Minutes of the 16th European Orthodontic Teachers' Forum Limassol, Cyprus, 31st May, 2022.

Chairpersons: Guy Willems, Andreu Puigdollers, Vaska Vandevska-Radunovic, Agneta Karsten
Attendees: 60 people registered, but only 37 signed the attendance list. Five more attendees were not previously registered.

TOPIC: ASSESSING COMPETENCE WITH HIGH STAKES EXAMS

1. Opening, announcements

Associate Professors Andreu Puigdollers and Agneta Karsten open the meeting and welcome the delegates.

The goal of the NEBEOP is to strengthen the training programs of postgraduate orthodontic education in Europe and is restricted to either provisional or full members. For more information on the provisional and full membership application procedure and bylaws: see http://www.nebeop.org.
The EOTF aims to be an open forum for everybody involved in strengthening orthodontic teaching.
In the past some interesting discussions on several pedagogical themes, trying to improve the quality of postgraduate education in orthodontics, have been held in both plenary and break-out sessions:
2006 Strengths and weaknesses of orthodontic education in Europe
2007 Development of NEBEOP and collaboration between programs
2008 Self-assessment for quality control
2009 The minimum requirements for the undergraduate orthodontic curriculum
2010 How do we learn? Adult learning
2011 New ways of learning
2012 E-learning
2013 Competency based postgraduate education
2014 Quality assurance in postgraduate education
2015 Final examination assessment procedures
2016 Adopting a coaching approach to teaching
2017 Short courses leading to a "Master". A new era in orthodontic education? .
2018 Inter-Professional Collaboration and Education'.
2019 Communication between postgraduate schools in Europe.
2020 Cancelled
2021 Adult learning in blended learning environments

This year the theme of the EOTF is ´ **ASSESSING COMPETENCE WITH HIGH STAKES EXAMS**´.

2. Charles Bolender Award

Dr. Melissa Disse, President of EFOSA presents the Charles Bolender Award. Charles Bolender was a founding member and the first President of EFOSA. The award is given to a teacher or teachers, involved with orthodontic education, who have been identified with specific educational skills enhancing the delivery of specialist orthodontic knowledge. This year's award goes to Associate Professor, Dr. Magnus Hultin, from Umeå University, Sweden.

3. ASSESSING COMPETENCE WITH HIGH STAKES EXAMS

The topic of the 16th EOTF is ´**Assessing Competence with High Stakes Exams**´ and is presented by Associate Professor, Dr. **Magnus Hultin.** Anesthesiology and Critical Care Medicine. Associate Dean of Clinical Education. 7 years as Program Director of Umeå Medical Program iin Sweden. Chair Swedish Proficiency test for Medical Doctors. Umeå University, Umeå , Sweden.

Introduction

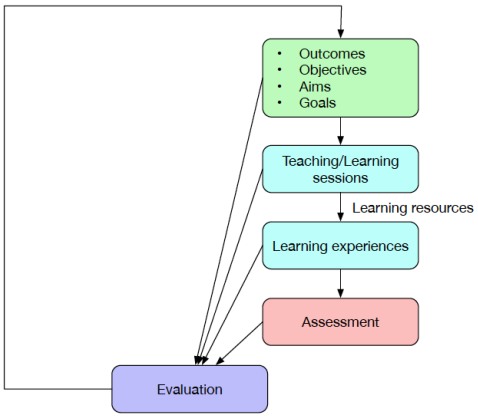
The objective of the presentation was to give an overview of how to set up a high stakes assessment measuring both theoretical knowledge and practical skills, and to exemplify with six years experiences of setting up and running the Swedish Proficiency Test for Medical Doctors. The different topics of the lecture was divided in different parts along with practical exercises done by the attendees in preformed groups.

As a background, Prof. Hultin started explaining the experience with the Proficiency Test done in Sweden to Medical Doctors trained outside UE/EES. Specifically, the experience of admission in Sweden the Medical Doctors that came to the Syrian refugee crisis.

Test design and blueprinting

The **curriculum cycle of continuous development** follows the steps of ´CONSTRUCTIVE ALIGNMENT´, that ends up in the ´Evaluation´ that, in turn, feed-back again to continue the development.

CONSTRUCTIVE ALIGNMENT

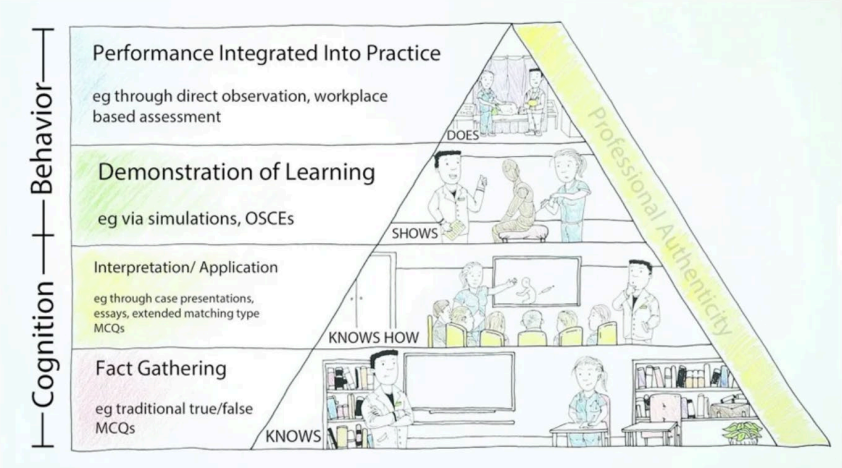


Constructive alignment within the curriculum cycle of continous development (Biggs & Tang 2007)

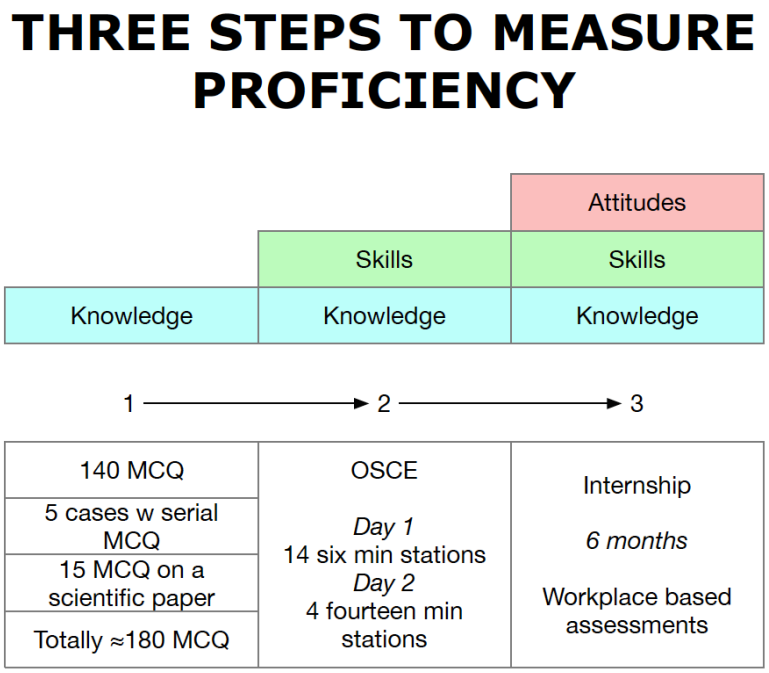
The **blueprinting** consists of a six-year program to become a Medical Doctor that examine 23 specific Knowledge, Skills and Attitudes with different tests. The test-takers should demonstrate broad and specialized knowledge in medicine; knowledge of the disciplinary foundation of the field and insight into current research and development work; demonstrate specialized skills for professional interaction with patients and their families, with respect for the integrity, needs, knowledge and experience of the patients and their families; demonstrate the ability to autonomously diagnose and start the treatment for acute life-threatening conditions; demonstrate specialized skills in autonomously diagnosing the most frequent illnesses and in treating them in collaboration with the patients.

In ascending order, ´WHICH TEST FOR WHICH COMPETENCE? ´: the ´Cognition´ competence of the test-takers are checked with theoretical exams (Multiple Choice Questions -MCQ-), and their ´Behavior´ via simulations, Objective Structured Clinical Examination -OSCE-, and with direct observation their performance at the workplace.

WHICH TEST FOR WHICH COMPETENCE?



A fair test should cover the curriculum and check the precision in answering the purpose of the test. The test-takers are examined for the three specific levels (Knowledge, Skills and Attitudes), by means of MCQ with theoretical questions and questions of clinical cases, OSCE tests and internship in the workplace (THREE STEPS TO MEASURE PROFICIENCY´).



Writing Well-Written Single -Best-Answer Questions (MCQ)

Prof. Hultin refers to the book: National Board of Medical Examiners, NBME, and provides, as a reference, examples of questions, and correct answers.
A comparison of one-best-answer item with true-false-item was shown. Next, a summary of some tips for both:
General Rules for **One-Best-Answer Items** (Recommendation of these questions is **preferable!**).

- Item and option text must be clear and unambiguous.
- Avoid imprecise phrases such as “is associated with” or “is useful for” or “is important”; words that provide cueing such as “may” or “could be”; and vague terms such as “usually” or “frequently.”
- The lead-in should be closed and focused and ideally worded in such a way that the test-taker can cover the options and guess the correct answer. This is known as the “cover-the-options” rule.
- All options should be homogeneous so that they can be judged as entirely true or entirely false on a single dimension.
- Incorrect options can be partially or fully incorrect.

General Rules for **True-False Items** (the Recommendation is to **avoid** these questions!).
Because test-takers are required to select all the options that are “true,” true-false items must satisfy the following rules:

- Item and option text must be clear and unambiguous. Avoid imprecise phrases such as “is associated with” or “is useful for” or “is important”; words that provide cueing such as “may” or “could be”; and vague terms such as “usually” or “frequently.”
- The lead-in should be closed and focused.
- Options must be absolutely true or false; no shades of gray are permissible.
- Options should be homogeneous so that they can be judged as entirely true or entirely false on a single dimension.

Many examples of technical flaws were provided. Next, a summary of technical item flaws was shown:

FLAWS	SOLUTIONS
Long, complex options	- Put common text in stem. - Use parallel construction in options. - Shorten options.
Tricky, unnecessarily complicated stems	- Include content that is necessary to answer the question or to make distractors attractive. - Avoid teaching statements.
Inconsistent use of numeric data	- Avoid overlapping options. - Ask for minimum or maximum value to avoid multiple correct answers.
Vague terms	- Avoid frequency terms, like usually and often. Such terms are interpreted differently by different people.
“None of the above” option	- Replace “None of the above” with specific action (eg, No intervention needed).
Nonparallel options	- Edit options to be parallel in grammatical form and structure.
Negatively structured stem (eg, “Each of the following EXCEPT”)	- Revise lead-in to have a positive structure. - If possible, use correct options to create a scenario.

FLAWS	SOLUTIONS
Collectively exhaustive options (subset of options cover all possibilities)	- Replace at least one option in subset. - When revising, avoid creating option pair.
Absolute terms (“always,” “never”) in options	- Eliminate absolute terms. - Use focused lead-in and short homogeneous options.
Grammatical clues	- Make all options singular or all options plural. - Use closed lead-ins.

Correct answer stands out	- Revise options to equal length. Remove language used for teaching points and rationales.
Word repeats (clang clue)	- Replace repeated word in either stem or option. OR - Use repeated word in all options.
Convergence	- Revise options to balance use of terms.

An exercise of writing and MCQ question was done by the attendees (in *italics*).
Example Group 2:
Stem
A 15 year old patient with only a persistent deciduous 65 is coming for first consultation. The patient has mild pain symptoms of this 65. At oral inspection, a class I molar relationship is seen with mild crowding of 2-3 mm.
Question

- *What is the next step to make an orthodontic diagnosis and treatment plan?*

Correct answer

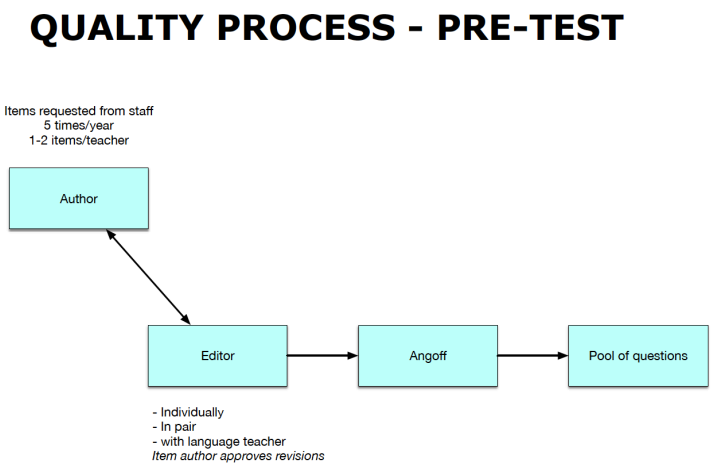
- *make an x-ray*

3-4 distractors

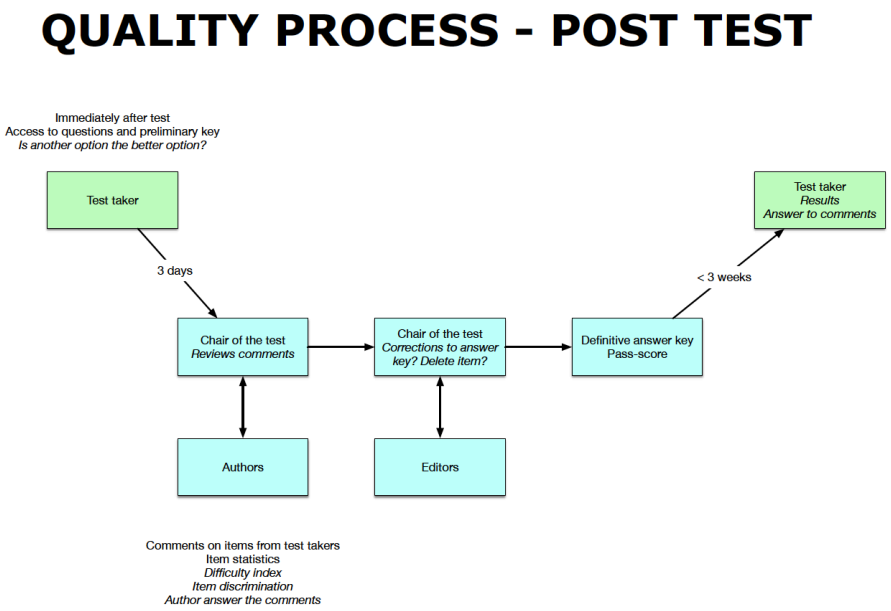
- *wait 6 months*
- *extract the 65*
- *do not treat*
- *start a pulpotomy on 65*

Ensuring question quality, pre- and post-test

Prof. Hultin explains that is recommended to evaluate the quality of the exams before and after the test is done.
To **prepare the test**, ´QUALITY PROCESS - PRE-TEST´, the staff is requested to prepare questions, these are revised individually or in pairs, a group of experienced teachers' give their professional judgement -Angoff value-, and from all of this sequence a pool of questions is generated.



To assess the question quality **after the test** a Qualitative and Quantitative analysis is necessary.
The Qualitative analysis seeks to know the feedback of the test-takers.
On the other hand, Quantitative analysis check de ´Difficulty Index´ (percent of all test takers that got the item correct), quantify the ´Item Discrimination´ (the items ability to discriminate between those who score high on the exam *vs* those who score low), and the ´Indices of Difficulty and Discrimination´ (for instance, Index of 0.30 - 0.70 means Moderate difficulty, from the Discrimination point of view are Very Good items).



Measuring practical skills

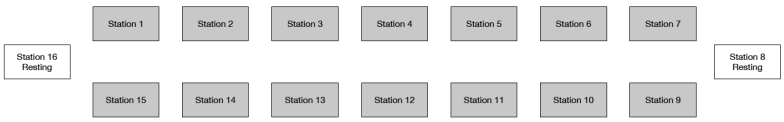
The practical skills measure the professional activities that are entrust able. One way to measure the practical skills is with one observer, one patient and one examinee. This is an authentic way, but it can be more difficult, depends on the patient's behavior and it is not standardized. It is good for application of skills and for formative feedback.

A second way is by means of an Objective Structured Clinical Examination (OSCE). The OSCE consists of 10-20 standardized stations (rooms where the exam-taker consecutively put into practice their knowledge, skills and attitudes in clinical situations -patients-). The stations are standardized, with patients or actors as patients, all the examinees have the same opportunities which is fair for all the examinees. On the contrary, the OSCE is difficult to organize and synchronized.

Delivering an OSCE test Scoring systems for practical tests

Prof. Hultin shows a sequence of ´active stations´ ´OSCE-EXAMPLE´ where the exam-takers should do the different steps in diagnosis, treatment plan and prescription of treatment.

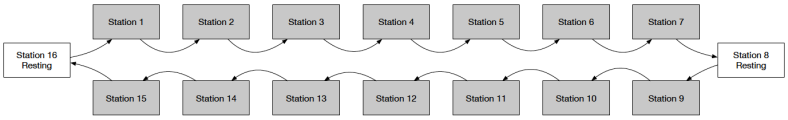
OSCE-EXAMPLE



14 active stations
- History taking
- Information
- Clinical examination
- Interprofessional skills
- Prescriptions
- Practical skills testing

2 passive stations
- *Chair*
- *WC*
- *Water*

SYNCHRONOUS ROTATION AT THE SOUND OF A WHISTLE…



E.g. 6 min inside station
- *Perform tasks at the station*

2 min transport time
- *Read instructions*

Starts with 2 min in front of instruction
—> 16 * (2+6 min) = 128 min (2 h 8 min)

Writing an OSCE station

Before writing an OSCE station is important to ask what skill is going to be tested, and the subject to be examined (a student, a licensing test, …). Besides, it is important to take into consideration what is the equipment needed for the exam, the score sheet, and the instructions both for the test taker and for the test examiner.

An exercise of writing an OSCE station (in *italics*).
Example Group 2:

Which level of performance?
undergraduate (becoming one and we are starting from a point that all OSCE stations are the EPAs from EPA1 to … 9)
Student? Licensing test?
Student
Equipment needed
-xrays
-study models
-intra and extra oral photographs
Instructions to test taker

-study the records; annotate whether you need more records? (1 points)
-make a summary of the records; make a diagnosis (5 points)
-make a treatment plan. (4 points)
Score sheet (0-10 p?)
10 points total
Instructions to test examiner
He should have a precorded answering document on which the points are divided. So also what is a minimum or maximum answer to get points.
Cues?
no

Scoring systems for practical test

The examiner has a checklist for the professional judgement on how well the exam- taker performed at this station. The grades are Excellent - Good pass - Clear pass - Borderline - Clear fail.

Quality assurance of an OSCE-station

As the MCQ tests, an OSCE station test needs, **before the test**, to work on the test development, and to standardize the information to test takers. **After the test**, it is important to check if the station has worked well, to ask the test takers for their experience, and finally to ask for the examiners about problems, quality of the stations, pass or fail criteria, and to collect all the information taken from the test takers.

Setting pass-scores on practical tests

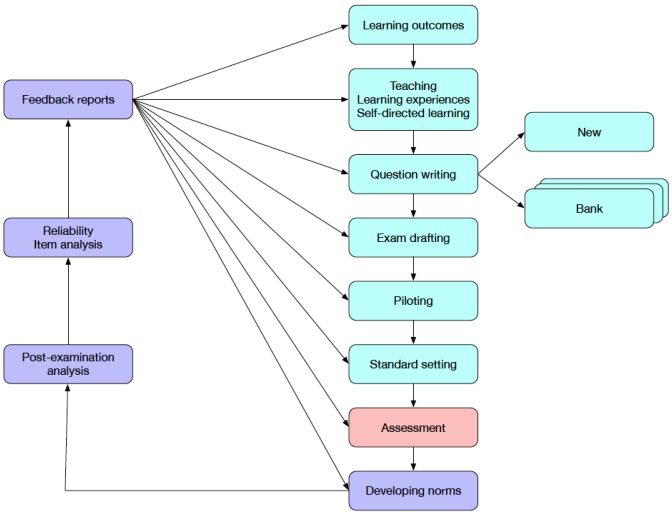
It is difficult to establish criteria for the pass-score on practical test. Prof. Hultin mentions several possibilities: the reference of others test-takers *vs* a criteria reference from an external standard. To compare a ´golden standard´, for instance to compare with a group that should pass or that should not pass. To use the ´Angoff´ criteria, which is the professional judgement of a group of experienced teachers.
As OSCE is a practical exam, and the attitudes are scored, if a examinee is rude or can be dangerous to a patient more than two ´red flags´ will be flagged as **Clear fail** on a station. A source of more qualitative information from the notes of the examiners can be added from the teacher´s discussion after the exam, the analysis of borderline students, or the limits of how large deficits are acceptable.

Test Questions and OSCE-Stations as a Source for Learning

There are some technical options to keep test questions in exams after they have been in a test. For instance, on one hand, the exam can be returned only to whom have a pass-fail decision, or to return the full test for the participants to learn from it. But, on the contrary, examinations can be done securely online, photos or videos of the exam cannot be allowed, be sure that all the exams are collected after the test, use a large bank of items to avoid reconstructing of the test after the exam, or not to do the same test to all test takers.
A **Source of Learning** is to consider that questions get old, and that a newly constructed question for each test is needed. Of course, it is possible to use old questions in formative test banks to train with (helps to learn important concepts!), and to change OSCE-stations.

The Cycle to do a continuous improvement of the test, ´THE EXAMINATION CYCLE´ is a sequence of steps that can help us to progress in our evaluations.

THE EXAMINATION CYCLE



The examination cycle (Tavakol & Dennick 2011)

Prof. Hultin ends his presentation suggesting two recommend courses:

- AMEE, An International Association for Medical Education. https://amee.org/courses
- Health Professional Assessment Consultancy (HPAC). https://www.hpac.sg

After a questions and answers session, Dr Hultin was thanked by the organizers and Dr. Melissa Disse, President of EFOSA handed over the Charles Bolender award 2022 to Dr Magnus Hultin.