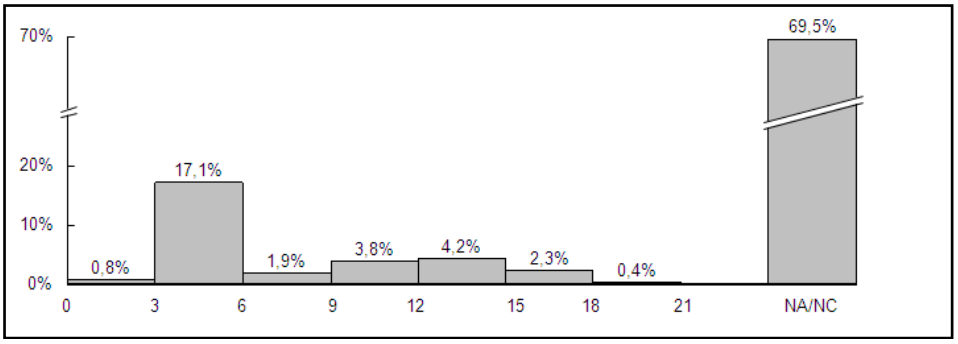

Figure 6. Distribution of OSCE exams per station time in minutes (n= 263 exams)

- *Examiners*

In papers where the exact number of examiners was available (n=79 studies/30%) we found 4 exams where as few as 2 examiners were involved, while in one study 189 were reported corresponding to several examinations (possibly 8 but information is not clear and we do not know if this corresponds to an OSCE performed in several days, in parallel circuits, in different cycles, etc.). (Figure 7).

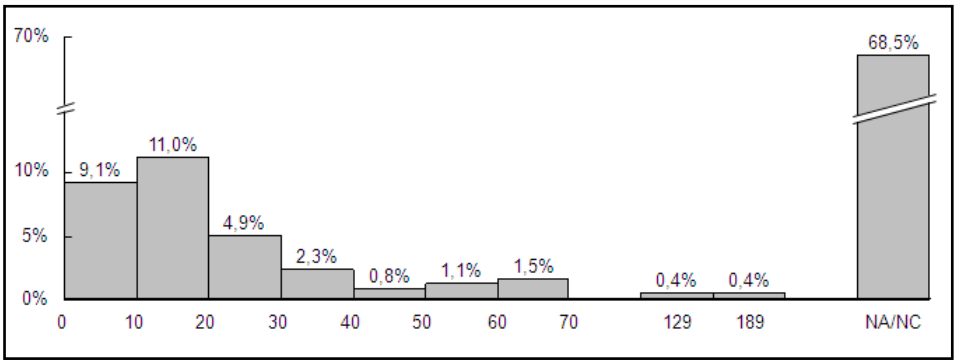

Figure 7. Distribution of OSCE exams per number of examiners (263 exams)

- *Staff and patients*

Only 20 papers (8%) gave information on the total number of staff engaged in an OSCE varying from 1 up to 129, and here again we are reporting the total number of staff per OSCE.

In 18 studies (7%) it was showed that it is feasible to organise an OSCE using real patients from 1 up to 10 or, alternatively, simulated patients from 1 up to 93, reported in 86/33% studies (Figure 8).

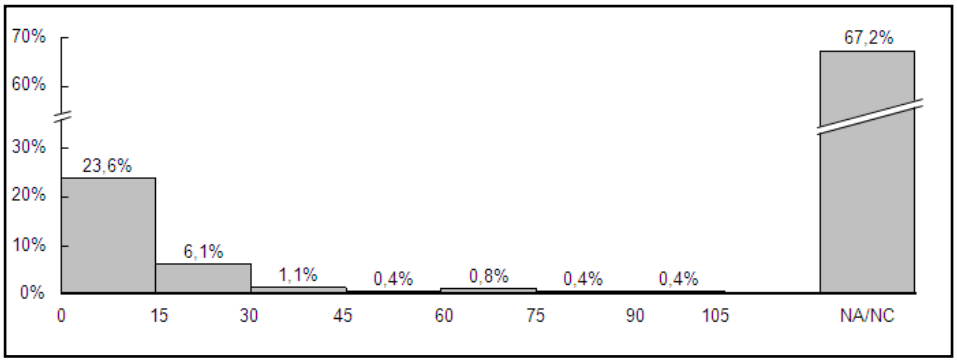

Figure 8. Distribution of OSCE exams per number of simulated patients (263 exams**)**

It would be of interest to note how many studies did not use simulated patients but this was not possible because, although 67% of the papers do not report information on the number of simulated patients, this does not imply they were not used. Information can be just missing. This is again an example justifying the need for more detailed reports on the OSCE exam.

Evidence on economic feasibility

Evidence based on all studies reporting data on OSCE costs and/or time (n=19)

Concerns have been expressed about the cost of OSCE implementation. Only a small number of studies (n=19/4%) quantified costs, reporting a wide range. Fourteen papers provide evidence, some reporting on 'overall OSCE cost', some on 'cost per candidate' and others on 'cost per candidate per station'.

Results in US$ are presented in Table 1. For Feather & Kopelman (1997) who reported the cost in Pounds, and for Kelly & Murphy (2004) who gave it in Euros the currency converter rates reported by the authors were used for conversion into US$. For Bembow et al. (1998), who reported the costs in Pounds, Poenaru et al. (1997) and Cusimano et al. (1994) who reported costs in Canadian dollars, we used the 'World's favourite Currency Site' (2011) for conversion. The remaining studies in the table below do not report the overall costs of the exam (only SPs or examiners' costs). Distinction was made between direct costs i.e. (out of pocket money) and indirect costs (hidden costs) because as stated by Carpenter (1995) and Kelly & Murphy (2004) '*the difference can be very high attaining respectively 80% and 74% of total OSCE budget'*.

The OSCE cost per candidate per station, probably the best indicator to allow comparisons between studies, varied from US$ .88 to 6.9 for direct costs and US$ 13.61 to US$ 50 for all costs. However, as Benbow et al. pointed out already in 1998, the '*usefulness of estimating costs is limited by the lack of consensus of what should be included in costing*'. When looking at the type of exams (high *vs.* low stakes) our results confirm Carpenter's (1995) statement (except for Feather et al. 1997) that direct costs are substantially higher for high stakes OSCE studies).

Table 1 - OSCE costs in US$ (n= 19 studies)

Information on studies			Costs			Authors' views		
First Author	Year	Type of Exams	Cost Per station per student	Cost per station for all students	Cost for all stations for all students	OSCE is a complex & expensive exam	OSCE costs can be reduced	Worthwhile due to unique benefits in spite of being expensive to administer
Studies reporting only on Direct Costs								
Feather A[1]	1997	High St	0.88-2.04	176 - 220 [7] 408 - 510 [8]	4400 - 5500 [7] 10200 - 12750 [8]	Yes	Yes	Yes
Hanson M[1]	1998	Low St	1,09	438	1752	Yes	Na	Yes
Petrusa ER[4]	1987	Low St.	1,24	253	4300	Na	Na	Yes
Frye AW[2]	1989	Low St	1,45	170	3723	Yes	Yes	Yes
Joorabchi B[3]	1991	High St	I,67	58,45	1987	Na	Na	Yes
Benbow EW[4]	1998	No Grade	2,35	188	3760	Na	Na	Yes
Poenaru D	1997	High St	3,75	67,5	4860	Yes	yes	Yes
Carpenter JL	1995	High St	5,63	1126	18016	Na	yes	Na
Cusimano MD	1994	High St	6,90	276	1656	Yes	yes	Na
Studies reporting on direct plus indirect costs								
Gilson GC	1998	High St	13,61	2695	16169	Yes	Yes	Yes
Szerlip H[5]	1998	Low St	18,66	2800	3600	Na	Na	Yes

Azcuenaga J	1998	Low St	22.57	519	17130	Na	Yes	Na
Kelly M	2004	Low St	25.82	1833	9166	Yes	Na	Na
Heard Jk[3]	1998	High St	50	1000	150000	Yes	Yes	Yes
Studies reporting only on SPs or Examiners data								
Grand'Maison P	1985	No Grade	only information on cost for SPs			Na	NA	Yes
Battles JB	1992	High St	only information on cost for SPs			Yes	Na	Na
Sloan PA	2001	No Grade	only information on cost for SPs			Na	Na	Yes
Young WW	1995	Low St	only information on cost for examiners			Na	NA	yes
Hodges B[6]	1997	High St	only information on time for examiners			Yes	Yes	Yes

[1]Approximate cost was based on average of SP range costs; [2]Cost estimated by Carpenter (1995); [3]Rest stations included; [4]Couplets not included; [5]Costs based only on two invasive stations; [6]Mini OSCE with only 4 stations; [7] Feather's cost for 200 students; [8] Feather's cost for 250 students; NA = not available.

Information on costs for staff time is scarce (7 studies) and not systematized. Battles et al. (1992) reported on 10 weeks of clinical assessment team effort for setting up an OSCE, Cusimano et al. (1994) considered the OSCE as labour-intensive (starting 6 months ahead plus 16 faculty hours in the 8 weeks prior to examination) and Heard et al. (1998) reported on approximately 5 hours for a team of six people. In 1998 Gilson et al. report on 8.5 hours per 'student group tested' which means 2.1 hours per student (including development, production administration and grade) for a 6 station OSCE, Hodges & Lofchy (1997) reported on 3.75 hours per student in a 4 station OSCE and 42 students which is significantly lower than the 8.2 hours reported by Cusimano et al. in 1994 for an OSCE with 6 stations and 40 students including development and administration. Azcuenaga et al. (1998) report costs of US$ 6.07 and US$ 23 for design and examination and Kelly & Murphy (2004) a cost of US$ 6.7 regarding overall personnel costs.

The same occurs with standardized patients, with limited data on time and rates of pay in only 9 studies. The lower examining costs per hour (a maximum of US$ 10) are reported by Battles et al. (1992), Carpenter (1995) and Gilson et al. (1998). Sloan et al. (2001), in the context of a Cancer pain management, used standardized volunteer patients paid approximately US$ 20 per hour. All these figures are significantly lower than costs given by Joorobachi (1991) and Szerlip et al. (1998) respectively of US$ 33 and US$ 50 per hour per SP. In terms of training, Petrusa et al. (1987) and Battles et al. (1992) report on US$ 10. Kelly & Murphy (2004) report on US$ 45,29 and US$ 81 per training or examining day. Three OSCEs with invasive stations report higher costs for SPs namely US$ 13 for a pelvic station (Battles et al.1992) and US$ 25 for breast or pelvic station (Gilson et al. 1998) with both indicating a cost of US$ 10 for standard stations per SP/hour. In terms of overall OSCE cost, Szerlip et al. (1998) reported on US$ 2800 if an invasive station is used i.e. more US$ 800 than the cost for a standard station.

Data regarding examiners is again scarce, with a first study from Newble & Swanson in 1988 reporting on raters using from 8.00 to 16.00 hours and markers from 10.00 to 17.00. In 1990 Rutala reported on 5.00 hours testing which increased the costs in 20% and Kelly & Murphy (2004) indicated an individual cost of US$ 45 per examiner with an overall cost of US$ 2582.

Data on administrative costs was based only in 4 studies: US$ 314 for Cusimano et al. (1994), US$ 443 for Kelly & Murphy (2004), US$ 1150 for Carpenter (1995) and US$ 8406 for Azcuenaga et al. (1998). More comparisons are difficult because information is not given or it is unclear. Data on supplies was based in 3 papers and some studies do not clarify what was considered supplies. Frye et al. (1989) report on US$ 1300 and Heard et al. (1998) on US$ 35000 for all costs. Curiously, Azcuenaga et al. (1998) reported no expenses when inquired on 'other costs'.

We are aware that costs limit the use of OSCE and this was why we looked for evidence on two questions: What is the real difference when comparing OSCE with other exams? Can the

costs be reduced? Evidence is not consensual, with some authors considering the OSCE more expensive while others estimating costs comparable to other formats. On one hand we have Cusimano et al. (1994) concluding that OSCE is substantially more expensive in terms of both human and material direct costs (OSCE US$ 6.9 and 8.2 hours for OSCE and no expenses and 2.75 hours for Structured Oral Examination per student), Frye et al. (1989) stating that in spite of actions to decrease costs they continue to exceed those from other testing method and Azcuenaga et al. in 1998 reconfirming those statements, reporting on OSCE US$ 745 compared with two CSAs (Clinical Skill Assessments) of US$ 141 and US$ 180. On the other hand, Kelly & Murphy (2004) and Young et al. (1995), stated that OSCE costs are comparable to cost of other methods of clinical assessment. Feather & Kopelman (1997) go further, stating that although expensive, OSCE cost is modest compared with overall costs of clinical training. Joorabchi (1991) reported a cost of US$ 57 per examinee compared with US$ 140 in a performance based assessment of clinical skills but, only direct costs were reported. The problem with these studies comparing OSCEs with other formats is the lack of underlying information on the specific learning outcomes assessed by each type of exam.

In terms of diminishing OSCE costs, evidence points to significant lower costs when it is used for subsequent examinations. First implementations are reported as more difficult, requiring greater investment in time and costs. Nine studies report measures to decrease OSCE costs. A summary is presented in Table 2.

Kelly & Murphy (2004) alerted to the need of considering ways of rendering the OSCE less expensive, yet maintaining its essential integrity. If the proposal for reducing the number of SPs could raise some concerns regarding OSCE overall quality, Poenaru et al. (1997) confirmed they found identical performance and no fatigue when instead of two SPs only one was used per station.

So far we reported evidence on OSCE costs, on alternative ways to decrease them and on comparing OSCEs with other type of exams. Below we report the evidence on the crucial question regarding economic viability: is it worthwhile to use the OSCE? Results suggest the OSCE should be used in spite of associated costs because of the range of achievable objectives within a single exam. In 14 out of the 19 studies reporting on costs, authors stated that the OSCE '*is worth to be used in spite of costs because of benefits, some unobtainable with other type of exams'* (Frye et al. 1989; Heard et al. 1998; Sloan et al. 2001; Poenaru et al. 1997; Joorabchi 1991; Feather & Kopelman 1997; Gilson et al. 1998; Szerlip et al. 1998; Grand' Maison et al. 1985; Hodges & Lofchy, 1997; Petrusa 1987; Young et al. 1995; Benbow et al. 1998; Hamson et al. 1998).

Table 2 - Proposals to decrease OSCE costs (n=14)

Authors	Year	Suggestions to decrease costs
Cusimano MD et al.	1994	Use volunteer examiners
Joorabchi B. Cusimano MD et al.	1991. 1994.	Use volunteer patients,
Grand' Maison P et al.	1985	Use volunteer pregnant women
Joorabchi B. Poenaru D et al. Young RC et al. Cusimano MD et al.	1991 1997 1995 1994	Use volunteer faculty
Frye AW et al. Poenaru D et al. Kelly M, Murphy A	1989 1997 2004	Use students as raters
Poenaru D et al.	1997	Use only one SP per station Condense total examination time
Young RC et al. Poenaru D et al.	1995 1997	Share space with other departments Share skill standardized patients with other faculties Share coordinator secretary jobs with other units
Frye AW et al.	1989	Bank and share OSCE questions Assemble experienced observers/administrators over time Reuse materials
Poenaru D et al.	1997	Reduce cost for meals and coffee breaks

Among them: Grand' Maison et al. (1985) stated that '*information given by OSCE is more valuable*…,' Joorabchi (1991) reported the OSCE as '*a practical feasible and highly desirable (exam)*', Young et al. (1995) highlighted the '*OSCE provided an useful insight into domain specific training deficits and can diagnoses program and students deficiencies'*, Poenaru et al. (1997) were '*convinced of the usefulness*

of the OSCE', Hanson et al. (1998) mentioned that '*accompanying benefits warranted OSCE expenses*', Gilson et al. (1998) reported that '*cost on time and money makes OSCE feasible and are worthwhile because of information not demonstrated by the usual assessment methods*', Szerlip et al. (1998) defended the use of OSCE '*should be encouraged*', Benbow et al. (1998) admitted that '*pathologists fall to realise that OSCE is ideal for assessment of certain competencies which are sadly lacking in current examinations'*, Heard et al. (1998) reported on '*great benefit for students and positive impact in curriculum adding that the positive endeavour should not be underestimated*' and Sloan et al. (2001) classified the OSCE as '*an useful performance based assess tool allowing faculty to test individual skills in the essential*', concluding the *'OSCE is feasible and consensus reached of introducing it in a large scale'*.

Discussion

As shown above, the number of reports describing the use of the OSCE continues to rise over the past 37 years and there is no sign that there is a decrease interest.

The objective of present study was to make available the underlying elements/conditions of a feasible OSCE and its costs. When analysing every abstract from the 1065 retrieved papers there were no references to non-feasible OSCEs and evidence was found on:

- OSCE universal expansion since it was introduced indicating its worldwide use
- OSCE technical feasibility demonstrating its versatility and multiplicity of designs
- OSCE economic viability although pointing to an expensive exam as its major drawback
- Alternative ways to decrease OSCE costs
- A range of unique educational benefits only achievable with OSCE approach which justifies its use.

If some teachers may be uncertain about using the OSCE as an assessment tool and may be discouraged because of its complexity, associated costs and staff time required, this research demonstrated that the OSCE is feasible when implemented in a wide range of contexts and geographical cultures, to assess different learning outcomes, in a variety of settings and in different phases of education for a broad group of professions, with multiple designs in situations within human and financial constraints. What was impressive was the wide range of contexts reported where OSCE can be used and the different formats in which it can be presented. Should this flexibility and adaptation to local situations be seen as one of the major advantages of this approach and a reason for its adoption?

This is particularly relevant, if we think, for example, on its potential application in the context of outcome-based education (OBE), a model '*providing a compelling statement of significant exit outcomes which may be adapted to suit the local context*' (Harden et al. 1999). By its characteristics the OSCE appears as an ideal method to cope with the exigencies of OBE assessment, namely when asking the students to demonstrate the achievement of learning outcomes at different levels, i.e. showing 'what to do', 'how to do it' and 'what to be'.

As previously highlighted not everything is positive with OSCE use. There are also potential concerns, encompassing the OSCE expansion, that have largely focused on the resources required and the associated costs. One of the problems highlighted by this BEMER is the fact that, in many situations, the costs are hidden and there is no actual information on additional cost with the examiners, patients, and the venue.

The question is prioritising the staff time. It may be felt that time spent on the OSCE is important and can be justified on the need to assess students' clinical competence, and also on being an important learning experience for students with feedback being given. In some situations, particularly in the United States, where significant sums are paid for standardised patients, then additional charges may be incurred.

Also highlighted by this BEMER are the examples where OSCE was feasible with limited resources, suggestions being made by authors on the ways to reduce costs. The words by Cusimano et al. (1994) almost twenty years ago '*diminishing budgets and competitiveness for what funds do exist are two important factors forcing medical educators to examine OSCE costs'* gain particular relevance.

Nowadays, more than ever, when world is living an economic crisis, it is not enough to have authors publishing on the educational benefits of

the OSCE but it is crucial to have a clear report on inherent costs. Therefore, the distinction on direct and indirect cost is essential as well as agreeing on a comprehensive set of transparent categories to report associated costs. Some models were reported:

- Reznick et al. in 1993 proposed costs to be allocated to four phases: development, production, administration and post-examination-reporting
- Carpenter in 1995 suggested the use of three categories to report costs: personnel, standardized patients and administrative costs
- Poenaru et al. in 1997, recommended to divide costs into direct (material honoraria) and indirect (hours of salaried work)
- Azcuenaga et al. in 1998 proposed costs to be allocated to four phases: design, training, examination setting, administration and other costs.

What matters is to explicit clearly what are the items within each rubric. Associated time for each item should be given with indication of overall cost in addition to the cost per hour for SPs, real patients, examiners and faculty, when applicable. The fundamental argument to inform schools' decision regarding OSCE implementation implies evidence on the question 'why should the OSCE be used if less expensive formats are available?'

Where it was possible to find evidence on the ways to decrease costs - namely by recruiting volunteers - OSCE use is advisable in spite of being expensive due to its unique educational benefits. When a school decides on OSCE implementation, what seems important is not to estimate the OSCE cost alone, but also consider what can only be achieved with an OSCE exam. We agree with Cusimano et al. who already in 1994 said that those responsible for deciding if OSCE is *'worth it'* will have to weight factors such as *'available technical expertise, faculty support, space materials and funds'* against factors like '*whether assessment of data gathering and interviewing skills are considered important from an education and evaluation perspective'*.

Looking at the whole BEME systematic review we acknowledge intrinsic and external limitations to this study. The quality of some OSCE reports, with unclear or missing data, and lack of human resources, were major limitations. Finance constraints determined that only the papers published in English were included.

The problem regarding non usable or missing data is quite significant if we consider the high percentages of studies with missing data in important fields. No doubt the quality of reports will continue to evolve and improve and editors of journals can play an important role by requesting structured, transparent and more detailed information from authors. This would allow further statistical analysis which, as previously highlighted, was not completely possible in this systematic review.

If as documented in this systematic review, the evidence already exists what we may see are studies reporting refinements in the administration of the OSCE, for example with the use of new technology for scoring or different forms of patient representation with simulators and computers.

We are also aware of a bias when significant OSCE costs were incurred. Costs are more likely to be reported and considered when values are significant, because when there are no incurred costs (for example with employment of SPs) they are not reported. In many situations costs are hidden and there is no actual additional cost with the examiners, the patients, and the venue all being made available from existing resources.

Finally a word of caution is also needed in terms of comprehensiveness of findings, because not only have the studies a tendency to report successful achievements but journals also have tendency to publish mostly positive studies.

Balancing pro's and con's, we conclude that the evidence brought up with this BEMER helps to understand why already ten years ago Norman stated '*the objective structured clinical examination, with its multiple samples of performance has come to dominate performance assessment'* (Norman, 2002).

This research shows that the Objective Structured Clinical Examination is a feasible, assessment method for undergraduate medical education.

Acknowledgement

The authors wish to warmly thank for their invaluable contributions to this report: Alex Haig

(PhD, Information scientist, NHS Education for Scotland) for literature search, Madalena Mexia (MA in Statistics) for data analysis, Dr. Rui Fonseca and Dr. Ana Catarina Silva (from the Department of Techonology and Informatics of the Faculty of Medicine of the University of Lisbon) for implementation of Lotus database.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on Contributors

MADALENA PATRICIO (MA) is president of the Association for Medical Education in Europe (AMEE), assistant professor at the Institute of Introduction to Medicine and scientific ´consultant at the Center for Evidence Based Medicine both at the Faculty of Medicine, University of Lisbon, Portugal.
MIGUEL JULIÃO is a resident in Family Medicine and scientific consultant at Center for Evidence Based Medicine, Faculty of Medicine of the University of Lisbon.
FILIPA FARELEIRA is a resident in Family Medicine and scientific consultant at Center for Evidence Based Medicine, Faculty of Medicine of the University of Lisbon.
ANTÓNIO VAZ CARNEIRO, MD, PhD, FACP is a professor of Medicine and the head of the Center for Evidence Based Medicine at the Faculty of Medicine University of Lisbon, and head of the Cochrane Coordinating Center Portugal.

References

Akici A, Kalaca S, Goren MZ, Akkan AG, Karaalp A, Demir D, Ugurlu U, Oktay S. 2004. Comparison of rational pharmacotherapy decision-making competence of general practitioners with intern doctors European Journal of Clinical Pharmacology 60(2):75-82.

Azcuenga J, Valls M, Martinez-Carretero JM. 1998. Cost Analysis of Three Clinical Skills Assessment (CSA) Projects. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1.

Battles JB, Sprankell SJ, Carpenter JL, Bedford JA, Kirk LM. 1992. Developing a support system for teaching and assessing clinical competence. Journal of Biocommunication 19(4):19-25.

BEME: Best Evidence Medical Education. Retrieved March 10, 2012. Available from: http:// www.bemecollaboration.org

Benbow EW, Harrison I, Dornan TL, ONeill PA. 1998. Pathology and the OSCE: insights from a pilot study. Journal of Pathology 184(1):110-114.

Biran LA. 1991. Self-assessment and learning through GOSCE (group objective structured examination). Medical Education 25(6):475-479.

Boursicot K, Etheridge L, Setna Z, Sturrock A, Ker J, Smee S, Sambandam E. 2011. Performance in assessment: consensus statement and recommendations from the Ottawa conference. Medical Teacher 33(5):370-83.

Carpenter JL. 1995. Cost analysis of objective structured clinical examinations. Academic Medicine 70(9):828-833.

Cusimano MD, Cohen R, Tucker W, Murnaghan J, Kodama R, Reznick R. 1994. A comparative analysis of the costs of administration of an OSCE. Academic Medicine 69(7):571-576.

Elliot DL, Fields SA, Keenan TL, Jaffe AC, Toffler WL. 1994. Use of a group objective structured clinical examination with first-year medical students. Academic Medicine 69(12):990-992.

Feather A, Kopelman PG. 1997. A practical approach to running an objective structured clinical examination (OSCE) for medical undergraduates. Education for Health 10(3):333-350.

Fields SA, Toffler WL, Elliot DL et al. 1995. Principles of clinical medicine: an interdisciplinary integrated 2-year longitudinal course. Medical Education 29(1):53-57.

Frye AW, Richards BF, Philp EB, Philp JR. 1989. Is it worth it? A look at the costs and benefits of an OSCE for second-year medical students. Medical Teacher 11(3/4):291-293.

Geddes EL, Crowe J. 1998. Peer-Related Objective Structured Clinical Examination. Physioterapy Canada 50(4):269-275.

Gilson GJ, George KE, Qualls CM, Sarto GE, Obenshain SS, Boulet J. 1998. Assessing clinical competence of medical students in women's health care: use of the Objective Structured Clinical Examination. Obstetrics and Gynecology 92(6):1038-1043.

Grand' Maison P, Blouin D, Briere D. 1985. Utilization of the Objective Structured Clinical Examination OSCE in Gynecology/Obstetrics. Proceedings of the Annual Conference on Research in Medical Education 24:65-71.

Hanson M, Hodges B, McNaughton N, Regehr G. 1998. The integration of child psychiatry into a Psychiatry Clerkship OSCE. Canadian Journal of Psychiatry 43(6):614-618.

Harden RM, Stevenson M, Downie WW, Wilson GM. 1975. Assessment of clinical competence using objective structured examination. British Medical Journal 1:447-451.

Harden RM, Crosby JR, Davis MH, Friedman M. 1999. AMEE Guide No. 14: Outcome-based education: From competency to meta-competency: a model for the specification of learning outcomes. Medical Teacher 21(6).

Heard JK, Allen RM, Cason GJ, Cantrell M, Tank PW. 1998. Practical issues in developing a program for the objective assessment of clinical skills. Medical Teacher 20(1):15-21.

Hill DA, Guinea AI, McCarthy WH. 1994. Formative assessment: a student perspective. Medical Education 28(5):394-399.

Hodges B, Lofchy J. 1997. Evaluating psychiatric clinical clerks with a mini-objective structured clinical examination. Academic Psychiatry 21(4):219-225.

Humphris GM, Kaney S. 2000. The objective structured video exam for assessment of communication skills. Medical Education 34(11):939-945.

Joorabchi B. 1991. Objective Structured Clinical Examination in a pediatric residency profram. American Journal of Diseases in Children 145(7):757-762.

Kelly M, Murphy A. 2004. An evaluation of the cost of designing, delivering and assessing an undergraduate communication skills module. Medical Teacher 26(7):610-614.

Kenkel P. 2004. Feasibility Assessment Templates for producers and Cooperatives. 2004 Triennial Conference Change in the rural America Social and Management Changes. Four Points Sheraton Hotel Lexington.

Newble DI, Swanson DB. 1988. Psychometric characteristics of the objective structured clinical examination. Medical Education 22(4):325-334.

Norcini JJ, McKinley DW. 2007. Assessment methods in medical education. Teaching and Teacher Education 23:239-250.

Norcini J, Anderson B, Bollelas V, Burch V, Costa MJ, Duvivier J, Galbraith R, Hays R, kent A, Perrott V, Roberts T. 2011. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010. Medical Teacher 33: 206-214.

Norman G. 2002.Research in medical education: three decades of progress BMJ 29:324:1560-2.

Petrusa ER, Blackwell TA, Rogers LP, Saydari C, Parcel S, Guckian JC. 1987. An Objective measure of clinical performance. American Journal of Medicine 83(7):34-42.

Poenaru D, Morales D, Richards A, OConnor M. 1997. Running an objective structured clinical examination on a shoestring budget. American Journal of Surgery 173(6):538-541.

Reznick RK, Smee S, Baumber JS, Cohen R, Rothman A, Blackmore D, Berard M. 1993. Guidelines for estimating the real cost of an Objective Structured Clinical Examination. Academic Medicine 68(7):513-517.

Robins LS, White CB, Alexander GL, Gruppen LD, Grum CM. 2001. Assessing medical students' awareness of and sensitivity to diverse health beliefs using a standardized patient station. Academic Medicine 76(1):76-80.

Rutala PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ. 1990. Student fatigue as a variable affecting performance in an objective structured clinical examination. Academic Medicine 65(9 Suppl):s53-s54.

Shallaly GH, Ali EA. 2004. Use of Video-Projected Structured Clinical Examination (ViPSCE) instead of the traditional oral (VIVA) examination in the assessment of final year medical students. Education for Health 17(1):17-26.

Singleton A, Smith F, Harris T, Ross-Harper R, Hilton S. 1999. An evaluation of the Team Objective Structured Clinical Examination (TOSCE) Medical Education 33,1 34-41.

Sloan PA, Plymale MA, Johnson M, Vanderveer B, LaFountain P, Sloan DA. 2001. Cancer pain management skills among medical students: the development of a cancer pain objective structured clinical examination. Journal of Pain and Symptom Management 21(4):298-306.

Szerlip H, Anderson DS, Garris JB, Stanton M. 1998. Development and Implementation of a Pelvic Examination Station for a High-Stakes OSCE. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia;1.

van Dalen J, Bartholomeus P, Kerkhofs E, Lulofs R, van Thiel J, Rethans JJ, Scherpbier A, van der Vleuten C. 2001. Teaching and assessing communication skills in Maastricht: the first twenty years Medical Teacher 23(39) 245-251.

van der Vleuten CP, Schuwirth LW. 2005. Assessing professional competence: from methods to programmes. Medical Education 39(3):309-17.

Vooijs MEEC, Scherpbier A, van Dalen J, Ramsay G, Kootstra G. 1997. Group objective structured clinical examination in an Inter-European accident and emergency course. Education for Health 10(1):69-78.

Young WW, Barthold JC, Birenbaum D, Long P, Dion M, Hamilton LA. 1995. An objective structured clinical exam in multisite obstetrics and gynecology clerkship. Teaching and Learning in Medicine 7(3):177-181.

World's Favourite Currency Site. Available from http://www.xe.com/ucc/ Accessed December 12, 2011.

Appendices

Appendices 1, 2 and 3 regarding Chapter 4 will be available as Supplemental Material at the web version of Medical Teacher and they are also presented in the last chapter (see ANNEXES)

Appendix 1 BEME Methodology

Appendix 2 OSCE studies included in the systematic review

Appendix 3 OSCE electronic coding sheet

CHAPTER 5
Evidence on OSCE assessment criteria. Is the OSCE meeting the requirements for a good assessment tool?

Is the OSCE a reliable and valid tool for accessing competencies in undergraduate medical education? Evidence from a BEME Systematic Review

MADALENA PATRÍCIO[1], MIGUEL JULIÃO[1], FILIPA FARELEIRA[1], MEREDITH YOUNG[2], GEOFFREY NORMAN[3] & ANTÓNIO VAZ CARNEIRO[1]

[1]Faculty of Medicine of the University of Lisbon, Portugal, [2]McGill University Faculty of Medicine, Canada, [3]McMaster University, Canada

Abstract

Background: The Objective Structured Clinical Examination (OSCE) was introduced in 1975 by Harden et al. as an attempt to ameliorate clinical assessment. Traditionally, literature has focused on the feasibility, reliability and validity of the assessment tool. In recent years, however, the importance of alternate roles for assessment such as assessment-enhanced learning and assessment as motivation for learning have also been added.

Objectives: To review and synthesize evidence on the reliability, validity, acceptability, fairness and educational impact of the OSCE when used in undergraduate medical education.

Methods: Best Evidence Medical Education (BEME) methodology was applied by two independent coders who examined OSCE literature from 1975 until the end of 2008.

Results: Higher average inter-station reliability was found when global ratings were used. Standardized patients' (SPs) ratings were as reliable as ratings from expert examiners. Studies have shown a positive relation to a number of other measures including multiple choice and short answer texts and clinical ratings. The OSCE is perceived almost universally as a fair and acceptable test with impact on education.

Conclusions: The OSCE is a reliable tool when used with standardized patients or examiner raters, checklists or global ratings. Some validity evidence exists as well as evidence on established fairness, acceptability and educational value.

Introduction

In 1975, following a period of intense criticism of current assessment formats, Harden et al. reported on a novel assessment tool, the Objective Structured Clinical Examination (OSCE). The OSCE was developed as an adaptable clinical examination, consisting of a circuit of multiple stations designed to be 'objective' (all candidates examined in the same stations) 'structured' (all candidates performing the same tasks, with same patients, during same time) and 'standardized' (all candidates assessed by same examiners using same criteria).

Despite the wide adoption of the OSCE, several questions still persist regarding its value and its most potential effective use. Traditionally, the literature on assessment has focused on psychometric properties such as reliability,

Practice Points

- Results on OSCE reliability should assist evaluators when designing OSCEs
- Evidence suggests value in the use of global ratings and the co-ordinated use of standardized patients and expert evaluators
- Evidence points to the OSCE as a fair and relevant exam, well received by teachers, students, examiners and patients which impacts positively on students' learning and teachers' teaching priorities
- Editors should request better quality of studies reporting on OSCE assessment criteria namely on its reliability and validity

feasibility and validity of the assessment tool. In recent years, however, the means and ways of evaluating assessment strategies has broadened, including examination of the relationship between assessment with learning and teaching (van der Vleuten 1996; Norcini 2011; General Medical Council 2011).

The objective of this paper is to describe how the OSCE examination meets relevant assessment criteria as defined by the General Medical Council (GMC) in 2011. In order to describe the alignment between the OSCE and recent assessment standards, we examined the published literature reporting on OSCE examinations when used in undergraduate medical education.

The study is a component of a Best Evidence Medical Education (BEME) systematic review (BEMER) (http://www.bemecollaboration.org) undertaken at the Center for Evidence Based Medicine of the Faculty of Medicine of the University of Lisbon. The objective was to comprehensively address issues surrounding 1) reliability 2) validity 3) acceptability 4) fairness and 5) educational impact of the OSCE exam in the context of undergraduate medical education. The remaining requirements stated by the GMC (2011) namely, feasibility, cost effectiveness and opportunities for formative feedback are object of another study also related to the same BEMER (Patricio et al. *Is OSCE a feasible tool for assessing competencies? Evidence from a BEME Systematic Review.* Medical Teacher. In press).

Methods

BEME methodology as described in BEME protocol (http://www.bemecollaboration.org) was applied by two independent coders who collected and examined literature from 1975 until the end of 2008. Given the available resources, it was decided to limit the analysis to studies in English, reporting on OSCEs that use the traditional definition (Harden et al. 1975) and utilization of OSCEs performed in undergraduate medical education. A study was coded as 'non-classical' when it did not conform in general terms to the classical approach of the OSCE, as described by Harden et al. in 1975. Studies coded as 'non classical' include those in which the candidate was a 'team' or a 'group' instead of an individual, where assessment was based on video instead of direct observation, where there were only written stations, tasks were only peer-rated, and the exam contained only one or two stations. 'non-classical' studies were not included in our analysis.

From the 1065 studies identified in the initial literature review, 366 reported on OSCEs used in undergraduate medical education. From them we considered 212 papers (58%) to be relevant to the analysis included in this report with evidence surrounding reliability (60 studies, 16%), validity (200, 55%), fairness (34, 9%), acceptability (78, 21%) based on OSCE relevance (31, 8%) and/or satisfaction (47, 13%) and educational impact (55, 15%) based on information on OSCE steering effect on learning (21, 6%) and/or teaching (34, 9%).

Since most reported OSCEs consisted of multiple stations with one rater each (examiner or SP), the most commonly reported reliability index was inter-station reliability. Generally this was reported as the reliability of the entire test (the reliability of the OSCE given 'n' stations included in the examination). Since the number of stations varied, we used a variant on the Spearman-Brown formula to return the total test reliability to an average inter-station reliability. Statistical tests were performed both as an 'unpaired tests' using all available data and as 'paired tests' where individual studies did, for example, report on comparable data sources (such as examinations that used both global and checklist evaluation grids). Evidence for validity in OSCEs most typically included concurrent validity evidence, where OSCE grades were correlated with a variety of other assessment methods, but also with multiple choice examinations. Studies were also scrutinised on face/content validity, with specific focus on the process of station design as a measure of whole test.

Evidence on OSCE requirements namely on fairness, acceptability and educational impact was based on published feedback provided by teachers, students, patients and examiners, as reported in individual publications. Opinions were collected through informal or formal feedback (when based on a structured questionnaire complemented or not by a subsequent interview). Feedback was

analysed and categorized as positive or negative accordingly to their content using *'a posteriori'* content analysis technique (Bardin 1998) i.e. the analysis was made without the definition of a prior conceptual framework.

A summary of findings is presented for each requirement with some quotations given as examples. When the results were based on surveys, the relevant percentage of respondents is indicated.

Consistency between the coders was checked throughout the entire process, with agreement of 97% for reliability and of 98% for validity.

Results

Results are presented for OSCE reliability, validity, fairness (as defined by feedback of students, teachers and examiners), acceptability (students, teachers, examiners and patients' feedback on satisfaction and relevance regarding the OSCE) and educational impact (OSCE steering effect on students' learning and teachers' teaching). The remaining assessment requirements stated by the GMC (2011) - namely the OSCE feasibility, cost effectiveness and opportunities for formative feedback - were object of another publication also related to the same BEMER.

Reliability

From the 366 accepted studies, 60 (16%), reporting on 78 OSCE exams, were included in this analysis. Data on reliability was presented in terms of inter-station reliability, inter-rater reliability and expected overall reliability (with a varying number of stations). The references concerning the papers contributing to the reliability analysis are presented in Appendix 1.

- *Inter-station reliability*

As the Cronbach's alpha coefficient (overall reliability) is directly related to the number of stations included in any assessment, longer OSCEs will have higher reliability coefficients. In order to control for this, we computed the average inter-station reliability. By using this metric as our data point of interest, we then render reliability coefficients comparable between studies (i.e. OSCE exams) regardless of the number of the stations. Through the Spearman-Brown formula it is possible to convert the test reliability *R* to the average inter-station reliability *r* if the number of stations is known:

$$r = \frac{R}{n-(n-1)R}$$

As an example, for a 12-station OSCE (average number of stations observed) with overall reliability of 0.74 (weighted average of the observed overall reliability) the average inter-station reliability would be 0.19.

Mean inter-station reliability values and sample sizes (meaning the number of students included in the analysis) for OSCE examinations can be found in Table 1. In order to compare different examiner groups (expert examiners and standardized patients) and different types of evaluation approaches (global ratings and station-specific checklists), the 4 variants of OSCE formats (examiner/checklist, examiner/global rating, standardized patient/checklist, standardized patient/global rating) are presented in Table 1. When two methods were calculated in the same study, paired comparisons were made. The most common form of evaluation used in OSCE examinations was a station-specific checklist, completed by an examiner.

Statistical tests were performed both as 'unpaired tests' and as 'paired tests'. Several comparisons are of interest in the application and use of OSCEs, and were tested for significance:

- *Reliability of global ratings compared to station-specific checklists*

Global ratings showed significantly higher average inter-station reliability than station-specific checklists (.207 *vs.* .168 for examiners, .233 *vs.* .168 for Standardized patients (SPs); t (82) = 8.84, $p<.0001$; t (12) = 3.99, $p<.005$ paired). The higher average inter-station reliability for global ratings was consistent for both examiners and SPs.

- *Reliability of expert examiners compared to standardized patients evaluators*

The common concern regarding the use of SPs is that they may not be as reliable as examiners, particularly when global ratings are

used. In terms of average inter-station reliability, there was no significant difference between SPs and expert examiners, when checklists were used (0.168 *vs.* 0.168). When using global ratings, the reliability of SPs was slightly higher than the expert examiner's (0.233 *vs.* 0.207; t (17) = 3.16, p <.01).

Table 1. Inter-station reliability*

OSCE design		Mean	S.D	Number of OSCE exams
Examiner Checklist	**All**	**.168**	**.108**	**60**
	Paired	.134	.078	11
Examiner Global rating	All	.207	.121	14
	Paired	.184	.096	11
Standardized Patient Checklist	All	.168	.146	5
	Paired	.290	.183	2
Standardized Patient Global rating	All	.233	.178	5
	Paired	.220	.042	2

**Statistical tests were performed both as unpaired tests' using all available data, and as 'paired' tests, where individual studies did, for example, report both global and checklist.*

Additionally, we compared the average inter-station reliability between OSCEs conducted with different underlying purposes, namely comparing OSCEs performed specifically to evaluate students, curriculum or intervention with OSCEs implemented to examine the psychometric properties of the OSCE itself. Furthermore, we compared the average inter-station reliability for summative *vs.* formative OSCEs and high stakes *vs.* non-high stakes OSCEs. These analyses were made because higher reliability was expected in OSCE exams with higher consequences in terms of students' assessment.

Due to missing or unclear data, these comparisons were only performed using studies that reported on examiners using checklists, in order to ensure data remained comparable across the above mentioned contexts (60, 23%). Mean values, sample sizes, *F*-statistics and *p*-values for each variable of interest are reported in Table 2.

There were no significant differences for average inter-station reliability (for examiners using checklists) for any of the analysed contrasts.

- *Inter-rater reliability*

Relatively few studies reported inter-rater reliability, since this requires multiple raters per station (by definition). For those studies that did report inter-rater reliability, results are summarized in Table 3. Data is presented separately for examiners and SPs, and for checklists and global ratings.

The inter-rater reliability of examiners using global ratings was significantly lower than using checklists (t = 2.59; p=0, 01) which contrasts with findings related to average inter-station reliability.

- *Predicted overall reliability with varying station number*

Predicted reliability for a various number of stations was estimated and is presented in Figure 1.

Table 2. Effect of context variables on inter-station examiner/checklist reliability

Context Setting		Mean	SD	Number of OSCE Exams	ANOVA test F	ANOVA test p-value
OSCE aim	To evaluate	0,15	0,0090	7	0,683	0,509
	To examine the OSCE	0,18	0,0129	47		
	Both	0,13	0,0052	5		
OSCE Purpose	Summative	0,16	0,0132	18	2,647	0,090
	Formative	0,12	0,0047	9		
	Both	0,30	0,0113	2		
Type of exam	High stakes	0,18	0,01	21	0,467	0,499
	Not high stakes	0,15	0,01	14		

Table 3. Inter-rater Reliability

OSCE design	Mean	S.D	Number of OSCE exams
Examiner Checklist	.716	.126	18
Examiner Global rating	.577	.104	7
SP Checklist	.655	.007	2
SP Global rating	n/a	n/a	-
Examiner – SP Checklist	.595	.148	2
Examiner – SP Global rating	.505	.106	2
Examiner checklist – SP Global rating	.510	.0233	3
Examiner Global rating –SP checklist	n/a	n/a	-

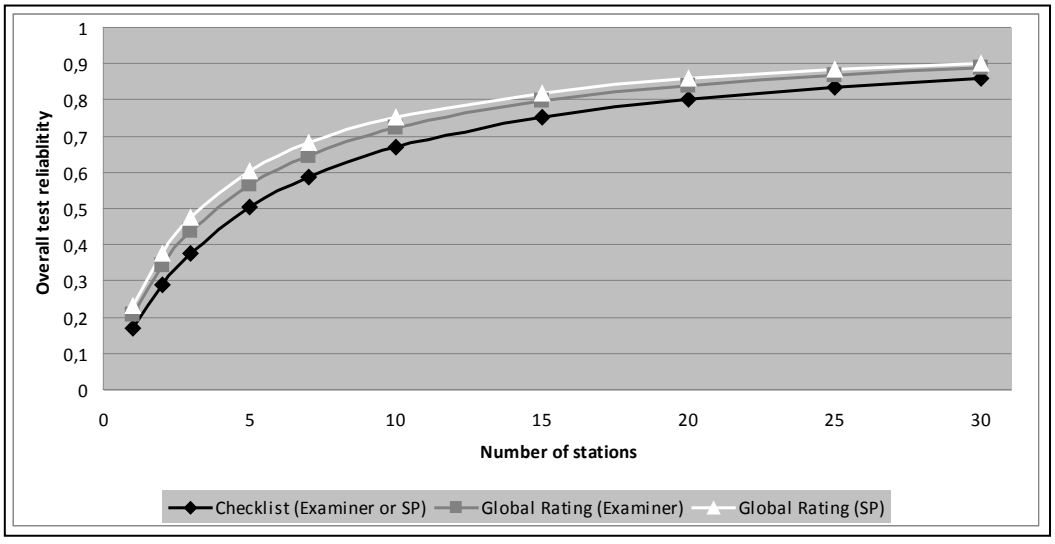

Figure 1. Expected overall reliability *vs.* number of stations

Due to lower average inter-station reliability, OSCEs using checklist-based assessment grids would require a higher number of stations than identical OSCE using global rating-based assessment grids to result in the same expected test reliability. As an example, for an inter-station reliability of .80, fifteen stations would be required when using global ratings (for examiner or SPs) and twenty when using checklists (Figure 1).

Validity

From the 366 studies included in our analysis, 200 (55%) present data to support the analysis of validity of the OSCE. Some papers were excluded (7, 2%), when authors made statements on the validity of the study without presenting supporting data, or where validity of the OSCE was assumed based on other studies (4, 1%).

- *Validity against criterion measure (concurrent validity)*

Evidence on validity was collected from studies where performance on the OSCE was compared to a variety of criterion measures. Those with multiple comparisons of criterion measures (30, 8%), reporting on a total of 38 OSCE exams, are summarized in Table 4. The references concerning papers included in this analysis are reported in Appendix 2.

Although correlations with multiple choice questions (MCQ) were in the low to medium range, the five studies that reported disattenuated correlations showed higher values, near 0.70. The highest correlation (0.82) was found in the comparison with short answer-style exams. The low correlation between OSCE and clinical grades may be explained by the poor reliability of the latter (Streiner et al. 1986).

Table 4. Validity against criterion measure

Criterion	Checklist			Global Ratings		
	n	Mean correlation with criterion	SD	n	Mean correlation with criterion	SD
Multiple choice questions	24	0.415	.20	4	0.245	.107
(disattenuated)	4	0.685	.14	1	0.770	
Short answer	2	0.820	.085	-	-	
Clinical grade	13	0.367	.13	-	-	
NBMECSE	2	0.09	.14	1	0.100	
Written Licensing Exam	-	-	-	1	0.39	
Patient satisfaction	-	-	-	1	0.24	

- *Face/content validity*

A test has face validity when it appears to measure what it is supposed to measure i.e. when 'it appears as a valid test in the absence of empirical testing' (Cook et al. 2006). Face validity is a type of content validity, determining the '*suitability of a given instrument as a source of data on the subject under investigation using common-sense criteria*' (Saunders 2003). When an examination is carefully designed through good selection and weighting of the topics to be assessed it is described as having content validity (van der Vleuten 2000).

References to face and content validity were found in 82 reports of OSCE exams (22%). Compared to other measures, face and content validity are usually not considered a very robust criterion for validity, but, as stated by Schuwirth and van der Vleuten in 2011 the '*validity of an assessment instrument is generally not determined by its format but by its content',* we decided to scrutinize the process of designing the OSCE stations.

Evidence was searched in terms of 1) who was involved in designing the station 2) if stations were blueprinted for selection of content and 3) how was agreement reached for decision on final content. A summary of findings is found in Table 5.

Table 5. Station design

Who was involved in the process of station design?	n	%
OSCE designers + other teachers (from same or other departments)	47	18
OSCE designers + other teachers + other experts	29	11
OSCE designers + other experts	11	4
Only OSCE designers (staff appointed to design the OSCE)	7	3
OSCE designers + other teachers + other experts + students	3	1
Total	**97**	**37**
How was final content determined??	**n**	**%**
Incorporation of feedback from other teachers without establishment of consensus	46	17
Final decision reached after consensus meetings	40	15
No incorporation of feedback from others than OSCE designers (just information on the content)	5	2
Total	**91**	**35**
How was station content generated?	n	%
Proceeding from a clear blueprint for the content for selection.	54	21
No blueprint	16	6
Total	**70**	**27**

Very few studies (7, 2%) reported on stations designed without the contribution of other teachers and experts. Decisions on content were based most commonly on a process of incorporating contributions and feedback from other teachers and/or other experts, and 47 studies (13%) report a formal process (including formal meetings) to reach consensus on station content. Interestingly, 3 studies (1%) reported that students' contributed to stations design, which supports the educational value of the OSCE and may contribute to OSCE fairness.

In 24 studies (7%), stations were blueprinted, contributions from other teachers and/or other experts were incorporated and final decisions regarding station content were taken by consensus. These results show that medical schools appear to make substantial investments to support the process of station design. This may be related to the complexity of designing and implementing an OSCE and, simultaneously, the recognition of the importance of the process for the validity of the OSCE.

The evidence on other assessment requisites, as proposed by the GMC (2011), implied to report on OSCE fairness, acceptability (relevance or satisfaction) and educational impact (OSCE effect on steering learning or steering teaching). The major difficulty found when doing this analysis was the existence of a considerable overlap between those categories, namely in terms of relevance and satisfaction with the OSCE. To overcome this problem each category was defined with examples of the comments to be considered. When results were based on a survey or focus groups (formal feedback) the relevant percentage of respondents is indicated. For informal feedback we transcribe the feedback reported by authors regarding the different participants.

OSCE Fairness

Feedback was considered when reports clearly mentioned OSCE fairness or something similar (for example, OSCE evaluated what it should be evaluating, reflected course objectives, corresponded to curriculum, to what was taught,

to what was expected or was based on realistic and appropriated stations.

Thirty four studies (9%) contributed to the analysis of OSCE fairness. Evidence included feedback received from students (18 studies, 5%), students and examiners (9, 2%), students and faculty (5, 1%) and from faculty (2, 1%). A summary of findings is reported in Table 6 with quotations presented as examples. Some studies contributed to more than one category.

Table 6. OSCE fairness*

Feedback from Students (n= 18 studies)	Study
'Is a fairest exam when compared with other methods (80%)'	Pierre et al. 2004
'The OSCE shows a high degree of fairness'; 'globally rated as highly fair'	Hart et al. 1987
[*in the study from Rahman 2001, students rated the OSCE as 4.7 in a 1-5 scale*]	Malik et al. 1988
	Rahman 2001
'Is a fair exam'	Allen et al. 1998
	Durak et al. 2007
	Frost et al. 1986
[*Percentages were indicated in the following studies: Allen et al.1998 (61%); Hodges et al. 1998 (93%); Grand'Maison et al. 1996 (89%); Pierre et al. 2004 (68%); Thomson, 1987 (43%)*]	Grand'Maison et al. 1996
	Hodges & Lofchy 1997a
	Hodges et al. 1998
	Jewell 1988
	Martin et al. 2000
	Newble et al. 1978
	Petrusa et al. 1991
	Pierre et al. 2004
	Thomson 1987
	Wilkinson et al. 2000
'Is as fairer as other traditional methods'	Raga & Coovadia 1985
'It adequately reflect course objectives '	Murray et al. 1997
'It corresponds to what was expected ... we feel confident and able to perform the required tasks'	Allen et al. 1998
Feedback from Students and Examiners (n= 9 studies)	**Study**
'Students (80%) felt the OSCE to be a fairer system and external examiners commented favourably ... OSCE is perceived as more fair and less stressful than other formats'	Smith et al. 1984
'The majority of students saw the OSCE as a fairer method of assessing clinical competence. Examiners preferred the OSCE than the long case (reduced examiner bias , similar test to all students)	McFaul & Howie 1993
	Searle 2000
'Fair exam appropriated to assess students' skills'	Ainsworth et al. 1995
'Students considered the OSCE as a fair exam and 13% firer than Long case, short answer and MCQ; examiners are supportive and initial scepticism was removed'	Collins & Gamble 1996
'Students (94%) and Faculty observers (97%) agreed the OSCE measured the skills third year students should possess'	Prislin et al. 1998
'Survey show a remarkable degree of confidence on OSCE (based on surveys over 3 years and five years later) by students and examiners [one of the reasons is OSCE fairness]Examiners considered the OSCE highly fair (30%) or fair (54%) which contrast with ward ratings respectively 9/% and 47%'	Newble 1988
'Scenarios very realistic coded by 70% students and 94% of examiners and 70% students and 88% of examiners felt that each station difficulty was appropriated'	Hodges 1997b

'Students consistently appraised the OSCE as a fair assessment. They commented that it reflected their course and curriculum and was well organized, clinically relevant and more equitable than other forms of assessment.'	Walters 2005
Feedback from Students and Faculty (n= 5 studies)	**Study**
'Highly relevant level of acceptance by students and staff. The reason seems to be the highlevel of perceived fairness' Provide addicional information on studennts deficiencies not available in PE No differences in relevance to course material, , level of examination difficulty and enjoyment Students felt the OSCE as a high degree of objectivity and simulation when compared with PE'	Lunenfeld 1991
'It is perceived as a fair method of assessing. The use of checklist ensures its fairness'	Adeyemi 2001
'90% of students and 91% of faculty agreed that the OSCE represented an appropriate and fair evaluation method'	Simon 2002
'The Majority of students (generally) considered the OSCE as 'more fair and more objective when compared to other assessment methods: (Mean scores in a scale from 1 to 6: OSCE= 5.02 ; MCQ= 4,01; Oral=2,18; Essays=3,07; Clinicals 3,29; In Course= 4,48). Faculty considered the OSCE adequate to the course'	Lazarus & Kent 1983
'The OSCE Gives an objective measure of the skills we try to teach' and students considered the OSCE as a 'fair exam'	Johnson & Reynard 1994
Feedback from Faculty (n= 2 studies)	**Study**
'It is accurate and effective for assessing what is taught by different tutors and how this is learned'	Troncon 2004
'OSCE was fair and appropriate in assessing clinical ability was confirmed'	Rudland et al.2008

**Percentages are reported when feedback is based in a survey.*

Despite the global tendency, documented in the above table, of reporting the OSCE as a fair or highly fair exam, in a study from Brazil (Troncon et al. 2004) a contradictory view was obtained with faculty reporting the OSCE as an *'accurate and effective [exam] for assessing what is taught'* but, simultaneously, as an '*inconsistent exam because the examinees were excessively stressed'*.

OSCE Acceptability

According to the GMC's criteria (2011), evidence on OSCE acceptability should be based on studies reporting on perceptions of OSCE as a relevant exam and satisfaction with the OSCE. Sixty two studies (17%) contributed to this analysis, with 16 studies (4%) reporting simultaneously on OSCE relevance and satisfaction.

- *OSCE Relevance*

Thirty one studies (8%) were included in the analysis when reporting on OSCE as a relevant, effective, important, valid, objective exam, worth-while to be used despite its disadvantages or associated effort, to be adopted in other disciplines or areas, better than other assessment formats.

Feedback was given by students in 15 studies (4%), students and faculty (5, 1%), students and examiners (4, 1%), examiners (3, 1%), faculty (3, 1%) and students, examiners and SPs' (1, 0,2%). A summary of findings is presented on Table 7 where quotations from different participants were reported as examples.

OSCE acceptance from students, teachers, examiners and SPs appeared as very positive globally, with high relevance levels. However, low percentages of less positive comments were identified, namely in the study from Roy et al. (2004) reporting that 3% of the students considered the OSCE as *'not able to test adequately either the knowledge or practical skills'*. Newble in 1988 reported that 13% of the examiners saw the *'OSCE as less appropriated than traditional formats*' and Lazarus & Kent, in 1983, reported '*a moderate to low support of the OSCEs' capacity to evaluate essential skills'* (with OSCE rated 0.82 in a scale from -2 to +2) and finally in a paper from Troncon (2004) one faculty member criticized the 'limitation of the OSCE for assessing the integrated approach to patients, emphasizing that this aspect might represent a dissociation between the examination and the objectives of the main course'.

Table 7. OSCE Relevance*

Feedback from Students (n=15 studies)	Study
' The most relevant and stimulating aspect of their training'	Knowles et al. 2001
'An important component of overall measurement of clinical competence'	A-Latif 1992
'More accurate assessment'	Newble et al. 1978
'An important part of assessment' (80%)	Watson et al. 1982
'Most relevant; high relevance' (Wilkinson /95%) or high-moderate (Newble 95%)	Elnicki et al. 1993
	Jewell 1988
	Newble et al. 1981
	Wilkinson et al. 2000
'An outstanding method, more relevant than ABSITE [American Board of Surgery In-Training Examination] and Mock Oral Examinations' ... also rated as highly overall	Sloan et al. 1996
'A worthwhile method' (86 %)	Hoole et al. 1987
'A comprehensive exam'	Pierre et al. 2004
'When compared with other formats the OSCE was considered to be 'most objective/ most effective with advantages over long cases' and MCQ'	Cuschieri et al. 1979
'Recognized the OSCE's for its objectivity and effectiveness'; 'Characteristics of OSCE were considered to be objectivity'	Adeyemi 2001
	A-Latif 1992
'The students felt OSPE to be a better method of assessment 64%). The main reason given for this was great objectivity and more uniform evaluation than the conventional'.	Roy et al. 2004
'Students report advantages of OSCEs compared to 'Long case [examinations]: [as OSCEs evaluate a] wide range of Knowledge and Skills, comparable test for all students, reduced examiner bias, opportunity for feedback and they see the OSCE as an excellent alternative'	Lazarus & Kent. 1983
'A valid test'	Martin et al. 2000
'It measured important outcomes not measured by other tests (4.1 in a 1-5 scale)	Joorabchi 1991
Feedback from Students and Faculty (n=5 studies)	**Study**
'Highly relevant level of acceptance by students and staff.	Lunenfeld 1991
'A Relevant exam ...for teachers and students Consensus was reached among faculty that the use of OSCE is worth wile ... acknowledge its usefulness raised the question of incorporating the OSCE in the National Medical Licensing	Ban et al. 1997
'Majority of students prefer the OSCE' when compared to other assessment methods: (Mean score in a scale from 1 to 6 were: OSCE= 4.65 ; MCQ= 2,43; Oral=2,85; Essays=2,75; Clinicals 4,17 and In Course= 4,19). Teachers reported the OSCE as 'an excellent alternative'	Lazarus & Kent. 1983
'Students (435) considered that other clerkships should adopt the OSCE exam with the majority of students and teachers see the OSCE is a valuable exam	Petrusa et al. 1991
'Students stated 'They all had the feeling that the evaluation they received with OSCE was more objective than the one usually received with the summative evaluation form filled-in by teachers at the end of the rotation' and 'Teachers also agree with last point'	Grand'Maison et al. 1985
'For teachers 'Consensus was reached that it is worth wile to use the OSCE. For students 'the OSCE should be part of skills training usually poorly evaluated'	Ban et al. 1997
Feedback from Students and Examiners (n=4 studies)	**Study**
'Surveys carried over 3 years and again five years later to test stability with students and examiners showing a remarkable level of acceptance and support (namely due to OSCE relevance).. All examiners had experience with traditional formats: the majority saw the OSCE as more appropriated measured than traditional formats (55%). OSCE was seen as 'very appropriate' (42%) or 'appropriate' (53%) of measuring competences. In terms of students' feedback – when excluding the year of implementation, 50% considered the OSCE as high relevant to intern practice which contrasts with MCQs (high relevant for less than 5% as high relevant and only 30-40% as moderate relevant'	Newble 1988.

'For students (except for the first implementation where OSCE was felt as more artificial and less likely to provide a valid measure) the OSCE was seen as valid as written or oral tests). Examiners rate the OSCE as better than other examination formats'	Kirby & Curry 1982
'The majority of students saw prefer the OSCE when compared to the Long case exam Examiners considered OSCE as a major improvement regarding traditional methods (advantage= range of skills and knowledge tested, reliable, valid, practical and flexible)'	McFaul & Howie 1993
'The effort [implied by OSCE] is worthwhile..... In 1987, 1988 and 1989 surveys: - 90%.98% and 92% of the evaluators considered the OSCE as an effective way to evaluate second year students' and worthwhile to be implemented (97%, 98%, 97%); - 92%,95% and 99% of the students positively rated the OSCE as a method of evaluation 100%, 100% and 100% of the students reported it should be1 administered again'	Kowlowitz et al. 1991
Feedback from examiners (n=3 studies)	**Study**
'Examiners and raters (who are students of other curriculum years) or faculty raters found OSCE as a relevant /effective examination and an effective way of assessment'	Feickert et al. 1992
'There was some variation in the opinion of examiners on the relevance of individual items [of the score sheet system]'	Johnson & Reynard 1994
'Favoured OSCE when compared to other assessments'	Hart et al. 1987
Feedback from Faculty (n= 3 study)	**Study**
'The technique is highly relevant and more accurate than previous examination methods ….is an effective [exam]''	Troncon 2004
'A worthwhile' exam'	Adeyemi 2001
'The use of OSCE is confirmed in spite of disadvantages'	Smith et al. 1984
Feedback from Students, Examiners and SPs (n= 1 study)	**Study**
'More relevant than other formats'	Walters et al. 2005

**Percentages are reported when feedback is based in a survey. Some studies contributed to more than one category.*

- *Satisfaction with OSCE*

Evidence on satisfaction was based on 47 studies (13%) where participants welcomed the OSCE as an enjoyable experience (felt enthusiastic about the OSCE, considered the OSCE as a valuable, favourable, useful, fun experience, which should be continued or repeated in a more regular basis) or express satisfaction with what resulted from the OSCE exam (for instance OSCE feedback).

Findings were globally very positive, with evidence based on support drawn from students (28 studies, 8%), examiners (7, 2%), faculty (6, 2%), SPs (2, 0.2%), faculty, examiners and simulated patients (1, 0.2%), and real patients (1, 0.2%). A summary of comments is presented in Table 8 where quotations from participants were reported as examples.

Five studies reported mixed acceptability evidence, with students from the same study expressing contradictory attitudes, namely rating the OSCE as 'poor' and simultaneously as an 'excellent exam' (Newble et al. 1978; Elnicki et al. 1993), '*as an exam creating strong anxiety, intimidation and stress*' and simultaneously as a *'well-structured and well administered exam*' (Pierre et al. 2004), as an exam '*better than the conventional exams* (70% of the students) when simultaneously 3% criticized this style of exam' (Roy et al. 2004), as a 'useful exam' and simultaneously expressing '*dissatisfaction with organization, station time available and degree of emotional stress elicited by examination'* (Troncon 2004).

In a very few number of studies, the OSCE was perceived less positively: 'the majority of the students do not like to do the OSCE again' (Allen et al. 1998), 'the OSCE is too hard' (Thistlethwaite 2002), and 'there is pre and intra examination emotional tension' (Kirby and Curry 1982). Finally, for 5% of the 69% students surveyed in a study from Watson et al. 1982, the OSCE was found as a 'waste of time'. It is of interest to observe that if in the study from Malik

et al. (1988) the OSCE is seen as 'mentally taxing' this was not confirmed in the study by Rahman et al. 2001 (where identical dimension the OSCE was rated 2.5 in a scale from 1 to 5).

Those less positive remarks highlight changes in attitude from early days to current opinions.

Table 8. Satisfaction regarding the OSCE*

Feedback from Students' (n=28 studies)	Study
'A high degree of satisfaction [with OSCE] '	Murray et al. 1997
'The majority of the students welcome the OSCE'	Cuschieri et al. 1979
	Hodges et al. 1997b
Students 'receive the [OSCE] exam enthusiastically'/'well received'	Adeyemi 2001
	Feickert et al. 1992
	Hodges et al. 1997b
	Hoole et al. 1987
	Kent & Lazarus 1983
	Smith et al. 1984
	Tervo et al. 1997
	Volkan et al. 2004
'A valuable, favourable experience /very positive attitudes towards OSCE' *In the study from Malik et al.1988 high rank students show higher intensity in positive attitudes towards OSPE*	Adeyemi 2001
	Lazarus & Kent 1983
	Malik et al. 1988
	Nalesnick et al. 2005
	Raga & Coovadia 1985
	Rahman 2001
	Verma & Singh 1993a
'Globally rated as highly effective, exciting, varied, active good, skills oriented as a interesting and challenging examination'	Malik et al. 1988
'An enjoyable experience' (4.1 in a 1-5 scale/ Joorabchi. 1991); most enjoyable'	Hodges et al. 1997b
	Joorabchi 1991
	Sloan et al. 1996
'An acceptable exam'	Walters 2005
'A fun experience'	Hodges & Lofchy 1997a
	Jewell 1988
'An useful[exam]'	Malik et al. 1988
	Troncon 2004
'Should 'be continued'	Collins & Gamble 1996
'Students expressed 'they wish to repeat a drill every Friday'	Durak et al. 2007
'To be run again next year ....is a valuable method (90%) a worthwhile method (86%)'	Hoole et al. 1987
'The OSCE exam should be extended to other courses'	Feickert et al. 1992
OSCE 'as an improvement' [regarding other methods of assessment'	Cuschieri et al. 1979
'The easiest format when compared with MCQ,'	Johnson & Reynard 1994
	Pierre et al. 2004
Students expressed satisfaction with 'faculty observation and feedback' providing 'valuable feedback' ; 'feedback is more valid than with other exams'	Hoole et al. 1987
	Grand'Maison et al. 1985
	Morag et al. 2001

Students were satisfied (95%) or very satisfied (25-50%) with OSCE. Satisfaction is higher than with MCQ (satisfied 70% and very satisfied 5-10%) or ward ratings (with barely 50% even being satisfied). Over 4 years students start by considering MCQ more influent to their study habits than the OSCE but over four years this influence is decreasing simultaneously with increase of OSCE Influence'	Newble 1988.
Feedback from Examiners (n=7 studies)	**Study**
'[The OSCE is] better than traditional exams due to the variety of skills and similar test to all students'	McFaul & Howie 1993
'[The OSCE was reported] 'as enthusiastic'; 'enthusiastic about OSCE concept and practice'	Smith et al. 1984
	Kent & Lazarus 1983
'The OSCE advantages were recognized'	Feickert et al. 1992
	Smith et al. 1984
'Acceptable exam ……' positive about the content of stations at the interactive stations'	Walters 2005
The OSCE should 'be continued'	Mossey & Newton 2001
'The level of stress induced by participation in an OSCE as adequate'	Collins & Gamble 1996
Feedback from Faculty (n=6 studies)	**Study**
'Acceptability was high amongst studies that reported on teachers' perceptions of the OSCE'	Hodges et al. 1997b
'With OSCE being considered 'an enjoyable experience'; 'Most enjoyable'	Hodges et al. 1997b
	Tervo et al. 1997
'Received with enthusiasm and recognition of its advantages'	Kent & Lazarus 1983
	Newble et al. 1978
	Tervo et al. 1997
'A great exam'	Ainsworth et al. 1995
'A highly appropriate and acceptable method of education and evaluation'	Hamann et al. 2002
Feedback from Sp's (n=2 studies)	**Study**
'The feedback received from SPs was generally positive'	Walters et al. 2005
	Sloan et al. 2001
'Acceptable exam ……' at the interactive stations SPs were positive about the content of stations'	Walters 2005
Feedback from Students and Faculty (n=2 studies)	**Study**
'Teachers considered the OSCE as relevant and 90% of the students rated the 90% as a valuable tool (very valuable 28% and valuable 62%)	Ban et al. 1997
'Teachers and students expressed interest in 'participating in a [OSCE] blueprint exercise' [i.e. to design stations according to course objectives]	Newble et al. 1978
Feedback from Students Faculty, Examiners, Simulated Patients (n=1 study)	**Author(s) /Date**
'They enjoyed the experience' ... Teachers and students agree that the content validity of the OSCE was high in spite of its limitations'	Grand'Maison et al. 1985
Feedback from Real Patients (n=1 study)	**Author(s) /Date**
'The feedback received from Real patients was generally positive'	Sloan et al. 2001.

**Percentages are reported when feedback is based in a survey. Some studies contributed to more than one category.*

OSCE Educational impact

Educational impact, recognized as one of the most valuable characteristics of any assessment (Roediger et al. 2011; McDaniel et al. 2011), was analysed in terms of OSCE impact on students' learning priorities namely on their habits of study. Forty six studies (13%) were examined in terms of the reported ability of the OSCE to support/steer learning and to support/steer teaching.

- *Steering effect on learning*

Twenty one studies (6%) supported the conclusion that OSCE exams steer learning, with the authors

using a variety of terms to express this effect (how students learn better the subjects in which they are examined, the strong messages students got from OSCE on important learning, the environment created by OSCE to facilitate learning, the impact of OSCE in students' habits of studies, i.e. on how OSCE provides a focus for relevant learning). A summary of results is reported in Table 9 where quotations illustrate the feedback given by the different participants namely by faculty (15 studies, 4%), students and faculty (3, 1%), students (2, 1%), and examiners (1, 0.2%).

Table 9. OSCE steering effect on learning*

Feedback from Faculty (n=15 studies)	Study
'The major effect of the OSCE on learning behaviours of students was mainly to encourage them to practise clinical skills on each other, to rehearse routines, and to work in groups'	Rudl et al. 2008
'Probably OSCE students would have spent more time improving clinical skills… (a desirable effect)'	Wilkinson & Frampton 2004
' OSPE helps better learning ...OSPE made students more attentive in practice classes and encourages them to practice skills and steps of a procedure thoroughly which helped them in better learning ... A distinct advantage was in the change of students learning behaviour'	Malik et al. 1988
'[the OSCE] resulted in a focusing of the students upon the importance of clinical skills'	Ribin & Philip 1998
[the OSCE] motivates [students]to participate more in problems of patient management' ... Students learn better the subjects in which they are examined'	Afroza 1985
	Peden et al. 1985
'[the OSCE)Encourages emphasis on learning practical skills (rather than the acquisition from books)'	Afroza 2000
	Johnson & Reynard 1994
'The great advantage is the strong messages conveyed to learners of what programs values as important and desirable outcomes'	Joorabchi 1991
'Students were encouraged to draw on existing knowledge and direct further learning towards solve the clinical problem presented'	Knowles et al. 2001
'[the OSCE sent a] message on what knowledge they [students] should acquire'	Hoole et al. 1987
'Implementing the OSCE has created an environment in which the learning of basic skills becomes important.....'	Kowlowitz et al. 1991
'We believe this study has contributed to students self assessment and foster self directed learning...'	Pierre et al. 2005
'[the OSCE] contributes to students self assessment and fostered self directed learning'	Weinreb et al. 1998
'[OSCE] has potential impact to change..... students habits'	Ban et al. 1997
'The OSCE brought students to realize how important are basic clinical skills'	Grand'Maison et al. 1985
Feedback from Students and Faculty (n= 3 studies)	**Study**
'The [OSCE] examination may be useful for transmitting to students what skills are important to learn. Faculty members also reported 'the OSCE possibly had a positive effect on students' drive to actually study and practice…. This effect was felt to be greater than what could be associated with the previous examination method'.	Troncon 2004
Teachers stated that students learn better the subjects in which they are examined ….. [the OSCE] encourages practice skills' while students found the OSCE as 'a useful learning experience' (69%)	Watson et al. 1082
'[The OSCE] has influenced study patterns of three-quarters of the student s[75%] who replied to the survey' Teachers stated that '[the OSCE] helped direct student learning'	Collins & Gamble 1996
Feedback from Students (n=2 studies)	**Study**
'[The OSCE] helps motivation, allows most useful learning when compared with other exams (60%)'	Pierre et al. 2004
'it can act as a guidance examination (4.5 in a 1-5 Likert scale)'	Rahman 2001
Feedback from Students and Examiners (n= 1study)	**Study**
'Results of the survey show remarkable degree of confidence [on OSCE] over 8 year period by students and examiners and direct learning [is among the reasons for such evaluation]'	Newble 1988

**Percentages are reported when feedback is based in a survey. Some studies contributed to more than one category.*

It is worth to mention that although Rudland et al. (2008) reported on OSCE steering learning, as documented above, the authors highlighted 'that performance on the OSCE can not necessarily be regarded as a direct marker of clinical experience' and that 'the assumption that an OSCE will stimulate students to spend more time seeing patients was not supported [by their study]'. Simultaneously they reported an unexpected effect of the OSCE namely a '*beneficial side-effect to drive more collaborative learning'*. Of interest is also to notice that their findings were based on a survey followed by focus groups.

In addition to the steering effect of the OSCE on students' habits, 11 studies (3%) also pointed to the value of the OSCE as an educational activity. Results are presented in Table 10.

Table 10. OSCE educational value*

Feedback from Students (n=7 studies)	Study
'Was a good and helpful learning experience'	Allen et al. 1998
'Valuable learning experience for students; provide a teaching learning experience	Harris et al. 1997
'95%, 98% and 94% of the students positively rated the OSCE educational value'	Kowlowitz et al. 1991
'Justified in terms of opportunity for learning'	Long 1997
'Help motivation / most useful learning (62%) when compared with other exams'	Pierre et al. 2004
'Opportunity for learning' (Students: 78% to 94.5%)	Wilkinson et al. 2000
'Provide useful feedback'	Ytterberg et al. 1998
Feedback from Faculty (n=2 studies)	**Study**
'Participants enjoyed the OSCE and rated the OSCE highly overall with 74% of them believing that it was above average or outstanding as an educational method Although the OSCE is fundamentally an examination it does provide excellent opportunities for clinical teaching'	Sloan et al. 1996
'Students meet with their small-group faculty leader to bring failed skill up to a passing performance'	Duerson et al. 2000
Feedback from Students and Faculty (n=1 study)	**Study**
'Teachers stated the ' OSCE can promote relevant learning in paediatrics' while students rated the OSCE in a very positive way, helpful for learning skills (mean= 4.5) and helpful for learning attitudes (mean 4.6)' *(1-5 Likert scale with 5 as maximum positive)*	Rahman 2001
Feedback from Examiners and Students (n=1 study)	**Study**
'Examiners and students found the OSCE very formative'	Grand'Maison et al. 1985

**Percentages are reported when feedback is based in a survey. Some studies contributed to more than one category.*

- *Steering effect on teaching*

OSCE was described as influencing and supporting teaching in 34 studies (9%). Some reported the OSCE influenced teaching in a very general manner, others report the structure and content of OSCE can influence teaching in specific ways (examples were given) and some even refer to OSCE as a teaching technique due to its impact on learning. The most common reported influence of the OSCE on teaching was as a diagnostic tool for strengths and weaknesses, with several studies reporting the use of students' performance on the OSCE to redesign teaching and curriculum.

A summary of results is reported in Table 11 where quotations from different participants were presented to document the OSCE steering effect on teaching.

Table 11. OSCE steering effect on teaching*

Feedback from Faculty (n=34 studies)	Study
'Suitable to improve undergraduate education'	Johnson & Reynard 1994
'OSCE have been useful in helping to identify areas of weaknesses that could benefit from remediation …. Have has also helped identify those parts of the curriculum students have difficulty mastering Station scores identified specific content needing improvement in students skills and in teaching those skills'; '[the OSCE allows to] review scores to identify weaknesses and strengths, provide guidance for remediation'	Morag et al. 2001
	Pierre et al. 2005
	Hamann et al. 2002
'Proved to be an efficient and effective means of improving cancer education'	Battles et al. 1997
'The impact of the OSCE on the content of medical education in Japan may be the more important effect of the OSCE ... OSCEC has potential impact to change faculty teaching ...'	Ban et al. 1997
'The experience change their ways of teaching ... the criteria used now serving as reference points for their teaching and direct observation of students in action' 'The experience with the OSCE prompted the Department to define more clearly the learning objectives to plan activities and to change the evaluation system'	Grand'Maison et al. 1985
'One of its most valuable application is as a teaching technique'	Thomson 1987
'It allows to identify weaknesses in our teaching '.... 'to identify areas of deficiencies'	Adeyemi 2001
	Sloan et al. 1996
	Szerlip et al. 1998
'The OSCE can identify areas of strengths in terms of students'	Watson et al. 1982
	Schenk et al. 1999
	Humphrey-Murto et al. 2005
	Feather & Kopelman 1997
	Ainsworth et al. 1995
'The OSCE can 'identify areas of strengths in terms of curriculum or site characteristics '	Young et al. 1995
	Duerson et al. 2000
	Singer et al. 1996
'The OSCE can 'identify areas of strengths in terms of faculty teaching'	Cuschieri et al. 1979
	Peden et al. 1985
	Watson et al. 1982
	Tervo et al. 1997
	Collins et al. 1994
'[The OSCE shows] how to redesign teaching and curriculum'	Singer et al. 1994
'A tool where poor performances were investigated and causative factors are identified, identification of areas where methods and content are deficient, review scores to identify weaknesses and strengths'	McFaul & Howie 1993
'Awareness of students abilities[and in addition] the results of the OSCE affect curriculum'	Hoole et al. 1987
'[The OSCE provides] specific information on each student's clinical weaknesses which could guide an educational plan'	Petrusa et al. 1987
'[The OSCE] motivates changes in curriculum... modules are being developed'	Kowlowitz et al. 1991
'[The OSCE provides] relevant information is given to course director that inform him in curriculum change'	Heard et al. 1996
'[The OSCE provides] guidance for curriculum planning'; 'the OSCE provided the impetus for curriculum change'	Duerson et al. 2000
	Stillman et al. 1991
'Observing stations deleted by low reliability we understood that they can be due to student who didn't acquire or were not taught... content was then examined ... '[The OSCE] indicates where there should be a remediation of the station and further change in the curriculum'	Auewarakul et al. 2005b
'[The OSCE is an opportunity of] concerted effort to improve aspects of students´ education [and] specific teaching changes has been made due to OSCE'	Elnick et al. 1993
'[The OSCE] examination may help to achieve more uniform teaching'	Ainsworth et al. 1995
	Troncon 2004
'OSCE information provides a foundation on which structured teaching will be based'	Sloan et al. 2001

**Percentages are reported when feedback is based in a survey. Some studies contributed to more than one category.*

Discussion

The intention of this BEMER was to provide evidence on OSCE reliability, validity and other assessment requisites namely, OSCE fairness, acceptability and educational impact.

Underlying these broad research questions are the multiple specific questions teachers face in their daily practice: Is the OSCE feasible to accommodate low and high number of students?, Could standardized patients be as effective as experts in assessing students?, Should global ratings be used instead of checklists?, How many stations do we need to achieve a reliability of 0.80?, How does the OSCE correlate with other assessment methods?, How do students and teachers evaluate the OSCE?, How fair is the OSCE in students' perspectives?, How does the OSCE correlate with other assessment formats?, Is the OSCE a relevant exam capable of educational impact?

These are just some examples that still persist and this study hopefully introduced evidence to support some answers. Below we present a summary of main results.

Evidence on OSCE reliability

- Evidence on inter-rater reliability was found and previous concerns, regarding the reliability of the exam when using SPs (instead of expert examiners) and global ratings (instead of checklists) were not confirmed in our study
- Evidence suggests the use of global ratings and the coordinated use of standardized patients and expert evaluators
- Although expected, no differences were found when comparing the average inter-station reliability between OSCEs conducted with different underlying purposes in terms of consequences of students' assessment:
  - Highs takes *vs.* low stakes
  - Summative *vs.* formative
  - OSCEs implemented to evaluate students, curriculum or an intervention *vs.* to examine/appraise the OSCE itself
- The inter-rater reliability of examiners using global ratings was significantly lower than the checklists, which contrasts with findings related to average inter-station reliability.
- To result in the same expected test reliability, the OSCEs using checklist-based assessment grids would require a higher number of stations than identical OSCE using global rating-based assessment grids.

Evidence on OSCE validity

- OSCE has established criterion validity showing higher correlations with studies that reported disattenuated correlations
- Evidence on face/content validity was found suggesting
  - the OSCE stations are sampled against blueprinting and course objectives
  - stations design incorporates contributions from other teachers and experts
  - final decisions on content of stations are reached through consensus meetings

Evidence on OSCE fairness

Results suggest the OSCE is a fair exam with students globally reporting the OSCE as such. When compared with other traditional methods, the OSCE is generally perceived as a fairer exam adequate to the course objectives and curriculum, corresponding to what was expected with students feeling confident they are able to perform the required tasks. Similar positive comments were found in studies reporting on students and examiners or students and faculty feedback.

Evidence on OSCE acceptability

The OSCE was generally reported as a relevant and satisfactory exam, supporting its acceptability. Examiners perceived the OSCE as being a *'valid examination and an effective way of assessment'*. Students' acceptance of the OSCE also appeared globally positive, with the majority welcoming the OSCE and receiving the exam with enthusiasm and confidence. When comparing the exam with other assessment formats, OSCE is frequently reported as being better and an important component of overall measurement of clinical competence.

Evidence on OSCE educational impact

Students and teachers perceive the OSCE as an exam capable of steering learning and teaching.

The OSCE is recognized as providing a focus for relevant and useful learning, because it conveys strong messages to learners on what should be valued in terms of curriculum and learning outcomes. It is also seen as a source of motivation, creating an environment which favours students' attitudes and approaches to learning, as well as being capable of changing studying habits.

OSCE is also described as influencing and supporting teaching, with some studies reporting capacity to redesign methods and curriculum based on students' strengths and weaknesses. The most commonly reported influence of the OSCE on teaching was as a 'diagnostic tool' capable of identifying areas of students' deficiencies or strengths, and thus become a starting point to reformulate curriculum and teaching methods.

The findings reported above are globally positive (or very positive) with few studies reporting less positive comments, usually based on feedback given by a minority of participants.

However and as highlighted in a previous study (Patricio and Carneiro, 2012) the need for multiple approaches when dealing with scientific research in medical education implies that knowledge should take into account what is already known.

This was why we searched for other perspectives on the OSCE criteria as an assessment tool, not only during the period covered by this study but also until April 2012, to investigate the existence of other systematic reviews on OSCE requisites as an assessment method (namely on its reliability, validity, feasibility, fairness, acceptability and educational impact.). Two hundred and ninety three studies (corresponding to 74 studies in 2009; 73 studies in 2010; 105 studies in 2011; 41 studies in 2012) were retrieved and scrutinized to identify the existence of any review and/or systematic review. Some studies were retrieved (for example Barman 2005) reporting on OSCE reliability and validity but as they were not primary studies, neither systematic reviews, they were not considered in present study.

Only one systematic review was identified - 'The reliability of the objective structured clinical examination scores' - from Brannick et al. (2011). Taking this review into consideration we now present an analysis comparing the two studies. A summary is reported in Table 12.

Table 12. Comparison the reliability results with Brannick et al. 2011

Dimension under analysis	A systematic review of the reliability of objective structured clinical examination scores (Brannick et al. 2011)	Is the OSCE a reliable and valid tool to assess learning competencies in undergraduate medical studies Evidence from a BEME Systematic Review
Scope of studies included	OSCE studies	OSCE performed in undergraduate medical education
Number of studies and reliability values for Cronbach's alpha coefficient considered for analysis	39 studies with 188 reliability values *Note* - There was 64 studies with 457 reliability values, but most reported on reliability values other than Cronbach's alpha coefficient	60 studies with 127 reliability values
Characterization of Reliability values (Cronbach's alpha coefficient)	100 values – for inter-station reliability 53 values – on across-item reliability 35 values – n/c what type is reported	91 values – on inter-station reliability 36 values – on inter-rater reliability
Regarding the inter-station reliability		
Reliabilities considered for analysis *Note* – inter-station Reliability was the only type of reliability that is comparable across the 2 papers)	Overall reliability (corrected for the number of stations through weighting)	Inter-station reliability for 4 variants of OSCE formats: - examiner /checklist - examiner / global rating, - standardized patient /checklist - standardized patient /global rating
Reliability values	On average the overall alpha coefficients reported was of 0,66	On average the inter-station reliability was 0,19 (which represents an overall reliability of 0,74)

Other effect analysed	Analysis of the effect on the overall alpha coefficient for: - Type of examiner (SP, faculty, etc…) - Number of examiner - Type of scale (check-list *vs.* gl. rating) - Content (communication or clinical) - Context (research or high-stakes	Analysis of the effect on inter-station reliability of OSCE with examiners/checklist for: - OSCE aim (evaluate /examine OSCE) - OSCE station (history-taking Yes/no) - Type of exam (high / low-stakes) - OSCE purpose (formative /summative)
Effect of type of examiners (SPs or examiner)	No significant difference was identified. Note - for inter-item there was a significant difference, but one could not really identify if it was due to the type of exams or the content, as all examiners were evaluating clinical skills and most of the SPs communication skills)	No significant difference was identified when using a checklist, When using global ratings SPs reported to be more reliable than examiners (at a significance level of 10%)
Effect of type of scale (check-list or global rating)	No significant differences were identified NOTE – for inter-item there was a significant difference with higher reliability for global ratings, but as reported on the paper: Items on a checklist are often rather easily observed. Items on a Likert scale are subject to interpretation to a greater degree and call for graded responses to a set of behaviours observed over a longer period of time. Because a single judge typically rates all of the communication items in a station, any global impression of the examinee's performance in that station is likely to colour all the evaluative ratings for that examinee (…).There are other possible explanations for differences between checklist and Likert scale scores, including the occurrence of ceiling effects for some checklist items, as well as possible differences in underlying causes of behaviour. It may be, for example, that the clinical skills evaluated in the checklist depend upon a great number of underlying factors than the skills required for communication"	Global ratings showed consistently significant higher inter-station reliability than checklists (for both SPs and examiners) NOTE – The first explanation presented by Brannick et al. (2011) to this unexpected result (global ratings being more reliable that checklists) is the fact of having a single judge for which "any global impression on the examiner performance is likely to colour all the evaluative ratings". This explanation was on an inter-item context, which is not observed here. In our perspective, this result can be explained by the existence of an underlying effect, for example sympathy/empathy with examiner or the overall personality of the examinee, which might influence ("colour" in the same sense used by Brannick et al., (2011) the global rating outcome but is not possible to report on a checklist.
Effect of other variables	Content (evaluating communication or clinical skills) and number of raters (1 or 2) reported to be significant with - Evaluating clinical skills was significantly more reliable than evaluating communication skills (the inverse relationship was observed within the inter-item where evaluating communication skills was more reliable) - 2 raters were significantly more reliable than 1 rater	None of the 4 context (OSCE aim; Including an history-taking station; OSCE Purpose and Type of exam)analysed showed significant differences between its settings

Brannick's (2011) systematic review considered all OSCE studies, while our BEMER only considered the OSCE studies performed in undergraduate medical education. Regarding the reliability values analysed (Cronbach's alpha coefficient) Brannick's review reported 100 values for inter-station and 53 values for inter-items, while in our study we analysed 91 inter-station and 36 inter-rater values, therefore the only comparable analysis are those addressing the inter-station reliability or overall reliability.

Both studies reported a positive relationship between the number of stations and the overall reliability.

While we tested whether the type of examiner (SPs or examiners) and type of scale (global rating or check-list) were significantly different, Brannick et al. tested each of these variables separately. In term of conclusions we identified that SPs were signifycantly more reliable than examiners when using global ratings, and with any type of examiner, global ratings were significantly more reliable than check-lists. None of these variables reported to be significantly different in Brannick's test.

Analysing the effect of other variables (high

stakes *vs*. non high stakes OSCEs, summative *vs.* formative, implemented to evaluate *vs.* to analyse the OSCE itself) we did not identify any other significant effect on reliability. However, due to information limitations (namely missing or unusable data) we could only study the effects on inter-station reliability of OSCEs with examiners and checklist while Brannick et al. reported that clinical skills were significantly more reliable than communication skills and two raters were significantly more reliable than one rater.

Conclusions

Evidence was obtained showing the OSCE has high reliability and established validity. This was expected, since by its nature the OSCE should be more reliable and more valid to assess clinical outcomes than the 'traditional methods', due to its standardized format with students required to perform tasks, not only to report knowledge on how those tasks should be performed.

Findings also pointed to the OSCE as an exam meeting the GMC (2011) assessment requirements namely as a fair, acceptable (relevant and satisfactory) exam with educational value.

The evidence on OSCE educational impact - namely its capacity to steer students' learning priorities, suggests the OSCE meets the highest level of practical use of an evaluation process, which according to Prislin et al. (1988) depends on how an exam is capable of influencing subsequent learning.

Of interest is the fact that the few negative comments are from quite old studies. The Troncon' paper (2004) for example assumes that OSCE stations are short and not integrated (history, examination, diagnosis and management) but they can be - especially as developed in the last 5 years - longer stations allowing more authentic examination.

Limitations of the Study

We acknowledge intrinsic and external limitations to this study. Technical constraints determined that only the papers published in English were included in the analysis and coders' constraints determined the period covered by literature search.

The major difficulty found when analysing validity and reliability of the OSCE relates to the quality of primary reports, where data is sometimes missing or unusable. In terms of reliability, for example, only 60 studies (16%) contributed to this analysis. In addition to 7 studies (2%) that have statements on reliability without supportive data, there were two main reasons for exclusion:

- Use of a non standard reliability metric (for example not using the Cronbach's alpha coefficient as a measure of reliability)
- Use of a non standard Cronbach's alpha coefficient (i.e. computing a Cronbach's alpha using all the checklist items for all the stations together, instead of computing it by looking at the Cronbach's alpha coefficient of each station based on the individual checklist items). See Appendix 3 for more detailed information.

Similar difficulties occurred when analysing validity. As for fairness, relevance, satisfaction and educational impact of OSCE, information is usually clear but only a small number of papers contribute to those criteria. Another limitation is due to the nature of data, since for the latest criteria (fairness, relevance, satisfaction and educational impact) we rely on self-perceptions from students, faculty, examiners and SPS.

Finally, a word of caution is also needed in terms of potential publication bias. The studies show a tendency to report successful achievements and journals have tendency to publish mostly successful studies. These may explain how comments are in general so favourable.

Acknowledgement

We would like to extend our appreciation and gratitude to Alex Haig (PhD, Information scientist, NHS Education for Scotland) for literature search.

Declaration of interest: The authors report no declarations of interest.

Notes on Contributors

MADALENA PATRICIO (MA) is president of the Association for Medical Education in Europe (AMEE), assistant professor at the Institute of Introduction to Medicine and scientific consultant at the Center for Evidence Based Medicine both at the Faculty of Medicine, University of Lisbon, Portugal.

MIGUEL JULIÃO is a resident in Family Medicine and scientific consultant at Center for Evidence Based Medicine, Faculty of Medicine of the University of Lisbon, Portugal.

FILIPA FARELEIRA is a resident in Family Medicine and scientific consultant at Center for Evidence Based Medicine, Faculty of Medicine of the University of Lisbon, Portugal.

MEREDITH YOUNG, PhD is an assistant professor in the Department of Medicine, and an education research scientist at the Centre for Medical Education, McGill University, Canada.

GEOFFREY NORMAN, PhD is a professor in the Department of Clinical Epidemiology and Biostatistics, McMaster University

ANTÓNIO VAZ CARNEIRO, MD, PhD, FACP is a professor of Medicine and the head of the Center for Evidence Based Medicine at the Faculty of Medicine University of Lisbon, and head of the Cochrane Coordinating Center Portugal.

References

A-Latif A. 1992. An examination of the examinations: the reliability of the objective structured clinical examination and clinical examination. Medical Teacher 14: 179-183.

Adeyemi EO. 2001. The performance profile of medical students in the mock objective structured clinical examination. Saudi Medical Journal 22: 504-507.

Afroza A. 2000. Use of a PMP manual as a teaching tool to accelerate paediatric teaching in Bangladesh. Medical Teacher 22: 365-369.

Ainsworth MA, Callaway MR, Perkowski L. 1995. An OSCE assessment of fourth-year students. Academic Medicine 70: 444-445.

Ali J, Cohen R, ReznickR. 1995. Demonstration of acquisition of trauma management skills by senior medical students completing the ATLS program. Journal of Trauma, Injury, Infection and Critical Care 38: 687-691.

Allen R, Heard J, Savidge M, Bittengle J, Cantrell M, HuffmasterT. 1998. Surveying students' attitudes during the OSCE. Advances in Health Sciences Education 3: 197-206.

Auewarakul C, Downing SM, Jaturatamrong U, Praditsuwan R. 2005a. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. Medical Education 39: 276-283.

Auewarakul C, Downing SM, Praditsuwan R, Jaturatamrong U. 2005b. Item analysis to improve reliability for an internal medicine undergraduate OSCE. Advances in Health Sciences Education 10: 105-113.

Ban N, Hatao M, Ohtaki J, Saito T, Matsueda K, Izumi Y, Kusumoto M, Hosoda S. 1997. An OSCE Trial in Japan: Feasibility and Correlation with a Written Test. Advances in Medical Education 7th Ottawa Conference 1996, Maastricht, Netherlands, 404-406.

Bardin L. 1998. L'analyse du contenu. 9th. Ed. Paris, Presses Universitaire de Paris.

Barman A. Critiques on the objective structured clinical examination. Annals of the Academy of Medicine, Singapore 2005;34(8):478-82.

Basco WT, Gilbert GE, Chessman AW, Blue AV. 2000. The ability of a medical school admission process to predict clinical performance and patients' satisfaction. Academic Medicine 75: 743-747.

Battles JB, Sprankell SJ, BakerL, Becker M, Camp,G, Browne,C, Morell S, Bowling JR, Shores J, Buck E, Frates CO, Green JB, Smith G, Pence BC, Weinberg A, Laufman L, Philips BU. 1997. OSCE Stations for the Longitudinal Assessment of Cancer Screening and Detection. Advances in Medical Education 7th Ottawa Conference 1996, Maastricht, Netherlands, 407-409.

Bell SK, Krupat E, Fazio SV, Roberts DH, Schwartzstein RM. 2008. Longitudinal Pedagogy: A Successful Response to the Fragmentation of the Third-Year Medical Student Clerkship Experience. Academic Medicine 83: 467-675.

Benbow EW, Harrison I, Dornan TL, ONeill PA. 1998. Pathology and the OSCE: insights from a pilot study. Journal of Pathology 184: 110-114.

Blue AV, Stratton TD, Plymale M, DeGnore LT, Schwartz RW, Sloan DA. 1998. The effectiveness of the structured clinical instruction module. The American Journal of Surgery 176: 67-70.

Brannick MT, Erol-Korkmaz HT, Prewett M. 2011. A systematic review of the reliability of objective structured clinical examination scores. Medical Education 45: 1181-1189.

Campos-Outcalt D, Rutala PJ, Witzke DB, Fulginiti JV. 1994. Performances of underrepresented-minority students at the university of Arizona College of Medicine 1987-1991. Academic Medicine 69: 577-582.

Campos-Outcalt D, Watkins A, Fulginiti J, Kutob R, Gordon P. 1999. Correlations of Family Medicine Clerkship Evaluations and Objective Structured Clinical Examination Scores and Residency Directors' Ratings. Family Medicine 31: 90-94.

Carpenter JL, Battles JB, McIntire D, Sprankell SJ. 1992. Assessing the usefulness of using standardized patients in a clinical medicine course. Academic Medicine 67: 286.

Carpenter JL, McIntire D, Battles J, Wagner JM. 1993. Administration of parallel simultaneous objective structured clinical examination to accommodate a large class of students. Teaching and Learning in Medicine 5: 79-85.

Chenot JF, Simmenroth-Nayda A, Koch A, Fischer T, Scherer M, Emmert B, Stanske B, Kochen M, Himmel W. 2007. Can student tutors act as examiners in an objective structured clinical examination? Medical Education 41: 1032-1038.

Chessman, AW, Blue AV, Gilbert GE, Carey M, Mainous AG. 2003. Assessing students' communication and interpersonal skills across evaluation settings. Family Medicine 35: 643-648.

Chesser AM, Laing MR, Miedzybrodzka ZH, Brittenden J, Heys SD. 2004. Factor analysis can be a useful standard setting tool in a high stakes OSCE assessment. Medical Education 38: 825-831.

Collins JP, Tregonning GD, Gamble GD. 1994. Uniform experience and assessment during a multisite surgical clerkship. Australian and New Zealand Journal of Surgery 64: 506-511.

Collins JP, Gamble GD. 1996. A multi-format interdisciplinary final examination. Medical Education 30: 259-265.

Cook DA, Beckman TJ. 2006. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. The American Journal of Medicine 119: 166.e7-166.e16.

Coutts L, Rogers J. 1999. Predictors of student self-assessment accuracy during a clinical performance exam: comparisons between over-estimators and under-estimators of SP-evaluated performance. Academic Medicine 74:S128-S130.

Coutts-van Dijk LC, Bray JH, Moore S, Rogers J. 1997. Prospective study of how students humanism and psychosocial beliefs relate to speciality matching. Academic Medicine 72:1106-1108.

Cunnington JPW, Neville AJ, Norman GR. 1997. The Risks of Thoroughness: Reliability and Validity of Global Ratings and

Checklists in an OSCE. Advances in Medical Education, 7th Ottawa Conference 1996, Maastricht, Netherlands, 143-145.

Cuschieri A, Gleeson FA, Harden RM, Wood RAB. 1979. A new approach to a final examination in surgery. Annals of the Royal College of Surgeons of England 61: 400-405.

Dacre JE. 1993. The effect of formal instruction in ophthalmoscopy on medical student performance. Medical Teacher 15:321-325.

Doig CJ, Harasym PH, Fick GH, Baumber JS. 2000. The effects of examiner background, station organization and time of exam on OSCE scores assessing undergraduate medical students' physical examination skills. Academic Medicine 75: S96-S98.

Duerson MC, Romrell LJ, Stevens BC. 2000. Impacting faculty teaching and student performance: nine years experience with the objective structured clinical examination. Teaching and Learning in Medicine 12: 176-182.

Durak H, Caliskan S, Bor S, van der Vleuten C. 2007. Use of Case-Based Exams as an instructional teaching tool to teach clinical reasoning. Medical Teacher 29: e170-e174.

Elnicki DM, Shockcor WT, Morris DK, Hallbritter KA. 1993. Creating an objective Structured Clinical Examination for the internal medicine Clerkship: Pitfalls and benefits. American Journal of Medical Sciences 306: 94-97.

Elzubeir MA, Rizk DEE. 2001. Assessing confidence and competence of senior medical students in an obstetrics and gynaecology clerkship using an OSCE. Education for Health 14: 373-382.

Famuyiwa OO, Zachariah MP, Ilechukwu STC. 1991. The objective structured clinical examination in undergraduate psychiatry. Medical Education 25: 45-50.

Feather A, Kopelman PG. 1997. A practical approach to running an objective structured clinical examination (OSCE) for medical undergraduates. Education for Health 10: 333-350.

Feickert JAD, Harris IB, Anderson DC, Bland CJ, Allen S, Poland GA, Satran L, Miller WJ. 1992. Senior medical students as simulated patients in an objective structured clinical examination: motivation and benefits. Medical Teacher 14: 167-177.

Fox R, Dacre J, Mclure C. 2001. The impact of formal instruction in clinical examination skills on medical student performance - the example of peripheral nervous system examination. Medical Education 35: 371-373.

Frost GJ, CaterJI, Forsyth JS. 1986. The use of an Objective Structured Clinical Examination (OSCE) in paediatrics. Medical Teacher 8:261-269.

General Medical Council 2011. Assessment in undergraduate medical education. Advice Supplementary to Tomorrows' Doctors 2009. Retrieved July 3, 2012. Available at: http://www.gmc-uk.org/Assessment_in_undergraduate_web.pdf_38514111.pdf

Gilson GJ, George KE, Qualls CM, Sarto GE, Obenshain SS, Boulet J. 1998. Assessing clinical competence of medical students in women's health care: use of the Objective Structured Clinical Examination. Obstetrics and Gynecology 92: 1038-1043.

Grand'Maison P, Blouin D, Briere D. 1985. Utilization of the Objective Structured Clinical Examination OSCE in Gynecology/Obstetrics. Proceedings of the Annual Conference on Research in Medical Education 24: 65-71.

Grand'Maison P, Brailovsky CA, Lescop J. 1996. Content validity of the Quebec licensing examination (OSCE). Canadian Family Physician 42:254-259.

Gruppen LD, Davis WK, Fitzgerald JT, McQuillan. 1997. Reliability, number of stations, and examination length in an Objective Structured Clinical Examination. Advances in Medical Education, 7th Ottawa Conference 1996, Maastricht, Netherlands, 441-442.

Hamann C, Volkan K, Fishman MB, Silvestri RC, Simon SR, Fletcher SW. 2002. How well do second-year students learn physical diagnosis? Observational study of an Objective Structured Clinical Examination (OSCE). BMC Medical Education 2: 1-11.

Hanson M, Hodges B, McNaughton N, Regehr G. 1998. The integration of child psychiatry into a Psychiatry Clerkship OSCE. Canadian Journal of Psychiatry 43: 614-618.

Harden RM, Stevenson M, Downie WW, Wilson GM. 1975. Assessment of clinical competence using objective structured examination. British Medical Journal 1: 447-451.

Harris I, Ytterberg S, Anderson D, Kofron P, Kvasnicka J, Moller JH. 1997. Tailored Response Test: a new approach for teaching in medical education. Medical Teacher 19: 194-199.

Hart IR, Woodward CA, Poewles P. 1987. Symposium: Use of the OSCE Objective Structured Clinical Examination to Assess Clinical Competence: Predictive Validity and Acceptability of the Objective Structured Clinical Examination. Proceedings of the Annual Conference on Research in Medical Education 23: 407-410.

Heard JK, Allen R, Tank PW, Cason GJ, Cantrell M, Wheller RP. 1996. Assessing clinical skills of medical students. Journal of the Arkansas Medical Society 93:175-179.

Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G. 1996. Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. Medical Education 30: 38-43.

Hodges B, Lofchy J. 1997a. Evaluating psychiatric clinical clerks with a mini-objective structured clinical examination. Academic Psychiatry 21: 219-225.

Hodges B, Regehr G, Hanson M, McNaughton N. 1997b. An objective structured clinical examination for evaluating psychiatric clinical clerks. Academic Medicine 72: 715-721.

Hodges B, Hanson M, McNaughton N, Regehr G. 1998. The Objective Structured Clinical Examination in Psychiatry: a Validation Study. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998, Philadelphia, USA.

Hoelker M, Breukelmann D, Saur M, Nippert RP. 2004. Process Evaluation of a Course in Basic Clinical Skills: Implementing an OSCE. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998, Philadelphia, USA.

Hoole AJ, Kowlowitz V, McGaghie WC, Sloane PD, Colindres RE. 1987. Using the objective structured clinical examination at the University of North Carolina Medical School. North Carolina Medical Journal 48: 463-467.

Humphrey-Murto S, Smith CD, Touchie C, Wood TC. 2004. Teaching the musculoskeletal examination: are patient educators as effective as rheumatology faculty? Teaching and Learning in Medicine 16:175-180.

Humphrey-Murto S, Wood TJ, Touchie C. 2005. Why do physicians volunteer to be OSCE examiners? Medical Teacher 27:172-174.

Humphris GM, Kaney S. 2001. Examiner fatigue in communication skills objective structured clinical examinations. Medical Education 35: 444-449.

Jefferies A, Simmons B, Regehr G. 2007. The effect of candidate familiarity on examiner OSCE scores. Medical Education 41: 888-891.

Jewell D. 1988. Learning through examinations: use of an objective structured clinical examination as a teaching method in general practice. Journal of the Royal College of General Practitioners 38: 506-508.

Johnson G, Reynard K. 1994. Assessment of an objective structured clinical examination (OSCE) for undergraduate students in accident and emergency medicine. Journal of Accident and Emergency Medicine 11: 223-226.

Jolly BC, Jones A, Dacre JE, Elzubeir M, Kopelman P, Hitman G.

1996. Relationships between students' clinical experiences in introductory clinical courses and their performances on an objective Structured Clinical Examination (OSCE). Academic Medicine 71: 909-916.

Joorabchi B. 1991. Objective Structured Clinical Examination in a pediatric residency program. American Journal of Diseases in Children 145: 757-762.

Junger J, Schafer S, Roth C, Schellberg D, Friedman Ben-David M, Nikendei C. 2005. Effects of basic clinical skills training on objective structured clinical examination performance. Medical Education 39: 1015-1020.

Kent AP, Lazarus J. 1983. An objective medical student examination in obstetrics. South African Medical Journal 64: 388-389.

Kilminster S, Roberts T. 2004. Standard setting for OSCEs: trial of borderline approach. Advances in Health Sciences Education 9: 201-209.

Kirby RL, Curry L. 1982. Introduction of an objective structured clinical examination (OSCE) to an undergraduate clinical skills programme. Medical Education 16: 362-364.

Knowles C, Kinchington F, Erwin J, Peters B. 2001. A randomized controlled trial of the effectiveness of combining video role play with traditional methods of delivering undergraduate medical education. Sexually Transmitted Infections 77: 376-380.

Kowlowitz V, Hoole AJ, Sloane PD. 1991. Implementing the Objective Structured Clinical Examination in a traditional medical school. Academic Medicine 66: 345-347.

Lazarus J, Kent A. 1983. Student attitudes towards the objective structured clinical examination (OSCE) and conventional methods of assessment. South African Medical Journal 64: 390-394.

Lloyd JS, Williams RG, Simonton DK, Sherman D. 1990. Order effects in standardized patient examinations. Academic Medicine: 65, S51-S52.

Long EM. 1997. An Evaluation of Osces for Final-Year Students. Advances in Medical Education, 7th Ottawa Conference 1996, Maastricht, Netherlands, 448-451.

Lunenfeld E, Weinreb B, Lavi Y, Amiel GE, Friedman M. 1991. Assessment of emergency medicine: a comparison of an experimental objective structured clinical examination with a practical examination. Medical Education 25: 38-44.

Malik SL, Manchanda SK, Deepack KK, Sunderam KR. 1988. The attitudes of medical students to the objective structured practical examination. Medical Education 22: 40-46.

Mann KV, MacDonald C, Norcicni JJ. 1990. Reliability and objective structured clinical examinations: four years of experience in a surgical clerkship. Teaching and Learning in Medicine 2: 219-224.

Mariolis A, Mihas C, Alevizos A, Papathanasiou M, Mariolis-Sapsakos T, Marayiannis K, Koutsilieris M. 2008. Evaluation of a clinical attachment in Primary Health Care as a component of undergraduate medical education. Medical Teacher 30: e2002-e2007.

Martin IG, Stark P, Jolly B. 2000. Benefiting from clinical experience: the influence of learning style and clinical experience on performance in an undergraduate objective structured clinical examination. Medical Education 34: 530-534.

Matsell DG, Wolfish NM, Hsu E. 1991. Reliability and validity of the objective structured clinical examination in paediatrics. Medical Education 25: 293-299.

Mazor KM, Ockene JK, Rogers HJ, Carlin MM, Quirk ME. 2005. The relationship between checklist scores on a communication OSCE and analogue patients' perceptions of communication. Advances in Health Sciences Education 10: 37-51.

McDaniel MA, Agarwal PK, Huelser BJ, McDermott KB, Roediger HL. 2011. Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. Journal of Educa Ref Man tional Psychology 103: 399-414.

McFaul PB, Howie PW. 1993. The assessment of clinical competence in obstetrics and gynaecology in two medical schools by an objective structured clinical examination. British Journal of Obstetrics and Gynaecology 100: 842-846.

McGaghie WC, Renner BR, Kowlowitz V, Sauter SVH, Hoole AJ, Schuh MS. 1994. Development and evaluation of musculoskeletal performance measures for an objective structured clinical examination. Teaching and Learning in Medicine 6: 59-63.

McKinley RK, Strand J, Gray T, Schuwirth L, Alun-Jones T, Miller H. 2008. Development of a tool to support holistic generic assessment of clinical procedure skills. Medical Education 42: 619-627.

Keane M. 2003. Encyclopedia and Dictionary of Medicine, Nursing, and Allied Health. O'Toole MT (Ed) 7th Ed., Saunders, Elsevier, Inc.

Morag E, Lieberman G, Volkan K, Shaffer K, Novelline R, Lang EV. 2001. Clinical competence assessment in radiology: introduction of an objective structured clinical examination in the medical school curriculum. Academic Radiology 8:74-81.

Mossey PA, Newton JP. 2001. The structured clinical operative test SCOT in dental competency assessment. British Dental Journal 190: 387-390.

Murray E, Jolly B, Modell M. 1997. Evaluation of the Effectiveness of Clinical Skills Teaching in the Community. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht, Netherlands, 564-566.

Nalesnick SW, Mills CS, Olsen CH, Haffner WH, Zahn CM. 2005. Creating an ideal objective structured clinical exam for an obstetrics and gynecology medical student clerkship. American Journal of Obstetrics & Gynecology, 193: 1544-1550.

Newble DI, Elmslie RG, Baxter A. 1978. A problem based criterion referenced examination of clinical competence. Journal of Medical Education 53: 720-726.

Newble DI, Hoare J, Elmslie RG. 1981. The validity and reliability of a new examination of the clinical competence of medical students. Medical Education 15: 46-52.

Newble DI, Swanson DB. 1988. Psychometric characteristics of the objective structured clinical examination. Medical Education 22: 325-334.

Newble D, Swanson D. 1998. Improving the quality of a multidisciplinary test of clinical competence: a longitudinal study. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998, Philadelphia, USA.

Norcini J, Anderson B, BollelaV, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V, Roberts T. 2011. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. Medical Education 33: 206-214.

OConnor HM, McGraw R. 1997. Clinical skills training: developing objective assessment instruments. Medical Education 31:359-363.

Park RS, Chibnall JT, Blaskiewicz RJ, Furman GE, Powell JK, Mohr CJ. 2004. Construct validity of an objective structured clinical examination (OSCE) in psychiatry: associations with the clinical skills examination and other indicators. Academic Psychiatry 28:122-128.

Patricio M, Carneiro AV. 2012. Systematic Reviews of evidence in medical education and vlinical medicine: is the nature of evidence similar? Medical Teacher 34: 474-482.

Peden NR, Cairncross RG, Harden RM, Crooks J. 1985. Assessment of clinical competence in therapeutics: the use of he Objective Structured Clinical Examination. Medical Teacher 7: 217-223.

Petrusa ER, Blackwell TA, Rogers LP, Saydari C, Parcel S,

Guckian JC. 1987. An Objective measure of clinical performance. American Journal of Medicine 83: 34-42.

Petrusa ER, Blackwell TA, Carline J, Ramsey PG, McGaghie W, Colindres R, Kowlowitz V, Mast TA, Soler,N. 1991. A multi institutional trial of an objective structured clinical examination. Teaching and Learning in Medicine 3:86-94.

Pierre RB, Wierenga A, Barton M, Branday JM, Christie CDC. 2004. Student evaluation of an OSCE in paediatrics at the University of the West Indies, Jamaica. BMC Medical Education 4: 1-7.

Pierre RB, Wierenga A, Barton M, Thame K, Branday JM, Christie CD. 2005. Student self-assessment in a paediatric objective structured clinical examination. West Indian Medical Journal 54: 144-148.

Prislin MD, Fitzpatrick CF, Lie D, Giglio M, Radecki S, Lewis E. 1998. Use of an Objective Structured Clinical Examination in Evaluating Student Performance. Family Medicine 30: 338-344.

Raga S, Coovadia HM. 1985. Experience with three identical objective structured clinical examinations (OSCEs) conducted simultaneously for an entire final-year class. South African Medical Journal 68: 819-820.

Rahman SA. 2001. Promoting learning outcomes in paediatrics through formative assessment. Medical Teacher 23: 467-470.

Regehr G, Freeman R, Hodges B, Russell L. 1999a. Assessing the generalizability of OSCE measures across content domains. Academic Medicine 74:1320-1322.

Regehr G, Freeman R, Robb A, Missiha N, Heissey R. 1999b. OSCE Performance evaluations made by standardized patients: comparing checklists and global ratings scores. Academic Medicine 74: S135-S137.

Remmen R, Scherpbier A, Denekens JO, Derese A, Hermann I, Hoogenboom R, Van der Vleuten C, Van Royen P, Bossaert L. 2001. Correlation of a written test of skills and a performance based test:a study in two traditional medical schools. Medical Teacher 23: 29-32.

Roediger HL, Agarwal PK, McDaniel MA, McDermott KB. 2011. Test-enhanced learning in the classroom: Long-term improvements from quizzing. Journal of Experimental Psychology Applied 1: 382-395.

Ribin NJ, Philip EB. 1998. Health care perceptions of the standardized patient. Medical Education 32: 538-542.

Rosebraugh CJ, Speer AJ, Solomon DJ, Szauter KE, Ainsworth MA, Holden MD, Lieberman SA, Clyburn EB. 1997. Setting standards and defining quality of performance in the validation of a standardized-patient examination format. Academic Medicine 72: 1012-1014.

Roy V, Tekur U, Prabhu S. 2004. A comparative study of two evaluation techniques in pharmacology practicals: Conventional practical examination *vs.* objective structured practical examination Indian Journal of Pharmacology 36: 386-388.

Rudland J, Wilkinson T, Smith-Han K, Thomson-Fawcett M. 2008. "You can do it late at night or in the morning. You can do it at home, I did it with my flatmate." The educational impact of an OSCE. Medical Teacher 30: 2006-2011.

Rutala PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ. 1990. Student fatigue as a variable affecting performance in an objective structured clinical examination. Academic Medicine, 65: S53-S54.

Rutala PJ, Fulginiti JV, McGeagh AM, Leko EO, Koff NA, Witzke DB. 1992. Predictive validity of a required multidisciplinary standardized-patient examination. Academic Medicine, 67: S60-S62.

Schenk M, Popps S, Bridge P, Gallagher R, Petrusa ER, Frank RR. 1999. Effectiveness of an occupational and environmental medicine curriculum as indicated by evaluation of medical student performance on an Objective Structured Clinical Examination. Journal of Occupational and Environmental Medicine 41: 954-959.

Schuwirth LW, van der Vleuten CP. 2011. General overview of the theories used in assessment: AMEE Guide No. 57 Medical Teacher 33: 783-707.

Schwartz RW, Donnelly MB, Sloan DA, Young B. 1994. Knowledge gain in a problem-based surgery clerkship. Academic Medicine 69: 148-151.

Searle J. 2000. Defining competency - the role of standard setting. Medical Education 34: 363-366.

Shatzer JH, Darosa D, Colliver JA, Barkmeier L. 1993. Station-length requirements for reliable performance-based examination scores. Academic Medicine 68: 224-229.

Simon SR, Volkan K, Hamann C, Duffey C, Fletcher SW. 2002. The relationship between second-year medical students' OSCE scores and USMLE Step 1 scores. Medical Teacher 24: 535-539.

Singer PA, Robb A, Cohen R, Norman G, Turnbull J. 1994. Evaluation of a multicenter ethics objective structured clinical examination. Journal of General Internal Medicine 9: 690-692.

Singer PA, Robb A, Cohen R, Norman G, Turnbull J. 1996. Performance-based assessment of clinical ethics using an objective structured clinical examination. Academic Medicine 71: 495-498.

Sloan DA, Donnelly MB, Johnson SB, Schwartz RW, Strodel WE. 1994. Assessing surgical residents' and medical students' interpersonal skills. Journal of Surgical Research 57: 613-618.

Sloan DA, Donnelly MB, Schwartz RW, Felts JL, Blue AV, Strodel W E. 1996. The use of the Objective Structured Clinical Examination (OSCE) for evaluation and instruction in graduate medical education. Journal of Surgical Research 63: 225-229.

Sloan PA, Plymale MA, Johnson M, Vanderveer B, LaFountain P, Sloan DA. 2001. Cancer pain management skills among medical students: the development of a cancer pain objective structured clinical examination. Journal of Pain and Symptom Management 21: 298-306.

Smith LJ, Price DA, Houston IB. 1984. Objective Structured Clinical Examination compared with other forms of student assessment. Archives of Disease in Childhood 59: 1173-1176.

Stern D, Committee C, Fitzgerald JT. 1998. A Patient Satisfaction Survey as a Valid Adjunct to Simulated Patient Exercises. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998, Philadelphia, USA.

Stillman PL, HaleyHL, Regan MB, Philbin MM. 1991. Positive effects of a clinical performance assessment program. Academic Medicine 66: 481-483.

Streiner DL. 1986. Global Ratings. In Neufeld V, Norman GR. (Edts) Assessing Clinical Competence. New York. Springer. USA.

Szerlip H, Anderson DS, Garris JB, Stanton M. 1998. Development and Implementation of a Pelvic Examination Station for a High-Stakes OSCE. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998, Philadelphia, USA.

Tann M, Amiel GE, Bitterman A, Ber R, Cohen R. 1997. Analysis of the Use of Global Ratings by Standardized Patients and Physicians. Advances in Medical Education, 7th Ottawa Conference 1996, Maastricht, Netherlands, 191-192.

Tervo RC, Dimitrievich E, Trujillo AL, Whittle K, Redinius P, Wellman L. 1997. The Objective Structured Clinical Examination (OSCE) in the Clinical Clerkship: an overview. South Dakota Journal of Medicine, 50: 153-156.

Thistlethwaite JE. 2002. Developing an OSCE station to assess the ability of medical students to share information and decisions with patients: Issues relating to interrater reliability and the use of simulated patients. Education for Health, 15:170-179.

Thomson, DM. 1987. The objective structured clinical examination for general practice: design, validity and reliability. Journal of

the Royal College of General Practitioners 37: 149-153.
Troncon LE. 2004. Clinical skills assessment: limitations to the introduction of an "OSCE" (Objective Structured Clinical Examination) in a traditional Brazilian medical school. Sao Paulo Medical Journal 122: 12-17.
Verhoeven BH, Hamers JGHC, Scherpbier A, Hoogenboom R, van der Vleuten C. 2000. The effect on reliability of adding a separate written assessment component to an objective structured clinical examination. Medical Education 34: 525-529.
van der Vleuten C. 1996. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ 1: 41-67.
van der Vleuten C. 2000. Validity of final examinations in undergraduate medical training. BMJ 11; 321(7270): 1217-1219.
Verma M, Singh T. 1993a. Attitudes of medical students towards Objective Structured Clinical Examination (OSCE) in Pediatrics. Indian Pediatrics 30: 1259-1261.
Verma M, Singh T. 1993b. Experiences with Objective Structured Clinical Examination (OSCE) as a tool for formative evaluation. Indian Pediatrics, 30: 699-703.
Volkan K, Simon SR, Baker H, Todres ID. 2004. Psychometric structure of a comprehensive objective structured clinical examination: a factor analytic approach. Advances in Health Sciences Education 9: 83-92.
Walters K, Osborn D, Raven P. 2005. The development, validity and reliability of a multimodality objective structured clinical examination in psychiatry. Medical Education 39:292-298.
Wass V, Jolly B. 2001a. Does observation add to the validity of the long case? Medical Education 35: 729-734.
Wass V, Jones R, van der Vleuten C. 2001b. Standardized or real patients to test clinical competence? The long case revisited. Medical Education 35: 321-325.
Wass V, McGibbon D, van der Vleuten C. 2001c. Composite undergraduate clinical examinations: how should components be combined to maximise reliability? Medical Education 35: 326-330.
Watson AR., Houston IB, Close GC. 1982. Evaluation of an objective structured clinical examination. Archives of Disease in Childhood 57: 390-398.
Weinreb B, Bentov Y, Burger S, Gvili B, Margolis CZ, Panso-Zablodowsky E, Schwartzman P, Solomon F. 1998. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998, Philadelphia, USA.
Wilkinson TJ, Newble DI, Wilson PD, Carter JM, Helms RM. 2000. Development of a three centre simultaneous objective structured clinical examination. Medical Education 34: 798-807.
Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. 2003. Objectivity in objective structured clinical examinations: Checklists are no substitute for examiner commitment. Academic Medicine 78: 219-223.
Wilkinson TJ, Frampton CM. 2004. Comprehensive undergraduate medical assessments improve prediction of clinical performance. Medical Education 38: 1111-1116.
Wong M, Fones C, Aw M, Hoon C, Low P, Amin Z, Wong P, Goh P, Wai C, Ong B, Tambyah P, Koh D. 2007. Should non-expert clinician examiners be used in objective structured assessment of communication skills among final year medical undergraduates? Medical Teacher 29: 927-932.
Woolf K, Haq I, McManus LC, Higham J, Dacre J. 2008. Exploring the underperformance of male and minority ethnic medical students in first year clinical examinations. Advances in Health Sciences Education 13: 607-616.
Young WW, Barthold JC, Birenbaum D, Long P, Dion M, Hamilton LA. 1995. An objective structured clinical exam in multisite obstetrics and gynecology clerkship. Teaching and Learning in Medicine 7: 177-181.

Appendices

Appendix 1
Studies contributing to reliability

A-Latif, 1992; Ali, Cohen, & Reznick, 1995; Auewarakul, Downing, Jaturatamrong, & Praditsuwan, 2005a; Auewarakul, Downing, Praditsuwan, & Jaturatamrong, 2005b; Basco, Gilbert, Chessman, & Blue, 2000; Bell, Krupat, Fazio, Roberts, & Schwartzstein, 2008; Benbow, Harrison, Dornan, & ONeill, 1998; Blue, Stratton, Plymale, DeGnore, Schwartz, & Sloan, 1998; Campos-Outcalt, Rutala, Witzke, & Fulginiti, 1994; Campos-Outcalt, Watkins, Fulginiti, Kutob, & Gordon, 1999; Carpenter, Battles, McIntire, & Sprankell, 1992; Carpenter, McIntire, Battles, & Wagner, 1993; Chenot, Simmenroth-Nayda, Koch, et al., 2007; Chesser, Laing, Miedzybrodzka, Brittenden, & Heys, 2004; Coutts-van Dijk, Bray, Moore, & Rogers, 1997; Coutts & Rogers, 1999; Cunnington, Neville, & Norman, 1997; Dacre, 1993; Doig, Harasym, Fick, & Baumber, 2000; Elzubeir & Rizk, 2001; Fox, Dacre, & Mclure, 2001; Gruppen, Davis, Fitzgerald, & McQuillan, 1997; Hanson, Hodges, McNaughton, & Regehr, 1998; Hodges, Turnbull, Cohen, Bienenstock, & Norman, 1996; Hodges & Lofchy, 1997a; Hodges, Regehr, Hanson, & McNaughton, 1997b; Hoelker, Breukelmann, Saur, & Nippert, 2004; Humphrey-Murto, Smith, Touchie, & Wood, 2004; Humphris & Kaney, 2001; Jefferies, Simmons, & Regehr, 2007; Jewell, 1988; Jolly, Jones, Dacre, Elzubeir, Kopelman, & Hitman, 1996; Junger, Schafer, Roth, Schellberg, Friedman Ben-David, & Nikendei, 2005; Kilminster & Roberts, 2004; Lloyd, Williams, Simonton, & Sherman, 1990; Long, 1997; Mann, MacDonald, & Norcicni, 1990; Mariolis, Mihas, Alevizos, et al., 2008; Martin, Stark, & Jolly, 2000; Mazor, Ockene, Rogers, Carlin, & Quirk, 2005; McGaghie, Renner, Kowlowitz, Sauter, Hoole, & Schuh, 1994; McKinley, Strand, Gray, Schuwirth, Alun-Jones, & Miller, 2008; Newble, Hoare, & Elmslie, 1981; Newble & Swanson, 1988; Newble & Swanson, 1998; OConnor & McGraw, 1997; Park, Chibnall, Blaskiewicz, Furman, Powell, & Mohr, 2004; Regehr, Freeman, Robb, Missiha, & Heissey. 1999b); Rutala, Witzke, Leko, Fulginiti, & Taylor. 1990; Shatzer, Darosa, Colliver, & Barkmeier. 1993; Schwartz, & Strodel, 1994; Sloan, Donnelly, Johnson, Schwartz, & Strodel. 1994; Sloan, Donnelly, Schwartz, Felts, Blue, & Strodel, 1996; Stern, Committee, & Fitzgerald, 1998; Tann, Amiel, Bitterman, Ber, & Cohen, 1997; Thomson, 1987; Verhoeven, Hamers, Scherpbier, Hoogenboom, & van der Vleuten, 2000; Verma & Singh, 1993b; Wass, Jones, & van der Vleuten, 2001b; Wilkinson, Newble Wilson, Carter, & Helms, 2000; Wilkinson, Frampton, Thompson-Fawcett, & Egan, 2003; Wong, Fones, Aw, et al., 2007; Woolf, Haq, McManus, Higham, & Dacre, 2008.

Appendix 2
Studies contributing to validity

A-Latif, 1992; Auewarakul, Downing, Jaturatamrong, & Praditsuwan, 2005a; Bell, Krupat, Fazio, Roberts, &

Schwartzstein, 2008; Benbow, Harrison, Dornan, & ONeill, 1998; Campos-Outcalt, Rutala, Witzke, & Fulginiti, 1994; Campos-Outcalt, Watkins, Fulginiti, Kutob, & Gordon, 1999; Chessman, Blue, Gilbert, Carey, & Mainous, 2003; Coutts-van Dijk, Bray, Moore, & Rogers, 1997; Cunnington, Neville, & Norman, 1997; Elnicki, Shockcor, Morris, & Hallbritter, 1993; Elzubeir & Rizk, 2001; Famuyiwa, Zachariah, & Ilechukwu, 1991; Gilson, George, Qualls, Sarto, Obenshain, & Boulet, 1998; Jolly, Jones, Dacre, Elzubeir, Kopelman, & Hitman, 1996; Joorabchi, 1991; Kirby & Curry, 1982; Matsell, Wolfish, & Hsu, 1991; Mazor, Ockene, Rogers, Carlin, & Quirk, 2005; Newble & Swanson, 1988; Park, Chibnall, Blaskiewicz, Furman, Powell, & Mohr, 2004; Petrusa, Blackwell, Rogers, Saydari, Parcel, & Guckian, 1987; Regehr, Freeman, Hodges, & Russell, 1999a; Remmen, Scherpbier, Denekens, et al., 2001; Rosebraugh, Speer, Solomon, et al., 1997; Rutala, Fulginiti, McGeagh, Leko, Koff, & Witzke, 1992; Schwartz, Donnelly, Sloan, & Young, 1994; Simon, Volkan, Hamann, Duffey, & Fletcher, 2002; Smith, Price, & Houston, 1984; Thomson, 1987; Verhoeven, Hamers, Scherpbier, Hoogenboom, & van der Vleuten, 2000; Wass & Jolly, 2001a; Wass, McGibbon, & van der Vleuten, 2001c; Woolf, Haq, McManus, Higham, & Dacre, 2008.

Appendix 3
Major difficulties when reporting reliability and validity

Some problems were found in terms of evidence on reliability: Commonly, studies of OSCE will report the overall test reliability, frequently calling it a 'Cronbach's alpha' (i.e. Cronbach's alpha coefficient). The terminology is confusing. Typically, this results from treating each station score as an item and computing the reliability of the 'n' item (station) test. However in its original form, alpha was developed for personality and achievement tests, where it really was items. The problem arises because occasionally OSCE studies will actually look at an alpha for each station based on the individual checklist items. And one paper put all checklist items for all stations together and computed alpha. The primary problem with this loose terminology is that it confounds error variance from items within stations (which is typically small) with error from different stations (which is large).

A second problem is that alpha is directly related to the number of stations – longer OSCEs have higher reliability. This comes out of the Spearman-Brown formula which related the test reliability R to the average inter-station reliability r and the number of stations, n.

$$r = \frac{R}{n-(n-1)R}$$

In order to look at the reliability of a checklist, for example, we work backwards from the test reliability to the average inter-station reliability

$$r = \frac{R}{n-R(n-1)}$$

So for a 10 station OSCE with overall reliability of .8, the inter-station reliability is:

$$.8 / [10 - 9x.8] = .8 / 2.8 = .286$$

More information on difficulties found when looking for evidence on OSCE feasibility, reliability and validity after analysing the first 400 studies of this BEMME systematic review were reported in 2009 by Patricio et al.

CHAPTER 6
Concluding remarks

A Best Evidence Medical Education (BEME) Systematic Review on the feasibility, reliability and validity of the Objective Structured Clinical Examination (OSCE) in undergraduate medical studies.

Concluding Remarks

The structure of the final chapter of a dissertation reporting on a systematic review justifies the inclusion of a summary of the work done so far, from the research questions to results. Also mandatory is to reflect on the results of the critical appraisal of the BEMER process, the limitations of the study and a last section on the future 'looking forward to the OSCE exam as a 'five star exam'.

Summary of the BEMER process: from questions to results

The aim of the study was to produce scientific evidence about the OSCE, by means of a BEMER, on its suitability to assess learning outcomes in undergraduate medical studies. Two instrumental objectives were defined: to characterize OSCE technical and economic feasibility and to get evidence on the reliability, validity, fairness, acceptability and educational impact of the exam.

BEME methodology (full details in Chapter 1) was applied by two independent coders who scrutinized the 1065 primary studies retrieved from the 1083 identified.

A summary of main results is given below (complete description reported in Chapters 4 and 5).

- *Evidence on OSCE technical feasibility*

Evidence on OSCE as a technical feasible exam was based in 1065 studies reporting on the implementation of the OSCE to assess multiple learning outcomes, in a range of several contexts (hospital *vs.* community, high *vs.* low stakes exams, etc.) to serve multiple purposes (formative *vs.* summative exams), implemented to evaluate (students, curriculum or an intervention) or to examine the OSCE itself, used by a range of 26 professions.

OSCE flexibility in terms of underlying designs and formats (accommodating a high and low number of students, performed in one or several venues, with one or multiple circuits, with a high and low number of real patients, standardized patients and examiners) was also documented as one of the major OSCE advantages and a major reason for its adoption.

The flexibility of OSCE is probably the responsible factor for teachers' ownership since, in spite of being a standardized exam, teachers have the possibility of designing the OSCE according to their specific contexts and objectives. This means that, without losing its characteristics to increase reliability and validity of the exam, each OSCE can be designed to meet specific needs.

- *Evidence on OSCE economic feasibility*

Evidence was also obtained on OSCE economic feasibility, with the OSCE being recommended because of its unique benefits, in spite of, in some circumstances, being expensive to administer

Evidence also suggests that there are alternative ways to decrease costs and transparent categories are needed for reporting OSCE, namely to distinguish direct and indirect costs. Editors should request high quality of OSCE primary reports namely reporting on economic viability to support schools' decisions.

- *Evidence on OSCE reliability*

Throughout previous chapters we highlighted some questions regarding the reliability of the OSCE. Among them we were interested in the reliability of the OSCE when using standardized patients (SPs) instead of examiners and using global ratings instead of checklists. Our results show that previous concerns with lower reliability when using SPs and global ratings were not confirmed.

These findings should assist evaluators when designing OSCEs, namely in terms of background of examiners and type of scoring tools, with evidence supporting the use of global ratings and the co-ordinated use of standardized patients and experts when using evaluators.

- *Evidence on OSCE validity*

Evidence was found on face-content validity, with a significant percentage of OSCE exams having the stations sampled against blueprinting and course objectives, with contents based on contributions from other teachers and experts and where final decisions were reached through consensus meetings. If the way stations are designed cannot assure *per se* the validity of the OSCE, the evidence on a highly formal process for designing the stations shows how much some schools are investing on reaching OSCE validity through a demanding design of respective stations.

Evidence was found on validity against other criterion measures with correlations on the low to medium range. This area needs more research, namely to understand what we are correlating when doing such analyses, because we usually do not know exactly what kind of learning outcome are assessed by a 'short answer', a 'clinical grade', or any other format.

We expect that, when analysing OSCE exams performed in other phases of medical education (namely on postgraduate), we will found studies investigating construct and predictive validity which were not so frequent in undergraduate medical studies and when available do not report enough quality data to allow further analysis.

- *Evidence on OSCE fairness*

Results on OSCE fairness were based on feedback received from students' teachers and examiners, who reported the OSCE as a fair exam even when compared with traditional exams or other exams. Students express their confidence because 'the exam corresponds to what they expected and feel prepared to perform the OSCE'.

- *Evidence on OSCE acceptability*

Evidence on acceptability is based on OSCE relevance and satisfaction with the OSCE. The OSCE was reported by students, teachers and examiners as a relevant, effective exam, capable of measuring the skills the students are expected to have acquired.

In what concerns satisfaction, feedback was received from students, teachers, examiners, SPs and real patients who globally welcomed the OSCE, considered as a *'gold standard exam'* by Norman already in 2002, with favourable comments namely when compared with other formats.

- *Evidence on OSCE impact on learning and teaching*

Evidence also points to the OSCE being capable of educational impact, which is one of the most important findings of this BEMER. OSCE appears as capable of driving teaching and, above all, driving learning, which makes the OSCE not only as an '*assessment OF learning'* but also as an '*assessment FOR learning'* (Lorna & Katz 2006; Schuwirth & van der Vlauten 2011) and *'assessment AS learning'* (Lorna & Katz 2006).

Before concluding this short overview it is important to mention that less favourable comments on the OSCE fairness and acceptability were found only in a very few studies. Of interest is to highlight that

they appear in schools where other students rate the OSCE very favourably, so they never represent the global opinion of candidates performing a certain OSCE but just as the opinion of a small cohort.

Critical appraisal of the BEMER

Every systematic review must critically examine the quality of obtained evidence (Guyatt et al. 2008) which is done through examining the quality of the main methodological steps.

For this purpose results are presented below in terms of pertinence of the research question, rigour of localization and selection of the evidence, critical appraisal of the literature, integration of findings with educational judgment and transfer of results into practice.

- *How pertinent was the educational research question?*

Despite the extraordinary expansion of OSCE, many questions persist in teachers' daily practice in terms of OSCE being a valid, reliable and feasible approach to assess learning outcomes. At a time when it is well accepted that teacher's decisions should be informed by the best available evidence, the research question in this systematic review appears to be pertinent in looking for evidence on the OSCE assessment criteria for undergraduate medical studies.

Consisting of multiple objective 'stations' designed to assess a range of clinical and practical skills, under similar circumstances (same assessment, same patients, and same examiners), the OSCE was used all over the world and was immediately considered as a revolutionary exam - with students requested to demonstrate skills and not only knowledge.

Therefore, the decision for looking comprehensively at the OSCE feasibility, reliability, validity, fairness, acceptability and educational impact was justified by the existent lack of systematic reviews to encompass OSCE exponential use after its introduction. The arguments behind OSCE creation – namely its psychometrics qualities - have not been investigated in a systematic way until November 2011, when a first review on OSCE reliability was published (Brannick et al. 2011).

This is why the educational research question seemed to be highly pertinent, since using the OSCE all over the world during more than three decades does not guarantee it to be a feasible, reliable, valid, fair and acceptable exam with educational impact when assessing clinical competences.

The research question was even more pertinent if we consider that assessment is a topic of high importance on the medical education agenda, and within assessment the 'assessment of clinical skills' is probably the most challenging educational area with teachers, students and researchers involved in intense debates all over the world concerning what is known as a 'community of assessors'.

- *How exhaustive was the process of localizing and selecting the evidence?*

Literature was searched since 1975 (date OSCE was created) until the end of 2008 (see criteria for inclusion in Chapter 4) through a very exigent search process described in Chapter 1 (see methodology). Attention to cross referencing when coding the studies confirmed the search process was trustful to obtain exhaustive and sensitive data.

- *How rigorous was the critical appraisal of the evidence?*

The quality control of the evidence obtained in this systematic review implied three steps: a) examining the quality of the BEME as a credible approach to get evidence; b) examining the quality of the primary studies; c) examining the quality of the process to extract data from primary studies.

a) Examining the quality of BEME as a credible approach to obtain evidence

The quality of the BEME methodology as a credible process to obtain evidence was examined, to

see if BEME would be a trustful means to produce evidence on the psychometric characteristics of the OSCE.

The conclusion (reported in Chapter 2) was that differences between BEME and Cochrane reviews are perhaps more a matter of degree of the supporting evidence than the existence of fundamental differences. BEME and Cochrane reviews will remain, in the foreseeable future, very demanding tasks and the medical education community is aware of difficulties regarding BEME reviews - common to all human sciences – due to the holistic nature of the object of the study, in addition to the lack of resources associated with the expected financial constraints.

In a time when it is already accepted that teacher's decisions should be informed by the best available evidence, and not only by individual opinions, the crux of the question is that BEME evidence should translate scientific knowledge into practice. This was achieved in this BEMER namely through the practice points mentioned in chapters reporting on results (Chapters 4 and 5).

b) Examining the quality of primary studies in which the evidence is based

The quality of primary studies was also investigated, because they determine the quality of the obtained evidence namely in terms of its applicability. Major problems identified in primary studies were reported - lack of information, heterogeneity when reporting data, lack of standardized vocabulary and weak structure within reports - suggesting a significant concern for readers (if they wish to transfer the results to their daily practice), for researchers (when interpreting or replicating a study) and also for reviewers (when conducting a systematic review).

These problems, fully described in Chapter 3, prevented us from using all the studies selected to answer the research questions. Nevertheless, we believe that these difficulties do not compromise the overall quality of the findings. In chapters reporting the results of this BEMER (Chapters 4 and 5) information was provided in terms of missing or non-usable data.

In the future, attention needs to be paid to the way OSCEs are reported in the literature concerning the problems above referred. In Chapter 3 a checklist was proposed to assist authors in the preparation of OSCE reports. As already mentioned, it is important to have high quality of primary studies reporting more detailed information namely on OSCE economic viability.

c) Examining the quality of data analysis to extract the evidence

The process of extracting data was performed by two independent coders. There were three levels of analysis when scrutinising papers (full description given in Chapter 4).

A preliminary level was based in all retrieved studies (1065) to identify:

- 'When and where were the studies published' (date, country and continent)
- 'Who published the studies' (name and type of institution)
- 'Who used the OSCE '(phase of education and professional groups performing the OSCE).

A second level of analysis was performed in all accepted studies (n=366) to collect evidence on the purpose of the OSCE exam.

Finally, a third level was based in all studies with data on BEME questions, (n=212 studies representing 263 exams) which were scrutinized in terms of:

- Learning outcomes assessed by the OSCE (history taking, physical examination, etc.
- Subject areas under assessment (medicine, dentistry etc.)
- Underlying elements regarding OSCE design/format:
    - Type of exams (high stakes *vs.* non high stakes)
    - Purpose of exam (formative *vs.* summative)
    - Feedback provided (to whom, by whom, when and how)
    - Number of students (total number and maximum number per circuit)
    - Number of venues

- Number of parallel OSCEs
- Number of cycles
- Number of days
- Number of stations
- Total time
- Duration of individual stations
- Scoring tools (checklists *vs.* global ratings)
- Number and background of examiners
- Number and background of standardized patients (SPs)
- Number of real patients
- Number and background of staff involved
- Use of mannequins and videos
- Training process for real patients, SPs and examiners
- Existence of a pilot study
- Data on research questions (feasibility, reliability and validity).

The level of agreement for evidence on OSCE feasibility, reliability and validity between coders (reported in Chapter 4 and 5) attests to the reliability of the coding process.

Having examined all process through these steps – examining BEME as an approach to get evidence, examining the primary studies in which evidence is based and examining the way data was extracted from studies - the obtained evidence appears to be trustworthy.

- *How trustworthy was the integration of findings with educational judgment?*

For a high quality BEMER, integration of results should be made with careful judgement. The details of this integration were given in Chapters 4 and 5 and one example is that further research questions were added to the initial question. In fact we started this BEMER to answer the question **'**Is the OSCE a feasible, reliable and valid exam to assess learning outcomes in undergraduate medical studies' but at a later phase - according to educational trends regarding assessment criteria - other requisites were incorporated and a decision was taken to also examine 'OSCE fairness, acceptability (relevance and satisfaction with OSCE) and educational impact (steering effect on learning and teaching)'.

- *How comprehensive was the transfer of evidence to induce changes in practice?*

It is not enough for a BEME of high quality to answer a pertinent question, to perform an exhaustive and sensitive search, to implement a rigorous method of extracting the evidence, and to have a good integration with educational judgment. As previously mentioned, the crux of the BEMER is to translate evidence into practice and this was why particular attention was paid to practice points regarding the findings of this work (full description given in Chapters 4 and 5).

The working team tried to posit a hierarchy of evidence to guide educational decisions on the OSCE assessment criteria, being aware that evidence alone is never sufficient to make an educational decision.

Final considerations on BEMER

As Aristotle stated in Metaphysica 'the whole is greater than the sum of its parts' and this was why a global critical evaluation of this systematic review was performed at the end of the process to check 'BEMER validity' (i.e. its closeness to the truth), 'BEMER impact' (i.e. the size of its effects), 'BEMER reliability' (i.e. the precision of the findings) and 'BEMER applicability' (i.e. the potential for improving outcomes). This was performed using the guides presented on the table below.

Table 1. Overall quality control of the BEMER process

Validity of BEME process *
Was the BEME study well designed?
Was the BEME methodology well applied?
Was data available to support evidence on OSCE feasibility
Was data available reported to support evidence on OSCE reliability
Was data available reported to support evidence on OSCE validity
Was data available reported to support evidence on OSCE fairness
Was data available reported to support evidence on OSCE acceptability
Was data available reported to support evidence on OSCE drive learning
Was data available reported to support evidence on OSCE drive teaching
Was the analysis of primary studies made by independent coders
Was the analysis made by coders based on identical criteria
Was consensus reached through discussion between coders when needed
Reliability of BEME results *
Which was the precision of findings
Importance of BEME results
Which are the findings?
Which was the dimension of findings?
Were the benefits educationally significant?
Applicability of BEME results *
Are the OSCE exams under analysis similar to OSCES exams in current practice?
Are all educational important outcomes included in the final report?
Are the benefits bigger than the drawbacks?
Will BEME results impact in terms of changing daily practice ?
Will OSCE as a clinical assessment method be better as a result of this BEME

*It is important to highlight that here we are talking about the validity, reliability and applicability of BEME findings and not scrutinizing the OSCE reliability, validity and feasibility.

Results point to this BEMER as a credible method to search evidence, where primary studies showed enough quality to provide evidence, where a valid and reliable analysis was used to extract data and where evidence points to important findings which can be transferred into practice.

Limitations of the study

We acknowledge intrinsic and external limitations to this study. Poor quality of OSCE reports, with unclear or missing data, and lack of human resources, were major limitations.

In terms of problems with data, several analyses which were initially foreseen, for example 'the influence on OSCE reliability of training patients and training evaluators', were not made, due to missing or non-usable data. The same occurred in terms of reliability associated with different types of learning outcomes assessed (for example history taking, physical examination, diagnosis, management, etc.), number of venues, number of exams within the same OSCE or existence of a pilot, just to mention some of them. Since information on reliability of individual stations was insufficient to

perform such analyses. We also expected to have collected evidence on construct and predictive validity but this information was almost unavailable (and when available has not enough quality) on the studies analysed in the context of this BEMER.

Practical constraints determined that only the papers published in English were included in the analysis, because there were no resources for translating other studies.

As previously reported, the problem of finding coders in systematic reviews was a serious one, since coding of primary studies is very demanding in terms of time and resources, and coders' profile is also a very limiting factor.

The difficulties regarding reports of primary studies were mentioned in previous chapters (Chapter 4 and 5) but a final a word of caution is needed in terms of comprehensiveness of findings, because not only have the studies a tendency to report successful achievements, but also, journals have a tendency to publish mostly positive studies.

We are aware of this possible bias and this is why we dare suggest that criteria for accepting/rejecting studies submitted to publication also take into account the 'educational lessons to be taken from the study' and not only on the 'success of the study'. Important educational lessons can be learnt from a non-successful study.

The work done so far is the first step of a more comprehensive BEMER which should look at the 293 studies published after the end of 2008, at OSCEs performed in other phases of medical education (postgraduate and CME), and OSCES used in some of the other 25 health related professions identified in this systematic review.

This further analysis is extremely important to complete the evidence on the OSCE assessment criteria.

OSCE: The way forward

The evidence obtained with this BEMER points to the OSCE as a '*gold standard for clinical assessment'* as stated by Norman in 2002. Emerging from our results are some OSCE features which may justify Norman's statement. From them, it appears that OSCE has a more important role to play in the future, namely in terms of:

- *OSCE as an 'authentic exam'*

Evidence points to the OSCE as an 'authentic exam' and this is why its inception was a really 'new take' in terms of clinical skills assessment. OSCE is *per se* a valid exam because it requests the 'demonstration of students' competencies' and not only the 'demonstration of students' knowledge'. It certifies the clinical skills the student must acquire and this is of utmost importance, namely in the context of Outcome Based Education where clear learning outcomes are defined and where OSCE may play a major role assessing a broad range of competencies. This is essential, at a time where more accountability is needed from medical schools to reassure society that doctors have the necessary skills for practice. (Harden 1999).

- *OSCE as 'assessment for learning', not just 'assessment of learning'*

Evidence points to the OSCE as an 'assessment for learning' and not just an 'assessment of learning' since it tackles the full roles of assessment, namely in terms of having a steering effect on learning. Teachers may have different perceptions and different expectations regarding the educational impact of assessment, which can be incorporated in formative OSCEs, given at different stages (during the OSCE, immediately after and sometime later) encouraging and creating diverse opportunities to provide feedback to students (usually highly valued by them). This would respond to students' long-standing complaints about not receiving enough feedback while being assessed and would allow them to identify gaps and mistakes they made, showing what they should do to remedy their deficiencies.

- *OSCE as an 'exam with wide applicability'*

Evidence points to OSCE not just addressing a small niche of teachers' requests but meeting the needs of a large community of different professional bodies who are using it all over the world, with no apparent geographical limitations, for multiple purposes (formative and summative), to certify a range of multiple competencies, which is extremely relevant in a time when curricula and assessment are moving to 'Competency Based Education' and 'Competency Based Assessment' (Schumway & Harden 2003).

- *OSCE as a 'response to the continuum of education'*

Evidence points to OSCE being used at all levels of education (undergraduate, postgraduate and continuing professional development (CPD). We will witness in the future the design of more OSCE stations recognizing the *continuum* of education, since what a first year student is expected to do and achieve is very different from a final year student, a postgraduate or a specialist.

- *OSCE as a 'feasibility exam used in a wide range of contexts'*

Evidence points to the OSCE implemented in a wide range of contexts (community, hospital, medical schools), with-high or low-stakes exams, including high and low number of students, and allowing multiple formats (in terms of number of venues, circuits, type of stations, duration, number and background of examiners, use of real or standardized patients).

- *OSCE as a 'flexible exam contributing to teachers ownership'*

Evidence points to the OSCE having great flexibility. It is impressive how OSCE, being a highly standardized exam (same tasks, same examiners, same patients and same duration), allows teachers to adapt to the format they wish, with almost no limitations in terms of its design to fit their needs. This is probably why teachers feel as 'major stakeholders' when implementing an OSCE.

- *OCSE as an 'exam capable of assessing professionalization in a global world'*

The other important thing is that in the future we will be moving to professionalization, and the OSCE exam - designed to reflect 'team work' and 'different professions working together' - will be a means to that end.

We will assist to OSCE exams incorporating 'standard stations' to assess and recognize global standards (as specified by the 'Global Essential Minimum Requirements'), which will contribute to certification of the competencies needed by a doctor in a global world with its mobility that students and professionals have to face. (Core Committee Institute for International Medical Education 2002; Schwarz & Wojtczak 2002)

- *OCSE as an 'exam where new technologies will bring new improvements'*

Presently the OSCE patients can be represented in different ways (real, simulated, models and computer simulations) and, in spite of the undeniable importance of real patients, we have to acknowledge the future impact of new technologies and virtual reality will have on OSCE patient representation, as well as on recording and scoring students' performance.

- *OCSE as an 'exam where students will be used as examiners'*

In an educational environment where curriculum is a major topic in the medical education agenda, more attention is likely to be paid in the future to the role of students in this exam - not just as 'learners' or 'examinees' but also as 'examiners' with peer assessment being part of the procedure.

- *OSCE as an 'overall assessment'*

By its structure and format the OSCE, when assessing multiple competencies in a single exam, is contributing to students' overall examination (van der Vleuten et al. 2012), namely when providing relevant data related to different learning outcomes.

- *OSCE as an 'exam contributing to research in medical education'*

As research in medical education is also very much in the agenda, as well as the need for evidence to inform decisions, the OSCE can be used to assess not only 'students' competencies' but also 'interventions in the curriculum'. We anticipate the emphasis on research will even become more important in the future, and therefore the role of OSCE will also be emphasized in terms of nearcoming contributions to medical education research.

These perspectives when 'looking forward to the OSCE exam' allow us to anticipate its future potential role and impact in medical education. However, it is important to clarify that the OSCE also has its limitations, as previously reported, and should never be considered as the 'only exam'.

The evidence brought up by this BEMER on OSCE feasibility, reliability, validity, fairness, acceptability and educational impact justifies the OSCE being considered a **'6 Star Exam'**.

References

Brannick MT, Erol-Korkmaz HT, Prewett M. 2011. A systematic review of the reliability of objective structured clinical examination scores. Medical Education 45: 1181-1189.

Core Committee Institute for International Medical Education. 2002. Global minimum essential requirements in medical education Institute for International. Medical Education, Medical Teacher 24(2) 130-135.

Guyatt G, Drummond R, Maureen M, Cook D. 2008. User's Guides to the Medical Literature. 2nd ed. New York: McGraw-Hill.

Harden RM, Crosby JR, Davis MH, Friedman M. 1999. AMEE Guide No. 14: Outcome-based education: From competency to meta-competency: a model for the specification of learning outcomes. Medical Teacher 21(6).

Lorna E, Katz S. 2006. Rethinking classroom assessment with purpose in mind: assessment for learning, assessment as learning, assessment of learning. Manitoba Education, Citizenship and Youth. Retrived March 10, 2011, Available at http://www.edu.gov.mb.ca/ks4/assess/index.html>

Norman G. 2002. Research in medical education: three decades of progress BMJ 29:324:1560-2.

Schwarz RM, Wojtczak A. 2002. Global minimum essential requirements: a road towards competence-oriented medical education. Medical Teacher 24(2) 125-129.

Schuwirth LW, van der Vleuten CP. 2011. Programmatic assessment: From assessment of learning to assessment for learning. Medical Teacher 33(6):478-85.

van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. 2012. A model for programmatic assessment fit for purpose. Medical Teacher 34(3), 205-214.

ANNEXES

Chapter 4
Appendix 1 – BEME methodology

BEME methodology - as described in the BEME protocol (www.bemecollaboration.org) was undertaken based on the following steps:

(1) Establishment of a working Systematic Review Group
(2) Framing the research question(s)
(3) Defining inclusion/exclusion criteria
(4) Developing a search strategy
(5) Retrieving the material under analysis
(6) Developing an OSCE Reference Manager database
(7) Developing an OSCE electronic database
(8) Coders' training and pilot phase
(9) Analysing and coding of primary studies
(10) Establishment of consensus
(11) Analysing data
(12) Discussion and synthesis
(13) Conclusions and application to practice

1. Establishment of a working Systematic Review Group

A working group in Lisbon was constituted by the coordinator (a MD-PhD and MSc in Medical Education), a research director (Educationalist and MSc in Medical Education - also acting as a coder), two coders (final-year medical students) and two administrative assistants.

Included in the team there was two international consultants: a PhD from the Department of Clinical Epidemiology and Biostatistics at McMaster University, Canada, and a PhD from the Centre for Medical Education at McGill University, Canada.

2. Framing the research question(s)

Whether the OSCE is feasible, reliable and valid as a method of assessment of learning outcomes in undergraduate medical studies were the initial research questions for this BEME Systematic Review. Later, due to current educational developments, other questions were added concerning new assessment criteria requirements, introduced by the GMC in 2011: fairness, acceptability (i.e. OSCE relevance and satisfaction with the OSCE) and educational impact (i.e. OSCE capacity of steering learning and teaching).

3. Defining inclusion/exclusion criteria

Only English studies reporting on 'classical OSCE exams' performed in undergraduate medical education were included in the study. Therefore the following primary studies were excluded:

- Non undergraduate studies
- Non medical studies
- Non English studies
- Non 'classical' OSCEs
- Non primary studies

- OSCE studies for teaching students
- OSCE studies for training teachers.

A study was coded 'non-classical' when it did not conform in general terms with the classical approach of the OSCE, as described by Harden et al. in 1975. Among them we found studies where the candidate was a 'team' or 'group' instead of an individual - for example, TOSCE (Singleton et al. 1999), G-OSCE (Hill et al. 1994), GOSCE (Elliot et al. 1994; Fields et al. 1995; Vooijs et al. 1997), GOSPE (Biran 1991), when assessment was based on video instead of direct observation, for example VIPSCE (Shallaly & Ali 2004), OSVE (Humphris & Kaney 2000), where exams had only written stations (Akici et al. 2004), where exams were only peer-rated (Geddes & Crowe 1998) and 'non-classical' formats, for example OSCEs with only one or two stations (van Dalen et al. 2001; Robins et al. 2001). 'non-classical' studies were excluded, to avoid a bias when calculating the reliability of the OSCE.

Secondary studies were excluded because systematic reviews must be based on primary studies only. Also excluded were studies where OSCEs' objectives were 'to teach students or train teachers', because the objective of this BEMER concerns the feasibility, reliability and validity of the OSCE when implemented for assessing learning outcomes.

Long case exams and OSLERs which appear as a result of the search were also excluded from analysis.

4. Developing a search strategy

Literature was searched by a BEME information scientist, from 1975 (date of the first publication on the OSCE) until the end of 2008. All identified references were inserted into a Reference Manager database.

Two different phases were considered:

- Phase I - Literature search from 1975 to end of 2001
  We started with the OSCE database material published by Harden et al. in 2003, which covered the OSCE literature from 1975 till the end of 2001. The 712 references were identified through:
  - Electronic search of medical, educational & related databases;
  - Hand search on 6 key medical education journals;
  - Search of TIMELIT reference database;
  - Search of Gray Literature (for example the Proceedings of Ottawa Conference);
  - Search on specialised literature collections, at the Medical Education Centre, University of Dundee.

  The key journals selected for searching were: Academic Medicine, Medical Education, Medical Teacher, Teaching & Learning in Medicine, Advances in Health Sciences Education and Education for Health.

- Phase II - Literature Search from 2002 to the end of 2008
  Previous search was updated until the end of 2008 by the same BEME information scientist who made the initial search. These references were electronically identified (electronic searches have improved considerably since the initial run and there was no need to repeat the intensive hand searches labour) and TIMELIT (used in phase I) was abandoned later.
  The key words used in both phases were base terms which were tested and adapted: 'OSCE', 'OSPE', 'GOSPE', 'objective structured clinical exam$', 'objective structured practical exam$', 'structured clinical exam$', 'structured clinical interview$' (the truncation symbolic $ is fairly generic and is used to pick up all alternative endings).

5. Retrieving the material under analysis

When a reference was identified the process of retrieving the paper started immediately. This was easier in Phase I because the papers were sent by the Medical Education Centre at the University of Dundee. For Phase II the process was more difficult: the majority was obtained through the libraries of the University of Lisbon and the University of Columbia. Finally, the editors of journals of non-retrieved papers were also approached to obtain missing papers.

6. Developing an OSCE Reference Manager database

An OSCE Reference Manager database was created to include the list of identified references. The objective was to facilitate a quick identification of a study through its author(s), date, journal, title, etc., and the insertion of references in publications.

The software Reference Manager is one of the most reliable databases management programs available to the academic world and has the advantage to be compatible with most bibliographic databases. All elements of the working team were trained in its development and use.

7. Developing an OSCE electronic database

A new 'online database' (Lotus software) was developed, since the existing BEME coding sheet was not applicable in our systematic review. Items for the new coding sheet were defined by the whole team according to research questions. Literature was blueprinted and the new database served as a coding sheet supporting the coding and establishment of consensus on line. The database was structured upon four main sections: 1) information on publication, 2) background of OSCE exam, 3) results on OSCE feasibility, reliability and validity, and 4) study problems, solutions and conclusions, each of them including several fields. Full description of this software is presented in Chapter 4.

One of the most important characteristics of the new OSCE database was its dynamic structure i.e. a structure, which could be modified during the coding process by adding or reformulating a field. Moreover, the majority of the fields were 'open fields' which could be fed by new items when they show up during the coding process. These features were extremely important since when a systematic review starts its impossible for researchers to have the full picture of what is under investigation.

Fields such as 'existence of a previous pilot', 'number of sub-stations under analysis', 'sub-sample of students under analysis', 'total number of students in the course', 'relevance', 'fairness', are examples of fields inserted later during the coding process. They implied a second review of the papers analysed until that date. In what concerns the number of options within the same field, we found, for instance, 266 different types of OSCE aims, 273 Institutions responsible for OSCE publications and 45 different stations organized in 156 different combinations (depending on the studies).

These are just some examples showing that it would be extremely difficult to code such a complex exam with the traditional paper based coding sheet, as the fields of the coding sheet are defined before the coding process starts.

This new database allowed independent coding by each coder and establishing of consensus online.

8. Coders' training and pilot phase

Two coders were trained by the research director in the BEMER systematic methodology. Background literature was made available to them and several meetings took place to discuss the process of coding and how to reach consensus.

After the initial 'theoretical training period', a 'pilot phase' started concerning the coding of the first 75 papers. During the 'pilot phase' each paper was reviewed and inserted in the database by each independent coder. A discussion followed to justify each decision, before consensus was established leading to the coding of a 'consensus record'.

9. Analysing and coding of primary studies

Each paper was coded by two independent reviewers. As already stated, the new electronic OSCE database was used to support coding.

10. Establishment of consensus

A 'consensus meeting' occurred after independent coding of a certain number of records (usually no more than twenty) was made. Consensus was reached by comparing the classification of the two coders and as a result of the discussion a 'consensus record' was created for each record in the online OSCE database. Consensus could be achieved face to face or electronically.

Disagreements were discussed until consensus was achieved. When disagreements occurred, a discussion took place and, frequently, this implied the coders had to reread the paper before consensus could be established at the next session. During the pilot phase the establishment of consensus for a single paper could easily take more than one hour. Progressively the time allocated to establish consensus diminished and within an hour it was possible to code 4-6 papers depending on its complexity.

Consistency among coders was established throughout the process excluding the pilot phase (see results in Chapters 3, 4 and 5).

11. Analysing data

The procedures for data analysis were determined by each research question. For detailed information on the different levels of analysis see Chapters 3, 4, and 5. The analysis was supported by the Department of Clinical Epidemiology and Biostatistics at McMaster University and the Centre for Medical Education & Department of Medicine, McGill University, both in Canada

12. Discussion and Synthesis

The discussion and synthesis of results was done in the context of each research question, involving the whole team. Results are presented in Chapters 3, 4, and 5.

13. Conclusions and application to practice

Conclusions were established in order to facilitate the transfer of results into practice when taking into account the limitations of the study.

Chapter 4
Appendix 2 – References of OSCE papers regarding this BEMER

A-Latif A. An examination of the examinations: the reliability of the objective structured clinical examination and clinical examination. Medical Teacher 1992;14(2/3):179-83.

Abe S, Kawada E. Development of computer-based OSCE re-examination system for minimizing inter-examiner discrepancy. Bulletin of Tokyo Dental College 2008;49(1):1-6.

Abo-Bakr AM. What makes OSCE a good adjunct to the conventional methods of assessing clinical competence? Annals of Saudi Medicine 1991;11(1):87-9.

Abouna G, Hamdy H. The Integrated Objective Direct Observation Clinical Encounter Examination: a Reliable Method of Assessing Students Clinical Competence in Problem-Based Learning Curriculum. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.

Aboutanos M, Rodas E, Aboutanos S, Mora F, Wolfe L, Duane TMA. Trauma education and care in the jungle of Ecuador, where there is no advanced trauma life support. Journal of Trauma, Injury, Infection and Critical Care 2007;62(3):714-9.

Adeyemi-Doro HO, Bamisaiye A. A new course in basic therapeutic skills for medical and dental students at the College of Medicine, University of Lagos: student evaluation. Medical Education 1983;17(6):354-9.

Adeyemi EO. The performance profile of medical students in the mock objective structured clinical examination. Saudi Medical Journal 2001;22(6):504-7.

Adeyemi SD, Omo-Dare P, Rao CR. A comparative study of the traditional long case with the objective structured clinical examination in Lagos, Nigeria. Medical Education 1984;18(1):106-9.

Adome R, Kitutu F. Creating an OSCE/OSPE in a resource-limited setting. Medical Education 2008;42:525-6.

Aeder L, Altshuler L, Kachur E, Barrett S, Hilfer A, Koepfer S, Schaeffer H, Shelov SP. The 'Culture OSCE' - Introducing a formative assessment into a postgraduate program. Education for Health 2007;20(1):1-11.

Afroza A. Use of a PMP manual as a teaching tool to accelerate paediatric teaching in Bangladesh. Medical Teacher 2000;22(4):365-9.

Ainsworth MA, Turner HE, Solomon DJ. Innovation in conducting and scoring a clerkship objective structured clinical examination. Teaching and Learning in Medicine 1994;6(1):64-7.

Ainsworth MA, Callaway MR, Perkowski L. An OSCE assessment of fourth-year students. Academic Medicine 1995;70(5):444-5.

Aithala R. Objective structured clinical examination in pediatrics. Archives of Pediatric and Adolescent Medicine 1998;152(7):715.

Akici A, Kalaca S, Goren MZ, Akkan AG, Karaalp A, Demir D, Ugurlu U, Oktay S. Comparison of rational pharmacotherapy decision-making competence of general practitioners with intern doctors. European Journal of Clinical Pharmacology 2004;60(2):75-82.

Ali J, Cohen R, Reznick R. Demonstration of acquisition of trauma management skills by senior medical students completing the ATLS program. Journal of Trauma, Injury, Infection and Critical Care 1995;38(5):687-91.

Ali J, Cohen R, Adam R, Gana TJ, Pierre I, Ali E, Bedaysie H, West U, Winn J. Attrition of cognitive and trauma management skills after the advanced trauma life support (ATLS) course. Journal of Trauma, Injury, Infection and Critical Care 1996;40(6):860-6.

Ali J, Cohen R, Adam R, Gana TJ, Pierre I, Bedaysie H, Ali E, West U, Winn J. Teaching effectiveness of the advanced trauma life support program as demonstrated by an Objective Structured Clinical Examination for practicing physicians. World Journal of Surgery 1996;20(8):1121-5.

Ali J, Adam R, Josa D, Pierre I, Bedaysie H, West U, Winn J, Haynes B. Comparison of performance of interns completing the old (1993) and new interactive (1997) advanced trauma life support courses. Journal of Trauma, Injury, Infection and Critical Care 1999;46(1):80-6.

Ali J, Adam R, Pierre I, Bedaysie H, Josa D, Winn J. Comparison of performance 2 years after the old and new (interactive) ATLS courses. Journal of Surgical Research 2001; 97(1):71-5.

Alinier G, Hunt WB, Gordon R. Determining the value of simulation in nurse education: study design and initial results. Nurse Education in Practice 2004;4(3):200-7.

Allen R, Heard J, Savidge M. Global ratings versus checklist scoring in an OSCE. Academic Medicine 1998;73(5):597-8.

Allen R, Heard J, Savidge M, Bittengle J, Cantrell M, Huffmaster T. Surveying students' attitudes during the OSCE. Advances in Health Sciences Education 1998;3(3):197-206.

Allen SS, Bland CJ, Harris IB, Andreson D, Poland G, Satran L, Miller W. Structured clinical teaching strategy. Medical Teacher 1991;13(2):177-84.

Alnasir FA. The watched structure clinical examination (WASCE) as a tool of assessment. Saudi Medical Journal 2004;25(1):71-4.

Altshuler L., Kachur E, Krinshpun S, Sullivan D. Genetics Objective Structured Clinical Exams at the Maimonides Infants & Children's Hospital of Brooklyn, New York. Academic Medicine 2008;83(11):1088-93.

Altshuler L, Kachur E. A culture OSCE: teaching residents to bridge different worlds. Academic Medicine 2001;76(5):514.

Amano H, Sano T, Gotoh K, et al. Strategies for training standardized patient instructors for a competency exam. Journal of Dental Education 2004;68(10):1104-11.

Amiel GE, Tann M, Krausz M, Bitterman A, Cohen R. A Combined Structured Oral Examination (SOE) and OSCE for the Assessment of Clinical Competence. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;402-3.

Amiel GE, Tann M, Krausz MM, Bitterman A, Cohen R. Increasing examiner involvement in an Objective Structured Clinical Examination by integrating a structured oral examination. American Journal of Surgery 1997;173(6):546-9.

Amiel GE, Ungar L, Alperin M. Using an OSCE to assess primary care physicians' competence in breaking bad news. Academic Medicine 2000;75(5):560-3.

Amiel GE, Ungar L, Alperin M, Baharier Z, Cohen R, Reis S.

Ability of primary care physician's to break bad news: A performance based assessment of an education intervention. Patient Education and Counseling 2006;60(1):10-5.

Ananthakrishnan N. Objective Structured Clinical/Practical Examination (OSCE/OSPE). Journal of Postgraduate Medicine 1993;39(2):82-4.

Anastakis DJ, Cohen R, Reznick RK. The structural oral examination as a method for assessing surgical residents. American Journal of Surgery 1991;162(7):67-70.

Anderson DC, Harris IB, Allen S, Satran L, Bland CJ, Davis-Feickert JA, Poland GA, Miller WJ. Comparing students' feedback about clinical instruction with their performances. Academic Medicine 1991;66(1):29-34.

Anderson M, Stickley T. Finding reality: the use of objective structured clinical examination (OSCE) in the assessment of mental health nursing students interpersonal skills. Nurse Education in Practice 2002;2(3):160-8.

Anderson MB, Stillman PL. Growing use of standardized patients in teaching and evaluation in medical education. Teaching and Learning in Medicine 1994;6(1):15-22.

Arnold RC, Walmsley AD. The use of the OSCE in postgraduate education. European Journal of Dental Education 2008;12(3):126-30.

Aso R, Yoshimura A, Shimura T, Takayanagi K, Iino Y, Kobayashi Y, Seino Y, Hidaka H. Special Training Course for Simulated Patients Who Participated in the Advanced OSCE at Nippon Medical School. Journal of Nippon Medical School 2008;75(1):46-7.

Auewarakul C, Downing SM, Jaturatamrong U, Praditsuwan R. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. Medical Education 2005;39(3):276-83.

Auewarakul C, Downing SM, Praditsuwan R, Jaturatamrong U. Item analysis to improve reliability for an internal medicine undergraduate OSCE. Advances in Health Sciences Education 2005;10(2):105-13.

Austin Z, Galli M, Diamantouros A. Development of a prior learning assessment for pharmacists seeking licensure in Canada. Pharmacy Education 2003;3(2):87-96.

Austin Z, Marini A, Croteau D, Violato C. Assessment of pharmacist's patient care competencies: Validity evidence from Ontario (Canada)'s quality assurance and Peer Review process. Pharmacy Education 2004;4(1):23-32.

Austin Z, Gregory PAM, Galli M. "I just don't know what I'm supposed to know": evaluating self-assessment skills of international pharmacy graduates in Canada. Research in Social & Administrative Pharmacy 2008;4(2):115-24.

Ayers WR, Boulet JR. Establishing the validity of test score inferences: performance of 4th year U.S. medical students on the ECFMG clinical skills assessment. Teaching and Learning in Medicine 2001;13(4):214-20.

Azcuenga J, Valls M, Martinez-Carretero JM. Cost Analysis of Three Clinical Skills Assessment (CSA) Projects. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.

Baerheim A, Malterud K. Simulated patients for the practical examination of medical students: intentions, procedures and experiences. Medical Education 1995;29(6):410-3.

Balslev T, Ostergaard JR, Nodgaard H, Hasle H, Schiotz PO. The objective structured clinical examination postgraduate pediatric training: the first Danish experiences. Ugeskrift for Laeger 2000;162(10):1383-7.

Ban N, Hatao M, Ohtaki J, Saito T, Matsueda K, Izumi Y, Kusumoto M, Hosoda S. An OSCE Trial in Japan: Feasibility and Correlation with a Written Test. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;404-6.

Ban N, Tsuda T, Hatao M, Ohtaki J, Fujisaki M, Nakamura C. Detailed Manuals can Accomplish Acceptable Inter-rater Agreement in an OSCE Assessing Basic Clinical Competence. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-2.

Bansal P, Saoji V, Gruppen LD. From a "generalist" medical graduate to a "specialty" resident: Can an entry-level assessment facilitate the transition? Assessing the preparedness level of new surgical trainees. Annals of the Academy of Medicine, Singapore 2007;36(9):719-24.

Barclay DM, McKinley D, Peitzman SJ, Burdick B, Curtis M, Whelan GP. Effect of training location on student's clinical skills. Academic Medicine 2001;76(4):384.

Barlow JW, Strawbridge JD. Teaching and assessment of an innovative and integrated pharmacy undergraduate module. Pharmacy Education 2007;7(2):193-5.

Barman A. Critiques on the objective structured clinical examination. Annals of the Academy of Medicine, Singapore 2005;34(8):478-82.

Barnhart AJ, Marcy ML, Colliver JA, Verhulst SJ. A comparison of second-and fourth-year medical students on a standardized-patient examination of clinical competence: a construct validity study. Teaching and Learning in Medicine 1995;7(3):168-71.

Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. Academic Medicine 1993;68(6):443-51.

Bartos R, Richards A. Effect of varying amounts of feedback on standardized patient checklist accuracy in clinical practice examinations. Teaching and Learning in Medicine 1999;11(3):148-52.

Barzansky B, Etzel SI. Educational programs in US medical schools, 2003-2004. Journal of the American Medical Association 2004;292(9):1025-31.

Basco WT, Gilbert GE, Chessman AW, Blue AV. The ability of a medical school admission process to predict clinical performance and patients' satisfaction. Academic Medicine 2000;75(7):743-7.

Bateman K, Menzies P, Sandals D, Duffield T, Leblanc S, Leslie K, Lissemore K, Swackhammer R. Objective Structured Clinical Examinations (OSCEs) as a Summative Evaluation Tool in a Ruminant Health Management Rotation for Final-Year DVM Students. Journal of Veterinary Medical Education 2008;35(3):382-8.

Battles J, Carpenter JL, Foster B, McIntire DD, Sprankell SJ. Assessing Basic Clinical Skills Using a Two-Part Iterative OSCE in a Second-Year ICM Course. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.

Battles JB, Sprankell SJ, Carpenter JL, Bedford JA, Kirk LM. Developing a support system for teaching and assessing clinical competence. Journal of Biocommunication 1992;19(4):19-25.

Battles JB, Carpenter JL, McIntire DD, Wagner JM. Analyzing and adjusting for variables in a large-scale standardized-patient examination. Academic Medicine 1994;69(5):370-6.

Battles JB, McIntire DD, Carpenter JL, Wagner JM. Using Algebraic Inversion of the Dependability Index to Determine Pass/Fail Criteria for Clinical Performance Examinations. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;652-5.

Battles JB, Sprankell SJ, Baker L, et al. OSCE Stations for the Longitudinal Assessment of Cancer Screening and Detection. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;407-9.

Battles JB, Wilkinson SL, Lee SJ. Using standardised patients in an objective structured clinical examination as a patient safety tool. Quality and Safety in Health Care 2004;13(Suppl 1):i46-i50.

Beck DE, Boh LE, Sullivan PS. Evaluating student performance in the experiential setting with confidence. American Journal of Pharmaceutical Education 1995;59 (3):236-47.

Beeson MS, Wilkinson LF. Evaluation of clinical performances of emergency medicine residents using standardized patients. Academic Medicine 1999;74(2):202.

Bell SK, Krupat E, Fazio SV, Roberts DH, Schwartzstein RM. Longitudinal Pedagogy: A Successful Response to the Fragmentation of the Third-Year Medical Student Clerkship Experience. Academic Medicine 2008;83(5):467-675.

Benbow EW, Harrison I, Dornan TL, ONeill PA. Pathology and the OSCE: insights from a pilot study. Journal of Pathology 1998;184(1):110-4.

Benning T, Broadhurst M. The long case is dead - long live the long case. Psychiatric Bulletin 2007;31:441-2.

Berg D, Sebastian J, Heudebert G. Development, implementation, and evaluation of an advanced physical diagnosis course for senior medical students. Academic Medicine 1994; 69(9):758-64.

Bergus GR, Kreiter CD. The reliability of summative judgements based on objective structured clinical examination cases distributed across the clinical year. Medical Education 2009;41(7):661-6.

Bernardin G, Mattei M. Assessment of Hospital training sessions using an objective structured clinical examination: results of a teaching experiment at Nice University Hospital. Reanimation Urgences 1998;7(6):621.

Bianchi F, Stobbe K, Eva K. Comparing academic performance of medical students in distributed learning sites: the McMaster experience. Medical Teacher 2008;30 (1):67-71.

Bienenstock JL, Tzou WS, Martin SA, Fox HE. Effect of student ethnicity on interpersonal skills and objective standardized clinical examination scores. Obstetrics and Gynecology 2000;96(6):1011-3.

Bingold F, Beckmann MW, Frobenius W. The new regulations for the licensing of physicians (arztliche approbationsordnung, AAppO): Part I: Implementation in gynecology up to 2006. Geburtshilfe Frauenheilkd 2007;67(4):341-7.

Biran LA. Self-assessment and learning through GOSCE (group objective structured examination). Medical Education 1991;25(6):475-9.

Black NMI, Harden RM. Providing feedback to students on clinical skills by using the Objective Structured Clinical Examination. Medical Education 1986;20(1):48-52.

Blackwell TA, Ainsworth MA, Dorsey NK, Callaway MR, Rogers LP, Collins KE. A comparison of short-answer and extended-matching question scores in an objective structured clinical exam. Academic Medicine 1991;66(9 Suppl):S40-S42.

Blanch D, Hall J, Rotter D, Frankel R. Medical student gender and issues of confidence. Patient Education and Counseling 2008;72(3):374-81.

Blaskiewicz RJ, Park RS, Chibnall JT, Powell JK. The influence of testing context and clinical rotation order on students' OSCE performance. Academic Medicine 2004;79(6):597-601.

Blatt B, Plack M, Maring J, Mintz M, Simmens SJ. Acting on reflexion: the effect of reflection on students'clinical performance on a standardized patient examination. Journal of General Internal Medicine 2007;22(1):49-54.

Blay C, Hidalgo F, Juncosa S, Mart MJ, Pila C, Serrallach S, Valls M. Performance-Based Assessment for Certification of Medical Transportation Professionals in Catalonia (Spain): Results of a Pilot Administration. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.

Blay C, Cots JM, Iruela A, Juncosa S, Morera R, Ros E, Sellares J. The First OSCE of Family Medicine in Catalonia (Spain): Results of a Pilot Administration. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-3.

Bleakley A, Brice J, Bligh J. Thinking the post-colonial in medical education. Medical Education 2008;42(3):266-70.

Blue AV, Stratton TD, Plymale M, DeGnore LT, Schwartz RW, Sloan DA. The effectiveness of the structured clinical instruction module. The American Journal of Surgery 1998;176(7):67-70.

Blue AV, Griffith CH, Wilson J, Sloan DA, Schwartz RW. Surgical teaching quality makes a difference. American Journal of Surgery 1999;177(1):86-9.

Blue AV, Chessman AW, Gilbert GE, Mainous AG. Responding to Patients' Emotions: Important for Standardized Patient Satisfaction. Family Medicine 2000;32(5):326-30.

Blue AV, Chessman AW, Gilbert GE, Shuman SH, Mainous III AG. Medical students' abilities to take an occupational history: use of the WHACS mnemonic. Journal of Occupational and Environmental Medicine 2000;42(11):1050-3.

Boehlechke B, Sperber AD, Kowlowitz V, Becker M, Contreras A, McGaghie WC. Smoking history-taking skills: a simple guide to teach medical students. Medical Education 1996; 30:283-9.

Boehler ML, Schwind CJ, Folse R, Dunnington G, Markwell S, Dutta S. An evaluation of study habits of third-year medical students in a surgical clerkship. American Journal of Surgery 2001;181(3):268-71.

Bogels SM, Van Mourik TGC, Van der Vleuten C. Authentic assessment of interviewing and counseling skills: effect of testing time per station on generalizability and validity. Teaching and Learning in Medicine 1995;7(3):155-62.

Boone WJ, McWhorter AG, Seale NS. Purposeful Assessment Techniques PAT Applied to an OSCE - Based Measurement of Competencies in a Pediatric Dentistry Curriculum. Journal of Dental Education 2001;65(11):1232-7.

Borbasi SA, Koop A. The Objective Structured Clinical Examination: it's Application in Nursing Education. The Australian Journal of Advanced Nursing 1993;11(2):45-50.

Bordage G, Williams R, Beaumont E, Goldyn S. Key Clinical Findings and Explanations Given by Medical Students after three Standardized-Patient Encounters. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-5.

Boudreau D, Tamblyn R, Defresne L. Evaluation of consultative skills in respiratory medicine using a structured medical consultation. American Journal of Respiratory and Critical Care Medicine 1994;150:1298-304.

Boudreau JD, Tamblyn R, Reid T. Physician volunteerism: thank you for the joy of being honored. Advances in Health Sciences Education 1999;4:223-31.

Bouhuijs PAJ, Van der Vleuten CPM, Van Luyk SJ. The OSCE as a part of a systematic skills training approach. Medical Teacher 1987;9(2):183-91.

Boulet J, Friedman Ben-David M, Hambleton R, Burdick WP, Ziv A. Assessing the Adequacy of the Post-Encounter Written Scores in Standardized Patient Exams. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;410-2.

Boulet JR, Friedman Ben-David M, Hambleton RK, Burdick W, Ziv A, Gary NE. An investigation of the sources of measurement error in the post-encounter written scores from standardized patient examinations. Advances in Health Sciences Education 1998;3(3):89-100.

Boulet JR, Friedman Ben-David M, Ziv A, Burdick WP, Gary NE. The Use of Holistic Scoring for Postencounter Written Exercises. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.

Boulet JR, Gimpel JR, Dowling DJ, Finley M. Assessing the ability of medical students to perform osteopathic manipulative treatment techniques. Journal of the American Osteopathic

Association 2004;104(5):203-11.
Boursicot KA, Roberts T, Pell G. Using borderline methods to compare passing standards for OSCEs at graduation across three medical schools. Medical Education 2007;41:1024-31.
Bradley P. A comparison of computerised MCQ marking and simulated patient assessment of 1st year medical student history-taking skills. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Bradley P, Bradley P, Johns V. Can we teach a gentler rectal examination? Medical Teacher 1999;21(2):207-8.
Bradley P, Humphris G. Assessing the ability of medical students to apply evidence in practice: the potential of the OSCE. Medical Education 1999;33(11):815-7.
Bradshaw A. Nursing competence 10 years on: fit for practice and purpose yet? Journal of Clinical Nursing 2008;17:1263-9.
Brailovsky C, Charlin B, Beausoleil S, Cote S, Van der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. Medical Education 2001;35(5):430-6.
Brailovsky CA, GrandMaison P, Lescop J. A large-scale multicenter Objective Structured Clinical Examination for Licensure. Academic Medicine 1992;67(10 Suppl):S37-S39.
Brailovsky CA, GrandMaison P, Miller F, Rainsberry P. Detection of Gender Differences in High-Stakes Performance SP-Based Examinations in Family Medicine. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;131-4.
Brailovsky CA, GrandMaison P, Lescop J. Construct validity of the Québec licensing examination SP-based OSCE. Teaching and Learning in Medicine 1997;9(1):44-50.
Brailovsky CA, GrandMaison P, Lescop J, Dandavino A. Bridging theory and practice: A Comprehensive Psychometric Assessment of an Sp-based OSCE Licensing Examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Brailovsky CA, GrandMaison P. Using evidence to improve evaluation: a comprehensive psychometric assessment of a SP-based OSCE licensing examination. Advances in Health Sciences Education 2000;5(3):207-19.
Brewlin J, Cantwell R. Implementing the OSCE in Nottingham. Psychiatric Bulletin 1997;21(1):30-2.
Bromley LM. The Objective Structured Clinical Exam - practical aspects. Current Opinon in Anaesthesiology 2000;13(6):675-8.
Brooks D. Objective Structured Clinical Examination assessment. Nursing Times 2007;103(43):30-1.
Brosnan M, Evans W, Brosnan E, Brown G. Implementing objective structured clinical skills evaluation (OSCE) in nurse registration programmes in a centre in Ireland: A utilisation focused evaluation. Nurse Education Today 2006;26(2):115-22.
Broudo M, White M, Rodenburg D, Arseneau R, Chalmers A, Wright J, Mizgala H, Lirenman D. The Effectiveness of Interactive Multimedia as an Instructional Aid for Learning Basic Clinical Skillds and Knowledge. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;321-6.
Brown G, Manogue M, Martin M. The validity and reliability of an OSCE in dentistry. European Journal of Dental Education 1999;3(3):117-25.
Buckey S, Zamora J. Effects of participation in a cross year peer tutoring programme in clinical examination skills on volunteer tutor's skills and atitudes towards teachers and teaching. BMC Medical Education 2007;7:1-9.
Bujack L, McMillan M, Dwyer J, Hazelton M. Assessing comprehensive nursing performance: the objective structured clinical assessment (OSCA) part 2 - report of the evaluation project. Nurse Education Today 1991;11:248-55.
Burch VC, Nash RC, Zabow T, Gibbs T, Aubin L, Jacobs B, Hift RJ. A structured assessment of newly qualified medical graduates. Medical Education 2005;39(7):723-31.
Burdick WP, Pujol-Farriols R, Martinez-Carretero JM, Descarrega R, Friedman Ben-David M, Boulet J. Reproducibility of Emergency Medicine Standardized Patient Cases from Philadelphia to Barcelona. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;413-5.
Burke J, Fayaz S, Graham K, Matthew R, Field M. Peer-assisted learning in the acquisition of clinical skills: a supplementary approach to musculoskeletal system training. Medical Teacher 2007;29:577-82.
Burrows P, Khan A, Bowden R, Jackson N. The 'Fresh Start' simulated surgery. Education for Primary Care 2004;15(3):328-35.
Burrows PJ, Bingham L, Brailovsky CA. A modified contrasting groups method used for setting the passmark in a small scale standardized patient examination. Advances in Health Sciences Education 1999;4(2):145-54.
Burrows SC, Tylman V. Evaluating medical student searches of MEDLINE for evidence-based information: process and application of results. Bulletin of the Medical Library Association 1999;87(4):471-6.
Bustamante M, Carvajal C, Gottlieb B, Contreras JE, Uribe M, Melkonian E, Cardenas P, Amadori A, Parra JA. A new instrument for the evaluation of the medical profession. Use of the OSCE method. Revista Medica de Chile 2000;128(9):1039-44.
Byrne E, Smyth S. Lecturers' experiences and perspectives of using an objective structured clinical examination. Nurse Education in Practice 2008;8(4):283-9.
Cadman D, Woodward CA. Symposium: Use of the OSCE Objective Structured Clinical Examination to Assess Clinical Competence: Use of the OSCE for student and program evaluation. Proceedings of the Annual Conference on Research in Medical Education 1987; 23:407-12.
Cadman SS, Boulet JR. Predicting holistic ratings of written performance assessments from analytic scoring. Advances in Health Sciences Education 2001;6(1):103-19.
Campos-Outcalt D, Rutala PJ, Witzke DB, Fulginiti JV. Performances of underrepresented-minority students at the university of Arizona College of Medicine, 1987-1991. Academic Medicine 1994;69(7):577-82.
Campos-Outcalt D, Watkins A, Fulginiti J, Kutob R, Gordon P. Correlations of Family Medicine Clerkship Evaluations and Objective Structured Clinical Examination Scores and Residency Directors' Ratings. Family Medicine 1999;31(2):90-4.
Cannic G, Horowitz A, Garr D, Reed S, Neville B, Day T, Woolson R, Lackland D. Use of the OSCE to Evaluate Brief Communication Skills Training for Dental Students. Journal of Dental Education 2007;71(9):1203-9.
Capt Zedek IJ, LTC Hines JF, MAJ Thomas AR, LTC Satin AJ, LTC Armstrong AY, CAPT Haffner WHJ. Development and use of military-unique standardized gynecology patients for military residents in Obstetrics and Gynecology training programs. Military Medicine 1998;163(11):767-9.
Carey JC. Integrating prevention into Obstetrics/Gynecology. Academic Medicine 2000;75(7 Suppl):S72-S76.
Carney PA, Ogrinc G, Harwood BG, Schiffman JS, Cochran N. The influence of teaching setting on medical students' clinical skills development: is the academic medical center the "gold standard"? Academic Medicine 2005;80(12):1153-8.
Carpenter JL, Battles JB, McIntire D, Sprankell SJ. Assessing the usefulness of using standardized patients in a clinical medicine course. Academic Medicine 1992;67(4):286.
Carpenter JL, McIntire D, Battles J, Wagner JM. Administration of parallel, simultaneous objective structured clinical

examination to accommodate a large class of students. Teaching and Learning in Medicine 1993;5(2):79-85.
Carpenter JL. Cost analysis of objective structured clinical examinations. Academic Medicine 1995;70(9):828-33.
Carracvio C, Englander R. The objective structured clinical examination: a step un the direction of competency-based evaluation. Archives of Pediatric and Adolescent Medicine 2000;154(7):736-41.
Caruso B, Nieman LZ, Gracely E. Developing and assessing the effectiveness of an HIV sexual history and risk assessment workshop for medical professionals. Journal of Sex Education and Therapy 1994;20(2):101-9.
Cass A, Regehr G, Reznick R, Rothman A, Cohen R. Sequential testing in the objective structured clinical examination: selecting items for the screen. Academic Medicine 1997;72(10 Suppl):S25-S27.
Castrén.M, Nurmi.J, Laakso.J-P, Kinnunen.A, Backman.R, Niemi-Murola.L. Teaching public access defibrillation to lay volunteers--a professional health care provider is not a more effective instructor than a trained lay person. Resuscitation 2004;63(3):305-10.
Cater JI, Forsyth JS, Frost GJ. The use of objective structured clinical examination as an audit of teaching and student performance. Medical Teacher 1991;13 (3):253-7.
Cave J, Washer P, Sampson P, Griffin M, Noble L. Explicitly linking teaching and assessment of communication skills. Medical Teacher 2007;29(4):317-22.
Cerille GJ, Merrick HW, Staren ED. Surgical educator preferences regarding key objective structured clinical examination topics. Journal of Surgical Research 2001;101(2):124-9.
Cerilli GJ, Merrick HW, Staren ED. Objective Structured Clinical Examination Technical Skill Stations Correlate More Closely with Postgraduate year level than do clinical skill stations. The American Surgeon 2001;67(4):323-7.
Cervera Alemany AM, Salva Casanovas A, Altimir Losada S, Miralles Basseda R, Yuste Marco A, et al.l. Objective and structured clinical evaluation in geriatrics. Results of the first project performed in Spain. Revista Espanola de Geriatria y Gerontologia 2002;37(6):298-303.
Chakravarty M, Latif NA, Abu-Hijleh MF, Osman M, Dharap AS, Ganguly PK. Assessment of anatomy in a problem-based medical curriculum. Clinical Anatomy 2005;18 (2):131-6.
Chalabian J, Garman K, Wallace P, Dunnington G. Clinical breast evaluation skills of house officers and students. The American Surgeon 1996;62(10):840-5.
Chalabian J, Dunnington G. Do our current assessments assure competency in clinical breast evaluation skills? American Journal of Surgery 1998;175(6):497-502.
Chambers KA, Boulet JR, Gary NE. The management of patient encounter time in a high-stakes assessment using standardized patients. Medical Education 2000;34(10):813-7.
Chambers KA, Boulet JR, Furman GE. Are interpersonal skills ratings influenced by gender in a clinical skills assessment using standardized patients? Advances in Health Sciences Education 2001;6(1):231-41.
Champlain AF, Macmillan MK, Margolis MJ, Klass DJ, Lewis E, Ahearn S. Modelling the effects of a test security breach on a large-scale standardized patient examination with a sample of international medical graduates. Academic Medicine 2000;75(10 Suppl):S109-S111.
Chateny M, Maguire T, Shakun E, Chang G, Cook D, Warnock GL. Does volume of clinical experience affect performance of clinical clerks on surgery exit examinations? American Journal of Surgery 1996;172(10):366-72.
Chenkin J, Lee S, Huyn T, Bandiera G. Procedures can be learned on the web: A randomized study of ultrasound-guided vascular access training. Academic Emergency Medicine 2008;15(10):949-54.
Chenot JF, Ehrhardt M. Objective structured clinical examination (OSCE) in medical training: An alternative to a written exam. ZFA Zeitschrift fur Allgemeinmedizin 2003;79(9):437-42.
Chenot JF, Simmenroth-Nayda A, Koch A, Fischer T, Scherer M, Emmert B, Stanske B, Kochen M, Himmel W. Can student tutors act as examiners in an objective structured clinical examination? Medical Education 2007;41:1032-8.
Cherry RA, Williams J, George J, Ali J. The effectiveness of a human patient simulator in the ATLS shock skills station. Journal of Surgical Research 2007;139(2):229-35.
Chesser AM, Laing MR, Miedzybrodzka ZH, Brittenden J, Heys SD. Factor analysis can be a useful standard setting tool in a high stakes OSCE assessment. Medical Education 2004;38(8):825-31.
Chessman AW, Blue AV, Gilbert GE, Carey M, Mainous AG. Assessing students' communication and interpersonal skills across evaluation settings. Family Medicine 2003;35(9):643-8.
Chibnall JT, Blaskiewicz RJ. Do Clinical Evaluations in a Psychiatry Clerkship Favor Students With Positive Personality Characteristics? Academic Psychiatry 2008;32(3):199-205.
Chipman JG, Beilman GJ, Schmitz CC, Seatter SC. Development and pilot testing of an OSCE for difficult conversations in surgical intensive care. Journal of Surgical Education 2007;64(2):79-87.
Chouinard A. The Oskie: poking and prodding in the name of medicine. Canadian Medical Association Journal 1986;135(9):517-9.
Chouinard A. OSCE is effective, efficient and not too painful. Canadian Journal of Surgery 1986;29(1):13-4.
Chumley HS. What Does an OSCE Checklist Measure? Family Medicine 2008;40(8):589-91.
Cichowski.E, Bell J, Huggett K. Tailoring the objective structured clinical examination to fit the Year one medical student. Medical Education 2007;41(5):515-6.
Clauser BE, Ross LP, Fletcher EA, Klass DJ, Finkbiner RG, King AM. Differential item functioning in checklist items from a standardized patient based examination. Academic Medicine 1994;69(10 Suppl):S72-S74.
Clauser BE, Clyman SG. A contrasting-groups approach to standard setting for performance assessments of clinical skills. Academic Medicine 1994;69(10 Suppl):S42-S44.
Clauser BE, Margolis MJ, Ross LP, Ningester RJ, Klass DJ. Regression-Based Weighting of Items on Standardized Patient Checklists. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;420-3.
Clauser BE, De Champlain AF, Nungester RJ. Applying Sequential testing Strategies to Performance Assessments of Clinical Skills. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Clauser BE. Further discussion of SP checklists and videotaped performances. Academic Medicine 2000;75(4):315-8.
Cleland JA, Milne A, Sinclair H, Lee AJ. Cohort study on predicting grades: is performance on early MBChB assessments predictive of later undergraduate grades? Medical Education 2008;42:676-83.
Cleries X, Borrel F, Epstein R, Juncosa S, Kronfly E, Martinez-Carretero JM, Roma J. Validation of an Assessment Instrument for Communication Skills in Catalonia (Spain). Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-2.
Cogbill KK, O'Sullivan PS, Clardy J. Residents' perception of effectiveness of twelve evaluation methods for measuring competency. Academic Psychiatry 2005;29(1):76-81.
Cohen DS, Colliver JA, Robbs RS, Swartz MH. A Large-Scale Study of the Reliabilities of Checklist Scores and Ratings of

Components of Interpersonal and Communication Skills Evaluated on a SP Examination. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;424-6.

Cohen R, Rothman AI, Poldre P, Ross J. Validity and generalizability of global ratings in an objective structured clinical examination. Academic Medicine 1991;66(9):545-8.

Cohen R, Singer PA, Rothman AI, Robb A. Assessing competency to address ethical issues in medicine. Academic Medicine 1991;66(1):14-5.

Cohen R, Rothman AI, Ross J, Poldre P. Validating an objective structured clinical examination (OSCE) as a method for selecting foreign medical graduates for a Pre-internship Program. Academic Medicine 1991;66(9 Suppl):S67-S69.

Cohen R, Rothman AI, Bilan S, Ross J. Analysis of the psychometric properties of eight administrations of an objective structured clinical examination used to assess international medical graduates. Academic Medicine 1996;71(1 Suppl):S22-S24.

Cohen R. Assessing professional behaviour and medical error. Medical Teacher 2001;23(2):145-51.

Collins JP, Tregonning GD, Gamble GD. Uniform experience and assessment during a multisite surgical clerkship. Australian and New Zealand Journal of Surgery 1994;64(7):506-11.

Collins JP, White GR, Kennedy JA. Entry to medical school: an audit of traditional selection requirements. Medical Education 1995;29(1):22-8.

Collins JP, Gamble GD. A multi-format interdisciplinary final examination. Medical Education 1996;30(4):259-65.

Collins JP, Harden RM. AMME Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations. Medical Teacher 1998;20(6):508-21.

Colliver JA, Verhulst SJ, Williams RG, Norcini JJ. Reliability of performance on standardized patient cases: a comparison of consistency measures based on generalizability theory. Teaching and Learning in Medicine 1989;1(1):31-7.

Colliver JA, Verhulst SJ, Williams RG. Using a standardized patient examination to establish the predictive validity of the MCAT and undergraduate GPA as admissions criteria. Academic Medicine 1989;64(8):482-4.

Colliver JA, Morrison LJ, Markwell SJ, Verhulst SJ, Steward DE, Dawson-Saunders E, Barrows HS. Three studies of the effect of multiple standardized patients on intercase reliability of five standardized-patient examinations. Teaching and Learning in Medicine 1990;2(4):237-45.

Colliver JA, Steward DE, Markwell SJ, Marcy ML. Effect of repeated simulations by standardized patients on intercase reliability. Teaching and Learning in Medicine 1991;3(1):15-9.

Colliver JA, Robbs RS, Vu NV. Effects of using two or more standardized patients to simulate the same case on case means and case failure rates. Academic Medicine 1991;66(10):616-8.

Colliver JA, Vu NV, Verhulst SJ. Effect of position - within sequence on case performance in a Multiple-Stations Examination using standardized-patient cases. Evaluation and the Health Professions 1991;14(3):343-55.

Colliver JA, Travis TA, Robbs RS, Barnhart AJ, Shirar LE, Vu NV. Test security in standardized-patient examinations: analysis with scores on working diagnosis and final diagnosis. Academic Medicine 1992;67(10 Suppl):S7-S9.

Colliver JA, Vu NV, Barrows HS. Screening test length for sequential testing with a standardized-patient examination: a receiver operating characteristic (ROC) analysis. Academic Medicine 1992;67(9):592-5.

Colliver JA, Vu NV, Robbs RS, Verhulst SJ, Travis TA, Barrows H. False-negative and false-positive rates resulting from measurement error for a mastery examination using standardized-patient cases. Teaching and Learning in Medicine 1992;4(4):238-42.

Colliver JA, Williams RG. Technical issues: test application. Academic Medicine 1993;68(6):454-63.

Colliver JA, Vu NV, Marcy ML, Travis TA, Robbs RS. Effects of examinee gender, standardized-patient gender, and their interaction on standardized patients' ratings of examinees' interpersonal and communication skills. Academic Medicine 1993;68(2):153-7.

Colliver JA, Barnhart AJ, Marcy ML, Verhulst SJ. Using a receiver operating characteristic (ROC) analysis to set passing standards for a standardized-patient examination of clinical competence. Academic Medicine 1994;69(10 Suppl):S37-S39.

Colliver JA, Marcy ML, Vu NV, Steward DE, Robbs RS. Effect of using multiple standardized patients to rate interpersonal and communication skills on intercase reliability. Teaching and Learning in Medicine 1994;6(1):45-8.

Colliver JA. Validation of standardized-patient assessment: a meaning for clinical competence. Academic Medicine 1995;70(12):1062-4.

Colliver JA, Willis MS, Robbs RS, Cohen DS, Swartz MH. Assessment of empathy in a standardized-patient examination. Teaching and Learning in Medicine 1998;10 (1):8-11.

Colliver JA, Swartz MH, Robbs RS, Lofquist M. The Interaction of Examinee Ethnicity and Standardized-Patient Ethnicity on Clinical Performance Assessed with Standardized Patients. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-5.

Colliver JA, Swartz MH, Robbs RS, Cohen DS. Relationship between clinical competence and interpersonal and communication skills in standardized-patient assessment. Academic Medicine 1999;74(3):271-4.

Colliver JA, Swartz MH, Robbs RS. The effect of examinee and patient ethnicity in clinical-skills assessment with standardized patients. Advances in Health Sciences Education 2001;6(1):5-13.

Conn HL, Cody RP. Results of the second clinical skills assessment examination of the ECFMG. Academic Medicine 1989;64(8):448-53.

Coombes JA, McGuire TM, Harrhy KL, McRobbie D, Davies JG, Fleming G. Piloting an Objective Structured Clinical Examination to Evaluate the Clinical Competency of Pre-Registration Pharmacists. Journal of Pharmacy Practice & Research 2003;33(3):194-8.

Coovadia HM, Moosa A. A comparison of traditional assessment with the objective structured clinical examination (OSCE). South African Medical Journal 1985;67 :810-2.

Cottel P. A Reflection on Objective Structured Clinical Examinations (OSCEs). Primary Health Care 2007;17(10):22-3.

Courteille O, Bergin R, Stockeld D, Ponzer S, Fors U. The use f a virtual patient in an OSCE-based exam- A pilot study. Medical Teacher 2008;30(3):e66-e76.

Coutts-van Dijk LC, Bray JH, Moore S, Rogers J. Prospective study of how students humanism and psychosocial beliefs relate to speciality matching. Academic Medicine 1997; 72(12):1106-8.

Coutts L, Rogers J. Predictors of student self-assessment accuracy during a clinical performance exam: comparisons between over-estimators and under-estimators of SP-evaluated performance. Academic Medicine 1999;74(10 Suppl):S128-S130.

Coutts LC, Rogers JC. Humanism: is its evaluation captured in commonly used performance measures? Teaching and Learning in Medicine 2000;12(1):28-32.

Cox K. No Oscar for OSCA. Medical Education 1990;24(6):540-5.

Cox K. Examining and recording clinical performance: a critique and some recommendations. Education for Health

2000;13(1):45-52.
Cox M, ODea D, Yee L. Use of an objective structured clinical examination to assess general practitioner competence in sexual and reproductive health. Education for General Practice 1999;10(1):62-8.
Critchley LAH, Short TG, Buckley T, OMeara ME, Gin T, Oh TE. An adaptation of the objective structured clinical examination to a final year medical student course in anaesthesia and intensive care. Anaesthesia 1995;50(4):354-8.
Croen LG, Moroff SV. Pilot-testing a holistic approach to scoring performance on standardized-patient examinations. Academic Medicine 1994;69(4):310-2.
Cunnington JPW, Neville AJ, Norman GR. The Risks of Thoroughness: Reliability and Validity of Global Ratings and Checklists in an OSCE. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;143-5.
Curry L, Woodward CA, Poewles P. Symposium: use of the OSCE Objective Structured Clinical Examination to Assess Clinical competence: Reliability and Concurrent Validity of the OSCE. Proceedings of the Annual Conference on Research in Medical Education 1987;23:407-11.
Curtis D, Lind S, Brear S, Finzen F. The correlation of student performance in preclinical and clinical prosthodontic assessments. Journal of Dental Education 2007;71(3):365-73.
Cuschieri A, Gleeson FA, Harden RM, Wood RAB. A new approach to a final examination in surgery. Annals of the Royal College of Surgeons of England 1979;61:400-5.
Cusimano M, Rothman A, Keystone J. Defining standards of competent performance on an OSCE. Academic Medicine 1998;73(10 Suppl):S112-S113.
Cusimano MD, Cohen R, Tucker W, Murnaghan J, Kodama R, Reznick R. A comparative analysis of the costs of administration of an OSCE. Academic Medicine 1994;69 (7):571-6.
Cusimano MD, Rothman AI. Consistency of standards and stability of pass/fail decisions with examinee-based standard-setting methods in a small-scale objective structured clinical examination. Academic Medicine 2004;79(10 Suppl):S25-S27.
Dacre JE. The effect of formal instruction in ophthalmoscopy on medical student performance. Medical Teacher 1993;15(4):321-5.
Dauphinee WD, Blackmore DE, Smee SM, Rothman AI, Reznick RK. Optimizing the Input of Physician Examiners in Setting Standards for a Large Scale OSCE: Experience with Part II of the Qualifying Examination of the Medical Council of Canada. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;656-8.
Dauphinee WD, Blackmore DE, Smee S, Rothman AI, Reznick R. Using the judgements of physician examiners in setting the standards for a national multi-center high stakes OSCE. Advances in Health Sciences Education 1997;2:201-11.
Dauphinee WD, Blackmore DE, Smee SM, Rothman AI, Des Marchais JE, Reznick R. Adaptative Testing: A Report on the Results and the Myths Arising from the use of a sequenced OSCE for National Lcensure. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-6.
Dauphinee WD, Boulais AP, Smee SM, Rothman AI, Reznick R, Blackmore DE. Examinations results of the licenciate of the Medical Council of Canada: trends, issues and future considerarions. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Davies H, Archer J, Bateman A, Dewar S, Crossley J, Grant J, Southgate L. Specialty-specific multi-source feedback: assuring validity, informing training. Medical Education 2008;42:1014-20.
Davis DH, Hill B. Clinical Assessment in an Academic Environment: To Do or Not To Do? Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;427-9.
Davis DH, Hill B. Academic Assessment of Clinical Skills Inferences of Predictive Validity. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;430-2.
Davis JEC, Fry H, Jones A. Developing Staff: Developing Stations. A dental Objective Structured Clinical Examination (OSCE) Activity Day. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;207-9.
Davis P, Andrews E, Donen N, et al. Case studies in osteoporosis: a problem based learning intervention for family physicians. The Journal of Rheumatology 1999;26(11):2418-22.
Davis P, Kvern B, Donen N, Andrews E, Nixon O. Evaluation of a problem-based learning workshop using pre-and post-test Objective Structured Clinical Examinations and standardized patients. The Journal of Continuing Education in the Health Professions 2000;20(3):164-70.
De Champlain AF, Fletcher EA, Macmillan MK, Klass DJ, Margolis MJ. Assessing the Reliability of Post-Encounter Note Scores in a Large-Scale Standardized Patient Examination: Comparing the Scoring Consistency of Medical Chart Abstractors and Physicians. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
De Champlain AF, Margolis MJ, King AM, Klass DJ. Investigating Halo Effects in a Nationally Administered Standardized Patient Examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
De Champlain AF, Macmillan MK, Margolis MJ, Klass DJ, Nungester RJ, Schimpfhauser F, Zinnerstrom K. Modeling the effects of security breaches on students' performances on a large-scale standardized patient examination. Academic Medicine 1999;74(10 Suppl):S49-S51.
De Champlain AF, Macmillan MK, King AM, Klass DJ, Margolis MJ. Assessing the impacts of intra-site and inter-site checklist recording discrepancies on the reliability of scores obtained in a nationally administered standardized patient examination. Academic Medicine 1999;74(10 Suppl):S52-S54.
De Champlain AF, Margolis ML, Macmillan MK, Klass DJ. Predicting mastery level on a large-scale standardized patient test: a comparison of case and instrument score based models using discriminant function analysis. Advances in Health Sciences Education 2001;6:151-8.
De Vries TPGM. Presenting clinical pharmacology and therapeutics: the course in pharamacotherapeutics. British Journal of Clinical Pharmacology 1993;35(6):587-90.
Dehaitem MJ, Ridley K, Kerschbaum WE, Inglehart MR. Dental Hygiene Education About Patients with Special Needs: A Survey of U.S. Programs. Journal of Dental Education 2008;72(9):1010-9.
DeLisa JA. Evaluation of clinical competency. American Journal of Physical Medicine Rehabilitation 2000;79(5):474-7.
Delva MD, Woodhouse RA, Birtwhistle RV, Knapper C, Kirby JR. Does PBL matter? Relations between instructional context, learning strategies, and learning outcomes. Advances in Health Sciences Education 2000;5(3):167-77.
Delzell Jr JE, Ringdahl EN, Kruse RL. The ACGME core competencies: A national survey of family medicine program director. Family Medicine 2005;37(8):576-80.
Denekens J. Education in basic clinical skills for doctors. Verhandelingen-Koninklijke Academie voor Geneeskunde van Belgie 2000;62(5):399-418.
Dennehy PC, Susarla SM, Karimbux NY. Relationship Between

Dental Students' Performance on Standardized Multiple-Choice Examinations and OSCEs. Journal of Dental Education 2008;72(5):585-92.
Des Marchais JE, Vu NV. Developing and evaluating the student assessment system in the preclinical problem-based curriculum at Sherbrooke. Academic Medicine 1996;71(3):274-83.
Descargues G, Sibert L, Lechevallier J, Weber J, Lemoine JP, Marpeau L. Evaluation of clinical competence in gynecology obstetrics: an innovative approach using the Objective Structured Clinical Examination. Journal of Gynecologie, Obstetrique et Biologie de la Reproduction 2001;30(3):257-64.
Descarrega-Queralt R, Barragan N, Blay C, Iruela A, Juncosa S, Kronfly E, Martinez-Carretero JM, Serrallach S. Accuracy of Standardized Patients in Clinical Skills Assessment. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-2.
Deveugele M, Derese A, De Maesschalck S, Willems S, Van Driel M, De Maeseneer J. Teaching communication skills to medical students, a challenge in the curriculum? Patient Education and Counseling 2005;58(3):265-70.
Diemer M, Cooper GS, Harvey J, Noel GL. What do second year students ask in the sexual history? Observations from an objective structured clinical examination. Proceedings of the Annual Conference on Research in Medical Education 1987;26 :217-21.
Dijchs R, Prince KJAH, Van der Vleuten CPM, Scherpbier AJJA. Validity of objective tests towards peer-rated competence by students. Medical Teacher 2003;25(3):273-6.
Dikici M, Yaris F. Standardized and simulated patient program in ondokuz mayis university school of medicine: Medical education. Turkiye Klinikleri Journal of Medical Sciences 2007;27(5):738-43.
Dissanayake AS, Ali BA, Nayar U. The influence of the introduction of objective structured practical examinations in physiology on student performance at King Faisal University Medical School. Medical Teacher 1990;12(3-4):297-304.
Dobay KJ, Nalesnik S. Lapar - OSCE: a laparoscopic observed structured clinical experience. Obstetrics and Gynecology 2001;87(4 Suppl):S8-S9.
Doig CJ, Harasym PH, Fick GH, Baumber JS. The effects of examiner background, station organization and time of exam on OSCE scores assessing undergraduate medical students' physical examination skills. Academic Medicine 2000;75(10 Suppl):S96-S98.
Dolenc T, Philbrick K. Achieving competency in electroconvulsive therapy: a model curriculum. Academic Psychiatry 2007;31(1):65-7.
Donnelly MB, Sloan D, Plymale M, Schwartz R. Assessment of residents' interpersonal skills by faculty proctors and standardized patients: a psychometric analysis. Academic Medicine 2000;75(10 Suppl):S93-S95.
Downing SM. Reliability: on the reproducibility of assessment data. Medical Education 2004;38(9):1006-12.
Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. Medical Education 2004;38(3):327-33.
Downing SM, Tekian A, Yudkowsky R. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. Teaching and Learning in Medicine 2006;18(1):50-7.
Doyle J. Correlation of Objective Structured Clinical Examination and National Board of Examiners part II scores. Surgery 1994;116(5):945-6.
Duerson MC, Romrell LJ, Stevens BC. Impacting faculty teaching and student performance: nine years experience with the objective structured clinical examination. Teaching and Learning in Medicine 2000;12(4):176-82.
Dupras DM, Li JTC. Use of an Objective Structured Clinical Examination to determine clinical competence. Academic Medicine 1995;70(11):1029-34.
Durak H, Caliskan S, Bor S, Van der Vleuten C. Use Of Case-Based Exams As An Instructional Teaching Tool To Teach Clinical Reasoning. Medical Teacher 2007;29 :e170-e174.
Edelstein RA, Reid HM, Ysatine R, Wilkes MS. A comparative study of measures to evaluate medical students' performances. Academic Medicine 2000;75(8):825-33.
Editorial. The place of Objective Structured Clinical Examinations in undergraduate medicine. Annals of Saudi Medicine 1991;11(1):106.
Edwards DR. The Effect of Changing Selection Procedures and the Pre-Entry Experience of Intending Students of Osteopathy. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;643-5.
Edwards M, Martin A. The objective structured clinical examination as a method of occupational therapy student evaluation. Canadian Journal of Occupational Therapy 1989;56(3):128-31.
Elizondo-Montemayor LL. How we assess students using an holistic standardized assessment system. Medical Teacher 2004;26(5):400-2.
Elliot DL, Fields SA, Keenan TL, Jaffe AC, Toffler WL. Use of a group objective structured clinical examination with first-year medical students. Academic Medicine 1994;69(12):990-2.
Elman D, Hooks R, Tabak D, Regehr G, Freeman F. The effectiveness of unannounced standardised patients in the clinical setting as a teaching intervention. Medical Education 2004;38(9):969-73.
Elnicki DM, Shockcor WT, Morris DK, Hallbritter KA. Creating an objective Structured Clinical Examination for the internal medicine Clerkship: Pitfalls and benefits. American Journal of Medical Sciences 1993;306(2):94-7.
Elnicki DM, Halbritter KA, Antonelli MA, Linger B. Educational and career outcomes of an internal medicine preceptorship for first-year medical students. Journal of General Internal Medicine 1999;14(6):341-6.
Elzubeir MA, Rizk DEE. Assessing confidence and competence of senior medical students in an obstetrics and gynaecology clerkship using an OSCE. Education for Health 2001; 14(3):373-82.
Endean ED, Sloan DA, Veldenz HC, Donnelly MB, Schwartz TH. Performance of the vascular physical examination by residents and medical students. Journal of Vascular Surgery 1994;19(1):149-56.
Ertl L, Christ F. Significant improvement of the quality of bystander first aid using an expert system with a mobile multimedia device. Resuscitation 2007;74 (2):286-95.
Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. Medical Education 2004;38(3):314-26.
Eva KW, Reiter HI, Rosenfeld J, Norman GR. The ability of the multiple mini-interview to predict preclerkship performance in medical school. Academic Medicine 2004;79(10 Suppl):S40-S42.
Evans DE, Wood DF, Robert CM. The effect of an extended hospital induction on perceived confidence and assessed clinical skills of newly qualified pre-registration house officers. Medical Education 2004;38(9):998-1001.
Evensen SA, Karterud SW, Marthisen P, Sekkelsten A. Assessment of medical students ability to communicate unpleasant news. Tidsskrift for Den Norske Laegeforening 1997; 117(19):2804-6.
Eyles M, DeLisa J, Jain SS, Kirshblum S, Nadler S, Smith A. Five Years of Experience in Developing a Comprehensive OSCE for Physical Medicine and Rehabilitation Residents. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa

Conference 1998 in Philadelphia 1998;1.
Fabiny A, McArdle P, Perls T, Inui T, Sheehan M. The Geriatric Objective Structured Clinical Exercise: A teaching tool in a Geriatrics Curriculum. Gerontology Geriatrics Education 1998;18(4):63-70.
Falck-Ytter Y. How to successfully approach a reform in medical education in Germany. Medizinische Ausbildung 2001;18(1):137.
Famuyiwa OO, Zachariah MP, Ilechukwu STC. The objective structured clinical examination in undergraduate psychiatry. Medical Education 1991;25(1):45-50.
Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. Academic Medicine 1996;71(12):1363-5.
Feather A, Kopelman PG. A practical approach to running an objective structured clinical examination (OSCE) for medical undergraduates. Education for Health 1997;10(3):333-50.
Feather A, McCrorie P, Kopelman P. Does Success at Graduation, Resulting from Intensive Remedial Skills Training Predict Future Clinical Performance. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Feather A, Stone SP, Wessier A, Boursicot KA, Pratt C. 'Now please wash your hands': the handwashing behaviour of final MBBS candidates. Journal of Hospital Infection 2000;45(1):62-4.
Feickert JAD, Harris IB, Anderson DC, Bland CJ, Allen S, Poland GA, Satran L, Miller WJ. Senior medical students as simulated patients in an objective structured clinical examination: motivation and benefits. Medical Teacher 1992;14(2/3):167-77.
Ferrel BG. Clinical performance assessment using standardized patients: a primer. Family Medicine 1995;27(1):14-9.
Ferrell BG, Thompson BL. Standardized patients: a long-station clinical examination format. Medical Education 1993;27(4):376-81.
Fields HW, Rowland ML, Vig KW. Objective structured clinical examination use in advanced orthodontic dental education. American Journal of Orthodontics & Dentofacial Orthopedics 2007;131(5):656-63.
Fields SA, Toffler WL, Elliot DL, Garland MJ, Atkinson RM, Kennan TL, Jaffe AC. Principles of clinical medicine: an interdisciplinary integrated 2-year longitudinal course. Medical Education 1995;29(1):53-7.
Finlay IG, Maughan TS, Webster DJT. A randomized controlled study of portfolio learning in undergraduate cancer education. Medical Education 1998;32(2):172-6.
Fitzgerald JT, Gruppen LD, White C, Davis WK, Barclay ML, Bergstorm TJ, Chamberlain KR, McQuillan MA, Zweifler AJ. Student and Faculty Ratings of Task Difficulty in Predicting Performance on a Clinical-Skills Exam. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;659-60.
Fitzpatrick CF, Radecki SE. Use of the OSCE in residency training. Family Medicine 1993;26(1):49-50.
Fletcher E, DeChamplain AF, Klass DJ, Macmillan MK. Surveying reactions of medical chart abstractors and physicians to the scoring process of post-encounter notes for an NBME standardized patient examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 im Philadelphia 1998;1.
Floreck LM, De Champlain AF. Assessing sources of score variability in a multisite medical performance assessment: an application of hierarchial linear modeling. Academic Medicine 2001;76(10 Suppl):S93-S95.
Forman GG, Ross LR, Galofre A, Heaney RM, Mootz WC. A standardized patient clinical examination to assess clinical performance of medical students in an ambulatory-care clerkship. Teaching and Learning in Medicine 1994;6(3):175-8.
Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? Medical Education 1999;33(4):276-81.
Fowell SL, Maudsley G, Maguire P, Leinster SJ, Bligh J. Student assessment in undergraduate medical education in the United Kingdom, 1988. Medical Education 2000;34(Suppl 1):1-49.
Fox R, Dacre J, Mclure C. The impact of formal instruction in clinical examination skills on medical student performance - the example of peripheral nervous system examination. Medical Education 2001;35(4):371-3.
Fox RA, Clark CJI, Scotland AD, Dacre JE. A study of pre-registration house officers' clinical skills. Medical Education 2000;34(12):1007-12.
Fox RA, Dacre JE, Clark CL, Scotland AD. Impact on medical students of incorporating GALS screen teaching into the medical school curriculum. Annals of Rheumatic Diseases 2000;59(9):668-71.
Fox RA, Dacre JA. Defining the content for the objective structured clinical examination component of the Professional and Linguistic Assessments Board examination: development of a blueprint. Medical Education 2000;34(7):566-72.
Franzese C. When to Cut? Using an Objective Structured Clinical Examination to Evaluate Surgical Decision-Making. Laryngoscope 2007;117(11):1938-42.
Franzese CB. Pilot study of an objective structured clinical examination ("the Six Pack") for evaluating clinical competencies. Otolaryngology Head and Neck Surgery 2008;138:143-8.
Fried E. Optimal time for a standardized patient encounter may be intermediate time. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 im Philadelphia 1998;1.
Friedlich M, MacRae H, Oandasan I, Tannenbaum D, Batty H, Reznick R, Regehr G. Structured assessment of minor surgical skills (SAMSS) for family medicine residents. Academic Medicine 2001;76(12):1241-6.
Friedman Ben-David M, Klass DJ, Boulet J, Champlain AD, King AM, Pohl HS, Gary NE. The performance of foreign medical graduates on the National Board of Medical Examiners (NBME) standardized patient examination prototype: a collaborative study of the NBME and the Educational Commission for foreign Medical Graduates (ECFMG). Medical Education 1999;33:439-46.
Friedman Ben-David M. The role of assessment in expanding professional horizons. Medical Teacher 2000;22(5):472-7.
Friedman M, Mennin SP. Rethinking critical issues in performance assessment. Academic Medicine 1991;66(7):390-5.
Frohna JG, Kalet A, Kachur E, Zabar S, Cox M, Halpern R, Hewson MG, Yedidia MJ, Williams BC. Assessing residents' competency in care management: report of a consensus conference. Teaching and Learning in Medicine 2004;16(1):77-84.
Frost GJ, Cater JI, Forsyth JS. The use of an Objective Structured Clinical Examination (OSCE) in paediatrics. Medical Teacher 1986;8(3):261-9.
Frye AW, Richards BF, Philp EB, Philp JR. Is it worth it? A look at the costs and benefits of an OSCE for second-year medical students. Medical Teacher 1989;11(3/4):291-3.
Furlong E, Fox P, Lavin M, Collins R. Oncology nursing students' views of a modified OSCE. European Journal of Oncology Nursing 2005;9(4):351-9.
Furman G, Colliver JA, Galofre A. Effects of student gender and standardized-patient gender in a single case using a male and a female standardized patient. Academic Medicine 1993;68(4):301-3.
Furman GE, Colliver JA, Galofre A, Reaka MA, Robbs RS, King A. The Effect of Formal Feedback Sessions on Test Security for a Clinical Practice Examination using Standardized Patients. Advances in Medical Education, 7th Ottawa

Conference, 1996 Maastricht 1997;433-6.
Furman GE, Coliver JA, Galofre A, Reaka MA, Robbs RS, King A. The effect of formal feedback sessions on test security for a clinical practice examination using standardized patients. Advances in Health Sciences Education 1997;2(1):3-7.
Furukawa K, Ogawa R, Norose Y, Tajiri T. A new surgical handwashing and hand antisepsis from scrubbing to rubbing. Journal of Nippon Medical School = Nihon Ika Daigahu Zasshi 2004;71(3):190-7.
Garde PM. Evaluation of the OSCE in the primary health care situation. South African Journal of Nursing 1984;7(4):36-7.
Gasquoine S. Objective structured clinical evaluation (OSCE), a report on the O.S.C.E for the third year nursing students at the School of Health Studies, Carrington Polytechnic. Nursing Praxis in New Zealand 1990;5(3):29-33.
Gaufberg E, Fitzpatrick A. The favour: a professional boundaries OSCE station. Medical Education 2008;42:513-43.
Geddes EL, Crowe J. Peer-Related Objective Structured Clinical Examination. Physioterapy Canada 1998;50(4):269-75.
Geiger DL, Heermann JA, Eilers J. Identification and validation of competencies for use in objective structured clinical examinations for lay caregivers. Cancer Nursing 2005;28(1):54-61.
George J, Taylor C, Conran P. The interdisciplinary generalist curriculum project at the medical college of Ohio. Academic Medicine 2001;76(4 Suppl):S100-S103.
George S. OSCEs in Psychiatry. British Journal of Psychiatry 2004;185(3):273.
Gerrow JD, Boyd MA, Duquette P, Bentley K. Results of the National Dental Examining Board of Canada Written Examination and Implications for Certification. Journal of Dental Education 1997;61(12):921-7.
Gilbart MK, Hutchison CR, Cusimano MD, Regehr G. A computer-based trauma simulator for teaching trauma management skills. American Journal of Surgery 2000;179(3):223-8.
Gilson GJ, George KE, Qualls CM, Sarto GE, Obenshain SS, Boulet J. Assessing clinical competence of medical students in women's health care: use of the Objective Structured Clinical Examination. Obstetrics and Gynecology 1998;92(6):1038-43.
Gispert R, Rue M, Roma J, Martinez-Carretero JM. Gender, sequence of cases and day effects on clinical skills assessment with standardized patients*. Medical Education 1999;33:499-503.
Gledhill RF, Capatos D. Factors affecting the reliability of an objective structured clinical examination (OSCE) test in neurology. South African Medical Journal 1985;67:463-7.
Gleeson F. Defects in postgraduate clinical skills as revealed by the objective structured long examination record (OSLER). Irish Medical Journal 1992;85(1):11-4.
Gleeson F. Postgraduate clinical skills standards: an analysis using the objective structured long examination record (OSLER). Postgraduate Medical Journal 1993;69(2):s65.
Gleeson F. AMME Medical Education Guide No. 9 Assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). Medical Teacher 1997;19(1):7-14.
Gomez JM, Prieto L, Pujol R, Arbizu T, Vilar L, Borrel F, Roma J, Martinez-Carretero JM. Clinical skills assessment with standardized patients. Medical Education 1997;31:94-8.
Goodyear HM. Problem based learning in a junior doctor teaching programme. Archives of Disease in Childhood 2005;90(3):275-8.
Gorter S, Rethans JJ, Scherpbier A, Van der Heijde D, Van der Vleuten C, Van der Linder S. Developing case-specific checklists for standardized-patient-based assessments in internal medicine: a review of the literature. Academic Medicine 2000;75(11):1130-7.
Govaerts MJB, Schuwirth LWT, Pin A, Clement MEJ, Van der Vleuten CPM. Objective assessment is needed to ensure competence. British Journal of Midwifery 2001;9(3):156-61.
Govindan VK. Enhancing communication skills using an OSCE and peer review. Medical Education 2008;42:513-43.
Goyal R. Examination reform: not only OSCE but reform selection of teachers also. Indian Pediatrics 1998;35(2):185-6.
GrandMaison P, Blouin D, Briere D. Utilization of the Objective Structured Clinical Examination OSCE in Gynecology/Obstetrics. Proceedings of the Annual Conference on Research in Medical Education 1985;24:65-71.
GrandMaison P, Lescop J, Rainsberry P, Brailovsky CA. Large-scale use of an objective, structured clinical examination for licensing family physicians. Canadian Medical Association Journal 1992;146(10):1735-40.
GrandMaison P, Lescop J, Brailovsky CA. Canadian experience with structured clinical examinations. Canadian Medical Association Journal 1993;148(9):1573-6.
GrandMaison P, Brailovsky CA, Lescop J. Content validity of the Quebec licensing examination (OSCE). Canadian Family Physician 1996;42:254-9.
GrandMaison P, Brailovsky CA, Lescop J, Rainsberry P. Using standardized patients in licensing/certification examinations: comparison of two tests in Canada. Family Medicine 1997;29(1):27-32.
GrandMaison P, Brailovsky CA, Lescop J. The Quebec Licensing OSCE: Modifications and Improvements over 6 Years of Experience. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;437-40.
GrandMaison P, Brailovsky CA, Miller F, Nasmith L, Emond JG, Rainsberry P, Frenette J, Dandavino A. Licensing examination and training programs in family medicine: a pilot-project to assess mututal influences. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-2.
Griffith CH, Haist SA, Blue AV. An interdisciplinary clinical performance examination for a third-year combines medicine-surgery clerkship. Academic Medicine 1998;73(5):597.
Gruppen LD, Davis WK, Fitzgerald JT, McQuillan. Reliability, Number of Stations, and Examination Length in an Objective Structured Clinical Examination. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;441-2.
Guiton G, Hodgson CS, Delandshere G, Wilkerson L. Communication skills in standardized-patient assessment of final-year medical students: a psychometric study. Advances in Health Sciences Education 2004;9(3):179-87.
Gupta P, Bisht HJ. A practical approach to running an objective structured clinical examination in neonatology for formative assessment of medical undergraduates. Indian Pediatrics 2001;38(5):500-13.
Haeney O. The objective structured clinical examination [5]. Psychiatric Bulletin 2004;28(10):383.
Hafler JP, Connors KM, Volkan K, Bernstein HH. Developing and evaluating a residents' curriculum. Medical Teacher 2005;27(3):276-82.
Halaas GW, Zink T, Brooks KD, Miller CJ. Clinical skills day: preparing third year medical students for their rural rotation. Rural & Remote Health 2007;7(4):788.
Hall-Turner WJA. An experimental assessment carried out in an undergraduate general practice teaching course (OSCE examination). Medical Education 1983;17(2):112-9.
Hall MJ, Adamo G, McCurry L, Lacy T, Waits W, Chow J, Rawn L, Ursano RJ. Use of standardized patients to enhance a psychiatry clerkship. Academic Medicine 2004;79(1):28-31.
Hamadeh G, Lancaster C, Johnson A. Introducing the Objective

Structured Clinical Examination to a Family Practice Residency Program. Family Medicine 1993;25(4):237-41.

Hamann C, Volkan K, Fishman MB, Silvestri RC, Simon SR, Fletcher SW. How well do second-year students learn physical diagnosis? Observational study of an Objective Structured Clinical Examination (OSCE). BMC Medical Education 2002;2(1):1.

Hanna MN, Donnelly MB, Montgomery CL, Sloan PA. Perioperative pain management education: a short structured regional anesthesia course compared with traditional teaching among medical students. Regional Anesthesia & Pain Medicine 2005;30 (6):523-8.

Hanson M, Hodges B, McNaughton N, Regehr G. The integration of child psychiatry into a Psychiatry Clerkship OSCE. Canadian Journal of Psychiatry 1998;43(6):614-8.

Haq C, Steele DJ, Marchand L, Seibert C, Brody D. Integrating the Art and Science of Medical Practice: Innovations in Teaching Medical Communication Skills. Family Medicine 2004;36(Suppl):S43-S50.

Haq I, Higham J, Morris R, Dacre J. Effect of ethnicity and gender on performance in undergraduate medical examinations. Medical Education 2005;39(11):1126-8.

Harasym PH, Mohtadi NG, Henningsmoen H. The Use of Critical Stations to Determine Clinical Competency in a 'High Stakes' OSCE. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;661-4.

Harasym PH, Mohtadi NG, Henningsmoen H. Construct Validity of 'High Stakes' OSCE Scores. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;443-5.

Harasym PH, Mohtadi NG, Henningsmoen H. A Comparison of Diagnostic Pattern Recognition Skills to Performance on a 'High Stakes' OSCE. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;152-5.

Harasym PH, Woloschuk W, Cunning L. Undesired variance due to examiner stringency/leniency effect in communication skill scores assessed in OSCEs. Advances in Health Sciences Education 2008;13(5):617-32.

Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. British Medical Journal 1975;1:447-51.

Harden RM. How to assess clinical competence - an overview. Medical Teacher 1979;1(6):289-96.

Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination OSCE. Medical Education 1979;13:41-54.

Harden RM, Cairncross RG. Self Assessment. Medical Teacher 1980;2(3):145-8.

Harden RM, Cairncross RG. Assessment of practical skills: the objective structured practical examination (OSPE). Studies in Higher Education 1980;5(2):187-96.

Harden RM. Editorial - 2: Assessment of clinical competence and the OSCE. Medical Teacher 1986;8(3):203-5.

Harden RM. What is an OSCE? Medical Teacher 1988;10(1):19-22.

Harden RM. Twelve tips for organizing an objective structured clinical examination (OSCE). Medical Teacher 1990;12(3/4):259-64.

Harden RM, Davis MH, McAleer S. The assessment of competence in history-taking stations in the Objective Structured Clinical Examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.

Hardison C, Fonken P, Chew T, Smith B. The emergence of family medicine in Kyrgyzstan. Family Medicine 2007;39(9):627-33.

Harris I, Ytterberg S, Anderson D, Kofron P, Kvasnicka J, Moller JH. Tailored Response Test: a new approach for teaching in medical education. Medical Teacher 1997;19(3):194-9.

Harris IB, Miller WJ. Feedback in an Objective Structured Clinical Examination by medical students serving as patients, examiners, and teachers. Academic Medicine 1990;65(7):433-4.

Hart IR, Cairncross RG, Harden RM. An objective structured examination of clinical competence. Annals of the Royal College of Physicians and Surgeons of Canada 1979;12(1):58.

Hart IR, Woodward CA, Poewles P. Symposium: use of the OSCE Objective Structured Clinical Examination to Assess Clinical Competence: Predictive Validity and Acceptability of the Objective Structured Clinical Examination. Proceedings of the Annual Conference on Research in Medical Education 1987;23:407-10.

Hasle JL, Anderson DS, Szerlip HM. Analysis of the costs and benefits of using standardized patients to help teach physical diagnosis. Academic Medicine 1994;69(7):567-70.

Hatala R, Issenberg SKBO, Cole G, Bacchus C, Scalese RJ. Assessing the relationship between cardiac physical examination technique and accurate bedside diagnosis during an objective structured clinical examination (OSCE). Academic Medicine 2007;82(suppl 10):s26-s29.

Hatala R, Issenberg S, Kassen B, Cole G, Bacchus C, Scalese R.J. Assessing cardiac physical examination skills using simulation technology and real patients: a comparison study. Medical Education 2008;42:628-36.

Hatch DJ. Specialist anaesthetic training and certification process in England. Annals of the Academy of Medicine, Singapore 1994;23(4):614-8.

Hauer KE, Hodgson CS, Kerr KM, Teherani A, Irby DM. A national study of medical student clinical skills assessment. Academic Medicine 2005;80(10 Suppl):S25-S29.

Heard JK, Cantrell M, Presher L, Klimberg VS, San Pedro GS, Erwin DO. Using standardized patients to teach breast evaluation to Sophomore Medical Students. Journal of Cancer Education 1995;10(4):191-4.

Heard JK, Allen R, Tank PW, Cason GJ, Cantrell M, Wheller RP. Assessing clinical skills of medical students. Journal of the Arkansas Medical Society 1996;93(4):175-9.

Heard JK, Allen RM, Cason GJ, Cantrell M, Tank PW. Practical issues in developing a program for the objective assessment of clinical skills. Medical Teacher 1998;20(1):15-21.

Heckmann JG, Dutscha M, Raucha C, Lang C, Weihb M, Schwab S. Effects of peer-assisted training during the neurology clerkship: a randomized controlled study. European Journal of Neurology 2008;15:1365-70.

Heermann JA, Eilers JG, Carney PA. Use of modified OSCEs to verify technical skill performance and competency of lay caregivers. Journal of Cancer Education 2001;16(2):93-8.

Hellewell SA. Machine check OSCE - Which guidelines are right? Anaesthesia 2001;56(5):487-8.

Hickling FW, Morgan KA, Abel W, Denbow CE, Ali Z, Nicholson GD, Sinquee C. A comparison of the objective structured clinical examination results across campuses of the University of the West Indies (2001 and 2002). West Indian Medical Journal 2005;54(2):139-43.

Hijazi M, Downing SM. Objective structured clinical examinations as an assessment method in residency training: Practical considerations. Annals of Saudi Medicine 2008;28(3):192-9.

Hill D, Stalley P, Pennington D, Besser M, McCarthy W. Competency based learning in traumatology. American Journal of Surgery 1997;173(2):136-40.

Hill DA, Lord RSA. Complementary value of traditional bedside teaching and structured clinical teaching in introductory surgical studies. Medical Education 1991;25(6):471-4.

Hill DA, Guinea AI, McCarthy WH. Formative assessment: a student perspective. Medical Education 1994;28(5):394-9.

Hilliard RI, Tallett SE. The use of an objective structured clinical examination with postgraduate residents in pediatrics. Archives

of Pediatric and Adolescent Medicine 1998;152(1):74-8.
Ho MJ, Yao G, Lee K, Beach MC, Green AR. Cross-cultural medical education: Can patient-centered cultural competency training be effective in non-Western countries? Medical Teacher 2008;30(7):719-21.
Ho MJ, Lee KL, Green AR. Can cultural competency self-assessment predict OSCE performance? Medical Education 2008;42(3):525.
Hodder RV, Rivington RN, Calcut LE, Hart IR. The effectiveness of immediate feedback during the Objective Structured Clinical Examination. Medical Education 1989;23(2):184-8.
Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G. Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. Medical Education 1996;30(1):38-43.
Hodges B, Lofchy J. Evaluating psychiatric clinical clerks with a mini-objective structured clinical examination. Academic Psychiatry 1997;21(4):219-25.
Hodges B, Regehr G, Hanson M, McNaughton N. An objective structured clinical examination for evaluating psychiatric clinical clerks. Academic Medicine 1997;72(8):715-21.
Hodges B, Regehr G, Hanson M, McNaughton N. Validation of an objective structured clinical examination in Psychiatry. Academic Medicine 1998;73(8):910-2.
Hodges B, Hanson M, McNaughton N, Regehr G. The Objective Structured Clinical Examination in Psychiatry: a Validation Study. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Hodges B, Hanson M, McNaughton N, Regehr G. What do Psychiatry residents think of an Objective Structured Clinical Examination? Academic Psychiatry 1999;23 (4):198-204.
Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. Academic Medicine 1999;74(10):1129-34.
Hodges B, McNaughton N, Regehr G, Tiberius R, Hanson M. The challenge of creating new OSCE measures to capture the characteristics of expertise. Medical Education 2002;36(8):742-8.
Hodges B. OSCE! Variations on a theme by Harden. Medical Education 2003;37(12):1134-40.
Hoelker M, Breukelmann D, Saur M, Nippert RP. Process Evaluation of a Course in Basic Clinical Skills: Implementing an OSCE. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 2004;1-5.
Hofer M, Schiebel B, Hartwig HG, Garten A, Modder U. Innovative concepts for small group teaching in clinical skills: diagnostic imaging procedures: results of a longitudinal two cohort study of medical education pilot project. Deutsche Medizinische Wochenschrift 2000;125(23):723.
Hofer M, Abanador N, Modder U. Effektive Didaktiktrainings fur Dozenten von CME-Fortbildungen. [Effective didactic skills training for teachers in continuing medical education]. [German]. Rofo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin 2005;177(9):1290-6.
Hofmeister M. The multiple mini interview in the assessment of international medical graduates for family medicine residency training in Canada: An examination of psychometric evidence. Dissertation Abstracts International Section B: The Sciences and Engineering 2008;69(4B):2198.
Hollingsworth MA, Richards BF, Frye AW. Description of observer feedback in an objective structured clinical examination and effects on examinees. Teaching and Learning in Medicine 1994;6(1):49-53.
Holm AL, Aspegren K. Objective Structured Clinical Examination--an established method of testing physicians' clinical competence. [Review] [16 refs] [Danish]. Ugeskrift for Laeger 2002;166(21):2002-4.
Holyfield LJ, Bolin KA, Rankin KV, Shulman JD, Jones DL, Eden BD. Use of computer technology to modify objective structured clinical examinations. Journal of Dental Education 2005;69(10):1133-6.
Hoole AJ, Kowlowitz V, McGaghie WC, Sloane PD, Colindres RE. Using the objective structured clinical examination at the University of North Carolina Medical School. North Carolina Medical Journal 1987;48(9):463-7.
Howley L, Szauter K, Perkowski L, Clifton M, McNaughton N. Quality of standardised patient research reports in the medical education literature: review and recommendations. Medical Education 2008;42:350-8.
Hudson JN, Vernon-Roberts JM. Assessment - putting it all together. Medical Education 2000;34(11):953-4.
Hull AL, Hodder S, Berger B, Ginsberg D, Lindheim N, Quan J, Kleinhenz ME. Validity of three clinical performance assessments of internal medicine clerks. Academic Medicine 1995;70(6):517-22.
Humphrey-Murto S, Smith CD, Touchie C, Wood TC. Teaching the musculoskeletal examination: are patient educators as effective as rheumatology faculty? Teaching and Learning in Medicine 2004;16(2):175-80.
Humphrey-Murto S, Wood TJ, Touchie C. Why do physicians volunteer to be OSCE examiners? Medical Teacher 2005;27(2):172-4.
Humphrey-Murto S, Smee S, Touchie C, Wood TJ, Blackmore DE. A comparison of physician examiners and trained assessors in a high-stakes OSCE setting. Academic Medicine 2005;80(10 Suppl):S59-S62.
Humphris GM, Kaney S. The objective structured video exam for assessment of communication skills. Medical Education 2000;34(11):939-45.
Humphris GM, Kaney S. Examiner fatigue in communication skills objective structured clinical examinations. Medical Education 2001;35(5):444-9.
Hussain I, Muzaffar F, Rashid T, Ahmed TJ, Haroon TS. Inter-rater variation and reliability of Objective Structured Clinical Examination. Journal of the College of Physicians and Surgeons Pakistan 1999;9(6):271-2.
Hymowitz N, Schwab J, Haddock C, Pyle S, Schwa L. The pediatric residency training on tobacco project: Four-year resident outcome findings. Preventive Medicine 2007;45:481-90.
Ichikawa A. Pharmaceutical Common achievement test: computer-based testing (CBT). Yakugaku Zasshi 2007;127(6):941-5.
Ichikawa A. Appropriate Usage of Antibiotics by Therapeutic Drug Monitoring. Yakugaku Zasshi 2007;127(6):923.
Iramaneerat C, Yudkwsky R. Rater errors in a clinical skills assessment of medical students. Evaluation and the Health Professions 2007;30(3):266-83.
Iramaneerat C. A validity study of a communication skills assessment of medical residents. Dissertation Abstracts International A: Humanities and Social Sciences 2008;68(12A):4976.
Iramaneerat C, Yudkowsky R, Myford C, Downing SM. Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. Advances in Health Sciences Education 2008;13(4):479-93.
Jacobs JC, Denessen E, Postma CT. The structure of medical competence and results of an OSCE. Netherlands Journal of Medicine 2004;62(10):397-403.
Jager J, Kollner V, Graf N. The (OSCE) Objective Structured Clinical Examination in general practice as an assessment of high competency after the block training period. ZFA Zeitschrift fur Allgemeinmedizin 2008;84(9):375-81.

Jain SS, DeLisa JA, Campagnolo DI. Methods used in evaluation of clinical competency of physical medicine and rehabilitation residents. American Journal of Physical Medicine and Rehabilitation 1994;73(4):234-9.

Jain SS, Nadler S, Eyles M, Kirshblum S, DeLisa JA, Smith A. Development of an Objective Structured Clinical Examination OSCE for Physical Medicine and Rehabilitation Residents. American Journal of Physical Medicine Rehabilitation 1997;76(2):102-18.

Jain SS, DeLisa JA, Eyles MY, Nadler S, Kirshblum S, Smith A. Further experience in development of an Objective Structured Clinical Examination for physical medicine and rehabilitation residents. American Journal of Physical Medicine Rehabilitation 1998;77(4):306-10.

Jamison. The fixed response Objective Structured Clinical Examinations: a useful adjunct for assessing competence in diagnostic decision making? Journal of Manipulative and Physiological Therapeutics 1992;15(4):261-6.

Jani SS, DeLisa JA, Nadler S, Kirshblum S, Banerjee SN, Eyles M, Johnston M, Smith AC. One program's experience of OSCE vs. written board certification results: a pilot study. American Journal of Physical Medicine Rehabilitation 2000;79(5):462-7.

Jansen JJM, Tan LHC, Van der Vleuten CPM, Van Luijk SJ, Rethans JJ, Grol RPTM. Assessment of competence in technical clinical skills of general practitioners. Medical Education 1995;29(3):247-53.

Jay A. Students' perceptions of the OSCE: a valid assessment tool? Br J Midwifery 2007;15(1):32-7.

Jefferies A, Simmons B, Tabak D, Mcllroy JH, Lee KS, Roukema H, Skinner BD. Using an objective structured clinical examination (OSCE) to assess multiple physician competencies in postgraduate training. Medical Teacher 2007;29(2-3):183-91.

Jefferies A, Simmons B, Regehr G. The effect of candidate familiarity on examiner OSCE scores. Medical Education 2007;41(9):888-91.

Jeffery HE, Henderson-Smart DJ, Hill DA. Competency-based learning in neonatology. Medical Education 1996;30(6):440-4.

Jewell D. Learning through examinations: use of an objective structured clinical examination as a teaching method in general practice. Journal of the Royal College of General Practitioners 1988;38(11):506-8.

Johnson G, Reynard K. Assessment of an objective structured clinical examination (OSCE) for undergraduate students in accident and emergency medicine. Journal of Accident and Emergency Medicine 1994;11(4):223-6.

Johnson SB, Donnelly MB, Sloan DA, Kearney PA, Strodel WE, Schwartz RW. Trauma and critical care clinical performance: impact of undergraduate education. Teaching and Learning in Medicine 1997;9(1):39-44.

Johnston BT, Boohan M. Basic clinical skills: don't leave teaching to the teaching hospitals. Medical Education 2000;34(9):692-9.

Jolly B, Newble D, Chinner T. Learning effect of reusing stations in an Objective Structured Clinical Examination. Teaching and Learning in Medicine 1993;5(2):66-71.

Jolly B, Cohen R, Newble D, Rothman A. Possible effects of reusing OSCE stations. Academic Medicine 1996;71(10):1023-4.

Jolly BC, Cohen R, Rothman AI, Ross J. Graduates of foreign medical schools: demographic and personal predictors of success on an OSCE-format internship programme entrance examination. Proceedings of the Annual Conference on Research in Medical Education 1988;27:234-9.

Jolly BC, Jones A, Dacre JE, Elzubeir M, Kopelman P, Hitman G. Relationships between students' clinical experiences in introductory clinical courses and their performances on an objective Structured Clinical Examination (OSCE). Academic Medicine 1996;71(8):909-16.

Jonassen JA, Pugnaire MP, Mazor K, Regan MB, Jacobson EW, Gammon W, Doepel DG, Cohen AJ. The effect of a domestic violence interclerkship on the knowledge, attitudes and skills of third-year medical students. Academic Medicine 1999;74(7):821-8.

Jones AR. Can a written test of clinical skills predict student performance in clerkship? Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.

Joorabchi B. Objective Structured Clinical Examination in a pediatric residency profram. American Journal of Diseases in Children 1991;145(7):757-62.

Joorabchi B, Devires JM. Evaluation of clinical competence: the gap between expectation and performance. Pediatrics 1996;97(2):179-84.

Joy R, Nickless LJ. Revolutionising assessment in a clinical skills environment - A global approach: The recorded assessment. Nurse Education in Practice 2008;8(5):352-8.

Julian TM. A graduating medical school class evaluates their educational experience. Wisconsim Medical Journal 1998;97(2):56-62.

Junger J, Schafer S, Roth C, Schellberg D, Friedman Ben-David M, Nikendei C. Effects of basic clinical skills training on objective structured clinical examination performance. Medical Education 2005;39(10):1015-20.

Kachur EK, Green S, Dennis C. Written comments on objective structured clinical examination rating forms: an exploration study. Teaching and Learning in Medicine 1990;2(4):225-31.

Kachur EK, Kachur EK. Multiple Station Exams and Teaching Exercises (MSEs) for Teaching About Teaching. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;226-7.

Kahn MJ, Merril WW, Anderson DS, Szerlip HM. Residency program director evaluations do not correlate with performance on a required 4th year objective structured clinical examination. Teaching and Learning in Medicine 2001;13(1):9-12.

Kalet A, Earp JA, Kowlowitz V. How well do faculty evaluate the interviewing skills of medical students? Journal of General Internal Medicine 1992;7(5):499-505.

Kalet A, Pugnaire MP, Cole-Kelly K, Janicik R, Ferrara E, Schwartz MD, Lipkin MJ, Lazare A. Teaching communication in clinical clerkships: models from the macy initiative in health communications. [Review] [58 refs]. Academic Medicine 2004;79(6):511-20.

Karaalp A, Akici A, Kocabasoglu YE, Oktay S. What do graduates think about a two-week rational pharmacotherapy course in the fifth year of medical education? Medical Teacher 2003;25(5):515-21.

Karani R, Leipzig RM, Callahan EH, Thomas DC. An Unfolding Case with a Linked Objective Structured Clinical Examination (OSCE): A Curriculum in Inpatient Geriatric Medicine. Journal of the American Geriatrics Society 2004;52(7):1191-8.

Karlberg L, Lindgren C. Att kunna samtala med patienten--aktuellt examensamne for lakarstudent. En utbildningsinvestering som lonar sig.[Communication skills in the encounter with patients--current examination subject for medical students. Beneficial educational investment]. [Review] [22 refs] [Swedish]. Lakartidningen 2004;101(40):3072-4.

Kassebaum DG. On standardized patients and clinical skills assessment. Academic Medicine 1990;65(5):307.

Kassebaum DG, Eaglen RH. Shortcomings in the evaluation of students clinical skills and behaviours in medical school. Academic Medicine 1999;74(7):842-9.

Kaufman DM, Mann KV, Muijtjens AMM, Van der Vleuten CPM. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. Academic

Medicine 2000;75(3):267-71.
Kaufman DM, Laidlaw TA, Macleod H. Communication skills in Medical School: Exposure, Confidence, and Performance. Academic Medicine 2000;75(10 Suppl):S90-S92.
Keely E, Myers K, Dojeiji S. Can written communication skills be tested in an objective structured clinical examination format? Academic Medicine 2002;77(1):82-6.
Keely EJ, Dojeiji S, Myers K, Faught W, Bonin B. Pre-pregnancy Counselling: What Do Residents Write in Their Consultation Letters? Canadian Journal of Diabetes 2004;28(1):15-9.
Kelly JM. OSCA: Assessing to Unite Theory and Practice. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;446-7.
Kelly M, Murphy A. An evaluation of the cost of designing, delivering and assessing an undergraduate communication skills module. Medical Teacher 2004;26(7):610-4.
Kent AP, Lazarus J. An objective medical student examination in obstetrics. South African Medical Journal 1983;64:388-9.
Kevelighan EH, Duffy S, Walker JJ. Innovations in teaching obstetrics and gynaecology - the Theme Afternoon. Medical Education 1998;32(5):517-21.
Khan J, Rooney K, Prosciak C, Javadpoor A, Rooney PJ. Effect of immediate feedback on performance at subsequent stations during an objective structured clinical examination. Education for Health 1997;10(3):351-7.
Khattab A, Rawlings B. Use of a modified OSCE to assess nurse practitioner students. British Journal of Nursing 2008;17(12):754-9.
Khattab AD, Rawlings B. Assessing Nurse Practitioner Students Using a Modified Objective Structured Clinical Examination OSCE. Nurse Education Today 2001;21(7):541-50.
Kigler B, Koithan M, Maizes V.Hayes M, Shneider C, Lebensohn P, Hadley S. Competency-based evaluation tools for integrative medicine training in family medicine residency: a pilot study. BMC Medical Education 2007;7:1-11.
Kilminster S, Roberts T, Morris P. Incorporating patients'assessments into objective structured clinical examinations. Education for Health 2007;20(1):1-5.
Kilminster S, Roberts T. Standard setting for OSCEs: trial of borderline approach. Advances in Health Sciences Education 2004;9(3):201-9.
Kilminster SM, Delmotte A, Frith H, Jolly BC, Stark P, Howdle PD. Teaching in the new NHS: the specialised ward based teacher. Medical Education 2001;35(5):437-43.
King A, DeChamplain AF, Duerson M, Rathe RJ. Use of hand-held computers by standardized patients for data recording during a clinical skills examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998.
King AM, Carr BA, Downing BK, Klass DJ. A description of National Board of Medical Examiners Training Processes for Standardized Patient Licensing Examinations. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Kirby RL, Curry L. Introduction of an objective structured clinical examination (OSCE) to an undergraduate clinical skills programme. Medical Education 1982;16(6):362-4.
Klass DJ. 'High stakes' testing of medical students using standardized patients. Teaching and Learning in Medicine 1994;6(1):28-32.
Klass DJ, De Champlain AF, Fletcher E, King AM, Macmillan MK. Development of a Performance-based test of Clinical Skills for the United States Medical Licensing Examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Kleinhenz ME, Hull AL. Standardized patients in evaluating medical students smoking cessation skills. Teaching and Learning in Medicine 1995;7(1):3-7.
Kligler B, Koithan M, Maizes V, Hayes M, Schneider C, Lebensohn P, Hadley S. Competency-based evaluation tools for integrative medicine training in family medicine residency: a pilot study. BMC Medical Education 2007;7:1-11.
Knowles C, Kinchington F, Erwin J, Peters B. A randomized controlled trial of the effectiveness of combining video role play with traditional methods of delivering undergraduate medical education. Sexually Transmitted Infections 2001;77(5):376-80.
Kogo M, Koyama N, Negoro T, et al. Analysis of students, achievement rate and contents of assessment for objective structured clinical examination (OSCE) attempted at the faculty of pharmaceutical sciencies, Showa University. Yakugaku Zasshi 2007;127(5):905-17.
Kokotailo PK, Fleming MF, Koscik RL. A model alcohol and other drug use curriculum for pediatric residents. Academic Medicine 1995;70(6):495-8.
Koktailo PK, Langhough R. Improving pediatric residents' alcohol and other drug use clinical skills: use of an experiential curriculum. Pediatrics 1995;96(1):99-104.
Kolsek M, Svab I, Pavlic DR, Bulc M. Introducing performance-based assessment of family physicians. Medical Teacher 2003;25(1):63-6.
Konje JC, Abrams KR, Taylor DJ. How discriminatory is the objective structured clinical examination (OSCE) in the assessment of clinical competence of medical students? Journal of Obstetrics and Gynaecology 2001;21(3):223-7.
Koop AJ, Borbasi SA. Towards enhanced OSCE in Australian nurse education: a contribution from South Africa. South African Journal of Nursing 1994;17(3):40-3.
Kopelow ML, Schnabi GK, Hassard TH, Tamblyn RM, Klass DJ, Beazley G, Hechter F, Grott M. Assessing practicing physicians in two settings using standardized patients. Academic Medicine 1992;67(10 Suppl):S19-S21.
Kothari N, Murthy M. Medical Residents as Students and Teachers in an Objective Structured Clinical Examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Kothari N, Murthy M, Abate H. An Objective Structured Clinical Examination for Internal Medicine Residency Training. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Kowlowitz V, Hoole AJ, Sloane PD. Implementing the Objective Structured Clinical Examination in a traditional medical school. Academic Medicine 1991;66(6):345-7.
Kramer A, Jansen JJM, Zuithoff P, Dusman H, Tan L, Grol R, Van der Vleuten CPM. Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice. Medical Education 2002;36:812-9.
Kramer A, Muijtjens A, Jansen K, Dusman H, Tan L, Van der Vleuten C. Comparison of a rational and an empirical standard setting procedure for an OSCE. Medical Education 2003;37(2):132-9.
Kramer AW, Jansen KJ, Dusman H, Tan LH, Van der Vleuten CPM, Grol RP. Acquisition of clinical skills in postgraduate training for general practice.[comment]. British Journal of General Practice 2003;53(494):677-82.
Kushnir T, Ehrenfeld M, Shalish Y. The effects of a coaching project in nursing on the coaches' training motivation, training outcomes, and job performance: An experimental study. International Journal of Nursing Studies 2008;45(6):837-45.
Kwolek CJ, Donnely MB, Sloan DA, Birrel SN, Strodel WE, Schwartz RW. Ward evaluations: should they be abandoned? Journal of Surgical Research 1997;69(1):1-6.
Kwolek CJ, Donnelly MB, Endean ED, Sloan DA, Schwartz RW.

Development of vascular surgery skills during general surgery training. Vascular Surgery 1999;33(2):129-35.
Kwolek DS, Witzke DB, Blue AV, Schwartz RW, Sloan DA. Using an OSCE to assess the ability of residents to manage problems in women's health. Academic Medicine 1997;72(10 Suppl):S48-S50.
Kwolek DS, Griffith CH, Blue AV. Using clinical skills workshops to teach complex assessment skills in women's health. Teaching and Learning in Medicine 1999;11(2):105-9.
Kyle V, Burns-Cox C, Box L, Burgess B. OSCEs for house officers LETTER. Postgraduate Medical Journal 1999;75:574.
Labelle M, Beaulieu M, Renzi P, Rahme E, Thivierge RL. Integrating clinical practice guidelines into daily practice: impact of an interactive workshop on drafting of a written action plan for asthma patients. Journal of Continuing Education in the Health Professions 2004;24(1):39-49.
Ladyshewsky R. Simulated patients and assessment. Medical Teacher 1999;21(3):266-9.
Lafave M, Katz L, Butterwick D. Development of a content-valid standardized orthopedic assessment tool (SOAT). Advances in Health Sciences Education 2008;13(4):397-406.
Laidlaw TS, Kaufman DM, Sargeant J, Macleod H, Blake K, Simpson D. What makes a physician an exemplary communicator with patients. Patient Education and Counseling 2007;68(2):153-60.
Lang F, Bennard B, Belanger A. Using standardized students to educate preceptors. Academic Medicine 1995;70(10):855-6.
Lang F, McCord R, Harvill L, Anderson DS. Communication assessment using the common ground instrument: psychometric properties. Family Medicine 2004;36(3):189-98.
Langford NJ, Landray M, Martin U, Kendall MJ, Ferner RE. Testing the practical aspects of therapeutics by objective structured clinical examination. Journal of Clinical Pharmacy & Therapeutics 2004;29(3):263-6.
Larsen T, Jeppe-Jensen D. The introduction and perception of an OSCE with an element of self and peer-assessment. European Journal of Dental Education 2008;12:2-7.
Lau E, Dolovich L, Austin Z. Comparison of self, physician, and simulated patient ratings of pharmacist performance in a family practice simulator. Journal of Interprofessional Care 2009;21(2):129-40.
Lauder W, et al.l. Measuring competence, self-reported competence and self-efficacy in pre-registration students. Nursing Standard 2008;22(20):35-43.
Lazarus J, Kent A. Student attitudes towards the objective structured clinical examination (OSCE) and conventional methods of assessment. South African Medical Journal 1983;64:390-4.
Lee JD, Erickson JC, Short MW, Roth BJ. Education research: evaluating acute altered mental status: are incoming interns prepared? Neurology 2008;71(18):e50-e53.
Leila N-M, Pirkko H, Eeva P, Eija K, Reino P. Training medical students to manage a chronic pain patient: both knowledge and communication skills are needed. European Journal of Pain 2006;10(2):167-70.
Lewin LO, Papp KK, Hodder SL, Workings MG, Wolfe L, Glover P, Headrick LA. Performance of third-year primary-care-track students in an integrated curriculum at case Western Reserve University. Academic Medicine 1999;74(1 Suppl):S82-S89.
Li JTC. Reassessing the OSCE. Academic Medicine 1996;71(11):1137.
Lind DS, Marum T, Ledbetter D, Flynn TC, Romrell LJ, Copeland EM. The effect of the duration and structure of a surgery clerkship on student performance. Journal of Surgical Research 1999;84(1):106-11.
Lind DS, Deladisma AM, Cue JI, Thomas AM, MacFadyen BV, Nesbit RR. Survey of student education in surgery. Journal of the American College of Surgeons 2007;204(5):969-74.
Lindsey AA, Stritter F. Use of an OSCE in Education for Radiation Therapists. Radiologic Technology 1990;62(1):36-41.
Lipkin M, Zabar S, Kalet A, Laponis R, Kachur E, Anderson M, Gillespie C. Two Decades of Title VII Support of a Primary Care Residency: Process and Outcomes. Academic Medicine 2008;83(11):1064-70.
Lloyd JS, Williams RG, Simonton DK, Sherman D. Order effects in standardized patient examinations. Academic Medicine 1990;65(9 Suppl):S51-S52.
Long EM. An Evaluation of Osces for Final-Year Students. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;448-51.
Loschen EL. Using the objective structured clinical examination in a psychiatry residency. Academic Psychiatry 1993;17(2):95-104.
Lown B, Sasson J, Hinrichs P. Patients as Partners in Radiology Education: An Innovative Approach to Teaching and Assessing Patient-Centered Communication. Academic Radiology 2008;15(4):425-32.
Lule GS, Cullinan T, Ngwale M. Assessment in community health: a village-based Objective Structured Clinical Examination in Malawi. Education for Health 1996;9(1):85-90.
Lumb AB, Vail A. Comparison of academic, application form and social factors in predicting early performance on the medical course. Medical Education 2004;38 (9):1002-5.
Lunenfeld E, Weinreb B, Lavi Y, Amiel GE, Friedman M. Assessment of emergency medicine: a comparison of an experimental objective structured clinical examination with a practical examination. Medical Education 1991;25(1):38-44.
Lypson ML, Frohna JG, Gruppen LD, Woolliscroft JO. Assessing residents' competencies at baseline: identifying the gaps. Academic Medicine 2004;79(6):564-70.
Macmillan CSA, Crosby JR, Wildsmith JAW. Skilled task teaching and assessment. Medical Teacher 2001;23(6):591-4.
Macmillan M, DeChamplain AF, Klass D. Assessing the comparability of checklist scores across standardized patients using travelling patients. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998.
Macmillan MK, De Champlain AF, Klass DJ. Using tagged items to detect threats to security in a nationally administered standardized patient examination. Academic Medicine 1999;74(10 Suppl):S55-S57.
Macmillan MK, Fletcher EA, Champlain AF, Klass DJ. Assessing post-encounter note documentation by examinees in a field test of a nationally administered standardized patient test. Academic Medicine 2000;75(10 Suppl):S112-S114.
MacNeily AE, Morales A. Initial assessment of a new preparatory tool for board certification in Urology. Urology 2000;55(5):647-51.
MacNeily AE, Baverstock RJ, Cole G, Morales A. Quantitative assessment of a new preparatory tool for board certification in urology. BJU International 2004;93(4):558-61.
MacRae HM, Vu NV, Graham B., Word-Sims M, Colliver JA, obbs R.S. Comparing checklists and databases with physicians' ratings as measures of students' history and physical-examination skills. Academic Medicine 1995;70(4):313-7.
Madan AK, Caruso BA, Lopes JE, Gracely EJ. Comparison of simulated patient and didactic methods of teaching HIV risk assessment to medical residents. American Journal of Preventive Medicine 1998;15(2):114-9.
Major DA. OSCEs - seven years on the bandwagon: the progress of an objective structured clinical evaluation programme. Nurse Education Today 2005;25(6):442-54.
Majumdar B, Roberts J, Knechtel R, Noesgaard C, Campbell K. Comparison of self-and faculty directed learning of

psychomotor skills. Advances in Health Sciences Education 1998;3(1):15-28.

Makinen M, Aune S, Niemi-Murola L, Herlitz J, Varpula T, Nurmi J, Axelsson AB, Thoren A.B., Castren M. Assessment of CPR-D skills of nurses in G¡§oteborg,Sweden and Espoo, Finland: Teaching leadership makes a difference. Resuscitation 2007;72(2):264-9.

Malik SL, Manchanda SK, Deepack KK, Sunderam KR. The attitudes of medical students to the objective structured practical examination. Medical Education 1988;22(1):40-6.

Malloy MH, Perkowski L, Callaway M., Speer A. The relationship between preceptor expectations and student performance on 2 pediatric objective structured clinical examination stations. Archives of Pediatric and Adolescent Medicine 1998;152(8):806-11.

Mann KV, MacDonald C, Norcicni JJ. Reliability and objective structured clinical examinations: four years of experience in a surgical clerkship. Teaching and Learning in Medicine 1990;2(4):219-24.

Manogue M, Brown G. Developing and implementing an OSCE in dentistry. European Journal of Dental Education 1998;2(2):51-7.

Manogue M, Brown G, Foster H. Clinical assessment of dental students: values and practices of teachers in restorative dentistry. Medical Education 2001;35(4):364-70.

Margolis MJ, De Champlain AF, Klass DJ. Setting Standards for a Performance-Based Assessment of Physicians' Clinical Skills. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-6.

Mariolis A, Mihas C, Alevizos A, Papathanasiou M, Mariolis-Sapsakos T, Marayiannis K, Koutsilieris M. Evaluation of a clinical attachment in Primary Health Care as a component of undergraduate medical education. Medical Teacher 2008;30(7):e2002-e2007.

Marshall G, Harris P. A study of the role of an objective structured clinical examination (OSCE) in assessing clinical competence in third year student radiographers. Radiology 2000;6(2):117-22.

Marshall G, Jones N. A pilot study into the anxiety induced by various assessment methods. Radiography 2003;9(3):185-91.

Marshall KG, Brailovsky CA, GrandMaison P. French-English, English-French translation process of an objective structured clinical examination (OSCE) used for licensing family physicians in Quebec. Teaching and Learning in Medicine 1995;7 (2):115-20.

Martin IG, Stark P, Jolly B. Benefiting from clinical experience: the influence of learning style and clinical experience on performance in an undergraduate objective structured clinical examination. Medical Education 2000;34(7):530-4.

Martin JA, Reznick RK, Rothman A, Tamblyn RM, Regehr G. Who should rate candidates in an Objective Structured Clinical Examination? Academic Medicine 1996;71 (2):170-5.

Martinez-Carretero JM, Roma-Milan J, Pujol-Farriols R, Blay-Pueyo C, Nogueras-Rimblas A, Brailovsky CA, GrandMaison P. Implementing a Multiple Station Clinical Examination: An International Collaboration. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;452-3.

Martinez-Carretero JM, Borrell-Carrio F, Blay-Puelo C, Roma-Milan J, Pujol-Farriols R, Gomez-Saez JM, Masana-Marin L, Sanchis-Aldas J. Clinical Skills Assessment (CSA) in Undergraduate Medical Education in Catalonia, 1995. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;454-7.

Martinez-Carretero JM, Blay C, Descarrega R, Iruela A, Kronfly E, Barragan N, Pujol R, Serrallach S, Sola M, Pulpon A. Clinical skills assessment: a strategy for improving undergraduate and postgraduate medical education in Catalonia (Spain). Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.

Martz AP, Gessaroli ME, Swanson DB, De Champlain AF. Equating Standardized Patient cases Using Structural Equation Modeling. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-9.

Mason S, Fletcher A, McCormick S, Perrin J, Rigby A. Developing assessment of emergency nurse practitioner competence--a pilot study. Journal of Advanced Nursing 2005;50(4):425-32.

Massey D, Papineau L. The Comprehensive Objective Examination (COE). Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;480-1.

Mathews L, Menon J, Mani NS. Micro-OSCE for assessment of undergraduates. Indian Pediatrics 2004;41(2):159-63.

Matsell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured clinical examination in paediatrics. Medical Education 1991;25(4):293-9.

Mau W, Kusak G. Umsetzung der neuen Approbationsordnung fur Arzte im Querschnittsbereich "Rehabilitation, Physikalische Medizin und Naturheilverfahren" an den medizinischen Fakultaten in Deutschland.[Implementation of the new Federal Medical Licensing Regulations for doctors in the interdisciplinary subject "Rehabilitation, physical medicine, naturopathic treatment" by the German medical faculties]. [German]. Rehabilitation 2005;44(3):129-33.

Maudsley R. Assessment of International Medical Graduates and Their Integration into Family Practice: The Clinician Assessment for Practice Program. Academic Medicine 2008;83:309-15.

Mavis B. Relationship between Confidence and Clinical Performance among Second Year Medical Students. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-5.

Mavis B. Self-efficacy and OSCE performance among second year medical students. Advances in Health Sciences Education 2001;6(1):93-102.

Mavis BE, Henry RC, Ogle KS, Hoppe RB. The emperor's new clothes: the OSCE reassessed. Academic Medicine 1996;71(5):447-53.

Mavis BE, Henry RC, Ogle KS, Hoppe RB. Reassessing the OSCE LETTER - REPLY. Academic Medicine 1996;71(11):1138.

Mavis BE, Lovell KL, Ogle KS. Why Johnnie can't apply Neuroscience: testing alternative hypotheses using performance-based assessment. Advances in Health Sciences Education 1998;3(3):165-75.

Mavis BE. Does studying for an objective structured clinical examination make a difference? Medical Education 2000;34(10):808-12.

Mazor K, Gammon W, Jacobson E, Regan MB. Reducing range restriction in standardized patients' ratings of interview skills. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.

Mazor K, Haley H, Sullivan K, Quirk ME. The video-based test of communication skills: description, development and preliminary findings. Teaching and Learning in Medicine 2007;19(2):162-7.

Mazor KM, Ockene JK, Rogers HJ, Carlin MM, Quirk ME. The relationship between checklist scores on a communication OSCE and analogue patients' perceptions of communication. [see comment]. Advances in Health Sciences Education 2005;10(1):37-51.

Mazor KM, Zanetti ML, Alper E, Hatem D, Barret S, Meterko V, Gammon W, Pugnaire MP. Assessing professionalism in the contect of an objective structured clinical examination: an indepth study of the rating process. Medical Education 2007;41(4):331-40.

McAleer S, Walker R. Objective structured clinical examination OSCE. Journal of the Royal College of General Practitioners 1990;46:3942.
McCann AL, Campbell PR, Schneiderman ED. A performance examination for assessing dental hygiene competencies. Journal of Dental Hygiene 2001;75(4):291-304.
McCauley M. The objective structured clinical examination. Psychiatric Bulletin 2005;29(6):232-3.
McClean KL, Card SE. Informed consent skills in internal medicine residency: how are residents taught, and what do they learn? Academic Medicine 2004;79(2):128-33.
McFaul PB, Howie PW. The assessment of clinical competence in obstetrics and gynaecology in two medical schools by an objective structured clinical examination. British Journal of Obstetrics and Gynaecology 1993;100(9):842-6.
McGaghie WC, Renner BR, Kowlowitz V, Sauter SVH, Hoole AJ, Schuh MS. Development and evaluation of musculoskeletal performance measures for an objective structured clinical examination. Teaching and Learning in Medicine 1994;6(1):59-63.
McGaughey J. Standardizing the assessment of clinical competence: an overview of intensive care course design. Nursing in Critical Care 2004;9(5):238-46.
McGill R, Fouracre R, Evans A, Bahrami J. A regional objective structured clinical examination-development and evaluation. Education for General Practice 1998;10(1):29-39.
McGowan JJ, Bangert M, Ballinger SH, Highbaugh S. Implementing Wireless Evaluation in a Hospital-based OSCE Center. Proceedings / AMIA Annual Symposium 2003;930.
McGraw RC, OConnor HM. Standardized patients in the early acquisition of clinical skills. Medical Education 1999;33(8):572-8.
McGuire C. Perspectives in assessment. Academic Medicine 1993;68(2 Suppl):S3-S8.
McKinley DW, Boulet JR. Detecting score drift in a high-stakes performance-based assessment. Advances in Health Sciences Education 2004;9(1):29-38.
McKinley DW, Boulet JR, Hambleton RK. A work-centered approach for setting passing scores on performance-based assessments. Evaluation and the Health Professions 2005;28(3):349-69.
McKinley DW, Boulet JR. Using factor analysis to evaluate checklist items. Academic Medicine 2005;80(10 Suppl):S102-S105.
McKinley RK, Strand J, Gray T, Schuwirth L, Alun-Jones T, Miller H. Development of a tool to support holistic generic assessment of clinical procedure skills. Medical Education 2008;42:619-27.
McKnight J, Rideout E, Brown B, Ciliski D, Patton D, Rankin J, Woodward C. The Objective Structured Clinical Examination: an alternative approach to assessing student clinical performance. Journal of Nursing Education 1987;26(1):39-41.
McLaughlin K, Heemskerk L, Herman R, Ainslie M, Rikers R, Schmidt H. Initial diagnostic hypotheses bias analytic information processing in non-visual domains. Medical Education 2008;42:496-502.
McLay RN, Rodenhauser P, Anderson DS, Stanton ML, Markert RJ. Simulating a full-length psychiatric interview with a complex patient. Academic Psychiatry 2002;26(3):162-7.
Mcllroy JH. The impact of an alternative approach to computing station cut scores in an OSCE. Academic Medicine 2000;75(10 Suppl):S18-S20.
McMahon RFT, Benbow EW. Designing assessment of pathology in the undergraduate curriculum. Diagnostic Histopathology 2008;14(9):453-8.
McManus IC, Thompson M, Mollon J. Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. BMC Medical Education 2006;6(42):1-22.
McNaughton N, Hodges B, Tiberius R. The Effects of Portraying Psychologically and Emotionally Complex OSCE Roles. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
McNaughton N, Tiberius R, Hodges B. Effects of portraying psychologically and emotionally complex standardized patient roles. Teaching and Learning in Medicine 1999;11(3):135-41.
McNaughton N, Ravitz P, Wadell A, Hodges BD. Psychiatric Education and Simulation: A Review of the Literature. The Canadian Journal of Psychiatry 2008;53(2):85-93.
Meadow R. The structured exam has taken over. British Medical Journal 1998;317:1329-30.
Melding P, Coverdale J, Robinson E. A 'fair play'? Comparison of an objective structured clinical examination of final year medical students training in psychiatry and their supervisors' appraisals. Australasian Psychiatry 2002;10(4):344-7.
Mercer C, Holsgrove G. Developing In-Course and Final Assessments in a New Dental Curriculum. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;374-6.
Merrick HW, Nowacek G, Boyer J, Robertson J. Comparison of the Objective Structured Clinical Examination with the performance of third-year medical students in surgery. American Journal of Surgery 2000;179(4):286-8.
Miller E, Green A. Student reflections on learning cross-cultural skills through a 'cultural competence' OSCE. Medical Teacher 2007;29(4):e76-84.
Miller GE. Commentary on 'assessment of clinical skills with standardized patients: state of the art'. Teaching and Learning in Medicine 1990;2(2):77-8.
Miller GE. AAMC consensus summary. Academic Medicine 1993;68(6):471-4.
Mohtadi NGH, Harasym PH, Pipe AL, Strother RT, Mah AF. Using an objective structured clinical examination to evaluate competency in sport medicine. Clinical Journal of Sport Medicine 1995;5(2):82-5.
Monaghan MS, Vanderbush RE, Allen RM, Heard JK, Cantrell M, Randall J. Standardized patient use outside of academic medicine: opportunities for collaboration between medicine and pharmacy. Teaching and Learning in Medicine 1998;10(3):178-82.
Mooney GA, Bligh JG, Leinster SJ. Some techniques for computer-based assessment in medical education. Medical Teacher 1998;20(6):560-6.
Morag E, Lieberman G, Volkan K, Shaffer K, Novelline R, Lang EV. Clinical competence assessment in radiology: introduction of an objective structured clinical examination in the medical school curriculum. Academic Radiology 2001;8(1):74-81.
Morgan PJ, Cleave-Hogg D, Guest CB. A comparison of global ratings and checklist scores from an undergraduate assessment using an anaesthesia simulator. Academic Medicine 2001;76(10):1053-5.
Morrison EH, Hafler JP. Yesterday a learner, today a teacher too: residents as teachers in 2000. Pediatrics 2000;105(1):238-41.
Morrison H, McNally H, Wylie C, McFaul P, Thompson W. The passing score in the Objective Structured Clinical Examination. Medical Education 1996;30(5):345-8.
Mortimer AM, Lunn B. Part I OSCE examinations. Psychiatric Bulletin 2004;28(12):458.
Moss F. Simulation and team training.[Special issue]. Quality and Safety in Health Care 2004;13(Suppl 1):i1-i104.
Moss JH. Evaluating a seminar designed to improve psychiatry skills of family medicine residents. Academic Medicine 1990;65(10):658-60.
Mossey P, Newton J, Stirrups D. The Authors respond. British

Dental Journal 2001;191(3):118-9.
Mossey PA, Newton JP, Stirruos DR. Scope of the OSCE in the assessment of clinical skills in dentistry. British Dental Journal 2001;190(6):323-6.
Mossey PA, Newton JP. The structured clinical operative test SCOT in dental competency assessment. British Dental Journal 2001;190(7):387-90.
Moule P, Wilford A, Sales R, Lockye L. Student experiences and mentor views of the use of simulation for learning. Nurse Education Today 2008;28(7):790-7.
Moule P, Wilford A, Sales R, Lockye L. Integrative Thinking and Learning in Undergraduate Nursing Education: Three Strategies. International Journal of Nursing Education Scholarship 2008;5(1):1-15.
Muijtjens AMM, Van Vollenhoven FHM, Van Luijk SJ, Van der Vleuten CPM. Sequential testing in the assessment of clinical skills. Academic Medicine 2000;75(4):369-73.
Mukohara K, Kitamura K, Wakabayashi H, Abe K, Sato J, Ban N. Evaluation of a communication skills seminar for students in a Japanese medical school: A non-randomized controlled study. BMC Medical Education 2004;4(24):1-6.
Murdoch ED, Cottrell D. Structured teaching methods enhance skill acquisition but not problem-solving abilities: an evaluation of the 'silent run through'. Medical Education 1999;33(1):19-23.
Murphy DJ, Bruce D, Eva KW. Workplace-based assessment for general practitioners: using stakeholder perception to aid blueprinting of an assessment battery. Medical Education 2008;42(1):96-103.
Murray E, Jolly B, Modell M. Evaluation of the Effectiveness of Clinical Skills Teaching in the Community. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;564-6.
Murray E, Jolly B, Modell M. Can students learn clinical method in general practice? A randomised crossover trial based on objective structured clinical examinations. British Medical Journal 1997;315(11):920-3.
Murray E, Todd C, Modell M. Can general internal medicine be taught in general practice? An evaluation of the University College London model. Medical Education 1997;31(5):369-74.
Murthy M, Kothari N, Abate H. A Women's Health Objective Structured Clinical Examination for Internal Medicine Residents. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Murthy M, Abate H, Kothari N. The Assessment of Nutritional Clinical Competence Using an Objective Structured Clinical Examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Naeem A, Rutherford J, Kenn C. The MRCPsych OSCE workshop: A new game to play? Psychiatric Bulletin 2004;28(2):62-5.
Nagoshi M, Williams S, Kasuya R, Sakai D, Masaki K, Blanchette PL. Using standardized patients to assess the geriatrics medicine skills of medical students, internal medicine residents, and geriatrics medicine fellows. Academic Medicine 2004;79(7):698-702.
Nalesnick SW, Mills CS, Olsen CH, Haffner WH, Zahn CM. Creating an ideal objective structured clinical exam for an obstetrics and gynecology medical student clerkship. American Journal of Obstetrics & Gynecology 2005;193(4):1544-50.
Narula A. The objective structured clinical examination. Psychiatric Bulletin 2005;29(2):72-3.
Nayar U, Malik SL, Bijlani RL. Objective structured practical examination: a new concept in assessment of laboratory exercises in preclinical sciences. Medical Education 1986;20(3):204-9.
Nayer M. An overview of the Objective Structured Clinical Examination. Physiotherapy Canada 1993;45(3):171-8.
Nayer M. The assessment of clinical competency: an overview and preliminary report of Canadian physiotherapy programs. Physioterapy Canada 1995;47(3):190-9.
Newble D, Dauphinee D, Macdonald M, Mulholland H, Dawson B, Page G, Swanson D, Thomson A, Van der Vleuten C. Guidelines for assessing clinical competence. Teaching and Learning in Medicine 1994;6(3):213-20.
Newble D, Swanson D. Improving the quality of a multidisciplinary test of clinical competence: a longitudinal study. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-5.
Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. Medical Education 2004;38(2):199-203.
Newble DI, Elmslie RG, Baxter A. A problem based criterion referenced examination of clinical competence. Journal of Medical Education 1978;53(720):726.
Newble DI, Hoare J, Elmslie RG. The validity and reliability of a new examination of the clinical competence of medical students. Medical Education 1981;15(1):46-52.
Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. Medical Education 1988;22(4):325-34.
Newble DI. Eight years' experience with a structured clinical examination. Medical Education 1988;22(3):200-4.
Newble DI. The observed long-case in clinical assessment. Medical Education 1991;25(5):369-73.
Nicol M, Freeth D. Assessment of clinical skills: a new approach to an old problem. Nurse Education Today 1998;18(8):601-9.
Niehaus AH, DaRosa DA, Markwell SJ, Folse R. Is test security a concern when OSCE stations are repeated across clerkship rotations? Academic Medicine 1996;71 (3):287-9.
Nikendei C, Schilling T, Nawroth P, et al. Integriertes Skills-Lab-Konzept fur die studentische Ausbildung in der Inneren Medizin. [Integrated skills laboratory concept for undergraduate training in internal medicine]. [German]. Deutsche Medizinische Wochenschrift 2005;130(18):1133-8.
No Authors Reported. Evaluation of students achievement. Academic Medicine 1993;68(6 Suppl):S23-S27.
Norman G. Checklists vs. Ratings, the Illusion of Objectivity, the Demise of Skills and the Debasement of Evidence. Advances in Health Sciences Education 2005;10(1):1-3.
Norman GR, Davis DA, Lamb S, Hanna E, Caulford P, Kaigas T. Competency assessment of primary care physicians as part of a peer review program. Journal of the American Medical School 1993;270(9):1046-51.
Nowotny RE, Grove DI. Description of an examination for the objective assessment of history-taking ability. Medical Education 1982;16(5):259-63.
Nuutinen M, Vainionpaa L, Kokkonen J, Rantala H, Tapanainen P, Mottonen M, Vayrynen M, Uhari M. Practice-orientated evaluation of medical students during a pediatric course. Medical Teacher 1997;19(3):200-4.
Nyquist JG, Naylor AJ, Woodward-Lopez G, Dixon S. Use of performance-based assessment to evaluate the impact of a skill-orientated continuing education program. Academic Medicine 1994;69(10 Suppl):S51-S53.
O'Sullivan P, Chao S, Russell M, Levine S, Fabiny A. Development and Implementation of an Objective Structured Clinical Examination to Provide Formative Feedback on Communication and Interpersonal Skills in Geriatric Training. Journal of the American Geriatrics Society 2008;56(9):1730-5.
OBrien MK, Feldman D, Alban T, Donoghue G, Sirkin J, Novack

DK. An innovative CME program in cardiology for primary care practitioners. Academic Medicine 1996;71(8):894-7.
OConnor HM, McGraw R. Clinical skills training*: developing objective assessment instruments. Medical Education 1997;31(5):359-63.
OConnor M, McGraw R, Killen L, Reich D. A computer-based self-directed training module for basic suturing [1]. Medical Teacher 1998;20(3):203-6.
Ogden GR, Green M, Ker JS. The use of interprofessional peer examiners in an objective structured clinical examination: can dental students act as examiners? British Dental Journal 2000;189(3):160-4.
OGorman EC, McBride M, McClure N. Communication skills training in the area of human sexuality. Sexual and Marital Therapy 1997;12(4):377-80.
Ogur B, Hirsh D, Krupat E, Bor D. The Harvard Medical School-Cambridge Integrated clerkship: an innovative model of clinical education. Academic Medicine 2007;82(4):397-404.
Ohyama A, Nitta H, Shimizu C, Ohara S, Araki K, Kurosaki N, Mataki S. [Educative effect of feedback after medical interview in objective structured clinical examination]. [Japanese]. Kokubyo Gakkai Zasshi - the Journal of the Stomatological Society, Japan 2005;72(1):71-6.
ONeill A, McCall JM. Objectively assessing nursing practices: a curricular development. Nurse Education Today 1996;16(2):121-6.
ONeill PN. Assessment of students in a problem based learning curriculum. Journal of Dental Education 1998;62(9):640-3.
Onishi H, Yoshida I. Rapid change in Japanese medical education. Medical Teacher 2004;26(5):403-8.
Orlander JD, Bor DH, Strunin L. A structured clinical feedback exercise as a learning-to-teach practicum for medical residents. Academic Medicine 1994;69(1):18-20.
Othman. The Teaching of Anatomy in the Integrated Medical Curriculum as Practised at the School of Medical Sciences, University Sains Malaysia. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;111-3.
Palarm, T.W, Griffiths M, Phillips R. The design, implementation and evaluation of electronic objective structured clinical examinations. Journal of Diagnostic Radiography & Imaging 2006;5(2):79-87.
Pangaro LN, Worth-Dickstein H, Macmillan MK, Klass DJ, Shatzer JH. Performance of standardized examinees in a standardized-patient examination of clinical skills. Academic Medicine 1997;72(11):1008-11.
Papp KK, Strohl KP. The effects of an intervention to teach medical students about obstructive sleep apnea. Sleep Medicine 2005;6(1):71-3.
Park RS, Chibnall JT, Blaskiewicz RJ, Furman GE, Powell JK, Mohr CJ. Construct validity of an objective structured clinical examination (OSCE) in psychiatry: associations with the clinical skills examination and other indicators. Academic Psychiatry 2004;28(2):122-8.
Park RS, Chibnall JT, Morrow A. Relationship of rotation timing to pattern of clerkship performance in psychiatry. Academic Psychiatry 2005;29(3):267-73.
Patel L, David TJ, Boshuizen HPA, Wolfhagen HAP. Implementation, students' perceptions and students' performance in problem-based learning and traditional paediatric clerkships. Education for Health 1998;11(2):215-23.
Patil JJP. Objective-Structured Clinical Examination. Canadian Medical Association Journal 1993;149(10):1376.
Patil NG, Saing H, Wong J. Role of OSCE in evaluation of practical skills. Medical Teacher 2003;25(3):271-2.
Paul CR, Devries J, Fliegel J, Van Cleave J, Kish J. Evaluation of a Culturally Effective Health Care Curriculum Integrated into a Core Pediatric Clerkship. Ambulatory Pediatrics 2008;8(3):195-9.
Paul VK. Assessment of clinical competence of undergraduate medical students. Indian Journal of Pediatrics 1994;61(2):145-51.
Payne NJ, Bradley EB, Heald EB, Maughan KL, Michaelsen VE, Wang X, Corbett EC. Sharpening the Eye of the OSCE with Critical Action Analysis. Academic Medicine 2008;83(10):900-5.
Peden NR, Cairncross RG, Harden RM, Crooks J. Assessment of clinical competence in therapeutics: the use of he Objective Structured Clinical Examination. Medical Teacher 1985;7(2):217-23.
Peeraer G, Scherpbier A, Remmen R, De winter B, Hendrickx K, van Petegem P, Weyler J, Bossaert L. Clinical Skills Training in a Skills Lab Compared with Skills Training in Internships: Comparison of Skills Development Curricula. Education for Health 2007;20(3):1-9.
Peeraer G, Muijtjens AMM, De Winter BY, Remmen R, Hendrickx K, Bossaert L, Scherpbier AJJA. Unintentional failure to assess for experience in senior undergraduate OSCE scoring. Medical Education 2008;42:669-75.
Peitzman SJ. Clinical skills assessment using standardized patients. American Journal of Physical Medicine Rehabilitation 2000;79(5):490-3.
Peitzman SJ. Physical Diagnosis findings among persons applying to Work as Standardized Patients. Academic Medicine 2001;76(4):383.
Pender FT, De Looy AE. The testing of clinical skills in dietetic students prior to entering clinical placement. Journal of Human Nutrition & Dietetics 2004;17(1):17-24.
Pereira J, Palacios M, Collin T, Wedel R, Galloway L, Murray A, Violato C, Lockyer J. The impact of a hybrid online and classroom-based course on palliative care competencies of family medicine residents. Palliative Medicine 2008;22:929-37.
Peskun C, Detsky A, Shandling M. Effectiveness of medical school admissions criteria in predicting residency ranking four years later. Medical Education 2007;41(1):57-64.
Petrocco-Napuli K, Merkle P. Preparing learners for National Board of Chiropractic Examiners (NBCE) Part IV Objective Structured Clinical Examination utilizing blended learning... ACC Conference. Journal of Chiropractic Education 2007;21 (1):77-8.
Petrusa ER, Blackwell T, Carline J, Ramsay P, McGaghie W, Colindres R, Mast TA, Soler NG. Symposium: improving clinical performance assessment: a multi-institutional trial of the OSCE. Proceedings of the Annual Conference on Research in Medical Education 1987;26:267-74.
Petrusa ER, Blackwell TA, Rogers LP, Saydari C, Parcel S, Guckian JC. An Objective measure of clinical performance. American Journal of Medicine 1987;83(7):34-42.
Petrusa ER, Blackwell TA, Ainsworth MA. Reliability and validity of an Objective Structured Clinical Examination for assessing the clinical performance of residents. Archives of Internal Medicine 1990;150(3):573-7.
Petrusa ER, Blackwell TA, Carline J, Ramsey PG, McGaghie W, Colindres R, Kowlowitz V, Mast TA, Soler N. A multi institutional trial of an objective structured clinical examination. Teaching and Learning in Medicine 1991;3(2):86-94.
Pfeiffer CA, Ardolino AJ, Madray H. The impact of a Curriculum Renewal Project on Students' Performances on a Fourth-year Clinical Skills Assessment. Academic Medicine 2001;76(2):173-5.
Philp E, Norton P, Hartman J. The effect of patient gender and race in counselling received during an OSCE. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-5.
Piccinato CE, Figueiredo JFC, Troncon LEA, Peres LC, Cianflone AR, Colares MDF, Rodrigues MDL. Assessment of educational objectives in two different curricular structures. Revista Da Associacao Medica Brasileira 2004;50(1):68-73.
Pierre RB, Wierenga A, Barton M, Branday JM, Christie CDC.

Student evaluation of an OSCE in paediatrics at the University of the West Indies, Jamaica. BMC Medical Education 2004;4:1-7.
Pierre RB, Wierenga A, Barton M, Thame K, Branday JM, Christie CD. Student self-assessment in a paediatric objective structured clinical examination. West Indian Medical Journal 2005;54(2):144-8.
Pocock I. A new route for dental graduates. Dental Update 2007;34(1):59.
Poenaru D, Morales D, Richards A, OConnor M. Running an objective structured clinical examination on a shoestring budget. American Journal of Surgery 1997;173(6):538-41.
Poenaru D, Davidson L, Donnely M, Tekian A. Is a mandatory general surgery rotation necessary in the surgical clerkship? American Journal of Surgery 1998;175(6):515-7.
Poldre P, Smee S, Reznick R, et al. The experience of thousands: the post examination OSCE station review process of the medical council of Canada. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-3.
Portanova R, Adelman M, Jollick JD, Schuler S, Modrzakowski M, Soper E, Ross-Lee B. Student assessment in the Ohio University College of Osteopathic Medicine CORE system: progress testing and objective structured clinical examinations. Journal of the American Osteopathic Association 2000;100(11):707-12.
Povar GJ. Evaluating competence in clinical ethics: is the OSCE the answer? Journal of General Internal Medicine 1994;9(12):709-10.
Prakash R. Undergraduate Nursing Evaluation: The OSCE Approach. The Nursing Journal of India 1999;90(5):101-4.
Price J, Byrne GJA. The direct clinical examination: an alternative method for the assessment of clinical psychiatry skills in undergraduate medical students. Medical Education 1994;28(2):120-5.
Prislin MD, Fitzpatrick CF, Lie D, Giglio M, Radecki S, Lewis E. Use of an Objective Structured Clinical Examination in Evaluating Student Performance. Family Medicine 1998;30(5):338-44.
Prislin MD, Fitzpatrick C, Giglio M, Lie D, Radecki S. Initial experience with a multi-station Objective Structured Teaching Skills Evaluation. Academic Medicine 1998;73(10):1116-8.
Prislin MD, Fitzpatrick CF, Radecki S. A comparison of Family Medicine Clerkship Student Performance Across Multiple Teaching Sites. Family Medicine 1998;30(4):279-82.
Prislin MD, Giglio M, Lewis EM, Ahearn S, Radecki S. Assessing the acquisition of core clinical skills through the use of serial standardized patient assessments. Academic Medicine 2000;75(5):480-3.
Prislin MD, Lie D, Shapiro J, Boker J, Radecki S. Using standardized patients to assess medical students professionalism. Academic Medicine 2001;76(10 Suppl):S90-S92.
Probert CS, Cahill DJ, McCann GL, Ben-Shlomo Y. Traditional finals and OSCEs in predicting consultant and self-reported clinical skills of PRHOs: a pilot study. Medical Education 2003;37(7):597-602.
Pryde I, Sachar A, Young S, Hukin A, Davies T, Rao R. Organising a mock OSCE for the MRCPsych Part I examination. Psychiatric Bulletin 2005;29(2):67-70.
Pujol-Farriols R, Nogueras-Rimblas A, Vilardell-Tarres M, Blay-Pueyo C, Roma-Milan J, Martinez-Carretero JM. Development of the Catalan OSCE of Internal Medicine. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;485-7.
Pujol R, Artigas V, Delas J, et al. Clinical Skills Assessment in Medical Schools in Catalonia (Spain), 1997. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-3.
Pulpon A, Caja C, Gispert R, Martinez-Carretero JM, Pedreny R, Sola M. Clinical Skills Assessment of Catalan Nursing: A Collaborative Project with Nursing Schools. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998.
Pulpon SAM, Wola PM, Martinez CJM, Gispert MR. Evaluating clinical competence: the immediate future for nursing. Revista Rol de Enfermeria 1999;22(6):475-8.
Qayumi AK, Kurihara Y, Imai M, et al. Comparison of computer-assisted instruction (CAI) versus traditional textbook methods for training in abdominal examination (Japanese experience). Medical Education 2004;38(10):1080-8.
Quero Munoz L, O'Byrne C, Pugsley J, Austin Z. Reliability, validity, and generalizability of an objective structured clinical examination (OSCE) for assessment of entry-to-practice in pharmacy. Pharmacy Education 2005;5(1):33-43.
Raga S, Coovadia HM. Experience with three identical objective structured clinical examinations (OSCEs) conducted simultaneously for an entire final-year class. South African Medical Journal 1985;68(11):819-20.
Ragucci KR, Fermo JD, Mazur JN. Objective structured clinical examinations for an ambulatory care pharmacy rotation. American Journal of Health System Pharmacy 2005;62(9):927-9.
Rahman SA, Islam N. Objective Structured Clinical Examination for the House Officers. Indian Journal of Pediatrics 1997;64(1):123-5.
Rahman SA. Promoting learning outcomes in paediatrics through formative assessment. Medical Teacher 2001;23(5):467-70.
Raj N, Badcock L, Brown GA, Deighton C, O'Reilly SC. Design and validation of 2 objective structured clinical examination stations to assess core undergraduated examination skills of the hand and knee. Journal of Rheumatology 2007;34(2):421-4.
Ram P, Van der Vleuten C, Rethans JJ, Schouten B, Hobma S, Grol R. Assessment in general practice: the predictive value of written-knowledge tests and a multiple - station examination for actual medical performance in daily practice. Medical Education 1999;33(3):197-203.
Ram P, Van der Vleuten C, Rethans JJ, Grol R, Aretz K. Assessment of practicing family physicians: comparison of observation in a multiple-station examination using standardized patients with observation of consultations in daily practice. Academic Medicine 1999;74(1):62-9.
Ramani S. Promoting the art of history taking. Medical Teacher 2004;26(4):374-6.
Ramasamy P, Osman A. The medical school curriculum at University Malaysia Sabah. Medical Journal of Malaysia 2005;60(Suppl D):58-65.
Ramchandani D. End of Third-Year Objective Structured Clinical Examination: Boon or Bane? Academic Psychiatry 2008;32(3):173-6.
Rao NG. OSPE - A proposal for innovative changes in UG practical forensic medicine examination. Medico-Legal Update 2005;5(1):11-5.
Rao SP, Bhusari PS. Evaluation of disability knowledge and skills among leprosy workers. Indian Journal of Leprosy 1992;64(1):99-104.
Rashid M. Performance of clinical clerks doing paediatric rotations in a community hospital versus a university hospital. Paediatric Child Health 2007;12(9):761-4.
Razali SM. Objective structured clinical examination in psychiatry: a video tape station. Education for Health 1998;11(3):391-2.
Reddy S, Vijayakumar S. Evaluating clinical skills of radiation oncology residents: part II and I. International Journal of Cancer 2000;90(1):1-12.
Reece A, Chung EMK, Gardiner RM, Williams SE. Competency domains in an undergraduate Objective Structured Clinical Examination: their impact on compensatory standard setting.

Medical Education 2008;42:600-6.
Reed S. Canadian Competence. Nursing Times 1992;88(3):57-9.
Regehr G, Szalay D, Reznick R. Forced choice rankings of clinical performance: A validation tool. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-6.
Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE - format examination. Academic Medicine 1998;73(9):993-7.
Regehr G, Freeman R, Robb A, Missiha N, Heissey R. OSCE Performance evaluations made by standardized patients: comparing checklists and global ratings scores. Academic Medicine 1999;74(10 Suppl):S135-S137.
Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalizability of OSCE measures across content domains. Academic Medicine 1999;74(12):1320-2.
Reiter HI, Rosenfeld J, Nandagopal K, Eva KW. Do clinical clerks provide candidates with adequate formative assessment during Objective Structured Clinical Examinations?. Advances in Health Sciences Education 2004;9(3):189-99.
Reiter HI, Eva KW, Rosenfeld J, Norman G. Multiple mini-interviws predict clerkeship and licensing examination performance. Medical Education 2007;41(4):378-84.
Remmen R, Scherpbier A, Derese A, Denekens J, Hermann I, Van der Vleuten C, Van Royen P, Bossaert L. Unsatisfactory basic skills performance by students in traditional medical curricula. Medical Teacher 1998;20(6):579-82.
Remmen R, Scherpbier A, Denekens JO, Derese A, Hermann I, Hoogenboom R, Van der Vleuten C, Van Royen P, Bossaert L. Correlation of a written test of skills and a performance based test:a study in two traditional medical schools. Medical Teacher 2001; 23(1):29-32.
Rentschlerd D, Eaton J, Cappielo J, McNally S, McWilliam P. Evaluation of undergraduate students using Objective Structured Clinical Evaluation. Journal of Nursing Education 2007;46(3):135-9.
Rezler AG, Woolliscroft JA, Kalishman SG. What is missing from patient histories? Medical Teacher 1991;13(3):245-52.
Reznick R, Smee S, Rothman A, et al. An objective structured clinical examination for the licentiate: report of the pilot project of the medical council of Canada. Academic Medicine 1992;67(8):487-94.
Reznick R, Regehr G, MacRae H, McCulloch-Leadbetter W. An Objective Structured Assessment of Technical Skills: Reliability and Validity of Testing Residents from Four Surgical Programs. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Reznick RK, Smee S, Baumber JS, Cohen R, Rothman A, Blackmore D, Berard M. Guidelines for estimating the real cost of an Objective Structured Clinical Examination. Academic Medicine 1993;68(7):513-7.
Reznick RK, Blackmore D, Dauphinee WD, Rothman AI, Smee S. Large-scale high-stakes testing with an OSCE: report from the medical council of Canada. Academic Medicine 1996;71(1 Suppl):S19-S21.
Reznick RK, Blackmore DE, Dauphinee WD, Smee SM, Rothman AI. An OSCE for Licensure: The Canadian Experience. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;458-61.
Reznick RK, Regehr G, Yee G, Rothman A, Blackmore D, Dauphinee D. Process-rating forms versus task-specific checklists in an OSCE for Medical Licensure. Academic Medicine 1998;73(10 Suppl):S97-S99.
Ribin NJ, Philp EB, Hartman JA. Health Care Perceptions: Do They Change for the Standardized Patient? Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;711-3.
Ribin NJ, Philip EB. Health care perceptions of the standardized patient. Medical Education 1998;32(5):538-42.
Richards B, Gray L, Rogers JC. Validity of performance assessments: an exploratory study of the relationships between perceptions of behaviour, similarity, preference, and satisfaction. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Richards BF, Philp EB, Philp JR. Scoring the objective structured clinical examination using a microcomputer. Medical Education 1989;23(4):376-80.
Ro YS, Shin YI. Erythema multiforme due to topical application of diphenycyclopropenone (DPCP). Korean Journal of Dermatology 2000;38(8):1130.
Robb A, Etchells E, Cusimano MD, Cohen R, Singer PA, McKneally M. A randomized trial of teaching bioethics to surgical residents. American Journal of Surgery 2005;189(4):453-7.
Robb KV, Rothman AI. The assessment of clinical skills in general medical residents - comparison of the objective structured clinical examination to a conventional oral examination. Annals of the Royal College of Physicians and Surgeons of Canada 1985;18(3):235-8.
Robbins L, Hoke M. Using Objective Structured Clinical Examinations to Meet Clinical Competence Evaluation Challenges With Distance Education Students. Perspectives in Psychiatric Care 2008;44(2):81-8.
Roberts J, Brown B. Testing the OSCE: A reliable measurement of clinical nursing skills. The Canadian Journal of Nursing Research 1990;22(1):51-9.
Roberts J, Norman G. Reliability and learning from the objective structured clinical examination. Medical Education 1990;24(3):219-23.
Roberts KC. A validity and reliability study of the objective structured clinical examination. Dissertation Abstracts International A: Humanities and Social Sciences 2008;68(8A):3362.
Robertson EM. The relationship between critical thinking, personality type, and performance on the Objective Structured Clinical Examination of senior medical students at Meharry Medical College (Tennessee). Dissertation Abstracts International A: Humanities and Social Sciences 2001;62(3A):869-A.
Robins LS, White CB, Alexander GL, Gruppen LD, Grum CM. Assessing medical students' awareness of and sensitivity to diverse health beliefs using a standardized patient station. Academic Medicine 2001;76(1):76-80.
Rogers JC, Dains JE. Can first-year students master clinical skills? Academic Medicine 2001;76(10):1065.
Rogers PL, Jacob H, Thomas EA, Harwell M, Willenkin RL, Pinsky MR. Medical students can learn the basic application, analytic, evaluative, and psychomotor skills of critical care medicine. Critical Care Medicine 2000;28(2):550-4.
Rogers PL, Jacob H, Rashwan AS, Pinsky MR. Quantifying learning in medical students during a critical care medicine elective: a comparison of three evaluation instruments. Critical Care Medicine 2001;29(6):1268-73.
Rose M, Wilkerson L. Widening the Lens on Standardized Patient Assessment: what the encounter can reveal about the development of clinical competence. Academic Medicine 2001;76(8):856-9.
Rosebraugh CJ, Speer AJ, Solomon DJ, Szauter KE, Ainsworth MA, Holden MD, Lieberman SA, Clyburn EB. Setting standards and defining quality of performance in the validation of a standardized-patient examination format. Academic

Medicine 1997;72(11):1012-4.
Rosen GM, Harris I, Mahovald MW. Objective structured clinical examinations (OSCE) for sleep. Sleep Medicine 2005;6(1):75-80.
Rosenfeld J, Reiter HI, Trinh K, Eva KW. A cost Efficiency Comparison Between The Muiltiple Mini-Interview and Tradicional Admissions Interviews. Advances in Health Sciences Education 2008;13(1):43-58.
Ross JR, Syal S, Hutcheon MA, Cohen R. Second-year students' scores improvement during an objective structured clinical examination. Journal of Medical Education 1987;62(10):857-8.
Ross L, DeChamplain AF, Margolis MJ. Examining fairness issues for a large-scale standardized patient examination using structural equating modelling. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Ross M, Carroll G, Knight J, Chamberlain M, Fothergill-Bourbonnais F, Linton J. Using the OSCE to measure clinical skills performance in nursing. Journal of Advanced Nursing 1988;13(1):45-56.
Rothman AI, Cohen R, Ross J. Evaluating the clinical skills of foreign medical school graduates participating in an internship preparation program. Academic Medicine 1990;65(6):391-5.
Rothman AI, Cohen EDR, Dirks FR, Poldre P, Ross J. Validity and Reliability of a domain-referenced test of clinical competence for foreign medical graduates. Academic Medicine 1991;66(7):423-5.
Rothman AI, Cohen R, Dawson-Saunders B, Poldre PP, Ross J. Testing the equivalence of multiple-station tests of clinical competence. Academic Medicine 1992;67(10 Suppl):S40-S41.
Rothman AI, Cohen R, Ross J, Poldre P, Dawson B. Station gender bias in a multiple-station test of clinical skills. Academic Medicine 1995;70(1):42-6.
Rothman AI. Understanding the objective structured clinical examination. Australian and New Zealand Journal of Surgery 1995;65(3):302-3.
Rothman AI, Blackmore D, Cohen R, Reznick R. The consistency and uncertainty in examiners' definitions of pass/fail performance on OSCE stations. Evaluation and the Health Professions 1996;19(1):118-24.
Rothman AI, Cohen R, Bilan S. A comparison of short and long-case stations in a multiple-station test of clinical skills. Academic Medicine 1996;71(1 Suppl):S110-S112.
Rothman AI, Cohen R, Blackmore D. Differences unrelated to clinical competence in the results of repeated multiple-station tests of clinical skills. Academic Medicine 1997;72(4):296-300.
Rothman AI, Blackmore DE, Dauphinee WD, Reznick RK. The Use of Global Ratings in OSCE Station Scores. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;181-3.
Rothman AI, Cusimano M. A comparison of physician examiners', standardized patients' and communication experts' ratings of international medical graduates' english proficiency. Academic Medicine 2000;75(12):1206-11.
Rothman AI, Cusimano M. Assessment of English proficiency in international medical graduates by physician examiners and standardized patients. Medical Education 2001;35(8):762-6.
Roy V, Tekur U, Prabhu S. A comparative study of two evaluation techniques in pharmacology practicals: Conventional practical examination versus objective structured practical examination [5]. Indian Journal of Pharmacology 2004;36(6):386-8.
Rubin N, Philp EB. How Does SP Experience Impact the Sps' Views of Their Own Physician/Patient Encounters? Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-4.
Rudland J, Wilkinson T, Smith-Han K, Thomson-Fawcett M. "You can do it late at night or in the morning. You can do it at home, I did it with my flatmate." The educational impact of an OSCE. Medical Teacher 2008;30(2):2006-11.
Ruíz IG, Florensa IC, Cots IY, Sellares J, Inuela LABPC, Morera CR, Martinez CJ. First experience in evaluation of the clinical competence of family doctors in Catalonia. Atencion Primaria 2001;28(2):105-9.
Rushforth HE. Objective structured clinical examination (OSCE): review of literature and implications for nursing education. Nurse Education Today 2007;27(5):481-90.
Rushforth HE. Reflections on a study tour to explore history taking and physical assessment education. Nurse Education in Practice 2008;8:31-40.
Rutala PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ. Student fatigue as a variable affecting performance in an objective structured clinical examination. Academic Medicine 1990;65(9 Suppl):s53-s54.
Rutala PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ. Sharing of information by students in an Objective Structured Clinical Examination. Archives of Internal Medicine 1991;151(3):541-4.
Rutala PJ, Witzke DB, Leko EO, Fulginiti JV. The influences of student and standardized patient genders on scoring in an objective structured clinical examination. Academic Medicine 1991;66(9 Suppl):S28-S30.
Rutala PJ, Fulginiti JV, McGeagh AM, Leko EO, Koff NA, Witzke DB. Predictive validity of a required multidisciplinary standardized-patient examination. Academic Medicine 1992;67(10 Suppl):S60-S62.
Ryan S, Stevenson K, Hassell A. Assessment of clinical nurse specialists in rheumatology using an OSCE. Musculoskeletal Care 2007;5(3):119-29.
Sachdeva AK, Loiacono LA, Amiel GE, Blair PG, Friedman M, Roslyn JJ. Variability in the clinical skills of residents entering training programs in surgery. Surgery 1995;118(2):300-9.
Sadeghi M, Taghva A, Mirsepassi G, Hassanzadeh M. How Do Examiners and Examinees Think About Role-Playing of Standardized Patients in an OSCE Setting? Academic Psychiatry 2007;31(5):358-62.
Saleem M. Editorial another step ahead... Introduction of OSCE in FCPS-II Exam. Journal of the Royal College of Physicians and Surgeons Pakistan 1998;8(3):100.
Sampson F, Hutton G. Assessment of Potential and Potential of Assessment. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-9.
Sanchez SAS, Millan HI, Rodriguez VJA, et al. El aprendizaje basado en problemas en la enseñanza de la urologia. Modelo de la Facultad de Medicina de la Universidad de Castilla-La Mancha. [Problem-based learning in urology training. The Faculty of Medicine of the Universidad de Castilla-La Mancha model]. [Spanish]. Actas Urologicas Espanolas 2005;29(1):8-15.
Sarid O, Anson O, Bentov Y. Students' reactions to three typical examinations in health sciences. Advances in Health Sciences Education 2005;10(4):291-302.
Sasson V, Kallenberg G, Blatt B, Delaney M, White FS. Teach one, do one better: Superior Communication Skills in Senior Medical Students Serving as Standardized Patient/Examiners for their Junior Peers. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Sasson VA, Blatt B, Kallenberg G, Delaney M, White FS. Teach 1, Do 1... Better": superior communication skills in senior medical students serving as standardized patient - examiners for their junior peers. Academic Medicine 1999;74(8):932-7.
Satran L, Harris IB, Allen S, Anderson DC, Poland GA, Miller WL. Hospital - based versus community-based clinical education: comparing performances and course evaluations by students in their second-year pediatrics rotation. Academic

Medicine 1993;68(5):380-2.
Sauer J, Hodges B, Santhouse A, Blackhood N. The OSCE has landed: One small step for British psychiatry? Academic Psychiatry 2005;29(3):310-5.
Scheffer S, Muehlinghaus I, Froehmel A, Ortwein H. Assessing students' communication skills: validation of a global rating. Advances in Health Sciences Education 2008;13(5):583-92.
Schenk M, Popps S, Bridge P, Gallagher R, Petrusa ER, Frank RR. Effectiveness of an occupational and environmental medicine curriculum as indicated by evaluation of medical student performance on an Objective Structured Clinical Examination. Journal of Occupational and Environmental Medicine 1999;41(11):954-9.
Schluter B, Hofmann M, Rutzler M. Concept and contents of an Objective Structured Clinical Examination (OSCE) for pediatric sleep medicine. Somnologie 2007;11(1):35-46.
Schmidts MB. OSCE logistics - handheld computers replace checklists and provide automated feedback. Medical Education 2000;34(11):957-8.
Schmitz CC, Chipman JG, Luxenberg M, Beilman GJ. Professionalism and communication in the intensive care unit: reliability and validity of a simulated family conference. Simulation in Healthcare: The Journal of the Society for Medical Simulation 2008;3(4):224-38.
Schol S. A multiple-station test of the Teaching Skills of General Practice Preceptors in Flanders, Belgium. Academic Medicine 2001;76(2):176-80.
Schoonheim-Klein M, Walmsley AD, Habets L, Van der Velden U, Manogue M. An implementation strategy for introducing an OSCE into a dental school. European Journal of Dental Education 2005;9(4):143-9.
Schoonheim-Klein M, Hoogstraten J, Habets L, Aartman I, Van der Vleuten C, Manogue M, Van der Velden U. Language background and OSCE performance: a study of potential bias. European Journal of Dental Education 2007;11:222-9.
Schoonheim-Klein M, Muijtjens A, Habets L, Manogue M, Van der Vleuten C, Hoogstraten J, Van der Velden U. On the reliability of a dental OSCE, using SEM: effect of different days. European Journal of Dental Education 2008;12(3):131-7.
Schwartz LR, Fernandez R, Kouyoumjian SR, Jones KA, Compton S. A randomized comparison trial of case-based learning versus human patient simulation in medical student education. Academic Emergency Medicine 2007;14(2):130-7.
Schwartz RW, Donnelly MB, Sloan DA, Young B. Knowledge gain in a problem-based surgery clerkship. Academic Medicine 1994;69(2):148-51.
Schwartz RW, Donnelly MB, Sloan DA, Johnson SB, Strodel WE. Assessing senior residents' knowledge and performance: an evaluation program. Surgery 1994;116(4):634-40.
Schwartz RW, Donnelly MB, Sloan DA, Johnson SB, Strodel WE. The relationship between faculty ward evaluations, OSCE and ABSITE as measures of surgical intern performance. American Journal of Surgery 1995;169(4):414-7.
Seager P. OSCE: experience as a simulated candidate. Psychiatric Bulletin 2007;31:435-6.
Seager P. Simulated Patients. Psychiatric Bulletin 2007;31(8):316.
Searle J. Defining competency - the role of standard setting. Medical Education 2000;34(5):363-6.
Selby C, Osman L, Davis M, Lee M. Set up and run an objective structured clinical exam. British Medical Journal 1995;310(5):1187-90.
Selby G, Walker V, Diwakar V. A comparison of teaching methods: interactive lecture versus game playing. Medical Teacher 2007;29:972-4.
Sensi S, Guagnano MT. Assessment of clinical competence. The state of the Art. Recenti Progressi in Medicina 1996;87(9):445-51.
Sethuraman KR. The use of objective structured clinical examination (OSCE) for detecting and correcting teaching-learning errors in physical examination. Medical Teacher 1993;15(4):365-8.
Shallaly GH, Ali EA. Use of Video-Projected Structured Clinical Examination (ViPSCE) instead of the traditional oral (VIVA) examination in the assessment of final year medical students. Education for Health 2004;17(1):17-26.
Shanley E. Misplaced confidence in a professional's ability to safeguard the public? Nurse Education Today 2001;21(2):136-42.
Shapiro J, Lie D. A comparison of medical students' written expressions of emotion and coping and standardized patients' ratings of student professionalism and communication skills. Medical Teacher 2004;26(8):733-5.
Shapiro J, Duke A, Boker J, Ahearn CS. Just a spoonful of humanities makes the medicine go down: introducing literature into a family medicine clerkship. Medical Education 2005;39(6):605-12.
Shatzer JH, Darosa D, Colliver JA, Barkmeier L. Station-length requirements for reliable performance-based examination scores. Academic Medicine 1993;68(3):224-9.
Shatzer JH, Wardrop JL, Williams RG, Hatch TF. Generalizability of performance in different-station-length standardized patient cases. Teaching and Learning in Medicine 1994;6(1):54-8.
Shirar LE, Vu NV, Colliver JA, Barrows HS. A survey of study methods, preparation time, test-taking strategies, and perceptions of test validity on a Clinical performance-based examination. Academic Medicine 1992;67(10 Suppl):S10-S12.
Show N, Kirwan J. Piloting the new specialist training year 1 (st1) interview process using the OSCE. Medical Teacher 2007;29(2-3):286.
Shriner CJ, Hickey DP. Assessing preclinical medical students' ability to communicate in writing [3]. Family Medicine 2005;37(3):159-60.
Shumway JM, Harden RM. AMEE guide no. 25: The assessment of learning outcomes for the competent and reflective physician. Medical Teacher 2003;25(6):569-84.
Sibert L, GrandMaison P, Charlin P, Grise P. Evaluation of clinical competence in urology: innovative approach based on performance observation. Progres en Urologie 1997; 7(4):581-9.
Sibert L, GrandMaison P, Doucet J, Weber J, Grise P. Initial experience of an Objective Structured Clinical Examination in evaluating urology residents. European Urology 2000;37(5):621-7.
Sibert L, Mairesse J-P, Aulanier S, Olombel P, Becret F, Thiberville J, Peron J-M, Doucet J, Weber J. Introducing the objective structured clinical examination to a general practice residency programme: results of a French pilot study. Medical Teacher 2001;23(4):383-8.
Sidhu RS, McIlroy JH, Regehr G. Using a comprehensive examination to assess multiple competencies in surgical residents: does the oral examination still have a role? Journal of the American College of Surgeons 2005;201(5):754-8.
Siker ES. Assessment of clinical competence. Current Opinon in Anaesthesiology 1999;12(6):677-84.
Simon SR, Volkan K, Hamann C, Duffey C, Fletcher SW. The relationship between second-year medical students' OSCE scores and USMLE Step 1 scores. Medical Teacher 2002;24(5):535-9.
Simon SR, Bui A, Day S, Berti D, Volka K. The relationship between second-year medical students' OSCE scores and USMLE Step 2 scores. Journal of Evaluation in Clinical Practice 2007;13:901-5.
Singer PA, Cohen R, Robb A, Rothman A. The ethics objective structured clinical examination. Journal of General Internal

Medicine 1993;8(1):23-8.
Singer PA, Robb A, Cohen R, Norman G, Turnbull J. Evaluation of a multicenter ethics objective structured clinical examination. Journal of General Internal Medicine 1994;9(12):690-2.
Singer PA, Robb A, Cohen R, Norman G, Turnbull J. Performance-based assessment of clinical ethics using an objective structured clinical examination. Academic Medicine 1996;71(5):495-8.
Singh R, Naughton B, Taylor JS, Koenigsberg MR, Anderson DR, McCausland LL, Wahler RG, Robinson A, Singh G. A comprehensive collaborative patient safety residency curriculum to address the ACGME core competencies. Medical Education 2005;39(12):1195-204.
Singleton A, Smith F, Harris T, Ross-Harper R, Hilton S. An evaluation of the Team Objective Structured Clinical Examination (TOSCE). Medical Education 1999;33(1):34-41.
Skinner BD, Newton WP, Curtis P. The educational value of an OSCE in a Family Practice Residency. Academic Medicine 1997;72(8):722-4.
Sliwa JA, Kowalske KJ. Assessing resident clinical competence. American Journal of Physical Medicine Rehabilitation 2000;79(5):468-73.
Sloan DA, Donnelly MB, Johnson SB, Schwartz RW, Strodel WE. Use of an Objective Structured Clinical Examination (OSCE) to measure improvement in clinical competence during the surgical internship. Surgery 1993;114(2):343-51.
Sloan DA, Donnelly MB, Johnson SB, Schwartz RW, Strodel WE. Assessing surgical residents' and medical students' interpersonal skills. Journal of Surgical Research 1994;57:613-8.
Sloan DA, Donnelly MB, Zweng TN, Lieber A, Yu G, Griffith C, Schwartz RW, Strodel WE. The structured clinical instruction module: a novel strategy for improving the instruction of clinical skills. Journal of Surgical Research 1995;58(6):605-10.
Sloan DA, Donnelly MB, Schwartz RW, Strodel VE. The objective clinical examination. The new gold standard for evaluating postgraduate clinical performance. Annals of Surgery 1995;222(6):735-42.
Sloan DA, Donnelly MB, Schwartz RW, Felts JL, Blue AV, Strodel WE. The use of the Objective Structured Clinical Examination (OSCE) for evaluation and instruction in graduate medical education. Journal of Surgical Research 1996;63(1):225-9.
Sloan DA, Donnelly MB, Schwartz RW, Plymale MA, Strodel WE, Kenady DE, McGrath PC, Edwards MJ. The multidisciplinary structured clinical instruction module as a cehile for cancer education. American Journal of Surgery 1996;172(3):220-5.
Sloan DA, Donnelly MB, Schwartz RW, McGrath PC, Kenady DE, Wood DP, Strodel WE. Measuring the ability of residents to manage oncologic problems. Journal of Surgical Oncology 1997;64(2):135-42.
Sloan DA, Donnelly MB, Schwartz RW, Vasconez HC, Plymale M, Kenady DE. Critical assessment of the head and neck clinical skills of general surgery residents. World Journal of Surgery 1998;22(3):229-35.
Sloan DA, Plymale MA, Donnelly MB, Schwartz RW, Edwards MJ, Bland KI. Enhancing the clinical skills of surgical residents through structured cancer education. Annals of Surgery 2004;239(4):561-6.
Sloan PA, Plymale MA, Johnson M, Vanderveer B, LaFountain P, Sloan DA. Cancer pain management skills among medical students: the development of a cancer pain objective structured clinical examination. Journal of Pain and Symptom Management 2001;21(4):298-306.
Sloan PA, Plymale M, LaFountain P, Johnson M, Snapp J, Sloan DA. Equipping medical students to manage cancer pain: a comparison of three educational methods. Journal of Pain & Symptom Management 2004;27(4):333-42.
Sloane PA, Donnelly MB, Vanderveer B, Delomas M, Schwartz RW, Sloan DA. Cancer pain education among family physicians. Journal of Pain and Symptom Management 1997;14(2):74-81.
Smee S. Comparing scoring instruments for the performance assessment of professional competencies. Dissertation Abstracts International A: Humanities and Social Sciences 2008;68(11A):4681.
Smee SM, Blackmore DE. Preparing Physician Examiners for a High Stakes, Multi-Site OSCE. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;462-4.
Smee SM, Blackmore DE, Rothman AI, Reznick R, Dauphinee WD. Pioneering a sequenced OSCE for the Medical Council of Canada: An Administrative Overview. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Smee SM, Blackmore DE. Setting standards for an objective structured clinical examination: the borderline group method gains ground on Angoff. Medical Education 2001;35(11):1009-10.
Smith CA, Hart AS, Sadowski LS, Riddle J, Evans AT, Clarke PM, Ganschow PS, Mason E, Sequeira W, Wang Y. Teaching cardiac examination skills. A controlled trial of two methods. Journal of General Internal Medicine 2006;21(1):7-12.
Smith CC, Newman L, Davis RB, Yang J, Ramanan R. A comprehensive new curriculum to teach and assess resident knowledge and diagnostic evaluation of musculoskeletal complaints. Medical Teacher 2005;27(6):553-8.
Smith J, Krabak BJ, Malanga GA, Moutvic MA. Musculoskeletal education in physical medicine and rehabilitation residency programs. American Journal of Physical Medicine and Rehabilitation 2004;83(10):785-90.
Smith JL. Cancer education for the generalist physician. Family Medicine 2001;33(5):371-5.
Smith LJ, Price DA, Houston IB. Objective Structured Clinical Examination compared with other forms of student assessment. Archives of Disease in Childhood 1984;59(12):1173-6.
Smith MD, Henry-Edwards S, Shanahan EM, Ahern MJ. Evaluation of patient partners in the teaching of the musculoskeletal examination. The Journal of Rheumatology 2000;27(6):1533-7.
Smith SR, Balint JA, Krause KC, Moore-West M, Viles PH. Performance-based assessment of moral reasoning and ethical judgement among medical students. Academic Medicine 1994;69(5):381-5.
Solomon DJ, Speer AJ, Perkowski LC, Dipette DJ. Evaluating problem solving based on the use of history findings in a standardized-patient examination. Academic Medicine 1994;69(9):754-7.
Solomon DJ, Szauter K, Rosenbraugh CJ, Callaway MR. Global ratings of student performance in a standardized patient examination: is the whole more than the sum of the parts? Advances in Health Sciences Education 2000;5(3):131-40.
Sood R. Objective Structural Clinical Examination. Journal of the Association of Physicians of India 1999;47(7):751-2.
Spatz ES, Gaaserud AM, Abrahamowicz M, Weinreb B, Wenger NS, Margolis CZ. A new approach to developing cross-cultural communication skills. Medical Teacher 2004;26(2):126-32.
Sprankell SJ. Using Standardized Patients to Assess Clinical Competence in the Primary Care of Pediatric Cancer Patients. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;717-8.
Stein MR, Parish SJ, Arnsten JH. The OSCE as a formative

evaluation tool for substance abuse teaching. Medical Education 2005;39(5):529-30.
Steiner BD, Cook RL, Smith AC, Curtis P. Does training location influence the clinical skills of medical students? Academic Medicine 1998;73(4):423-6.
Stern D, Committee C, Fitzgerald JT. A Patient Satisfaction Survey as a Valid Adjunct to Simulated Patient Exercises. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Stern DT, Wojtczak A, Schwarz MR. The assessment of global minimum essential requirements in medical education. Medical Teacher 2003;25(6):589-95.
Stern DT, Ben-David MF, De Champlain A, Hodges B, Wojtczak A, Schwarz MR. Ensuring global standards for medical graduates: a pilot study of international standard-setting. Medical Teacher 2005;27(3):207-13.
Stillman PL, Regan MB, Philbin M, Haley HL. Results of a survey on the use of standardized patients to teach and evaluate clinical skills. Academic Medicine 1990;65(5):288-92.
Stillman PL, Haley HL, Regan MB, Philbin MM. Positive effects of a clinical performance assessment program. Academic Medicine 1991;66(8):481-3.
Stillman PL, Regan MB, Haley HL, Norcini JJ, Friedman M, Sutnick AI. The use a patient note to evaluate clinical skills of first-year residents who are graduates of foreign medical schools. Academic Medicine 1992;67(10 Suppl):S57-S59.
Stillman PL. Technical issues: logistics. Academic Medicine 1993;68(6):464-70.
Stilman PL, Regan MB, Swanson DB, Case S, McCahan J, Feinblatt J, Smith SR, Willms J, Nelson DV. An assessment of the clinical skills of fourth-year students at four New England Medical Schools. Academic Medicine 1990;65(5):320-6.
Stimmel B. Reassessing the OSCE. Academic Medicine 1996;71(11):1137-8.
Stimmel B, Colliver JA, Cohen DS, Smith L, Swartz MH. Using SPS for Teaching and Assessing Clinical Skills for Encounters with Patients with Problems of Substance Abuse, Pain Control and AIDS. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;722-4.
Stratton TD, Elam CL, Murphy-Spencer AE, Quinlivan SL. Emotional intelligence and clinical skills: preliminary results from a comprehensive clinical performance examination. Academic Medicine 2005;80(10 Suppl):S34-S37.
Strickland-Hodge B. Continuing professional development goes hand-in-hand with pharmacy practice. Pharmacy in Practice 2008;18(5):173-7.
Supiano MA, Fiztgerald JT, Hall KE, Halter JB. A vertically integrated geriatric improves medical student knowledge and clinical skills. Journal of the American Geriatrics Society 2007;55(10):1650-5.
Sutnick AI, Friedman M, Stillman PL, Norcini JJ, Wilson MP. International use of standardized patients. Teaching and Learning in Medicine 1994;6(1):33-5.
Sutyak JO, Lebeau RB, ODonnell AM. Unstructured cases in case-based learning benefit students with primary care career preferences. American Journal of Surgery 1998;175(6):503-7.
Swanson DB. Factors influencing reproducibility of tests using standardized patients. Teaching and Learning in Medicine 1989;1(3):158-66.
Swanson DB, Clauser BE, Case SM. Clinical skills assessment with standardized patients in high stakes tests: a framework for thinking about score precision equating and security. Advances in Health Sciences Education 1999;4:67-106.
Swartz MH, Colliver JA, Bardes CL, Charon R, Fried ED, Moroff S. The Validity of Standardized Patient Assessment Using Faculty-Physician Global Ratings as the Gold-Standard Criterion. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;725-7.
Swartz MH, Colliver JA, Bardes CL, Charon R, Fried ED, Moroff S. Validating the standardized-patient assessment administered to medical students in the New York City Consortium. Academic Medicine 1997;72(7):619-26.
Swartz MH, Colliver JA, Lower T, Morrison LS. Reliability of Standardized-Patient Assessment Revisited: Results from Two Medical School Performance-Testing Programs. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998.
Swartz MH, Colliver JA, Robbs RS, Cohen DS. Effect of multiple standardized patients on case and examination means and passing rates. Academic Medicine 1999;74(10 Suppl):S131-S134.
Swartz MH, Colliver JA, Bardes CL, Gharon R, Fried ED, Moroff S. Global ratings of videotaped performance versus global ratings of actions recorded on checklists: a criterion for performance assessment with standardized patients. Academic Medicine 1999;74(9):1028-32.
Swartz MH, Colliver JA, Robbs RS. The interaction of examinees' ethnicity and standardized patients' ethnicity: an extended analysis. Academic Medicine 2001;76(10 Suppl):S96-S98.
Swarzkopf S, Morfeld M, Gulich M, Lay W, Horn K, Mau W. Current teaching, learning and examination methods in medical education and potential applications in rehabilitative issues. Rehabilitation 2007;46(2):64-73.
Syme-Grant J, Johnstone P. Even re-sit students find the OSCE fair! Medical Education 2004;38(2):223-5.
Symonds I, Cullen L, Fraser D. Evaluation of a formative interprofessional team objective structured clinical examination (ITOSCE): a method of shared learning in maternity education. Medical Teacher 2003;25(1):38-41.
Sytnick AI, Stillman PL, Norcini JJ, Friedman M, Williams RG, Trace DA, Schwartz MA, Wang Y, Wilson MP. Pilot study of the use of the ECFMG clinical competence assessment to provide profiles of clinical competencies of graduates of foreign medical schools for residency directors. Academic Medicine 1994;69(1):65-7.
Szalay D, MacRae H, Regehr G, Reznick R. Using Operative Outcome to Assess Technical Skill. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-8.
Szauter K, Ainsworth MA, Callaway MR, Rosebraugh CJ, Solomon DJ. Identifying students at risk for suboptimal performance on a senior-level standardized patient examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Szerlip H, Anderson DS, Garris JB, Stanton M. Development and Implementation of a Pelvic Examination Station for a High-Stakes OSCE. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Talente G, Haist S, Wilson J. The relationship between experience with standardized patient examinations and subsequent standardized patient examination performance: a potential problem with standardized patient exam validity. Evaluation and the Health Professions 2007;30(1):64-74.
Tamblyn RM, Abrahamowicz M, Berkson L, Dauphinee WD, Gayton DC, Grad RM, Isaac LM, Marrache M, McLeod PJ, Snell LS. First-visit bias in the measurement of clinical competence with standardized patients. Academic Medicine 1992;67(10 Suppl):S22-S24.
Tan CP, Rokiah P. The use of OSCE (Objective Structured Clinical Examination) as an assessment tool--initial experiences at the University of Malaya. Medical Journal of Malaysia 2005;60(Suppl D):48-53.

Tan CP, Azila N. Improving OSCE examiner skills in a Malaysian setting. Medical Education 2007;41(5):517.

Tanaka N, Maruyama M, Amagai A, Saotome K, Tomikawa E, Michiyama K, Kondo S, Tanaka H, Sato H. Development and evaluation of a communication-skills learning suitable for community pharmacy. Yakugaku Zasshi 2008;128(1):97-110.

Tandon SP, Chandrasekar R, Mohapatra TP, Vaidyanathan S, Rao MS. Objective Structured Clinical Examination with Application of Aggregate Scoring - Its application to Urology. Indian Journal of Urology 1986;3(1):16-9.

Tann M, Amiel GE, Bitterman A, Ber R, Cohen R. Analysis of the Use of Global Ratings by Standardized Patients and Physicians. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;191-2.

Taren DL, Thomson CA, Koff NA, Gordon PR, Marian MJ, Bassford TL, Fulginiti JV, Ritenbaugh CK. Effect of an integrated nutrition curriculum on medical education, student clinical performance, and student perception of medical nutrition training 1-3. American Journal of Clinical Nutrition 2001;73(6):1107-12.

Taylor A, Rymer J. The new MRCOG Objective Structured Clinical Examination - the examiners evaluation. Journal of Obstetrics and Gynaecology 2001;21(2):103-6.

Teo.A. The current state of medical education in japan: a system under reform. Medical Education 2007;41(1):302-8.

Tepper MS. Providing comprehensive sexual health care in spinal cord injury rehabilitation: implementation and evaluation of a new curriculum for health care professionals. Sexuality and Disability 1997;15(3):131-65.

Tervo RC, Dimitrievich E, Trujillo AL, Whittle K, Redinius P, Wellman L. The Objective Structured Clinical Examination (OSCE) in the Clinical Clerkship: an overview. South Dakota Journal of Medicine 1997;50(5):153-6.

The researchers on clinical skills assessment of the clinical skills assessment alliance. Consensus statement of the researchers in clinical skills assessment (RCSA) on the use of standardized patients to evaluate clinical skills APPENDIX 1 and 2 AAMC Conference. Academic Medicine 1993;68(6):475-7.

Thind CK, Ormerod AD. Recent Advances in Inflammatory Skin Diseases. Scottish Medical Journal 2008;53(2):30-4.

Thistlethwaite JE. Developing an OSCE station to assess the ability of medical students to share information and decisions with patients: Issues relating to interrater reliability and the use of simulated patients. Education for Health 2002;15(2):170-9.

Thomas PA, Shatzer JH. Standardized patient assessment of ambulatory clerks: effect of timing and order of the clerkship. Teaching and Learning in Medicine 2000;12(4):183-8.

Thomson DM. The objective structured clinical examination for general practice: design, validity and reliability. Journal of the Royal College of General Practitioners 1987;37:149-53.

Timmermann A, Roessler M, Barwing J, et al. IN Zentrum Anasthesiologie, Rettungs- und Intensivmedizin, Georg-August-Universitat, Gottingen. atimmer@web.de TI [New pathways in undergraduate medical education - first experiences with the cross section speciality emergency and intensive care medicine]. [German]. Anasthesiologie, Intensivmedizin, Notfallmedizin, Schmerztherapie 2005;40(9):536-43.

Timmermann A, Barwing J, Russo S, Roessler M, Schmid O, Eich C, Schwertfeger K, Klockgether-Radke A, Graf B. Planning and realization of an objective standard clinical evaluation (OSCE). Journal fur anasthesie und Intensivbehandlung 2007;14(1):207-9.

Tiong TS. Should clinical normality be examined in medical course? Singapore Medical Journal 2008;49(4):328-32.

Townsend AH, Mcllvenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. Medical Education 2001;35(9):841-6.

Tracey J, Rodwell P, Clearwater G. Development and evaluation of the rural trauma and emergency care roadshow. New Zealand Medical Journal 1999;112(1086):144-7.

Traina AD, Gour N, Scaringe JG. Multi-modal components of instruction in a preclinical educational program. Journal of Chiropractic Education 1995;9(2):63-71.

Traina AD, Goubran E, Gour NJ, Scaringe JG, Talmage DM, Wells K. Description of integrated competency examination: tools to assess the chiropractic curriculum effectiveness and students competency levels. Journal of Manipulative and Physiological Therapeutics 1996;19(7):463-8.

Troncon L, Dantas RO, Ferriolli E, Figueiredo JFC, Martinelli ALC, Moriguti JC, Voltarelli JC. A Modified OSLER (Structured Long Examination Record) for Summative Assessment of Medical Students Completing an Internship in Internal Medicine. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.

Troncon LE. Clinical skills assessment: limitations to the introduction of an "OSCE" (Objective Structured Clinical Examination) in a traditional Brazilian medical school. Sao Paulo Medical Journal 2004;122(1):12-7.

Troncon LEA, Rodrigues MLV, Piccinato CE, Figueiredo JFC, Peres LC, Cianflone ARL. Overcoming Difficulties in the Introduction of a Summative Assessment of Clinical Competence in a Brazilian Medical School. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;197-9.

Troncon LEA, Dantas RO, Figueiredo JFC, Ferriolli E, Moriguti JC, Martinelli ALC, Voltarelli JC. A standardized, structured long-case examination of clinical competence of senior medical students. Medical Teacher 2000;22(4):380-5.

Tsai TC. Using children as standardised patients for assessing clinical competence in paediatrics. Archives of Disease in Childhood 2004;89(12):1117-20.

Turner J, Dankoski ME. Objective Structured Clinical Exams: A Critical Review. Family Medicine 2008;40(8):574-8.

Turnwald G, Stone E, Bristol D, et al. Assessing clinical competency: reports from discussion groups. Journal of Veterinary Medical Education 2008;35(3):343-53.

Usatine RP, Edelstein RA, Yajima A, Slavin SJ, Wilkes MS. Medical Student Perceptions of the Accuracy of Various New Clinical Evaluation Methods. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;200-2.

Vadillo M, Pujol-Farriols R, Roma J, Martinez-Carretero JM. Predicting Performance Through Clinical Skills Assessment with Standardized Patients. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;732-3.

Vaidya NA. Psychiatry Clerkship Objective Structured Clinical Examination is Here to Stay. Academic Psychiatry 2006;32(3):177-9.

Valentino J, Donnelly MB, Sloan DA, Schwartz RW, Haydon RC. The reliability of six faculty members in identifying important OSCE items. Academic Medicine 1998;73(2):204-5.

Vallevand A. Reliability, validity and sources of errors in assessing physician performance in an Objective Structured Clinical Examination: A generalizability theory analysis. Dissertation Abstracts International A: Humanities and Social Sciences 2008;69(4A):1339.

Van Aswegen EJ, Basson AA. Clinical Evaluation and the OSCE. Curationis 1994;17(1):32-7.

Van Dalen J, Bartholomeus P, Kerkhofs E, Lulofs R, Van Thiel J, Rethans JJ, Scherpbier A, Van der Vleuten C. Teaching and

assessing communication skills in Maastricht: the first twenty years. Medical Teacher 2001;23(3):245-51.

Van der Vleuten C, Scherpbier A, Dolmans DHJM, Schuwirth LWT, Verwijnen GM, Wolfhagen HAP. Clerkship assessment assessed. Medical Teacher 2000;22(6):592-600.

Van Nulan M, Aertgeerts B, Goedhuys J, Hannes K. Educational Interventions to improve the communication skills og general practice trainees in the clinical consultation. Cochrane Database of Systematic Reviews 2006;(1):1-16.

Van Nuland M, Van Den Noortgate W, Degryse J, Goedhuys J. Comparison of two instruments for assessing communication skills ina general practice objective structured clinical examination. Medical Education 2007;41(7):676-83.

Van Vollenhoven FHM, Van Luijk SJ, Muijtjens AMM. Sequential Testing In A Standardized Patient Based Test. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;734-6.

Vargas AL, Boulet JR, Errichetti A., Van Zanten M, Lopez MJ, Reta AM. Developing performance-based medical school assessment programs in resource-limited enviroments. Medical Teacher 2007;29(2-3):192-8.

Varkey P, Gupta P, Bennet K. An Innovative Method to Assess Negotiation Skills Necessary for Quality Improvement. American Journal of Medical Quality 2008;23 (5):350-5.

Varkey P, Natt N, Lesnick T, Downing S, Yudkowsky R. Validity Evidence for an OSCE to Assess Competency in Systems-Based Practice and Practice-Based Learning and Improvement: A Preliminary Investigation. Academic Medicine 2008;83(8):775-80.

Vatsa M. OSCE - An experience in first-aid examination of nursing students. The Nursing Journal of India 1990;81(6):176-7.

Vento TM, Moya BM, Schiff D. Utilización de los exámenes clínicos estructurados objetivamente en la evaluación de pregraduado en pediatría. Anales Espanoles de Pediatria 1990;32(1):41-8.

Verhoeven BH, Hamers JGHC, Scherpbier A, Hoogenboom R, Van der Vleuten C. The effect on reliability of adding a separate written assessment component to an objective structured clinical examination. Medical Education 2000;34(9):525-9.

Verma M, Singh T. Experiences with Objective Structured Clinical Examination (OSCE) as a tool for formative evaluation. Indian Pediatrics 1993;30(5):699-703.

Verma M, Singh T. Attitudes of medical students towards Objective Structured Clinical Examination (OSCE) in Pediatrics. Indian Pediatrics 1993;30(10):1259-61.

Verma M, Singh T. Communication skills in clinical practice fad or necessity? Indian Pediatrics 1994;31(2):237-8.

Vetto JT, Elder NC, Toffler WL, Fields SA. Teaching medical students to give bad news: does formal instruction help? Journal of Cancer Education 1999;14(1):13-7.

Vincent DS, Cooper GS, Harvey J, Noel GL. Assessing medical student interviews of HIV patients: a randomized trial of a simple intervention. Proceedings of the Annual Conference on Research in Medical Education 1988;27:216-20.

Vivekananda-Schmidt P, Lewis M, Hassell AB, ARC Virtual Rheumatology CAL Research Group. Cluster randomized controlled trial of the impact of a computer-assisted learning package on the learning of musculoskeletal examination skills by undergraduate medical students. Arthritis and Rheumatism 2005;53(5):764-71.

Vivekananda-Schmidt P, Lewis M, Coady D, Morley C, Kay L, Walker D, Hassell A. Exploring the use of videotaped objective structured clinical examination in the assessment of joint examination skills of medical students. Arthritis and Rheumatism 2007; 57(5):869-76.

Vivekananda-Schmidt P, Lewis M, Hassell AB, Coady D, Walker D, Kay L, McLean M, Haq I, Rahman A. Validation of MSAT: an instrument to measure medical students' self-assessed confidence in musculoskeletal examination skills. Medical Education 2007;41(4):402-10.

Volkan K, Simon SR, Baker H, Todres ID. Psychometric structure of a comprehensive objective structured clinical examination: a factor analytic approach. Advances in Health Sciences Education 2004;9(2):83-92.

Vooijs MEEC, Scherpbier A, Van Dalen J, Ramsay G, Kootstra G. Group objective structured clinical examination in an Inter-European accident and emergency course. Education for Health 1997;10(1):69-78.

Vu NV, Marcy MM, Colliver JA, Verhulst SJ, Travis TA, Barrows HS. Standardized (simulated) patients' accuracy in recording clinical performance checklist-items. Medical Education 1992;26(2):99-104.

Vu NV, Barrows HS, Marcy ML, Verhulst SJ, Colliver JA, Travis T. Six years of comprehensive, clinical, performance-based assessment using standardized patients at the Southern Illinois University School of Medicine. Academic Medicine 1992;67(1):42-50.

Vu NV, Henkle JQ, Colliver JA. Consistency of pass-fail decisions made with clinical clerkship ratings and standardized-patient examination scores. Academic Medicine 1994;69(10 Suppl):S40-S41.

Vu NV, Marcy ML, Barnhart A, Colliver JA, Henkle JQ, Hodgson K, Schrage JP, Travis TA. Further evidence of construct validity of standardized patient-based performance examinations. Teaching and Learning in Medicine 1994;6(4):255-9.

Walker R, Walker B. Use of the objective structured clinical examination for assessment of vocational trainees for general practice. Journal of the Royal College of General Practitioners 1987;37(296):123-4.

Wallace J, Rao R, Haslam R. Simulated patients and objective structured clinical examinations: Review of their use in medical education. Advances in Psychiatric Treatment 2002;8(5):342-8.

Wallach PM, Elnick M, Bognar B, Kovach R, Papdakis M, Zucker S, Speer A. Standardized patient's perceptions about their own health care. Teaching and Learning in Medicine 2001;13(4):227-31.

Walmsley AD, Manogue M, Brown GA. Scope of the OSCE LETTER and REPLY to LETTER. British Dental Journal 2001;191(3):116-8.

Walters J, Adams J. A child health nursing objective structured clinical examination (OSCE). Nurse Education in Practice 2002;2(4):224-9.

Walters K, Osborn D, Raven P. The development, validity and reliability of a multimodality objective structured clinical examination in psychiatry. Medical Education 2005;39(3):292-8.

Wang M, Wen-Ling L. The nurturing and development of nurse practitioners. Hu Li Tsa Chih - Journal of Nursing 2007;54(6):11-5.

Wang Y, Stillman P, Stunick AI, Friedman Ben-David M, Williams RG. The effect of fatigue on the accuracy of Standardized patient's checklist recording. Teaching and Learning in Medicine 1996;8(3):148-51.

Wang Y, Gliva-McConvey G. Cognitive Strategies used in Standardized Patients' Checklist Recording. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;737-8.

Ward D. The new Membership of the Royal College of Surgeons (MRCS) examination. Surgery 2008;26(10):408-10.

Warf BC, Donnelly MB, Schwartz RW, Sloan DA. The relative contributions of interpersonal and specific clinical skills to the

perception of global clinical competence. Journal of Surgical Research 1999;86(1):17-23.
Warf BC, Donnelly MB, Schwartz RW, Sloan DA. Interpreting the judgement of surgical faculty regarding resident competence. Journal of Surgical Research 1999;86(1):29-35.
Wass V, Jolly B. The long case: what does it measure? a comparison of observation and presentation of history-taking in a traditional long case within a final year OSCE. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1.
Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001;357(9260):945-9.
Wass V, McGibbon D, Van der Vleuten C. Composite undergraduate clinical examinations: how should components be combined to maximise reliability? Medical Education 2001;35(4):326-30.
Wass V, Jones R, Van der Vleuten C. Standardized or real patients to test clinical competence? The long case revisited. Medical Education 2001;35(4):321-5.
Wass V, Jolly B. Does observation add to the validity of the long case? Medical Education 2001;35(8):729-34.
Wass V, Van der Vleuten C. The long case. Medical Education 2004;38(11):1176-80.
Wasserman HP, Slabbert BR, Van Zyl JJ. The objective structured clinical examination OSCE. South African Medical Journal 1982;61(9):325-50.
Waterson T, Cater JI, Mitchell RG. An objective undergraduate clinical examination in child health. Archives of Disease in Childhood 1980;55:917-22.
Watson AR, Houston IB, Close GC. Evaluation of an objective structured clinical examination. Archives of Disease in Childhood 1982;57(5):390-8.
Webb T, Duthie Jr E. Geriatrics for Surgeons: Infusing Life into an Aging Subject. Dissertation Abstracts International Section A: Humanities and Social Sciences 2008;65(2):91-4.
Weinreb B, Bentov Y, Burger S, Gvili B, Margolis CZ, Panso-Zablodowsky E, Schwartzman P, Solomon F. Determination of difficulty level of OSCE stations. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 im Philadelphia 1998.
Weinreb B, Benor D, Gunik N, Livshiz-Riven I, Malkin SM, Reizer H, Solomon F. Assessment of clinical competence of 1st year nursing students by Objective Structured Clinical Examination. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 im Philadelphia 1998.
Weinreb B, Bentov Y, Burger S, Gvili B, Margolis CZ, Reisenberg K, Solomon F. Using global ratings in Objective Clinical Skills Assessment. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 im Philadelphia 1998.
Weinreb B, Bentov Y, Burger S, Flechner N, Gvili B, Margolis CZ, Schwartzman P, Solomon F. Immediate feedback influence on clinical skills competence. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998.
Wenreb B, Gal N, Naimark S, Panso-Zablodowsky E, Pines A. Introducing Objective Clinical Skills Assessment to Physiotherapy Education. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998.
Wessel J, Williams R, Finch E, Gemus M. Reliability and validity of an objective structured clinical examination for physical therapy students. Journal of Allied Health 2003;32(4):266-9.
Wheeler DW, Degnan BA, Murray LJ, Dunling CP, Whittlestone KD, Wood DF, Smith HL, Gupta AK. Retention of drug administration skills after intensive teaching. Anaesthesia 2008;63(4):379-84.
Whelan GP, Friedman Ben-David M. Educational commission for foreign medical graduates clinical skills assessment prototype. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Whelan GP. Educational commission for foreign medical graduates: clinical skills assessment prototype. Medical Teacher 1999;21(2):156-60.
Whelan GP. Educational commission for foreign medical graduates: lessons learned in a high stakes, high volume medical performance examination. Medical Teacher 2000;22(3):293-6.
Whelan GP, Boulet JR, McKinley DW, Norcini JJ, Van Zanten M, Hambleton RK, Burdick WP, Peitzman SJ. Scoring standardized patient examinations: lessons learned from the development and administration of the ECFMG Clinical Skills Assessment (CSA). Medical Teacher 2005;27(3):200-6.
Whelan P, Church L. The first cohort of OSCE Part 1 candidates reaching Part 2 [5]. Psychiatric Bulletin 2005;29(9):353.
White CB, Ross PT, Purkiss JA, Hammoud MM. Improving medical students' competence at breast examination. International journal of gynaecology and obstetrics 2008;102(2):173-4.
Wilkinson LF, Pethtel LP. Clinical Skills Assessment at NEOUCOM: A Description. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;397-9.
Wilkinson LF, Cuevas NM. Variations on a Theme: Scoring the Clinical Skills Assessment. Advances in Medical Education, 7th Ottawa Conference, 1996 Maastricht 1997;671-4.
Wilkinson TJ, Newble DI, Wilson PD, Carter JM, Helms RM. Development of a three centre simultaneous objective structured clinical examination. Medical Education 2000;34(10):798-807.
Wilkinson TJ, Newble DI, Frampton CM. Standard setting in an objective structured clinical examination: use of global ratings of borderline performance to determine the passing score. Medical Education 2001;35:1043-9.
Wilkinson TJ, Fontaine S, Egan T. Was a breach of examination security unfair in an objective structured clinical examination? A critical incident. Medical Teacher 2003; 25(1):42-6.
Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations: Checklists are no substitute for examiner commitment. Academic Medicine 2003;78(2):219-23.
Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. Medical Education 2004;38(10):1111-6.
Willet L, Estrada C, Castiglioni A, Massie F, Heudebert G, Jennings M, Centor R. Does residency training improve performance of physical examination skills? American Journal of Medical Sciences 2007;333(2):74-7.
Williams M, Ambrose M, Carlin AM, Tyburski JG, Steffes CP. Evaluation of academic and community surgery clerkships at a Midwestern medical school. Journal of Surgical Research 2004;116(1):11-3.
Williams R, McLaughlin M, Eulenberg B, Hurm M, Nendaz M. Should the patient findings questionnaire replace the SP checklist? Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-3.
Williams RG, Barrows HS, Vu NV, Verhulst SJ, Colliver JA, Marcy M, Steward D. Direct, standardized assessment of clinical competence. Medical Education 1987;21:482-9.
Williams RG, Tafesse E, Sinacore JM, Valaski M. Are standardized patient ratings influenced by patient perceptions of the preceding examinee? Teaching and Learning in Medicine 1998;10(4):201-6.
Williams RG, McLaughlin MA, Eulenberg B, Hurm M, Nendaz

MR. The patient findings questionnaire: one solution to an important standardized patient examination problem. Academic Medicine 1999;74(10):1118-24.
Williams RG. Have standardized patient examinations stood the test of time and experience? Teaching and Learning in Medicine 2004;16(2):215-22.
Winckel CP, Martin JA, Reznick RK. The simulated surgical office: an innovation in ambulatory care teaching. Medical Teacher 1995;17(2):167-74.
Wiskin CM, Allan TF, Skelton JR. Gender as a variable in the assessment of final year degree-level communication skills. Medical Education 2004;38(2):129-37.
Wong M, Fones C, Aw M, et al. Should non-expert clinician examiners be used in objective structured assessment of communication skills among final year medical undergraduates? Medical Teacher 2007;29:927-32.
Woodburn J, Sutcliffe N. The Reliability, Validity and Evaluation of the Objective Structured Clinical Examination in Podiatry (Chiropdoy). Assessment and Evaluation in Higher Education 1996;21(2):131-46.
Woods SE, Francis BW. MEDLINE as a component of the objective structured clinical examination: the next step in curriculum integration. Bulletin of the Medical Library Association 1996;84(1):108-9.
Woodward CA, McConvey G. The effect of simulating on standardized patients. Academic Medicine 1995;70(5):418-20.
Woodward CA, McConvey GG. Children as standardized patients: initial assessment of effects. Teaching and Learning in Medicine 1995;7(3):188-91.
Woolf K, Haq I, McManus LC, Higham J, Dacre J. Exploring the underperformance of male and minority ethnic medical students in first year clinical examinations. Advances in Health Sciences Education 2008;13(5):607-16.
Wright C, Hogard E, Ellis R. Effect of PETTLEP imagery training on performance of nursing skills: pilot study. Journal of Advanced Nursing 2008;63(3):259-65.
Yak A, Abdul-Hamid WK, Chad R. The objective structured clinical examination (OSCE). Psychiatric Bulletin 2004;28(7):265-6.
Yelland MJ. Standardized patients in the assessment of general practice consulting skills. Medical Education 1998;32(1):8-13.
Young RC, Boweb RE. On and off the playing field: ethics in medicine. Journal of Health Care for the Poor and Underserved 1999;10(2):178-85.
Young WW, Barthold JC, Birenbaum D, Long P, Dion M, Hamilton LA. An objective structured clinical exam in multisite obstetrics and gynecology clerkship. Teaching and Learning in Medicine 1995;7(3):177-81.
Ytterberg SR, Harris IB, Allen SS, Anderson DC, Kofron PM, Kvasnicka JH, McCord JP, Moller JH. Clinical confidence and skills of medical students: use of an OSCE to enhance confidence in clinical skills. Academic Medicine 1998;73(10 Suppl):S103-S105.
Yudkowsky R, Alseidi A, Cintron J. Beyond fulfilling the core competencies: an objective structured clinical examination to assess communication and interpersonal skills in a surgical residency. Current Surgery 2004;61(5):499-503.
Zabar S, Hanley K, Stevens DL, Kalet A, Schwartz MD, Pearlman E, Brenner J, Kachur EK, Lipkin M. Measuring the competence of residents as teachers. Journal of General Internal Medicine 2004;19(5):530-3.
Ziv A, Friedman Ben-David M, Sutnick AI, Gary NE. Lessons learned from six years of international administrations of the ECFMGS SP - based clinical skills assessment. Academic Medicine 1998;73(1):84-91.
Ziv A, Boulet JR, Friedman Ben-David M, Curtis M, Burdick WP, Peitzman S, Gary NE. A Holistic Behaviourally Anchored Measure of Physician Communication Skills: Issues of Rater Consistency. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-6.
Ziv A, Boulet JR, Burdick WP, Friedman Ben-David M, Gary NE. The Use of National Medical Care Surveys to Develop and Validate test Content for Standardized Patient Examinations. Evolving Assessment: Protecting the Human Dimension: 8th Ottawa Conference 1998 in Philadelphia 1998;1-7.
Zonneveld P, Van Den Berg YWMM, Verhoeven WMA. Assessment of medical competence in the psychiatric training program. Tijdschrift voor Psychiatrie 1995;37(5):426.
Zuberi RW, Jafarey NA, Qureshi AF, Elahi F. The diploma in family medicine examination; a scientific exercise. Journal of the Pakistan Medical Association 1993;43(10):217-20.
Zyromski NJ, Staren ED, Merrick HW. Surgery residents' perception of the Objective Structured Clinical Examination (OSCE). Current Surgery 2003;60(5):533-7.

Chapter 4
Appendix 3 – Structure of OSCE electronic coding sheet

Section 1 Information on the record	
Paper Code	Reference attributed to each study under analysis
Paper Sub-code	Sub reference- when the study reports on more than one OSCE exam
Justification	Justification regarding sub-codes
Coder	Coder identification
Doubts	Aspects still to be decided
Reference Manager	Code number regarding the Reference Manager Database
Source	Searcher identification
Section 2 Information on the publication	
Date	Date of publication
Publication title	Title of publication
Publication Type	Description of publication format
Authors	Paper published by one or more authors:
Country	Country of authors
City	City of authors
Institution	Authors' affiliation
Section 3 Information on OSCE Background	
Course Area	Medicine, Nursing, Dentistry, etc.
Medical Education level	Undergraduate, Postgraduate and CME/PCD
Inclusion /Exclusion	Justification for accepting rejecting papers
Rejected due	Identifying the sources for rejection
Other Designation	Other designation than OSCE (TOSCE, GOSCE, and VOSPE) etc.
Aim of Paper	Description of the aim of the paper
Aim Quant.	Coding in terms of Detecting Results or Examining OSCE
Focus on Validity	Study focus (fully, incidental, none) mentioned in the Abstract & introduction
Focus on Reliability	Study focus (fully, incidental, none) mentioned in the Abstract & introduction
Focus on Feasibility	Study focus (fully, incidental, none) mentioned in the Abstract & introduction
Section 4 Information on OSCE Design	
Subjects	Subject and specialty areas being assessed (specific & transversal areas)
Pilot	Information on pre-test of stations (past or Present pilot)
Training	Information on Coders and patients training
Briefing	Information on Briefing the students / teachers on OSCE structure, process,
Reported elsewhere	When information on the OSCE was already given in other paper(s)
Assessment Type	Type of assessment (High stakes, Moderate, volunteer etc.)
Purpose	Role of assessment (formative, Summative, both
Feedback given TO	Feedback on performance was to Students, SPs, Examiners etc.
Description of feedback	Description of feedback given to students
Learning Outcomes/ station type	Learning outcomes being assessed (History-taking, Physical examination, etc.)

Coding	Recording on scoring students' performance (Checklist, Global ratings)
Venue	Number of different locations where the same OSCE exam takes place
Days	Number of days when the OSCE was performed
Parallel	Number of circuits to accommodate all students.
Parallel Description	Run simultaneously or sequentially
Cycles	How many circuits in each parallel
Station number	Number of stations included in the OSCE
Rest Stations	Number of rest stations included in the OSCE
Subset stations	When authors report results just for some stations (sub-set of stations)
Exam Number	Number of different OSCE exams reported in the paper
Students total	Total number of students in the course under assessment
Undergraduate students	Number of undergraduate students performing the OSCE
Postgraduate candidates	Number of postgraduate students performing the OSCE
CPD candidates	Number of CPD candidates performing the OSCE
Clerks	Information on OSCE performed in the context of a clerkship
Subset of students	When OSCE results are reported only for a subset of students
Maximum number of students	Maximum number of students performing the OSCE in each circuit
Duration	Information on the total time spent in each OSCE
Station Time	Duration of each station (identical or non-identical in all stations)
Station description	When details on non-identical stations are reported
Course year	'Course year' students attending
Curriculum	Number of curriculum years
Subset stations	When authors report results just for some stations (sub-set of stations)
Exam Number	Number of different OSCE exams reported in the paper
Students total	Total number of students in the course under assessment
Faculty Number	Staff involved in the planning / development / implementation of the OSCE
Faculty Description	Description of respective background
Examiner Number	Number of examiners in charge of assessing students' performance
Examiner description	Description of respective background, level of seniority (senior, junior)
Simulated Patients	Number of simulated patients
Standardized SP	Information regarding if they were standardized / Information on training process
Standardized SP description	Information on background (teachers, other students, nurses, other people)
Real Patients	Number of real patients involved in the OSCE
Real Patients standard.	Information regarding if they were standardized / training process
Real Patient description	Information on background (teachers, other students, nurses, other people)
Mannequin	Information on mannequins, models, etc.
Video	Case(s) presented on video or computer
Section 5 Evidence on OSCE Feasibility, Reliability and Validity	
Validity data	Information on Validity results and process
Reliability data	Information on Reliability results
Relevance data	Information on how students, teachers, patients etc. evaluate the OSCE relevance
Fairness data	Information on how students evaluate OSCE fairness according to what was taught in class, opportunities for training etc. during the course
Drive Learning	Information regarding drive learning: i.e. data concerning students and teachers report on how the OSCE determined specifically directed learning that occurred specifically focused on the OSCE objectives
Drive Teaching	Information regarding drive Teaching: i.e. data concerning teachers reporting that OSCE highlight students' weaknesses and strengths and from this information curriculum and or teaching methods were modified
OSCEE technical viability	Information from students, teachers, examiners, simulated patients, real patients
OSCEE economic viability	Information on 'cost' and 'time taken' (planning and development)

Section 6. Information on OSCE global results versus initial aims, problems, solutions	
OSCE global results	Results highlighted by authors concerning initial aims
OSCE problems /difficulties	Major difficulties related to the OSCE implementation
OSCE solutions	Possible, how to overcome them
Bibliography	Bibliographic references mentioned in the article which could be of special interest
7. Final section	
Section for notes	Open area for notes etc.
Study uploaded	Digital or scan version of the study