# How to remove the gender bias in multiple choice assessments in engineering education?

## Experimental validation and theoretical analysis using prospect theory

**T. De Laet**[1]
Professor, Head of Tutorial Services Engineering Science and Head of Leuven Engineering and Science Education Center (LESEC)
Tutorial Services, Faculty of Engineering Science, KU Leuven
Leuven, Belgium
Tinne.DeLaet@kuleuven.be

**J. Vanderoost**
Researcher
Tutorial Services, Faculty of Engineering Science and LESEC, KU Leuven
Leuven, Belgium
Jef.Vanderoost@kuleuven.be

**R. Callens**
Tutor
Tutorial Services, Faculty of Engineering Science and LESEC, KU Leuven
Leuven, Belgium
Riet.Callens@kuleuven.be

**J. Vandewalle**
Professor Emeritus
Faculty of Engineering Science and LESEC, KU Leuven
Leuven, Belgium
Joos.Vandewalle@kuleuven.be

Keywords: assessment, multiple choice, prospect theory, gender

## INTRODUCTION

In higher education, exams with multiple choice questions (MPQ) are very common since they allow testing large groups while providing fast feedback. Although multiple choice exams are considered to be an objective way of assessing knowledge and competences of students, their use raises some concerns. One important concern is that a student can obtain the correct answer by randomly guessing from the offered alternatives. A variety of marking methods are available that try to address this concern. Literature has however indicated that some marking methods disadvantage particular personality traits. The personality trait handled in this paper is risk-

[1] Corresponding Author
T. De Laet
Tinne.DeLaet@kuleuven.be

aversion. In this paper we investigate if elimination marking, an existing alternative for negative marking, reduces the impact of risk-aversion. To this end we use both an experimental validation and a theoretical analysis using prospect theory.

## 1 BACKGROUND

A variety of marking methods for multiple choice exams are available: negative marking, number right, elimination marking (also called elimination testin), etc. [1,2,3,4]. Methods such as negative marking, which is the standard method at KU Leuven, try to discourage students from gaining marks by guessing by introducing a penalty for a wrong answer. Other universities, e.g. UGent, prefer standard setting [4]. This method does not penalize a wrong answer, but corrects for guessing by increasing the threshold for passing.

Literature, both from the pedagogical and psychological field of "assessment" as from the economic research field of "decision making under uncertainty", has however indicated that these different marking methods can have a different influence on students depending on their personality traits. In particular it has been shown that negative marking disadvantages risk-averse students [3,4,5]. As female students are in general more risk-averse, marking methods can introduce an unwanted gender bias. Experience with previous negative marking at KU Leuven *(Table 1)* indicates that the female students leave, as predicted by literature, more questions blank. This has been picked up by students and media, who are strongly protesting against negative marking.

Bond et al. [8] showed that elimination marking, a method that rewards partial knowledge but still introduces a penalty for guessing, does not introduce a gender bias in life sciences. Moreover, they found that this method increases student performance and satisfaction and reduces anxiety.

Berbery-Meyer et al. [5,9] used prospect theory to analyse guessing in multiple choice tests. Their study is the first to show the usefulness of prospect theory in the non-financial context of multiple choice exams. They showed that the behaviour on multiple choice exams can be predicted by prospect theory.

In this paper, we take a step further by using prospect theory to analyse the predicted examination scores based on risk-aversion for two marking methods. Moreover, we use prospect theory to propose a new marking method that is neutral with respect to risk-aversion.

*Table 1.* Comparison of the number of blank answers for the last edition of the positioning test [6,7] between male and female students.

| | ***number of blank answers (out of 35 questions)*** | ***total*** | ***female*** | ***male*** |
|---|---|---|---|---|
| **September 2014** | average | 9,42 | 11,09 | 9,00 |
| | standard deviation | 4,89 | 4,86 | 4,82 |
| | **p-value T-test male vs female** | **3,57E-03** | **→ significant** | |

## 2 MARKING MULTIPLE CHOICE EXAMS

This paper compares two specific marking methods: negative marking, which is the standard method at KU Leuven and elimination marking. Throughout the paper it is assumed that every multiple choice question has N alternatives, only one alternative is correct, the remaining N-1 alternatives are wrong.

### 2.1 Negative marking

For a traditional multiple choice question, the student has to indicate one of the N alternatives he/she believes to be correct. Often, the student has the additional

possibility to leave the question blank, and even indicate this as such. *Table 2* shows an example of such a question with four alternatives (N=4).

*Table 2.* Multiple choice question for negative marking with 4 alternatives.

| **What is the temperature in the room (up to 1°C precise)?** | | | | |
|---|---|---|---|---|
| 10-14°C | 15-19°C | 20-24°C | 25-30°C | blank |
| ○ | ○ | ○ | ○ | ○ |

*Table 3*. Different types of knowledge and corresponding score for negative marking, including an example where A=correct answer. (1=chosen alternative)

| **[A* B C D]** | ***type of knowledge*** | ***score*** |
|---|---|---|
| [1 0 0 0] | perfect knowledge | 1 |
| [0 0 0 0] | no knowledge | 0 |
| [1 1 1 0] | misconception | $-1/3$ |

The scoring scheme of negative marking, shown in *Table 3*, rewards 1 point if the chosen alternative is the correct one, $\frac{-1}{N-1}$ ($\frac{-1}{3}$ for N=4) if the chosen alternative is not the correct one, and 0 if the blank option is chosen. The punishment of $\frac{-1}{N-1}$ is introduced to discourage random guessing.

## 2.2 Elimination marking

For a multiple choice question with elimination marking, the student has to indicate which of the alternatives can be eliminated. *Table 4* shows an example of a question with elimination marking with four alternatives (N=4).

*Table 4.* Example of a multiple choice question for elimination marking, for each of the alternatives the student has to indicate if it is impossible (eliminated) or possible

| **What is the temperature in the room (up to 1°C precise)?** | | | | |
|---|---|---|---|---|
| | 10-14°C | 15-19°C | 20-24°C | 25-30°C |
| impossible | Ⓘ | Ⓘ | Ⓘ | Ⓘ |
| possible | Ⓟ | Ⓟ | Ⓟ | Ⓟ |

*Table 5*. Different types of knowledge and corresponding score for elimination marking, including an example where A=correct answer. (1=alternative eliminated, 0=alternative not eliminated)

| **[A* B C D]** | ***type of knowledge*** | ***score*** |
|---|---|---|
| [0 1 1 1] | perfect knowledge | 1 |
| [0 0 1 1] | partial knowledge type 2 | $2/3$ |
| [0 0 0 1] | partial knowledge type 1 | $1/3$ |
| [0 0 0 0] | no knowledge | 0 |
| [1 1 1 0] | partial misconception type 2 | $-1/3$ |
| [1 1 0 0] | partial misconception type 1 | $-2/3$ |
| [1 0 0 0] | total misconception | -1 |

The scoring scheme of elimination marking, shown in *Table 5*, evaluates the answer for each of the N alternatives and adds the scores. For each wrong alternative that the student eliminates correctly, he gains $\frac{+1}{N-1}$ ($\frac{+1}{3}$ for N=4). In case the student eliminates the correct alternative a punishment of -1 is given (misconception). Elimination marking allows the student to indicate his doubts (partial knowledge) by eliminating fewer than N-1 alternatives. If the student does not indicate doubt (and eliminates all but one alternative), the scoring is exactly the same as negative marking.

# 3 RESEARCH METHODOLOGY

The aim of the research is to investigate if elimination marking is suited for grading multiple choice exams in engineering education. More specifically the goal of the research is to answer the following questions: a) Does elimination marking

discourage random guessing, just as negative marking? b) How do students perceive the elimination marking regarding expected grade, imposed stress, etc.? c) Is elimination marking less disadvantageous for risk-averse students, especially for female students, than the widely used negative marking? The research methodology uses a double approach: an experimental analysis and a theoretical analysis using prospect theory.

## 4 EXPERIMENTAL VALIDATION

The experimental validation of elimination marking was performed using a first year Engineering Science bachelor course “Electric circuits”. In the middle of the semester 331 students got a **trial exam** with 10 questions with elimination marking. This trial was executed during a lecture and was not announced beforehand. First, students got a brief explanation on the method. Second, students got one hour to solve the test. Third, students completed a questionnaire to ask them about their opinion on elimination marking. Questions concerned the expectation on their result with elimination marking compared to negative marking, the amount of stress induced by elimination marking and negative marking, and their preference for elimination marking or negative marking. It was announced to the students that the final examination would only change from negative marking to elimination marking if they showed their preference for the latter in the questionnaire. *Fig. 1* shows two questions of the questionnaire indicating that the students find elimination marking less stressful than negative marking and that they prefer elimination marking over negative marking. Based on the questionnaire it was decided that elimination marking would be used for the final examination.

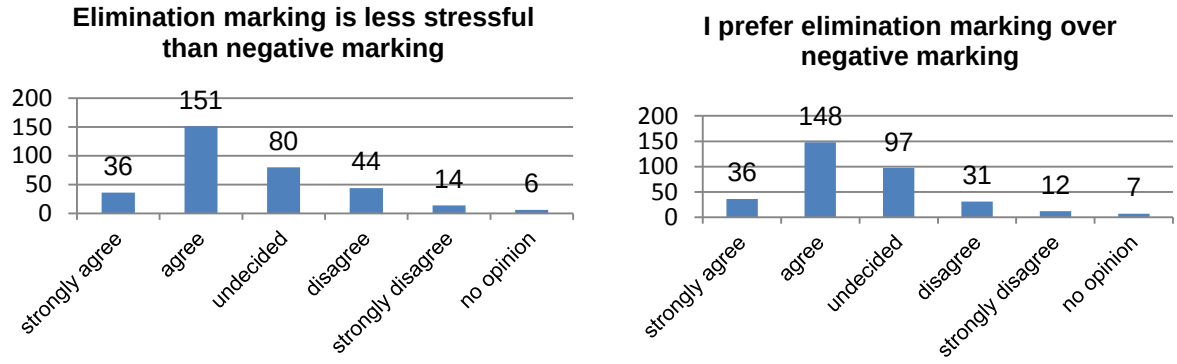



*Fig. 1*: Answer to two questions of the questionnaire taken after the elimination marking trial exam.

Next, the final (June) examination is studied in more detail. The exam consisted of 25 multiple choice questions with four alternatives. 425 students participated of which 68% passed and the average score was 12.1/20. With respect to previous years, when negative marking was used, the percentage of passed students and the average score increased significantly. This is caused by the reward that elimination marking gives to partial knowledge. *Table 6* shows that students use elimination marking to express their partial knowledge. *Table 7* shows that female students have on average more blank answers and more answers indicating doubt (not significant).

*Table 6*: June examination: % of all questions (25*425) with specific score and related type of knowledge

| analysis of June examination | | |
|---|---|---|
| score | % of questions | type of knowledge |
| 1,00 | 57% | perfect knowledge |
| 0,67 | 9% | partial knowledge |
| 0,33 | 4% | |
| 0,00 | 19% | no knowledge |
| -0,33 | 9% | partial |
| -0,67 | 2% | misconception |
| -1,00 | 0% | total misconception |

*Table 7:* Analysis of blank answers and answers indicating doubt

| | ***number of blank answers*** | | ***number of answers with doubt*** | |
|---|---|---|---|---|
| | **male** | **female** | **male** | **female** |
| **average** | 4,35 | 4,77 | 5,37 | 5,47 |
| **stdev** | 3,56 | 2,96 | 4,47 | 4,13 |
| **p-value T-test M vs F** | **0.15 not significant** | | **0.43 not significant** | |

## 5 THEORETICAL ANALYSIS USING PROSPECT THEORY

A theoretical analysis using prospect theory of "decision making under uncertainty" [5,9,10,11] was used to compare different scoring methods.

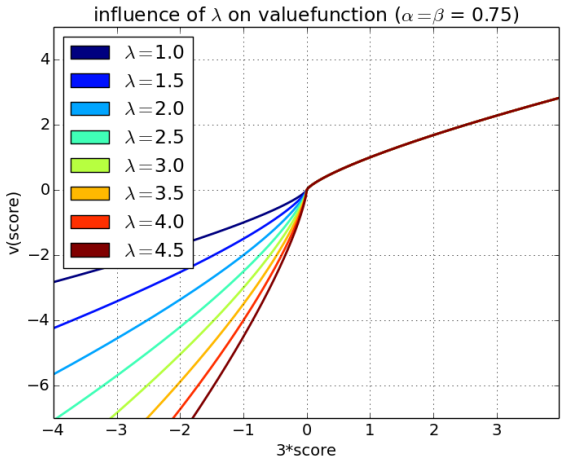



*Fig. 2.*Value function of prospect theory showing the value a person attaches to a certain score depending on his/her risk-aversion (λ).

### 5.1 Prospect theory

Prospect theory is a behavioural economic theory that describes the way people choose between probabilistic alternatives that involve risk, where the probabilities of outcomes are known. The theory states that people make decisions based on the potential value of losses and gains rather than the final outcome, and that people evaluate these losses and gains using certain heuristics. Depending on individual characteristics people attach a personal "value" $v(x_i)$ to an outcome $x_i$. In case of a multiple choice question the outcome $x_i$ is the scaled score on the question. The value function $v(x_i)$ is defined as:

$$v(x_i) = \begin{cases} x^{\alpha} \; when \; x \geq 0 \\ -\lambda x^{\beta} \; when \; x < 0 \end{cases}. \quad (1)$$

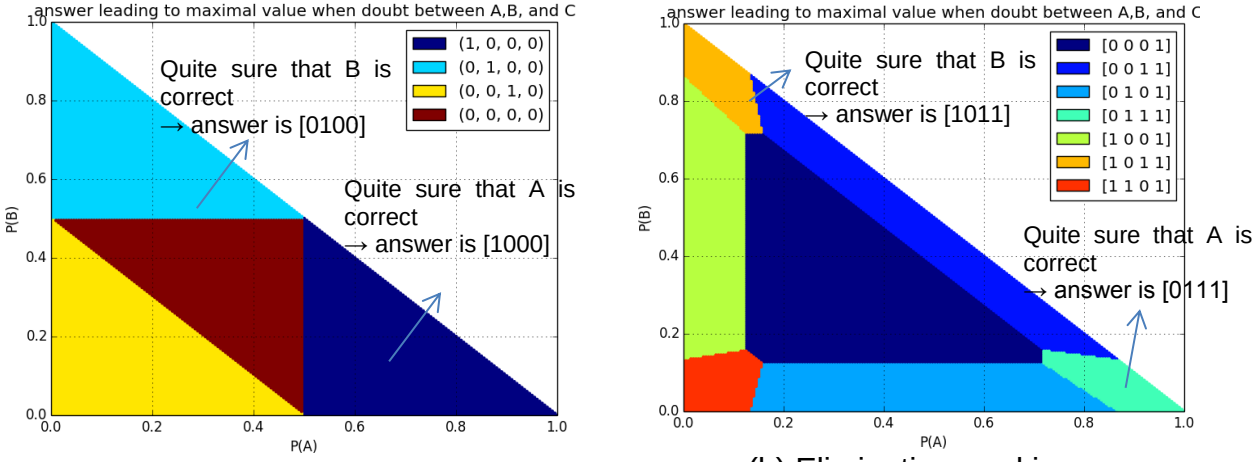
*Fig. 2* shows the s-shaped and asymmetrical value function: typically losses hurt more than gains feel good. The level of risk aversion can be modelled using the loss-aversion parameter λ.

When making a decision under uncertainty, people take into account the probabilities $p(x_i)$ they attach to the different outcomes $x_i$. In particular prospect theory assumes people take the decision that optimizes the expected utility:

$$U = \sum_{i=1}^{n} p(x_i)\, v(x_i). \quad (2)$$

## 5.2 Prospect theory and marking methods for multiple choice questions

Prospect theory can be used to model the decisions students make when confronted with multiple choice exams. As an example, *Fig. 3,* shows the optimal answer of a student according to prospect theory for both negative and elimination marking.




(a) Negative marking. (b) Elimination marking.

*Fig. 3.* Optimal answer according to prospect theory in case a student can exclude the fourth alternative p(D)=0 and is still in doubt between the three other alternatives, depending on the probabilities he attaches to these three alternatives p(A), p(B), and p(C)=1-p(A)-p(B)-p(D). The decision boundaries will depend on the risk-aversion λ (here λ=2.25). For example: when p(A)=p(B)=p(C)=1/3, the student should answer [0000], i.e. choose for a blank answer when using negative marking and [0001], i.e. only eliminate the fourth alternative when using elimination marking.

In order to analyse the influence of the type of correction (negative marking versus elimination marking), we perform a Monte Carlo simulation (M=500). In this Monte Carlo simulation random students are generated from the levels of knowledge observed at the June examination of Electric circuits (*Table 6*). Subsequently, for different levels of risk-aversion λ, the optimal decisions are calculated using prospect theory for 25 exam questions. Finally, the total exam score of this student is calculated from the optimal decisions. To compare the impact of the risk-aversion on the global score for the different multiple choice methods, the average score for the M=500 samples is analysed as a function of λ (*Fig. 4*).
From this analysis three observations are made. Firstly, negative marking is disadvantageous for risk-averse students as the average total exam score decreases as λ increases. Secondly, elimination marking is disadvantageous for risk-seeking students as the average total exam score decreases as λ decreases. Thirdly, the expected average exam score for elimination marking is higher than the one with negative marking.

## 5.3 Elimination marking with new scoring scheme

Based on the analysis using prospect theory, we propose a new scoring scheme for elimination marking that is neutral for the risk-aversion of a student and has no higher expected score with respect to negative marking.
The proposed scoring scheme is as follows (for N alternatives). If the student eliminates the correct alternative he always receives a punishment of $\frac{-1}{N-1}$ independently of the answers to the other alternatives; if the student eliminates no

answers he receives 0 points; for correctly eliminating N-x wrong alternatives: the student receives a reward of $+\frac{1}{(N-1)^{x-1}}$. *Table 8* shows the scoring method for N=4.

*Fig. 4* shows that according to the Monte Carlo simulation, the proposed scoring rule realizes the objectives: it is neutral with respect to risk-aversion and has almost the same average score (for non risk-averse students) with respect to negative marking.

*Table 8*. Different types of knowledge and corresponding score for the new scoring scheme, including an example where A=correct answer. (1=alternative eliminated, 0=alternative not eliminated)

| **[A* B C D]** | ***type of knowledge*** | ***score*** |
|---|---|---|
| [0 1 1 1] | perfect knowledge | 1 |
| [0 0 1 1] | partial knowledge type 2 | 1/3 |
| [0 0 0 1] | partial knowledge type 1 | 1/9 |
| [0 0 0 0] | no knowledge | 0 |
| [1 1 1 0] | partial misconception type 2 | −1/3 |
| [1 1 0 0] | partial misconception type 1 | −1/3 |
| [1 0 0 0] | total misconception | −1/3 |

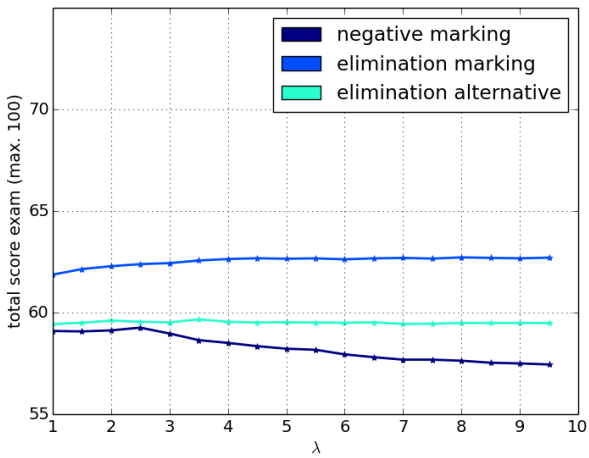



*Fig. 4.* Average total exam score for a Monte Carlo simulation of 500 exams based on the knowledge levels of *Table 6* in function of risk-aversion (λ).

## 6 SUMMARY

Elimination marking was found to be suited for practical use in multiple choice exams in a first year Engineering Science Bachelor course. After a short explanation students are able to apply the method. Even more, after this short explanation, students use the special features of elimination marking to express their partial knowledge. Students find elimination marking less stressful than negative marking, and they prefer elimination marking over negative marking. When applying elimination marking to a real examination, an increase in average score and percentage of passed students is observed. The theoretical analysis using prospect theory, a model for decision making under uncertainty, confirms the higher expected score for elimination marking for the same knowledge level. Additionally, the theoretical analysis shows that elimination marking, in contrast to negative marking, is not disadvantageous risk-averse students. Finally, the theoretical analysis was used to propose a new scoring rule that is neutral with respect to risk-aversion and has no higher expected score for the same knowledge level with respect to negative marking.

As future work we will perform an experimental validation for the newly proposed scoring method. Furthermore, we will extend the theoretical analysis and combine models from pedagogy to study the impact of the multiple choice scoring method for students with different risk-aversion ánd knowledge level.

## 7 ACKNOWLEDGMENTS

We gratefully acknowledge the support of the KU Leuven project OWP IJkingPRO 2013/17 and LESEC- the Leuven Engineering and Science Education Centre. We also appreciate the collaboration with the students and teacher Wim Dehaene of Electric circuits.

## REFERENCES


[1] Ben-Simon, A., Budescu, D. V, Nevo B. (1997), A comparative study of measures of partial knowledge in multiple-choice tests; *Applied Psychological Measurement*; 1997; Vol. 21; pp. 21-65.

[2] Kansup W., Hakstian A.R. (1975), A comparison of several methods of assessing partial knowledge in multiple choise tests: scoring procedures; *Journal of Educational Measurement*, Vol. 12, No. 5, pp. 219-230.

[3] Kansup W., Hakstian A.R. (1975), A comparison of several methods of assessing partial knowledge in multiple choise tests: testing procedures; *Journal of Educational Measurement*, Vol. 12, No. 5, pp. 231-239.

[4] Lesage, E., Valcke, M., Sabbe, E. (2013), Scoring methods for multiple choice assessment in higher education- is it still a matter of number right scoring or negative marking; *Studies in Educational Evaluation*; Vol. 39, pp. 188-193.

[5] Berbery-Meyer, Y., Meyer, J., and Flasher, O.M. (2002), Prospect theory analysis of guessing in multiple choice tests, *Journal of behavioral decision making*; Volume 15; pp. 313-327.

[6] Vandewalle, J. and Callens R. (2013), A positioning test mathematics in Flanders for potential academic engineering students, Proc. of the 41st SEFI Conference, Leuven, Belgium

[7] Vanderoost J, Vandewalle, J., Callens R., De Laet,T. (2014), Diversity and the academic engineering positioning test in Flanders: Impact on female students and students with disabilities, Proc. of the 42st SEFI Conference, Birmingham, England

[8] Bond, A. E., Bodger, O, Skibinski, D.O.F, Jones, D.H., Restall, C.J., D Dudly, E., van Keulen, G (2013) Negatively-marked MCQ assessments that reward partial knowledge to not introduce gender bias yet increase student performance and satisfaction and reduce anxiety, *Plos One,* Vol. 8; No. 2.

[9] Berbery-Meyer, Y., Meyer, J. Budescu, D. V. (2003), Decision making under internal uncertainty: the case of multiple-choice tests with different scoring rules; *Acta Psychologica*, Vol. 112, pp. 207-220.

[10] Barberis, N.C. (2013), Thirthy years of prospect theory in economics: a review and assessment; *Journal of Economic Perspectives*, Vol. 27, No. 1, pp. 173-196.

[11] Fennema, H., Wakker, P. (1997), Original and cumulative prospect theory: a discussion of emperical differences; *Journal of Behavioral Decision making*, Vol. 10, pp. 53-64.