




Evaluation of EBOD 2009

MSc. Danny G.P. Mathysen
EBOD Assessment Officer

Antwerp University Hospital
Department of Ophthalmology
Wilrijkstraat 10 – B-2650 Edegem
 danny.mathysen@uza.be



EBO Diploma Examination

- EBO Diploma Examination
 - Test designed to assess **knowledge** and **clinical skills** requisites to deliver a **European standard** of ophthalmologic care in hospital and private settings.
 - Organised on a yearly basis since its introduction in June 1995



Components of education

- **Knowledge**

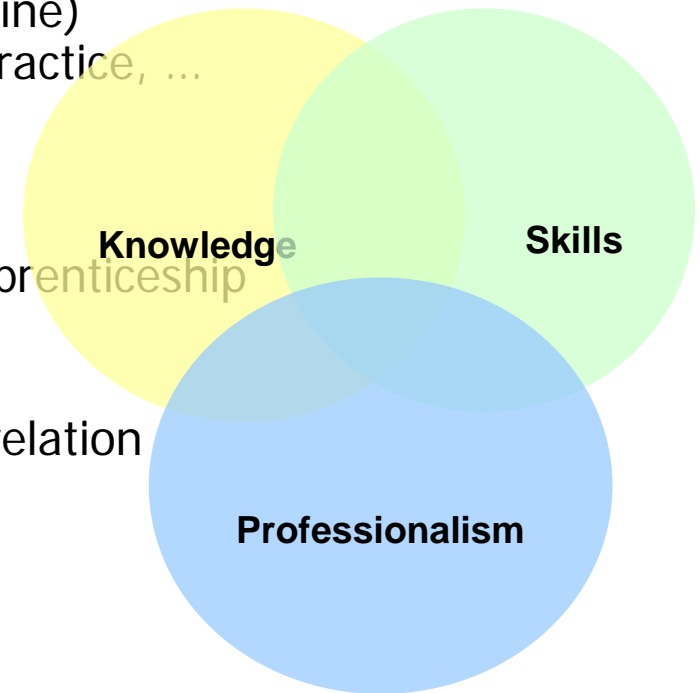
- Textbooks, guidelines, articles, (online) courses, clinical ophthalmological practice, ...

- **Skills**

- Practical/technical skills
→ measure of ophthalmological apprenticeship

- **Professionalism**

- Personal skills: empathy, attitude, relation with related paramedicals, ...



Structure of EBOD

■ Written paper

- 52 MCQs with 5 T/F items each (260 items)
- 10 topics (see EBO website)
- 40 percent of total candidate score
- Languages: English, French, German
- Assessment of knowledge

■ Oral examination (Viva Voce)

- 4 different topics (see EBO website)
- 60 percent of total candidate score
- Languages: English (native language)
- Assessment of knowledge, skills and professionalism

The image shows a sample 'Answer Sheet' from the European Board of Ophthalmology (EBO). The form includes fields for 'Candidate Name', 'VISA Address', and 'Country of residence' (with 'SWITZERLAND' entered). It also has a 'Date of Language' field and a 'SPONSOR' field. Below these are 'INSTRUCTIONS' and a 'QUESTIONS' section. The 'QUESTIONS' section is a grid with 52 rows and 5 columns (A, B, C, D, E) for multiple-choice questions, plus 5 columns for True/False (T/F) items. The grid contains various question numbers and options.





EBOD vs other (European) exams

- Written examinations exist in different formats
 - All of them use MCQs
 - (In Europe) there are several formats:
 - True/False (T/F) items (EBOD)
 - Single-Best-Answer (SBA) items
 - *(Extended Matching Indices (EMI) items)*



EBOD vs other (European) exams

- Advantages of T/F items
 - Easy to construct
 - Reliable in case of translation
 - Accessibility (e.g. dyslexia)
 - Duration of the examination
 - Relatively easy to process
 - No need to restructure current MCQ-bank
- Disadvantages of T/F items
 - Assessing mainly knowledge
 - Probability of guessing right = 50 %



EBOD vs other (European) exams

- How to overcome the disadvantages of T/F items?
 - Viva Voce examination
 - Covering all components: knowledge, skills, professionalism
 - Introduction of negative marking
 - Increase of discriminative power of examination
 - Reduction of guess factor
 - wild guesses will be punished
 - guesses by reasoning (partial knowledge) will be rewarded



Spread of total test scores with negative marking

Spread of total test scores
without negative marking

0

260



Statistical analysis of EBOD 2008

- C_iTO
 - Examination assessment company established in The Netherlands
 - Specialised in organising national examinations
 - Provided software tools
 - Design of the MCQ answer sheet
 - Design of the Viva Voce mark sheets
 - Statistical analysis output (TiaPlus[®])



Statistical analysis of EBOD 2009

- From C₂TO to SpeedWell
 - SpeedWell is specialised in organising medical examinations
 - Change to optical reader system
 - continuous and yearly increase of applications
 - Provided software tools
 - Design of the MCQ answer sheet
 - Design of the Viva Voce mark sheets
 - Statistical analysis output (MultiQuest[®]) based on similar statistical techniques



Demographics of EBOD

Country	2008	2009	Δ	Country	2008	2009	Δ
Austria	2	5	↑	Latvia	2	1	↓
Belgium	23	25	↑	Lithuania	1	1	→
Bulgaria		4	↑	Norway		1	↑
Czech Republic	2	2	→	Poland	1	2	↑
Denmark	4	6	↑	Slovakia	1	1	→
Estonia	3	2	↓	Slovenia	6	5	↓
Finland	7	2	↓	Spain	14	17	↑
France	92	96	↑	Sweden	6	5	↓
Germany	44	59	↑	Switzerland	32	29	↓
Greece	10	19	↑	The Netherlands	7	7	→
Hungary	1	2	↑	Turkey	11	5	↓
Ireland	5	5	→	United Kingdom	2	1	↓
Italy	4	6	↑	Total	284	308	↑



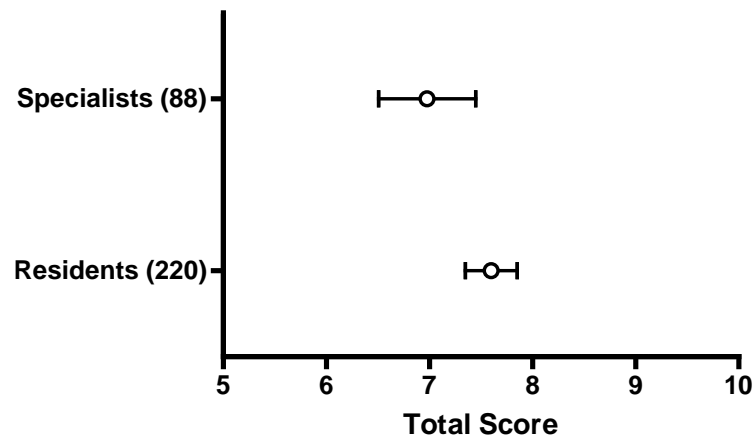
Demographics of EBOD

Country	MCQ	VV	Total	Country	MCQ	VV	Total
Austria	214.00	7.85	8.31	<i>Latvia</i>	<i>194.00</i>	<i>8.25</i>	<i>7.35</i>
Belgium	203.48	7.39	7.35	<i>Lithuania</i>	<i>200.00</i>	<i>7.25</i>	<i>7.15</i>
Bulgaria	208.00	8.19	8.21	<i>Norway</i>	<i>200.00</i>	<i>7.50</i>	<i>7.30</i>
Czech Republic	190.50	6.88	6.13	Poland	200.00	7.00	7.00
Denmark	212.50	8.13	8.48	<i>Slovakia</i>	<i>202.00</i>	<i>7.00</i>	<i>7.00</i>
Estonia	190.00	6.38	5.83	Slovenia	213.00	8.60	8.60
Finland	215.00	7.63	8.18	Spain	200.59	7.46	7.25
France	201.41	7.63	7.38	Sweden	192.60	6.75	6.29
Germany	208.46	7.70	7.87	Switzerland	207.97	7.81	7.86
Greece	200.16	7.64	7.32	The Netherlands	206.29	7.96	7.86
Hungary	188.50	6.00	5.60	Turkey	210.20	7.35	7.69
Ireland	200.20	7.35	7.13	<i>United Kingdom</i>	<i>198.00</i>	<i>8.00</i>	<i>7.60</i>
Italy	206.50	7.54	7.66	Total	204.11	7.62	7.54

EBOD 2009: Analysis of MCQs

- MCQ total scores
 - Range of total scores: 154 – 230
 - Mean \pm SD total score: 204.11 \pm 13.04

EBOD 2009 MCQ Scores
with 95 % Confidence Intervals

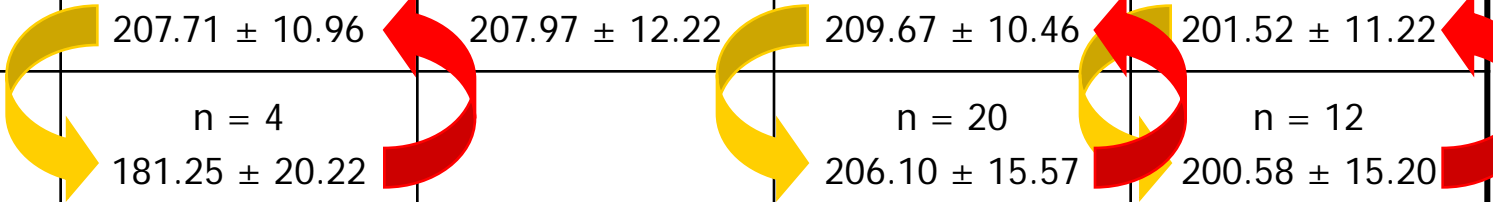


Residents	n = 220	205.40 \pm 12.18	Specialists	n = 88	200.91 \pm 14.41
-----------	---------	--------------------	-------------	--------	--------------------



EBOD 2009: Analysis of MCQs

	Belgium	Switzerland	Germany	France
Residents	n = 21 207.71 ± 10.96	n = 29 207.97 ± 12.22	n = 39 209.67 ± 10.46	n = 84 201.52 ± 11.22
Specialists	n = 4 181.25 ± 20.22		n = 20 206.10 ± 15.57	n = 12 200.58 ± 15.20
Total	n = 25 203.48 ± 16.14	n = 29 207.97 ± 12.22	n = 59 208.46 ± 12.54	n = 96 201.41 ± 11.80



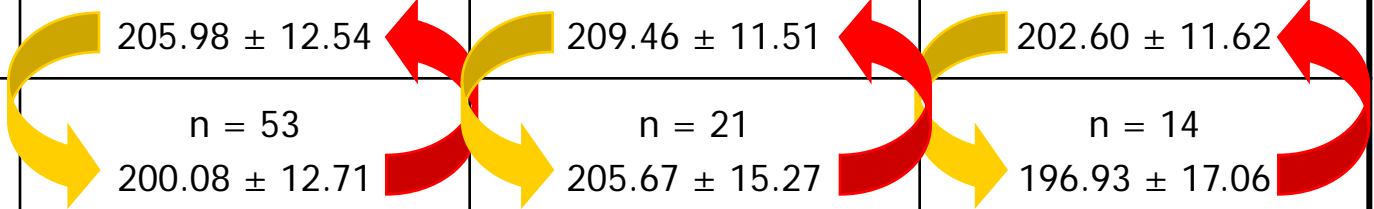
Residents have **higher total MCQ scores** with **lower standard deviations** when compared to specialists.

In general there are **no statistically significant differences** between **countries**.



EBOD 2009: Analysis of MCQs

	English	German	French
Residents	n = 58 205.98 ± 12.54	n = 61 209.46 ± 11.51	n = 101 202.60 ± 11.62
Specialists	n = 53 200.08 ± 12.71	n = 21 205.67 ± 15.27	n = 14 196.93 ± 17.06
Total	n = 111 203.16 ± 12.96	n = 59 208.46 ± 12.54	n = 115 201.91 ± 12.55



Residents have **higher total MCQ scores** with **lower standard deviations** when compared to specialists.

In general there are **no statistically significant differences** between **languages**.



EBOD 2009: Analysis of MCQs

- Cronbach's coefficient alpha (r) = 0.78
 - Estimator of the lower bound of the reliability of EBOD 2009 (95% CI: 0.75 – 0.81)

$$r = \frac{260}{260 - 1} \left[1 - \frac{\sum_{i=1}^{260} \sigma_i^2}{\left(\sum_{i=1}^{260} Rit_i \sigma_i \right)^2} \right] = 0.78$$



EBOD 2009: Analysis of MCQs

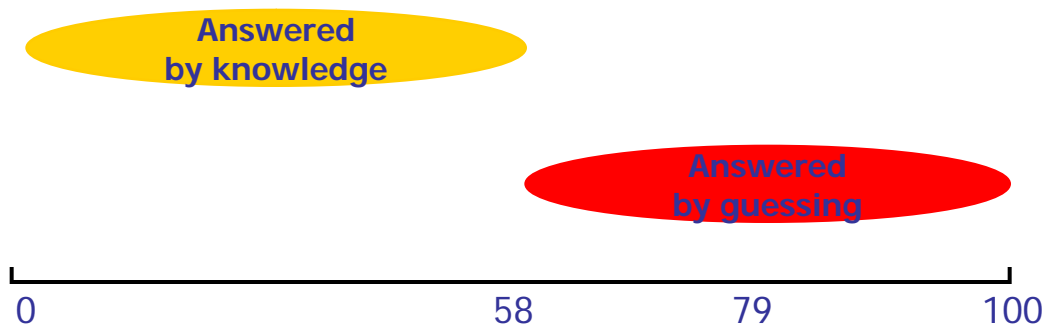
- Point biserial correlation coefficient (R_{it}) = 0.14
 - Estimator of the correlation between the **individual item scores** X_i (either 0 or 1) and the **total MCQ scores** Y_i (ranging from 154 to 230) of the candidates

$$R_{it} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s_X} \right) \left(\frac{Y_i - \bar{Y}}{s_Y} \right)$$



EBOD 2009: Analysis of MCQs

- Assessment of the degree of difficulty
 - Average P-value ≈ 0.79
 - Indication of items answered incorrectly by guessing ≈ 0.21
 - Estimation of items answered correctly by guessing ≈ 0.21
 - Estimation of percentage of candidates guessing ≈ 0.42
OR Estimation of percentage of candidates knowing ≈ 0.58





EBOD 2009: Secondary analysis

- Development of EBO analysis tool
 - Validation of MCQs
 - Assessment of degree of difficulty
 - Assessment of discriminative power
 - ...
 - Potential of personalisation of total test score composition by
 - rewarding correct responses of more difficult items
 - Reduction of items marked as being statistically sick items (mainly caused by guessing of candidates)



EBOD 2009: Secondary analysis

- Development of EBO analysis tool
 - Classical analysis techniques (Wilcoxon Rank-Sum)
 - Total test scores of candidates answering a specific item correctly
 - Total test scores of candidates answering a specific item incorrectly
 - Item-response analysis techniques
 - 1-parameter analysis technique (Rasch): item difficulty
 - (2-parameter analysis technique): idem + item discrimination
 - 3-parameter analysis technique: idem + guess factor



EBOD 2009: Secondary analysis

- Conclusions (classic analysis technique)
 - Reduction of sick items (199 \Rightarrow 103)
 - Thumb rules based on P-value and Rit-value
 - 75 items with P-value > 90 or < 10
 - 124 items with $10 \leq \text{P-value} \leq 90$ and a Rit-value < 20
 - Wilcoxon rank-sum test
 - 103 items remain statistically sick
 - Estimation of 42 percent of candidates answering on average by guessing, of which half (21 percent) answered correctly by guessing



EBOD 2009: Secondary analysis

- Conclusions Item-response analysis techniques vs Classical testing
 - 1-parameter analysis technique (Rasch)
 - correlation coefficient with P-value = 0.90
 - 3-parameter analysis technique
 - correlation coefficient with classical testing = 0.13
 - estimation of 49.9 percent of candidates answering correctly by guessing



EBOD 2009: MCQ analysis

- General conclusion on negative marking
 - Several **indications** to apply **negative marking**, which have been demonstrated by our analysis!
 - Introduction of negative marking
 - Increase of discriminative power of examination
 - Reduction of guess factor
 - wild guesses will be punished
 - guesses by reasoning (partial knowledge) will be rewarded



Spread of total test scores with negative marking

Spread of total test scores
without negative marking

0

260



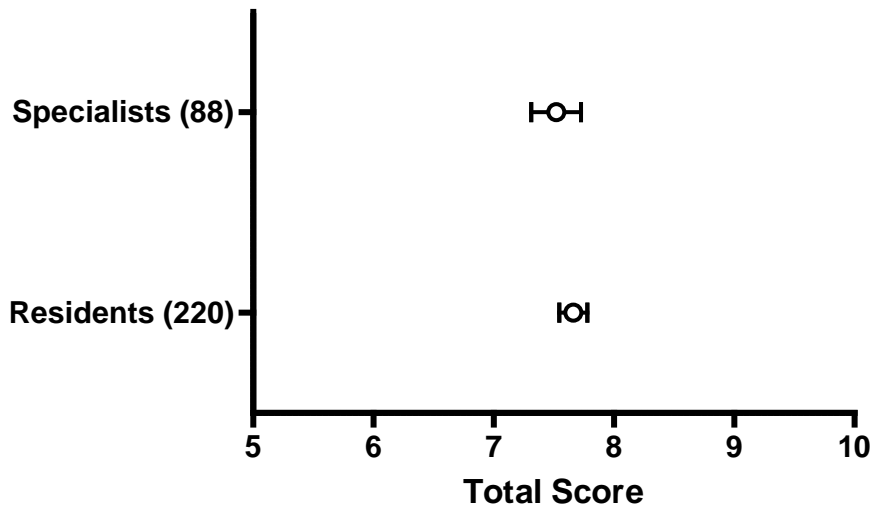
EBOD 2009: MCQ analysis

■ Conclusions

- Classical analysis techniques and item-response analysis techniques are complementary and correlated to a certain extent
- Correlation between these techniques will be better once negative marking will be introduced
- Statistical analysis tools
 - items to be verified on clinical content
 - time-consuming but worthwhile activity
 - better and validated MCQs

EBOD 2009: Viva Voce analysis

**EBOD 2009 Viva Voce Scores
with 95 % Confidence Intervals**



Topic	Score
A. Optics, Refractions, Strabismus and Neuro-ophthalmology	7.62 ± 1.32
B. Cornea, External diseases and Ocular adnexa	7.59 ± 1.29
C. Glaucoma, Cataract and Refractive surgery	7.45 ± 1.24
D. Posterior segment, Ocular inflammation and Uveitis	7.83 ± 1.28

Residents have **higher Viva Voce scores** with **lower standard deviations** when compared to specialists.

No statistically significant differences between the **topics**.



EBOD 2009: Viva Voce analysis

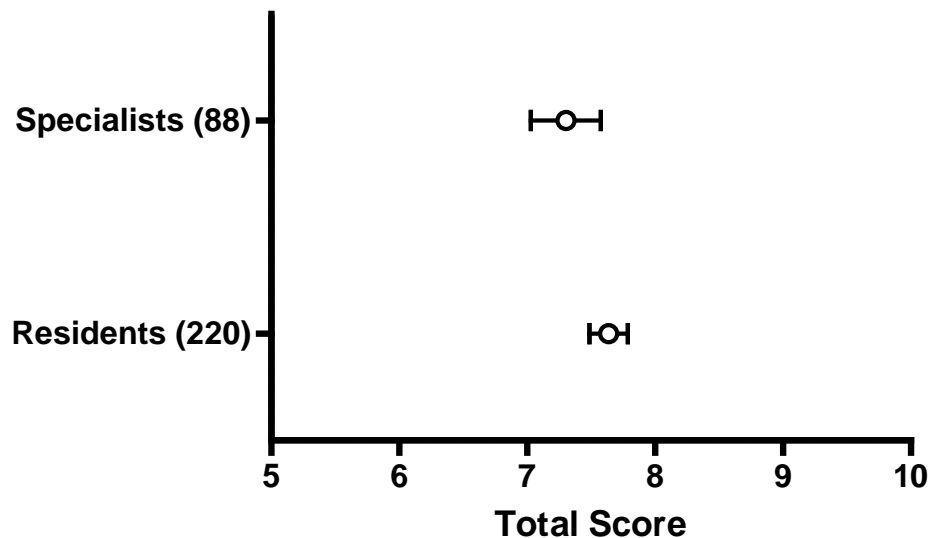
- Conclusions

- Differences between juries
 - standardisation of Viva Voce in the future
- Residents have higher Viva Voce scores with lower standard deviations as compared to specialists
 - same conclusion as for MCQs
- No statistically significant differences observed between the different topics.



EBOD 2009: General analysis

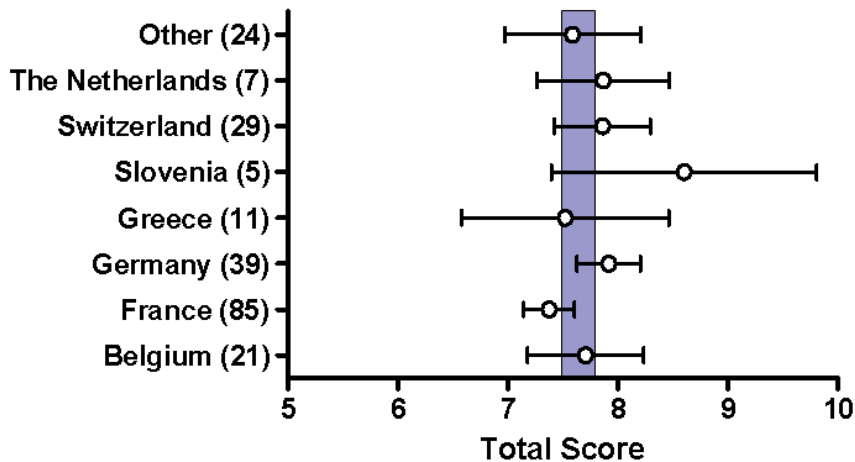
**EBOD 2009 Total Scores
with 95 % Confidence Intervals**



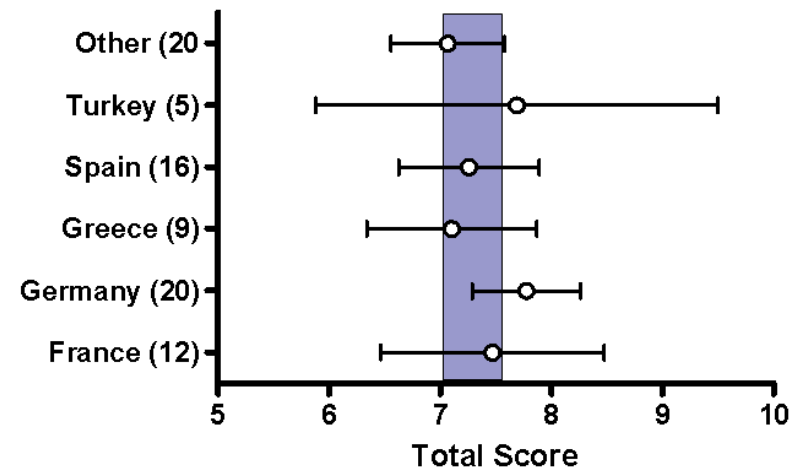
Residents have **higher Viva Voce scores** with **lower standard deviations** when compared to specialists.

EBOD 2009: General analysis

EBOD 2009 Total Scores Residents
with 95 % Confidence Intervals



EBOD 2009 Total Scores Specialists
with 95 % Confidence Intervals



Residents have **higher total scores** with **lower standard deviations** when compared to specialists.

No significant differences are observed between the **countries**.



EBOD 2009: Success rate

	2005	2006	2007	2008	2009
Success Rate	87.6%	88.1%	89.2%	90.8%	89.6 %

EBOD **success rate** is quite **stable over the years** and quite high as the level of candidates usually tends to be good.

18 Residents (out of 220: 8.2%) and **14 specialists** (out of 88: 15.9 %) failed at EBOD 2009. As there were 308 candidates the general failure rate was 10.4 %.



EBOD statistical analysis

- Items remaining for the future
 - Validation of MCQ-bank
 - Standardisation of oral examination
 - ...

... To be continued !