

## Abandoning negative marking

The European Diploma of Anesthesiology and Intensive Care was created in 1984. It has matured and developed over the last 24 yr into a multi-lingual examination, which takes place today in more than 20 centres in Europe. It consists of two parts – Part 1 (the primary examination) and Part 2 (the final examination).

The primary examination takes place in the form of a written multiple-choice examination. The 'True/False/Don't Know' format is used, where every candidate can mark one of three boxes on the computer-readable mark sheet corresponding to one of these three decisions. If none of the three boxes are marked for an individual question, the choice of 'Don't Know' is assumed. The number of correct responses is counted and each is given a score of +1. Likewise, each incorrect response is given a score of -1. A response of 'Don't Know' receives a score of 0. Subtracting the number of incorrect responses from the number of correct responses gives the candidate's score.

Until 1969, in the UK, the Part I of the Membership Examination of the Royal College of Physicians was scored by the above method (+1, -1 and 0) described by the Study Group of The London College in 1967 and modified in 1969 [1]. This method was called Formula Scoring, defined by the formula

$$S = R - W/C - 1,$$

where  $S$  is the so-called 'corrected score for guessing',  $R$  is the number of correct responses,  $W$  is the number of incorrect responses and  $C$  is the number of choices per item. Formula scoring was introduced in an attempt to correct for guessing, improving validity and reliability of test scores.

This method was adopted by a number of other Royal Colleges and was also used for the European

Diploma of Anesthesiology and Intensive Care. There are a number of studies [2–4] that have indicated that examinees tended either to answer 'Don't Know' or to omit items which they had a better than random chance of answering correctly and this behaviour was related to certain candidate-personality characteristics [5].

Hammond and colleagues [6] studied candidates attending a pre-examination revision course, encouraging them to attempt all questions and asking them to assign a level of certainty to each answer according to three alternatives:

- Positive: candidate is sure of the answer
- Educated guess: candidate knows something about the subject but is not absolutely sure of the answer
- Wild guess: candidate is taking a random guess

In this study, for questions assigned the positive option, candidates were correct 89.2% of the time, while 75% of educated guesses and 65% of wild guesses were correct. The authors stated that in their view there may be no such entity as the wild guess, as it is likely that candidates will have some experience of the subject of a question, however limited their knowledge.

As demonstrated in this study and in previous ones [7,8], the 'Don't know' option may conceal a small but very real residue of knowledge. In fact, examinees are more likely to be right than wrong if they answer a question for which they have a certain degree of uncertainty and overall they perform better in examinations if they act on this informed guess rather than answering with the 'Don't Know' option [9].

Harden and colleagues [5] asked a group of 99 students, immediately at the end of an examination and before the examination papers were collected, to go back over the examination paper and answer in red the questions which they had left blank. Both sets of responses were then scored. Of the 85 candidates who agreed to undertake that exercise, 62 gained marks and 23 lost marks. If the sign test is applied, it is found that the probability of this result being due to chance under the guessing

Correspondence to: Zeev Goldik, Department of Anesthesiology and Intensive Care, The Lady Davis Carmel Medical Centre, 7 Mihal Street, Haifa, Israel. E-mail: goldik@netvision.net.il; Tel: +972 54 4545793; Fax: +972 48 250761

On behalf of the Board of Examiners of the European Diploma of Anaesthesia and Intensive Care – ESA.

Accepted for publication 24 January 2008 EJA 4975  
First published online 13 March 2008

hypothesis is 0.00002. The authors concluded, 'this is very strong evidence against the hypothesis that all the students are guessing their responses to the 'Don't Know' questions'.

Elimination of the 'Don't Know' answers caused no obvious differences in rank position of students and the top 10 and bottom 10 candidates remained at exactly the same rank position.

### Negative marking: what's wrong?

The negative marks for wrong answers were adopted in order to discourage candidates from guessing and the 'Don't Know' option was introduced accordingly. As stated above, many authors have demonstrated that there is a danger that this 'Don't know' answer hides partial knowledge and even good candidates may not have willingness to commit themselves to an answer. Bliss [10] found that negative marking (formula scoring) tends to penalize the more able students.

The decision to omit questions is influenced by personality characteristics [5]. According to McGuire [11], The Royal College of General Practitioners in the UK discontinued negative marking many years ago when they demonstrated that it discriminated against female candidates because they tended to be more cautious with regard to guessing.

Candidates have been shown to interpret the instructions in relation to guessing in different ways [12]. This showed that candidates' reluctance to commit themselves to an answer in case of doubt is influenced by characteristics of instructions received before the examination. For example, when examiners write the description 'the penalty for a wrong answer is severe' of their marking scheme, then there is a clear preference to omitting answers.

In the Glasgow Surgical Fellowship Primary Examination, despite all having received instructions before the examinations, 66% of candidates indicated that they avoided guessing, while 33% did not. While only 55% of the candidates from countries where English was the official language avoided guessing, 80% from other countries did so. The author concluded, 'it appears likely that the latter group, whether because of cultural characteristics or test naivety, adopted a less than optimum strategy in the examination and this may at least in part account for their poorer performance'. As a multi-cultural examination, the European Diploma may well belong to this group.

During last years' examinations, a variable high rate of questions omitted was observed in our examinations (average 25%). This high rate of omitted questions affected the discriminator questions

as well. These questions were defined as 'good' questions in the sense that they acted well in discriminating between the good, the average and the poor candidates. Discriminator questions are used in many examinations (actually a set of them) as a benchmark comparing the level of candidates through consecutive years. The fact that many candidates omitted to answer these questions may affect reliability and validity of the examination when it comes to deciding the pass marks in the future.

### What do we conclude?

Muijtjens and colleagues [13] summarized the dilemma. On the one hand, 'medical professionals should not be stimulated to react with guessing when they are faced with a lack of knowledge' yet on the other hand, 'in medical practice the doctor has to make a decision on only partially complete information'.

If this is, as we understand it, the choice between absolute knowledge and partial knowledge, then the 'Don't Know' answer may hide partial knowledge. Changing to a system where we abolish negative marking and the 'Don't Know' option is then worthwhile.

If the purpose of the examination is to place the students in rank order (as is the case for most academic examinations), there is little merit in negative marking as it does not affect the rank order.

The essence of the art of medicine is decision making. The availability of a 'Don't Know' option and strict penalties for doubt and partial knowledge simply encourage only absolute knowledge. This might be, in our opinion, a wrong goal.

Taking all of the above reasons into account, the Examination Committee of the European Society of Anaesthesiology has decided to abolish negative marking and the 'Don't Know' option for both the in-training assessments and also Part I examinations of the European Diploma of Anaesthesia and Intensive Care starting at next examinations to be held in October 2008.

### References

1. Harden RMcG, Lever R, Wilson GM. Two systems of marking objective examination questions. *Lancet* 1969; I: 40.
2. Votaw DF. The effect of do-not-guess directions upon the validity of true-false or multiple-choice tests. *J Educ Psychol* 1936; 27: 698-703.
3. Sheriffs AC, Boomer DS. Who is penalized by the penalty for guessing? *J Educ Psychol* 1954; 45: 81-90.
4. Slakter MJ. The effect of guessing on objective test scores. *J Educ Meas* 1968; 5: 217-222.
5. Harden RMcG, Brown RA, Biran LA, Dallas Ross WP, Wakeford RE*Med Educ* 1976; 10: 27-32.

6. Hammond EJ, McIndoe AK, Sansome AJ, Spargo PM. Multiple-choice examinations: adopting an evidence-based approach to exam technique. *Anaesthesia* 1998; 53: 1105–1108.
7. Ebel RL. *Measuring Educational Achievement*. Englewood Cliffs, NJ: Prentice Hall, 1965.
8. Sanderson PH. The 'don't know' option in MCQ examinations. *Br J Med Educ* 1973; 7: 25.
9. Fleming PR. The profitability of guessing in multiple choice question papers. *Med Educ* 1998; 22(6): 509–513.
10. Bliss LB. A test of Lord's assumption regarding examinee guessing behavior on multiple-choice tests using elementary school students. *J Educ Meas Summer* 1980; 17(2): 147–152.
11. McGuire B. Multiple choice examinations. *Anaesthesia* 1999; 54(7): 720.
12. Young S, Gillespie G. Experience with the multiple choice paper in the Primary Fellowship Examination in Glasgow. *Br J Med Educ* 1972; 6: 44.
13. Muijtjens AMM, van Mameren H, Hoogenboom RJI, Evers JLH, van der Vleuten CPM. The effect of a 'don't know' option on test scores: number-right and formula scoring compared. *Med Educ* 1999; 33: 267–275.