

MCQ revision Workshop

In this document you will find the schedule for the revision workshop and some instructions for the revision procedure. The materials for discussion are sent to participants separately (please note these materials are confidential. We ask that you keep the folder closed and secure when you are not working on it)

We appreciate your serious preparation that will allow us to maximise our given time.

Contained within this document you will find:

1. Procedure for revision
2. Workshop schedule
3. Information on MCQ construction
4. Review article: Angoff method of standard setting

Procedure for the revision

Please note **in preparation** for each question and their corresponding notes:

- 1) a **global mark of 0, 1 or 2** in the lower right corner.
0=not suitable (too sophisticated/ professionally incorrect/ needs to be rewritten completely)
1=can be used but needs modifications
2=good, can be used without modifications
- 2) a judgement of phrasing
- 3) a judgement of the relation to the syllabus
- 4) a judgement of the scientific correctness
- 5) a judgement of the relevance

Mark your score in the appropriate box at the bottom of the form

Pre-review of phrasing	Pre-review of content Relation to HERMES Syllabus	Scientific	Relevance
<input type="checkbox"/> good	<input type="checkbox"/> relevant to Syllabus	<input type="checkbox"/> correct, definite	<input type="checkbox"/> high
<input type="checkbox"/> correction/s	<input type="checkbox"/> marginal	<input type="checkbox"/> in dispute, ambiguous	<input type="checkbox"/> medium
<input type="checkbox"/> unsuitable	<input type="checkbox"/> not relevant to Syllabus	<input type="checkbox"/> incorrect	<input type="checkbox"/> low
Final decisions	<input type="checkbox"/> Question accepted	<input type="checkbox"/> to be rephrased by author	<input type="checkbox"/> rejected

If you think that a question or comment needs to be modified, please **write out your proposal** and bring it to the revision meeting. You can refer to the literature reference cited in the MCQ if available. *As a review, please also read the handout on constructing MCQ's which is included in this folder.*

At the meeting, we will first listen to the marks of all participants for each question. Questions obtaining 2 points from all participants will be accepted without discussion. Questions obtaining 0 points from the majority will be rejected without discussion. As a guide for revision and feedback to the author, the most important reason of rejection will be noted.

6) **Angoff calibration method** for assessment of MCQ difficulty will be done. We kindly ask you to review the questions with this method. *As a guideline, please read the review article "The Angoff Method of Standard Setting for Licensure and Certification Examinations"*

In each MCQ, on the lower right corner, **Angoff Rating:** is indicated. **Please complete the cell with ratings in % with the following question:**

In your opinion, what percentage of minimally competent candidates will answer this item correctly?
In other words: Given 100 minimally competent candidates, how many will answer this item correctly?

Please do not be confused with giving a rating of the average score (percentage of all candidates). Focus on the minimally competent candidates, the ones who will just pass the exam, the "borderline candidate".

We define a **borderline candidate** for standard setting based on an understanding of the concept of the minimal acceptable level of competency for this qualification. Such candidates perform "on the borderline" between acceptable and unacceptable performance and thus have a 50:50 percent chance of passing the examination. The standard setting judges will thus have to assess the percentage of borderline candidates expected to answer each individual question correctly.

Regarding the **Kprime questions**, candidates will receive 1 point for 4 correct answers in one K' question and 0.5 points for 3 correct answers. This has to be rated differently from the A questions.

The judges have to estimate how many out of 100 "borderline candidates" would give 4 correct answers and therefore get the whole point (let's say 30) and add half the number of the borderline candidates who would give 3 correct answers and therefore get 0.5 points (let's say 10) -> in this example their rating should be 35 ($30 + (0.5 \times 10) = 35$).

Please do not hesitate to direct questions or issues, or to request some assistance to hermes@ersnet.org

Workshop Schedule:

Monday, March 23.03.2015	
When	What
08:30 - 09:00	Problematic questions 2014 examination
09:00 - 10:00	Review of MCQ handbook questions and contributions Angoff rating
10:00 - 10:30	Coffee Break
10:30-12:15	Review of MCQ handbook questions and contributions Angoff rating
12:15 - 13:30	Lunch
13:30-16:00	Review of MCQ handbook questions and contributions Angoff rating
16:00 - 16:30	Coffee Break
16:30-18:00	Review of MCQ handbook questions and contributions Angoff rating
18:00	END OF MEETING

Tuesday, March 24.03.2014	
When	What
08:30 - 10:00	Review of MCQ handbook questions and contributions Angoff rating
10:00 - 10:30	Coffee Break
10:30-12:15	Review of MCQ handbook questions and contributions Angoff rating
12:15 - 13:30	Lunch
13:30-16:00	Review of MCQ handbook questions and contributions Angoff rating
16:00 - 16:30	Coffee Break
16:30-18:00	Review of MCQ handbook questions and contributions Angoff rating
18 :00	END OF MEETING



CONSTRUCTING MULTIPLE CHOICE QUESTIONS

11 September 2013
BARCELONA ERS ANNUAL CONGRESS
European Respiratory Society
Julie-Lyn Noël
julie-lyn.noel@ersnet.org



ITEM TERMINOLOGY (1)

Stem

A 52-year old man comes to the office because of a 2-day history of increasing shortness of breath and cough productive of thick, yellow-green sputum. He has smoked one pack of cigarettes daily for 30 years. His temperature is 37.2 C. Breath sounds are distant with a few rhonchi and wheezes. Leukocyte count is 9000/mm³ with a normal differential. Gram's stain of sputum shows numerous neutrophils and gram-positive diplococci. Chest x-ray shows hyperinflation.

Lead-in

Which of the following is the most likely diagnosis?

european respiratory society every breath counts



OVERVIEW

- Item terminology
- HERMES examination item formats
- Technical Item Flaws
- Basic rules for writing MCQ items
- The clinical vignette
- Clinical content question examples

european respiratory society every breath counts



ITEM TERMINOLOGY (2)

Options

- (A) Asthma
- (B) Bronchoectasis
- (C) Bronchitis
- (D) Pulmonary embolus
- (E) Streptococcal pneumonia

An MCQ has body parts, so MCQ authors must complete the parts in order to have the MCQ function. The stem tells the story, the lead-in asks the question and the options are the possible answers.

european respiratory society every breath counts



ITEM TERMINOLOGY

european respiratory society every breath counts



HERMES EXAMINATION ITEM FORMATS

european respiratory society every breath counts

ERS

TYPE A+

- **Type A questions (positively formulated one-best-answer)**
 - For each question, there is only one correct or the most appropriate answer.

HERMES examinations have 5 options for this type of question.

european respiratory society every breath counts

ERS

TECHNICAL ITEM FLAWS

european respiratory society every breath counts

ERS

TYPE A-

- **Type A-negative (negatively formulated one-best-answer)**
 - For each question, there is only one incorrect answer, only one exception, and one answer that appears LEAST appropriate. The negation is written in bold

Which of the following antibiotics must **not** be given to an 18-month-old child with an acute middle ear infection?

A. amoxicillin
B. Ampicillin
C. Cefaclor
D. Co-trimoxazole
E. Doxycycline

1. Indicated when the candidate must identify an important exception or something that must not be done or given to a patient because it could cause serious harm.
2. *Be careful not to misuse this type of question, very often they are not well focused and have heterogeneous options.*

european respiratory society every breath counts

Technical Item Flaws

- Issues related to Testwiseness
- Issues related to irrelevant difficulty

ERS

TYPE K'

- **Type Kprime - K' (quadruple correct/incorrect decision)**
 - For each question or incomplete statement there are 4 options. Decide for each option whether they are correct or incorrect and mark them with a (+) or (-). Selecting four options correctly results in one full point. If 3 options are correctly marked, half a point is given.

The risk for malignancy of a solitary pulmonary nodule depends on:

A. The age of the patient
B. The gender of the patient
C. The smoking history of the patient
D. The size of the nodule

1. Stem and options must be phrased positively whenever possible
2. Stem must be phrased neutrally with regard to the number of correct options
3. Each option must unambiguously be correct or false
4. Each option must contain only 1 statement
5. Every option must be independent from the others
6. Correct and false options must be balanced throughout the exam
7. Avoid having options to be all true or false.

european respiratory society every breath counts

TESTWISENESS



GRAMMATICAL CUES

A 60-year-old man is brought to the emergency department by the police, who found him lying unconscious on the sidewalk. After ascertaining that the airway is open, the first step in management should be **intravenous administration** of

- (A) Examination of cerebrospinal fluid
- (B) Glucose with vitamin B1 (thiamine)
- (C) CT Scan of the head
- (D) Phenytoin
- (E) Diazepam

Intravenous gives the clue and immediately eliminates CSF examination, glucose and CT scan.

european respiratory society every breath counts



LONG CORRECT ANSWER

Secondary gain is

- (A) Synonymous with malingering
- (B) A frequent problem in obsessive-compulsive disorder
- (C) A complication of variety of illnesses and tends to prolong many of them
- (D) Never seen in organic brain damage

Often MCQ authors invest a lot on the correct answer and forget the other options. The correct answer therefore has the tendency to be longest. Testwise participants can sense this.

european respiratory society every breath counts



LOGICAL CUES

Crime is

- (A) Equally distributed among the social classes
- (B) Overrepresented among the poor
- (C) Overrepresented among the middle class and the rich
- (D) Primarily an indication of psychosexual maladjustment
- (E) Reaching a plateau of tolerability for the nation

Option A, B, C and E could not be logically correct answers because one cannot exactly know about distribution, overrepresentation and tolerability.

european respiratory society every breath counts



WORD REPEATS

A 58-year-old man with a history of heavy alcohol use and previous psychiatric hospitalisation is confused and agitated. He speaks of experiencing the world as **unreal**. The symptom is called

- (A) Depersonalisation
- (B) Derailment
- (C) Derealization
- (D) Focal memory deficit
- (E) Signal anxiety

The mention of the word real gives a clue to the testwise participant for the correct answer.

european respiratory society every breath counts



ABSOLUTE TERMS

In patients with advanced dementia, Alzheimer's type, the memory defect

- (A) Can be treated adequately with phosphatidylcholine (lecithin)
- (B) Could be a sequela of early parkinsonism
- (C) Is **never** seen in patients with neurofibrillary tangles at autopsy
- (D) Is **never** severe
- (E) Possibly involves the cholinergic system

We can never say NEVER.

european respiratory society every breath counts



CONVERGENCE STRATEGY

Local anaesthetics are most effective in the

- (A) Anionic form, acting from **inside** the nerve membrane
- (B) Cationic form, acting from **inside** the nerve membrane
- (C) Cationic form, acting from outside the nerve membrane
- (D) Uncharged form, acting from **inside** the nerve membrane
- (E) Uncharged form, acting from outside the nerve membrane

Inside is mentioned 3 times, Cationic 2 times and the correct answer converges to B where cationic and inside are mentioned.

european respiratory society every breath counts

ISSUES RELATED TO TESTWISENESS

- Grammatical cues
- Logical cues
- Absolute terms
- Long correct answer
- Word repeats
- Convergence strategy

NUMERIC DATA STATED INCONSISTENTLY

Following a second episode of infection, what is the likelihood that a woman is infertile

- (A) Less than 20%
- (B) 20 to 30 %
- (C) Greater than 50%
- (D) 90%
- (E) 75%

Numeric data must be consistent for example maybe 10, 20, 30, 40, 50. Otherwise, the question does not make sense.

IRRELEVANT DIFFICULTY

VAGUE FREQUENCY TERMS

Severe obesity in early adolescence

- (A) Usually responds dramatically to dietary regimens
- (B) Often is related to endocrine disorders
- (C) Has a 75% chance of clearing spontaneously
- (D) Shows a poor prognosis
- (E) Usually responds to pharmacotherapy and intensive psychotherapy

Even experts themselves do not agree to what is actually meant with usual, often etc. Avoid using such terms.

LONG, COMPLICATED OR DOUBLE OPTIONS

Peer review committees in HMO's may move to take action against a physician's credentials to care for participants of the HMO. There is an associated requirement to assure that the physician receives due process in the course of these activities. Due process must include which of the following?

- (A) Notice, an impartial forum, council, a chance to hear and confront evidence in defense
- (B) Proper notice, a tribunal empowered to make the decision, a chance to confront witnesses against him/her, and a chance to present evidence in defense.
- (C) Reasonable and timely notice, impartial panel empowered to make a decision, a chance to hear evidence against himself/herself and to confront witnesses, and the ability to present evidence in defense

The HERMES examination does not test reading comprehension.

NON-PARALLEL LANGUAGE, ILLOGICAL ORDER OF OPTIONS

In a vaccine trial, 200 2-year-old boys were given a vaccine against certain disease and then monitored for five years for occurrence of the disease. Of this group, 85% never contracted the disease. Which of the following statements concerning these results is correct?

- (A) No conclusion can be drawn, since no follow-up was made of nonvaccinated children
- (B) The number of cases (ie 30 cases over five years) is too small for statistically meaningful conclusions
- (C) No conclusions can be drawn because the trial involved only boys
- (D) Vaccine efficiency (%) us calculated as 85-15/100

Use parallel language in order for your stem to be understood. Order the options logically i.e. numeric options in ascending order, alphabetically arranged names of drugs etc.



NONE OF THE ABOVE OPTION

Which city is closest to New York City?

- (A) Boston
- (B) Chicago
- (C) Dallas
- (D) Los Angeles
- (E) None of the Above

For the case of clinical vignette questions, none of the above options can always be substituted with alternative actions i.e. discharge patient, observe patient, stop pharmacotherapy etc.

european respiratory society every breath counts



BASIC RULES FOR WRITING MCQ ITEMS

european respiratory society every breath counts

ERS TRICKY STEMS OR UNNECESSARILY COMPLICATED STEMS

Arrange the parents of the following children with Down's syndrome in order of highest to lowest risk of recurrence. Assume that the maternal age in all cases is 22 years and that a subsequent pregnancy occurs within 5 years. The karyotypes of the daughters are:

- I. 46, XX, -14, +T (14q21q)pat
- II. 46, XX, -14, +T (14q21q)de novo
- III. 46, XX, -14, +T (14q21q)mat
- IV. 46, XX, -14, +T (14q21q)pat
- V. 46, XX, -14, +T (14q21q) parents not karyotyped

- (A) III, IV, I, V, II
- (B) IV, III, V, I, II
- (C) III, I, IV, V, II
- (D) IV, III, I, V, II
- (E) III, IV, I, II, V

It is unfair to add unnecessary complexity and tricky stems for the examination. These kinds of questions do not have positive selection of good candidates.

european respiratory society every breath counts



COVER THE OPTIONS RULE

A 52-year old man comes to the office because of a 2-day history of increasing shortness of breath and cough productive of thick, yellow-green sputum. He has smoked one pack of cigarettes daily for 30 years. His temperature is 37.2 C. Breath sounds are distant with a few rhonchi and wheezes. Leukocyte count is 9000/mm³ with a normal differential. Gram's stain of sputum shows numerous neutrophils and gram-positive diplococci. Chest x-ray shows hyperinflation.

Which of the following is the most likely diagnosis?

- (A) Asthma
- (B) Bronchoectasis
- (C) Bronchitis
- (D) Pulmonary embolus
- (E) Streptococcal pneumonia

european respiratory society every breath counts



ISSUES RELATED TO IRRELEVANT DIFFICULTY

- Options are long, complicated or double
- Numeric data are not stated consistently
- Terms in the options are vague (e.g. rarely, usually, often)
- Language in the options is not parallel
- Options are in a nonlogical order
- None of the above is used
- The answer to an item is hinged to the answer of a related item

european respiratory society every breath counts



COVER THE OPTIONS RULE

A 52-year old man comes to the office because of a 2-day history of increasing shortness of breath and cough productive of thick, yellow-green sputum. He has smoked one pack of cigarettes daily for 30 years. His temperature is 37.2 C. Breath sounds are distant with a few rhonchi and wheezes. Leukocyte count is 9000/mm³ with a normal differential. Gram's stain of sputum shows numerous neutrophils and gram-positive diplococci. Chest x-ray shows hyperinflation.

Which of the following is the most likely diagnosis?

If all the options are covered, the possible options can still be identified. The lead-in is therefore very focused and the options will be homogenous.

european respiratory society every breath counts



SHAPELINESS

A 52-year old man comes to the office because of a 2-day history of increasing shortness of breath and cough productive of thick, yellow-green sputum. He has smoked one pack of cigarettes daily for 30 years. His temperature is 37.2 C. Breath sounds are distant with a few rhonchi and wheezes. Leukocyte count is 9000/mm³ with a normal differential. Gram's stain of sputum shows numerous neutrophils and gram-positive diplococci. Chest x-ray shows hyperinflation.

Which of the following is the most likely diagnosis?

- (A) Asthma
- (B) Bronchoectasis
- (C) Bronchitis
- (D) Pulmonary embolus
- (E) Streptococcal pneumonia

european respiratory society every breath counts



SAMPLE BAD QUESTION

Which of the following is true about pseudogout?

- (A) It occurs frequently in women.
- (B) It is seldom associated with acute pain in a joint
- (C) It may be associated with a finding of chondrocalcinosis.
- (D) It is clearly hereditary in most cases.
- (E) It responds well to treatment with allopurinol.

european respiratory society every breath counts



SHAPELINESS

A 52-year old man comes to the office because of a 2-day history of increasing shortness of breath and cough productive of thick, yellow-green sputum. He has smoked one pack of cigarettes daily for 30 years. His temperature is 37.2 C. Breath sounds are distant with a few rhonchi and wheezes. Leukocyte count is 9000/mm³ with a normal differential. Gram's stain of sputum shows numerous neutrophils and gram-positive diplococci. Chest x-ray shows hyperinflation.

Which of the following is the most likely diagnosis?

- (A) Asthma
- (B) Bronchoectasis
- (C) Bronchitis
- (D) Pulmonary embolus
- (E) Streptococcal pneumonia

Relevant data are covered in the stem. The lead-in is focused and the options are homogenous and short.

european respiratory society every breath counts



COVER THE OPTIONS

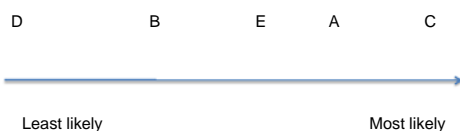
Which of the following is true about pseudogout?

We have no clue as to what the possible options could be.

european respiratory society every breath counts



RANK ORDERED OPTIONS

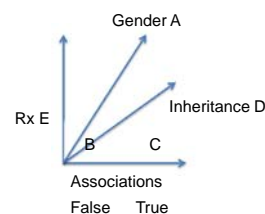


All options can be rank ordered from the least likely to the most likely.

european respiratory society every breath counts



RANK ORDER THE OPTIONS



Many aspects are given in the options i.e. drug therapy, epidemiology, genetics, association with clinical findings. Therefore, it is not possible to rank order the options from least likely to most likely.

european respiratory society every breath counts

ERS

SHAPELY

Which of the following is true about pseudogout?

- (A) It occurs frequently in women.
- (B) It is seldom associated with acute pain in a joint
- (C) It may be associated with a finding of chondrocalcinosis.
- (D) It is clearly hereditary in most cases.
- (E) It responds well to treatment with allopurinol.

No relevant stem is given. The lead-in is not focused. The options are long.

European Respiratory Society every breath counts

ERS

THE CLINICAL VIGNETTE

European Respiratory Society every breath counts

ERS

THEREFORE A FLAWED ITEM

Which of the following is true about pseudogout?

- (A) It occurs frequently in women.
- (B) It is seldom associated with acute pain in a joint
- (C) It may be associated with a finding of chondrocalcinosis.
- (D) It is clearly hereditary in most cases.
- (E) It responds well to treatment with allopurinol.

European Respiratory Society every breath counts

ERS

GUIDELINES FOR ITEM CONTENT

- Test application of knowledge using clinical vignettes to pose medical decisions in patient care situations
- Focus items on common or potentially catastrophic problems; avoid zebras and esoterica
- Pose clinical decision-making tasks that would be expected of a respiratory specialist (HERMES context)
- Avoid clinical situations that would be handled by another (sub)specialist

European Respiratory Society every breath counts

ERS

BASIC RULES

- Item stems must satisfy the Cover-The-Options Rule: Even without looking at the options, the lead-in gives a clue already to possible answers (i.e. diagnostic procedures, diagnosis, treatment etc)
- Items should have the right shape: long stems, and short options
- Lead-ins must be focused
- Options must be homogenous

European Respiratory Society every breath counts

ERS

CONSTRUCTING THE STEM

- A **Good stem** provides sufficient and relevant information and of appropriate length. It must not be too long. In a HERMES examination, candidates have only 2 minutes for each question.

A 52-year old man comes to the office because of a 2-day history of increasing shortness of breath and cough productive of thick, yellow-green sputum. He has smoked one pack of cigarettes daily for 30 years. His temperature is 37.2 C. Breath sounds are distant with a few rhonchi and wheezes. Leukocyte count is 9000/mm³ with a normal differential. Gram's stain of sputum shows numerous neutrophils and gram-positive diplococci. Chest x-ray shows hyperinflation.

European Respiratory Society every breath counts



COMPONENTS OF THE PATIENT VIGNETTE

- Age, gender (e.g. 45 year old man)
- Site of care (e.g. comes to the emergency department)
- Presenting complaint (e.g. due to dyspnoea)
- Duration (e.g. since 3 hours)
- Patient history
- Physical Findings
- +/- results of diagnostic studies
- +/- initial treatment, subsequent findings etc

european respiratory society every breath counts



EXAMPLE LEAD-INS: DIAGNOSIS

- Which of the following is the most likely diagnosis?
- Which of the following is the most appropriate next step in diagnosis?
- Which of the following is most likely to confirm the diagnosis?

european respiratory society every breath counts



FINE POINTS ON ITEM STEMS

- **Use of real patients**
 - Incidental findings can be done
 - Avoid red herrings
- **Use of reference materials**
 - Provide information if in real life, someone would be likely to refer to a reference source to obtain the information (guidelines, ERS handbook etc)
- **Use of patient's or physicians own words**
 - Not useful to use patient's own words
 - Useful to ask examinee to select the most appropriate physician response to a patient
- **Patients who lie**
 - All MCQ patients should always tell the truth

european respiratory society every breath counts



EXAMPLE LEAD-INS: MANAGEMENT

- Which of the following is the most appropriate initial or next step in patient care?
- Which of the following is the most effective management?
- Which of the following is the most appropriate pharmacotherapy?
- Which of the following is the first priority in caring for this patient?

european respiratory society every breath counts



EXAMPLE LEAD-INS: MECHANISMS OF DISEASE

- Which of the following is the most likely explanation for these findings?
- Which of the following is the most likely pathogen?
- Which of the following findings is most likely to be increased/decreased?
- A biopsy is most likely to show which of the following?

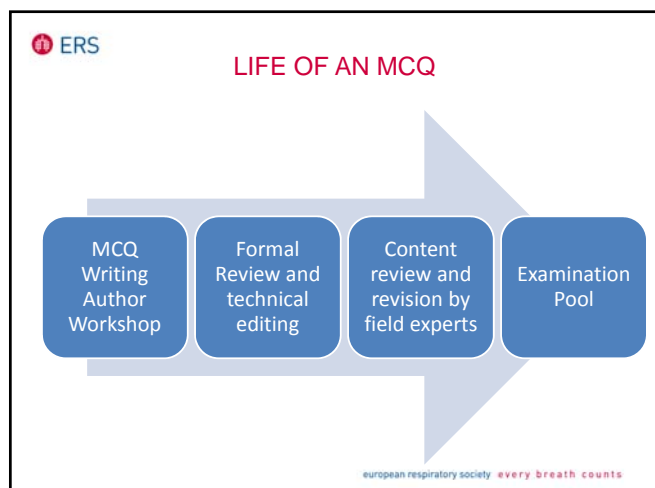
european respiratory society every breath counts



SUBMITTING AN ERS MCQ

Adult Respiratory Medicine
Paediatric Respiratory Medicine
Spirometry

european respiratory society every breath counts

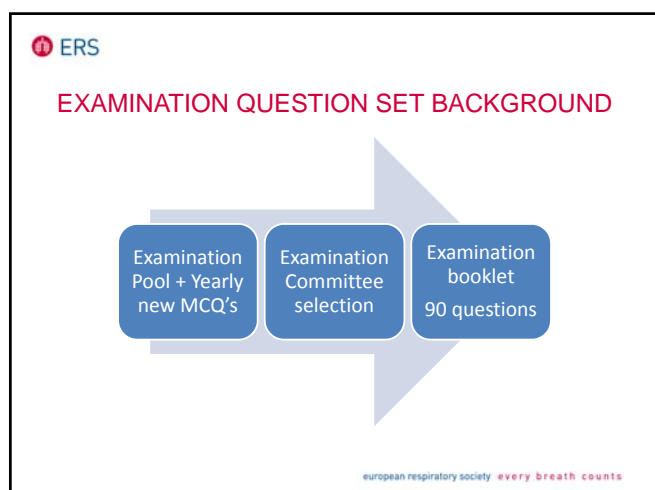


ERS

EXAMINATION BLUEPRINT

- [Blueprint Adult European Exam General Respiratory Medicine 2012.pdf](#)
- [PRM Blueprint revised.pdf](#)

european respiratory society every breath counts



ERS

PLAN AHEAD

- Do I have a good case that I can adapt for an MCQ clinical vignette?
- What will be my teaching point for the question? Use the examination blueprint.
 - Basic Science
 - Diagnosis
 - Diagnostic procedures
 - Therapy
 - Etc...
- What are the possible homogenous options?
 - Cover the options rule

european respiratory society every breath counts

ERS

BASIS: SYLLABI

- [Adult syllabus.pdf](#)
- [paediasyllabus HERMES Breathe Paediatric Syllabus.pdf](#)

european respiratory society every breath counts

ERS

USE REFERENCES

- Current ERS Guidelines
- 2nd edition of ERS Handbook
- Paediatric Handbook

european respiratory society every breath counts



BEGIN WRITING THE MCQ

- Reference your teaching point in the MCQ to the examination blueprint and the HERMES syllabus
- Have a complete literature citation
- Provide good images (at least 300 dpi) if relevant to your MCQ
- Provide tables or findings in English
- Give generic names of drugs
- Use well-known units or if possible give both conversions for example for pressures that must be given both in kPa and mmHg
- Use acronyms as frequently used in the HERMES examinations and the MCQ handbook. Refer to the [list of abbreviations](#)
- Use the 2014 template and complete accordingly [MCQ blank 2014.doc](#)

european respiratory society every breath counts



european respiratory society every breath counts

ersnet.org



MCQ SUBMISSION

- Strictly use the template, completing all the required data
- Reference to appropriate syllabus item
- Literature, guidelines, handbook chapters must be clearly cited as these are important metadata for revision and results processing of the examination
- SUBMISSION DEADLINE: **30 November 2013**
- Email to julie-lyn.noel@ersnet.org

european respiratory society every breath counts



BENEFITS FOR MCQ AUTHOR FOR ADULT AND PAEDIATRIC MEDICINE EXAMS

- CME credits: 3 points accepted questions
- Honorarium 50 EUR per accepted question
- Can be FUN!!!

european respiratory society every breath counts

**Canadian Association of Medical
Radiation Technologists**

**The Angoff Method of Standard
Setting for Licensure and
Certification Examinations**

Prepared by:



***Assessment Strategies
Stratégies en évaluation***

January 2006

*Assessment Strategies
Stratégies en évaluation*
1100 Blainville, Suite 210
Ottawa, ON K1M 9H3

Introduction

Many health and health related agencies require licensure for their professionals as one means of assuring the quality of practice. As a standardized examination is often a requirement for licensure, determination of an appropriate pass mark for the examination is essential to the effectiveness of the process.

Relevant Issues

Setting a pass mark for an examination is setting a standard of performance on which decisions will be made about an individual's level of competence in a given field of practice. The pass mark determination is a judgment made by informed individuals (i.e., experts in the field of practice). It is determined through a rational discussion of the field of practice as well as an awareness of the consequences involved when a decision affecting individuals is made.

The Pass Mark and Consequences

Whenever a pass mark is determined for a licensure examination, there are a number of potential consequences that must be anticipated: an inappropriately low pass mark will allow non-competent candidates to practice, perhaps at the expense of the public welfare; an unrealistically high pass mark will exclude competent candidates from being licensed.

The accuracy and precision of the measuring instrument (i.e., examination validity and reliability) must also be considered. Examinations are not perfect: they cannot include all the knowledge and skills in a given field of practice. An examination can only sample the field. Furthermore, if it were possible to repeatedly administer the same examination to a single candidate 100 times, the candidate's score would likely not be exactly the same each time. The inconsistency of the scores is a result of the reliability of the examination and the variables affecting candidate performance (such as anxiety level and health).

The Angoff Method

The Angoff method requires expert judges to discuss the issues involved in determining a pass mark and to evaluate the examination by using a well-defined and rational procedure.

1. Competence and the Borderline Candidate

The Angoff method is based on the concept of the borderline or minimally competent candidate. The minimally competent candidate can be conceptualized as the candidate possessing the minimum level of knowledge and skills necessary to perform at a licensure level. This candidate performs at a level "on the borderline" between acceptable and unacceptable performance. It is essential that each judge arrive at a clear and specific definition of the minimally competent candidate.

To better understand the concept of the minimally competent candidate, it is often helpful to think of the people you work with everyday; a few of them are the “superstars” performing at a level well above the majority, while others perform rather poorly and perhaps should not be practising. Somewhere between these two extremes is the group that performs at the level of minimum competence. The borderline candidate belongs to the group that just qualifies for licensure.

2. Rating the Items

The Angoff method requires the judges to independently rate each item in the examination in terms of the minimally competent candidate. For each item, each judge answers the question: “In your opinion, what percentage of minimally competent candidates will answer this item correctly?” Alternately phrased, “Given 100 minimally competent candidates, how many will answer this item correctly?” The judge then indicates the appropriate percentage on the rating form and proceeds with the next item.

One common error made when rating items is to base the rating on the average candidate or the exceptional candidate rather than on the borderline candidate. Another potential error involves the interpretation of the question asked for each item. The items are to be rated in terms of how many borderline candidates will answer the item correctly. In a large group of borderline candidates, only some may actually know the correct response. It should not be assumed that all borderline candidates will know the answer. Finally, although item statistics may be used to provide additional information to the judges, the ratings should not be based solely on these statistics; item statistics are calculated on the entire candidate population, not on the borderline group alone.

3. Determining the Pass Mark

Once all the judges have rated each item in the examination, the ratings are collated and tabulated. The ratings for any single item should be in agreement. By agreement it is meant that the ratings for an item must all be within a certain percentage range (e.g., a 30% range). If the range of the ratings is greater than the specified range, the judges providing the extreme ratings are asked to explain why they rated the item in that fashion. The other judges should explain why they rated the item as they did. If a brief discussion of these reasons does not cause any of the judges to change their ratings, the ratings should be left as they are. Once the discussion has ended, the average rating is calculated for each item and then for the total examination. This results in a percentage value which is the percentage score expected to be achieved by the borderline candidate. In addition to the expert ratings, a variety of relevant data is carefully considered to ensure that the standard that examinees will be required to achieve is valid and fair. This can include information on the preparation of new graduates, data on the performance of examinees on previously administered examinations, and pertinent psychometric findings. Based on all of this information, a point is set on a measurement scale that represents the minimum acceptable standard.

As an illustration of the rating process, consider a fictitious application of the Angoff method, with a panel of six judges setting the pass mark on a four-item exam. Following the orientation session, the judges provide independent ratings on each of the items in the exam. The ratings obtained are presented in Table 1. With a 30%-rule specified to define the target level of agreement, items 1, 2, and 4 require no post-rating discussion (i.e., for each of these items, the extreme ratings do not differ by more than 30%). Item 3, however, is identified as requiring discussion because of the 40% difference in the ratings of Judge 3 and Judge 5 on this item. As a result of the discussion, Judge 3 decreases her rating from 85% to 75% and Judge 5 increases his rating from 45% to 50%. The average ratings are then calculated for each item, and the average of these values is calculated, to arrive at an overall pass mark of 69%.

Table 1: Example of the Angoff Method

Item #	Judges' Ratings (%)						Average Rating
	Judge 1	Judge 2	Judge 3	Judge 4	Judge 5	Judge 6	
1	65	70	65	65	70	65	67
2	85	70	60	80	70	80	74
3	70	65	85→75	70	45→50	70	67
4	75	60	65	70	75	70	69
Overall Average Rating							69

4. Factors for Successful Implementation

A number of factors contribute to the successful implementation of the Angoff method. An effective training session is essential in orienting the judges to the concept of the minimally competent candidate. As well, discussion and modification of extreme ratings help ensure that a defensible and valid cutoff score is established.

Summary

The Angoff method allows expert judges to determine an appropriate pass mark for an examination, based on a discussion of the issues involved in licensure and their assessment of the examination. A major advantage to this methodology is that the determined pass mark is based on the content of the examination and not on group performance. The pass mark is set in direct reference to the candidates, the competence required of the candidates, and the difficulty of the items themselves and, as a result, it is considered fair and valid.

