BENEFITS OF NEGATIVE MARKING AT THE EUROPEAN BOARD OF OPHTHALMOLOGY DIPLOMA (EBOD) EXAMINATION, BOTH FOR ORGANISER & CANDIDATES

Danny G.P. Mathysen, Marie-José Tassignon, Catherine Creuzot-Garcher, Marko Hawlina, Peter J. Ringens, Wagih Aclimandos and Kristien Wouters

Correspondence: D.G.P. Mathysen, Antwerp University Hospital, Department of Ophthalmology Wilrijkstraat 10, B-2650 Edegem (Antwerp), Belgium, E-mail: danny.mathysen@uza.be
Disclosure of interest

• I, Danny G.P. Mathysen DO NOT have a financial interest/arrangement or affiliation with one or more organisations which could be perceived as a real or apparent conflict of interest in the context of the subject of this presentation.
INTRODUCTION AND RESEARCH QUESTIONS

Introduction:
The European Board of Ophthalmology Diploma (EBOD) examination consists of a written examination (presented in this poster), followed by an oral examination. The written part of EBOD contains 52 MCQs with 5 multiple true-false items each (260 answers to be given by the candidates), with a pre-defined distribution of 10 topics within ophthalmology (more details on EBO website: http://www.ebo-online.org).

Research questions:
1. Does the introduction of negative marking at EBOD (to avoid wild guesses with a probability as high as 50% to be correct) have a positive effect on the statistical performance parameters of all EBOD test items in total and test items individually and?

2. Does negative marking have a discriminative effect towards female candidates, (main argument against negative marking according to literature)?
### Study Population:

In 2009 a total of 308 candidates (185 males and 123 females) took part at EBOD **without negative marking**. In 2010 a total of 310 candidates (168 males and 142 females) took part at EBOD **with negative marking**.

### Data Analysis and Results (Statistical performance parameters):

<table>
<thead>
<tr>
<th>Statistical Performance Parameter</th>
<th>Parameter: Rule of thumb</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>General statistical performance of EBOD (i.e. all items)</td>
<td>Cronbach-α ≥ 0.80</td>
<td>0.78</td>
<td>0.87</td>
</tr>
<tr>
<td>• Cronbach-α value (internal consistency)</td>
<td></td>
<td>to be considered as the degree to which all test items are measuring the same (i.e. knowledge of candidates)</td>
<td></td>
</tr>
<tr>
<td>Statistical performance of individual EBOD items (average)</td>
<td>0.10 &lt; P-value &lt; 0.90</td>
<td>0.79</td>
<td>0.66</td>
</tr>
<tr>
<td>• P-value (percentage of correct answers)</td>
<td></td>
<td>to be considered as an estimation of the level of difficulty (or facility) of test items</td>
<td></td>
</tr>
<tr>
<td>• Rit-value (correlation of item score with EBOD score)</td>
<td>Rit-value ≥ 0.15</td>
<td>0.14</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>to be considered as the degree to which a test item has an added value to the total examination</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
DATA ANALYSIS AND RESULTS, DISCUSSION

Data analysis and Results (Male versus Female):

<table>
<thead>
<tr>
<th>2009 – Male versus Female candidates ($\chi^2$ test)</th>
<th>2010 – Male versus Female candidates ($\chi^2$ test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference between pass – fail? p = 0.909 (NS)</td>
<td>Difference between pass – fail? p = 0.286 (NS)</td>
</tr>
<tr>
<td>Difference between scores (1–10)? p = 0.430 (NS)</td>
<td>Difference between scores (1–10)? p = 0.264 (NS)</td>
</tr>
<tr>
<td>Difference between “Don’t know” p = 0.02 (S)</td>
<td></td>
</tr>
</tbody>
</table>

Discussion:

- The **rationale** behind negative marking relies upon the fact that with true-false test items the **probability** of a correct answer by guessing is as high as **50 %**, due to which the **level** of the non-able or **borderline candidates** is generally assumed to be **over-estimated**. Hence, the space available to **discriminate** able from borderline candidates is (too) **limited**.

- The main **argument against** negative marking as described in literature, is the **assumption** that negative marking would be **discriminative towards females**.
DISCUSSION AND CONCLUSION

Discussion:

- Cronbach-α value: has **improved** after introduction of negative marking
- P-value: was **no longer over-estimated** with negative marking
- Rit-value: has **improved** after introduction of negative marking
- Males vs. Females:
  - Female candidates are **less keen to guess** (significance) (female candidates choose more for “Don’t know”)
  - Female candidates have the **same chances to pass** EBOD as male candidates (no significant difference in scoring)

Conclusion:

The introduction of **negative marking** for EBOD has proven to be **very successful**, with **benefits** for both:

- the **organiser** of the examination (statistical performance parameters), and
- the **candidates** (better discrimination with borderline candidates).