

# Negative Marking and the Student Physician—A Descriptive Study of Nigerian Medical Schools

Ikenna Kingsley Ndu<sup>1</sup>, Uchenna Ekwochi<sup>1</sup>, Chidiebere DI Osuorah<sup>2</sup>, Isaac Nwabueze Asinobi<sup>1</sup>, Michael Osita Nwaneri<sup>3</sup>, Samuel Nkachukwu Uwaezuoke<sup>4</sup>, Ogechukwu Franscesca Amadi<sup>1</sup>, Ifeyinwa Bernadette Okeke<sup>1</sup>, Josephat Maduabuchi Chinawa<sup>4</sup> and Casmir James Ginikanwa Orjioke<sup>1</sup>

<sup>1</sup>Enugu State University Teaching Hospital, Parklane, Enugu, Nigeria. <sup>2</sup>Child Survival Unit, Medical Research Council UK, The Gambia Unit, Fajara, Gambia. <sup>3</sup>Omega Pediatrics, Roswell, Georgia, USA. <sup>4</sup>University of Nigeria Teaching Hospital, Enugu, Nigeria.

## ABSTRACT

**BACKGROUND:** There is considerable debate about the two most commonly used scoring methods, namely, the formula scoring (popularly referred to as negative marking method in our environment) and number right scoring methods. Although the negative marking scoring system attempts to discourage students from guessing in order to increase test reliability and validity, there is the view that it is an excessive and unfair penalty that also increases anxiety. Feedback from students is part of the education process; thus, this study assessed the perception of medical students about negative marking method for multiple choice question (MCQ) examination formats and also the effect of gender and risk-taking behavior on scores obtained with this assessment method.

**METHODS:** This was a prospective multicenter survey carried out among fifth year medical students in Enugu State University and the University of Nigeria. A structured questionnaire was administered to 175 medical students from the two schools, while a class test was administered to medical students from Enugu State University. Qualitative statistical methods including frequencies, percentages, and chi square were used to analyze categorical variables. Quantitative statistics using analysis of variance was used to analyze continuous variables.

**RESULTS:** Inquiry into assessment format revealed that most of the respondents preferred MCQs (65.9%). One hundred and thirty students (74.3%) had an unfavorable perception of negative marking. Thirty-nine students (22.3%) agreed that negative marking reduces the tendency to guess and increases the validity of MCQs examination format in testing knowledge content of a subject compared to 108 (61.3%) who disagreed with this assertion ( $\chi^2 = 23.0$ ,  $df = 1$ ,  $P = 0.000$ ). The median score of the students who were not graded with negative marking was significantly higher than the score of the students graded with negative marking ( $P = 0.001$ ). There was no statistically significant difference in the risk-taking behavior between male and female students in their MCQ answering patterns with negative marking method ( $P = 0.618$ ).

**CONCLUSIONS:** In the assessment of students, it is more desirable to adopt fair penalties for discouraging guessing rather than excessive penalties for incorrect answers, which could intimidate students in negative marking schemes. There is no consensus on the penalty for an incorrect answer. Thus, there is a need for continued research into an effective and objective assessment tool that will ensure that the students' final score in a test truly represents their level of knowledge.

**KEYWORDS:** multiple choice questions, negative marking, medical schools, South East Nigeria

**CITATION:** Ndu et al. Negative Marking and the Student Physician—A Descriptive Study of Nigerian Medical Schools. *Journal of Medical Education and Curricular Development* 2016;3:165–170 doi:10.4137/JMECD.S40705.

**TYPE:** Original Research

**RECEIVED:** August 5, 2016. **RESUBMITTED:** October 13, 2016. **ACCEPTED FOR PUBLICATION:** October 15, 2016.

**ACADEMIC EDITOR:** Steven R. Myers, Editor in Chief

**PEER REVIEW:** Four peer reviewers contributed to the peer review report. Reviewers' reports totaled 680 words, excluding any confidential comments to the academic editor.

**FUNDING:** Authors disclose no external funding sources.

**COMPETING INTERESTS:** Authors disclose no potential conflicts of interest.

**COPYRIGHT:** © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

**CORRESPONDENCE:** chidi.osuorah@yahoo.com

Paper subject to independent expert single-blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Published by Libertas Academica. Learn more about this journal.

## Background

Multiple choice questions (MCQs) are often used in medical education to assess knowledge of a subject or course content.<sup>1–3</sup> Apart from the ease of script grading, this form of assessment has the advantage of testing wide areas of knowledge encountered in medical education. Furthermore, MCQs that are well planned and delivered have desirable psychometric properties and are a viable and effective alternative to assess the critical-thinking skills of students.<sup>3–5</sup>

Despite its popularity, there are constant research efforts to exploit the strengths and mitigate the weaknesses of MCQs. One such weakness is the scoring system. There is considerable debate about the two most commonly used scoring systems. The first is the formula scoring method<sup>6</sup> popularly referred to as negative marking method in our

environment. The other is the number right (nonnegative marking) scoring method.<sup>7,8</sup>

In the number right scoring method, the correct answers are awarded a positive point, while incorrect and omitted answers are given no point.<sup>3</sup> The negative marking scoring system attempts to discourage students from guessing in order to increase test reliability and validity.<sup>9</sup> A correct response results in a positive score and omitted items result in no mark, while marks are lost for incorrect answers.<sup>10,11</sup> The number right scoring method was recommended by Frary<sup>12</sup> because he believed that penalizing students may induce psychological factors that could influence their decision to omit questions on which they have partial knowledge. Conversely, other authors suppose that with this scoring method, students can answer correctly through guessing, which introduces a random factor into test scores that



lowers reliability and validity.<sup>13,14</sup> Guessing the correct answer in the assessment can increase the total mark despite a candidate not possessing the required knowledge but good clinical practice does not include random guessing when unsure.<sup>15</sup>

The standard negative mark for an incorrect answer is one that will give an expected score of zero if the candidate ticks all answers in a test randomly, and for this to happen, the penalty for an incorrect answer should be  $1/(n - 1)$ , where  $n$  stands for the number of choices.<sup>16</sup> However, even among proponents of negative marking, there is confusion as to the amount of negative marks for an incorrect answer.<sup>16</sup> Ahmed and Michail<sup>17</sup> reported that a minus 1 mark penalty is a genuine reflection of inadequate performance rather than bad evaluation and recommended it over a minus 1/4 penalty for an incorrect answer. In the same vein, other authors are of the opinion that an appropriate penalty that would effectively discourage guessing should exceed the standard penalty of  $1/(n - 1)$ .<sup>7,8,18</sup>

Since the adopted methods for medical licensing examinations vary among countries, there is no consensus as to whether a penalty for wrong answers should be applied in medical examinations.<sup>8</sup> None of these methods can fulfill the requirements of a perfect scoring system for the assessment of knowledge, and available research hardly offers alternative methods.<sup>7</sup> Medical examinations in Nigeria and the West African subregion try to strike a balance by combining the two methods of MCQ marking.

Feedback from students is part of the education process, and there is a dearth of information concerning their perception of negative marking. Duffield and Spencer<sup>19</sup> commented that while students' views about the fairness of specific assessment tools may sometimes be at variance with published research on assessment, their perceptions are important because they influence the acceptability of an assessment instrument. They also recommended that students' views on assessment should be sought systematically as part of the process of course evaluation.

It has been suggested that different marking methods can influence students depending on their personality traits<sup>20</sup> and also that negative marking places risk-averse students at a disadvantage.<sup>7,21,22</sup> Gender preferences for particular types of assessment have produced some considerable debate, but there is little empirical research in the educational literature.<sup>23,24</sup> As female students are in general more risk averse, marking methods can introduce an unwanted gender bias.<sup>20</sup> Ng and Chan<sup>25</sup> reported that MCQ tests may also introduce gender bias in test performance dependent on subject test area, instruction/scoring condition, and question difficulty.<sup>25</sup> However, Bond et al<sup>26</sup> found no significant gender difference in performance between different MCQ formats under negative marking conditions. Further research into objective questioning, negative marking, and gender bias has been recommended in order to stimulate informed debate and reduce prejudicial attitudes.<sup>27</sup>

The research question asked in this study is whether in Enugu, South East Nigeria, medical students' test scores in MCQ examination format are affected by perception, gender,

and risk-taking behavior? Therefore, this study sought to assess the perception of medical students concerning negative marking scoring method and the effect of gender and risk-taking behavior on scores obtained from multiple choice assessment with the negative marking scoring method. It is hoped that the findings will stimulate more interest and informed debate concerning this assessment tool and enhance the student experience.

## Methodology

This was a prospective multicenter survey carried out among fifth year medical students in Enugu State University and the University of Nigeria. The study was carried out over a 2-month period from September to October 2015 at the two study centers. Ethical approval was obtained from the Research and Ethics Committee of Enugu State University before commencement of data collection. The ethical clearance obtained from Enugu State University was accepted by the University of Nigeria. Written informed consent was also obtained from the students before enrollment.

The study took place in two stages. In the first stage, a structured questionnaire was administered to 175 medical students. Thirty-seven out of a class of 42 students were enrolled from Enugu State University giving a response rate of 88.1%, while the remaining 138 were enrolled from a class of 194 in the University of Nigeria giving a response rate of 71.1%. Purposive sampling was used to recruit students who gave consent to participate in the study. Information on examination format preference by students and their experiences with negative marking were ascertained. The questions related to students opinion and perception of negative marking in MCQs examination were asked and response recorded using Likert 5-point scale, namely, strongly agree, agree, neutral, disagree, and strongly disagree. These responses were then recategorized into positive (strongly agree and agree), negative (strongly disagree and disagree), and not sure (neutral) appropriately depending on the way the question was structured.

In the second stage, a class test was administered to medical students from Enugu State University based on a series of pediatric lectures. Of the 37 students from Enugu State University who initially responded to the questionnaire interview, 28 were present on the day the test was administered. They were informed that number right scoring method would be used to evaluate their answers. Three weeks later the same test was administered to the same set of students without prior notice. For this test, the students were informed that negative marking method would be applied with a penalty of minus 1 mark for every wrong answer.

In order to minimize the effect of learning that would have occurred between the administration of the two tests, the following steps were taken: (1) The results were not discussed after the first test. (2) The questions were projected on a screen for the first test to eliminate later use. (3) The tests were not expected because they were not part of the curriculum, and there was no prior warning before any of the tests.



**Data analysis.** All the data obtained were recorded and analyzed using the Statistical Package for Social Sciences (SPSS) version 20.0. Qualitative statistical methods including frequencies, percentages, and chi square were used to analyze categorical variables. Quantitative statistics using analysis of variance was used to analyze continuous variables. Results were presented as percentages, and statistical significance was set at  $P$ -value  $<0.05$ .

## Results

One hundred and seventy-five medical students in their second clinical year (ie, fifth year in medical school) consented to participate and were enrolled in this study. Sixty-two percent of them were male and the remaining were female students with an overall mean age of  $25.55 \pm 3.61$  years.

Inquiry into assessment format revealed that 108/164 (65.9%) preferred multiple choice questions (MCQs), 30/164 (18.3%) preferred essay-type questions, while 26/164 (15.9%) had no particular preference (Table 1). Reasons given for

MCQs preference included the following: (a) it gives a broader knowledge of a course (23/108; 21.3%), (b) it helps understand a course better because it makes one read “in between lines” (21; 19.4%), (c) it covers a wider area of the course outline during assessment (58; 53.7%), (d) it gives the opportunity to make informed and/or random guesses (5; 4.6%), (e) it eliminates examination marker bias (2/108; 0.02%), and (f) it is more objective and everyone is treated equally during marking (12; 0.11%).

Eleven (36.7%) of 30 students who prefer the essay format assessment believed that it is a better way of assessing knowledge because it shows detailed knowledge of a subject content, eight of these students (26.7%) were of the opinion that MCQs can be sometimes confusing, and seven (23.3%) felt that because students memorize past MCQs, it is not a true test of knowledge.

Of the 175 students enrolled, 169 (96.6%) have taken MCQs examination where negative marking was applied (Table 1). Majority (138/169; 81.7%) of these students first experienced negative marking in their third year in medical school (second MBBS class). One hundred and thirty students (74.3%) had a negative perception of negative marking compared to 45 (25.7%) of the students who had a positive perception of negative marking. Out of 130 students with negative perception of negative marking, 80 (74.1%) were male and 50 (74.6%) were female, while among the 45 students with a positive perception, 28 (25.9%) were male and 17 (25.4%) were female. The proportions of male and female students with negative or positive perceptions of negative marking were not significantly different ( $P = 0.935$ ).

Thirty-nine students (22.3%) believed that negative marking reduces guess work and increases the validity of MCQs examination format in testing knowledge content of a subject compared to 108 (61.3%) who disagreed with this assertion ( $\chi^2 = 23.0$ ,  $df = 1$ ,  $P < 0.001$ ). It was also noted that well over half of students, 155 (88.6%) and 110 (62.8%), respectively, believed that negative marking application during MCQs increases anxiety and unfairly penalizes students. Majority of the students (110; 62.9%) agreed that negative marking system should be discarded in MCQs examination format, 42 (24.0%) believed it should be retained, while 21 (13.1%) were neutral in their opinion about negative marking in MCQs (Table 1).

When asked to list in order of appropriateness the type of MCQ without negative marking that is fairest in medical school examination, 98 (56.5%), 46 (26.3%), and 30 (17.1%), respectively, approved, disapproved, and were neutral about the “True or False” format. Similarly, 87 (49.9%), 44 (25.1%), and 44 (25.1%), respectively, approved, disapproved, and were neutral about the “Best option” MCQ format, and lastly, the “Matching type” format was approved by 53 (30.3%), while 60 (34.3%) and 62 (35.4%) disapproved and were neutral about it, respectively. Significantly more students preferred the True

**Table 1.** Description of study variables.

VARIABLES	N = 175
<b>Sex of respondents</b>	<b>N = 175</b>
Male	108 (61.7)
Female	67 (38.3)
<b>Exam format preference</b>	<b>N = 164</b>
Essay	108 (65.9)
Multi-choice questions (MCQs)	30 (18.2)
None	26 (15.9)
<b>MCQ-type preference<sup>†</sup></b>	<b>N = 238</b>
True or False	98 (38.7)
Best option	87 (36.6)
Matching type	53 (22.3)
<b>Perception of negative marking in MCQs</b>	<b>N = 175</b>
Fair	45 (25.7)
Unfair	130 (74.3)
<b>Experienced negative marking in exam(s)</b>	<b>N = 175</b>
Yes	169 (96.6)
No	6 (3.4)
<b>Approach of attempting MCQs with negative markings</b>	<b>N = 175</b>
Risk takers	15 (18.6)
Cautious	127 (72.6)
Conservatives	33 (18.9)
<b>Should negative marking be discarded in medical exams?</b>	<b>N = 173</b>
Yes	110 (62.9)
No	42 (24.0)
Not sure	21 (13.1)

**Note:** <sup>†</sup>More than one option permitted.



**Table 2.** Association between risk taking behavior and perception of negative marking.

RISK TAKING BEHAVIOUR	PERCEPTION		$\chi^2$	P-VALUE
	POSITIVE	NEGATIVE		
Risk taker (answer all questions both the ones I know and those am not sure about)	4 (26.7)	11 (73.3)	5.958	0.051
Cautious (answer those I know and make an educated guess for those am not sure about)	38 (29.9)	89 (70.1)		
Conservative (only answer questions am absolutely sure about their answers)	3 (9.1)	30 (90.9)		

and False MCQ format to the other two formats of MCQ surveyed in this study ( $\chi^2 = 10.2$ ,  $df = 2$ ,  $P = 0.006$ ).

The pediatrics test was marked with negative marking and number right methods. All 28 students passed the test (score above 50) with a median score of 75 without negative marking, while only six (0.03%) passed the examination when negative marking was applied (median score of 22). The median score of the students without negative marking was significantly higher compared to their median score with negative marking ( $P = 0.001$ ).

An assessment of risk-taking behavior with negative marking showed that 15 (8.6%) considered themselves “risk takers” (ie, attempts all questions irrespective of knowledge of the correct answers), the vast majority (127; 72.6%) considered themselves as “cautious” (ie, they will answer questions they are sure of and make an educated guess where possible for others), while 33 (18.9%) considered themselves as being “conservative” (ie, answer only questions they are absolutely sure of).

There was no significant differences in risk-taking behavior between male and female students in their MCQ answering patterns with negative marking method (risk takers [66.7% vs. 33.3%], cautious [63.0% vs. 37.0%], and conservative [54.5% vs. 45.5%],  $P = 0.618$ ).

The second pediatrics test that was scored with negative marking method (penalty of minus 1 mark for every wrong answer) was analyzed further. The mean comparison of test scores across the various risk-taking behavior groups revealed no significant difference (risk takers [ $8 \pm \text{nil}$ ], cautious [ $26.00 \pm 18.73$ ], and conservative [ $36.67 \pm 22.72$ ],  $F = 1.213$ ,  $P = 0.314$ ).

The risk-taking behavior of these students also had no significant association with perception or opinion of negative marking or their performance when negative marking method was employed (Tables 2 and 3).

### Discussion

This study assessed the impact and perception of negative marking among medical students in two medical schools in Enugu, South East Nigeria.

A significantly greater number of our students had a negative perception of negative marking and felt that it was an excessive and unfair penalty that increases anxiety. Goonewardene<sup>28</sup> also observed a consistent disapproval of negative marking among the same set of students he surveyed when they were in the second year and again when they were in the fourth year.

With negative marking applied, the majority of students in this study (72.6%) self-identified their guessing patterns as “cautious” (ie, they will answer questions they are sure of and make an educated guess if possible for others). On the other hand, 15 (8.6%) consider themselves “risk takers” (ie, they will attempt all questions). However, when negative marking was not applied, all students admitted being risk takers. This phenomenon in which students who should not have been selected get selected was described as “gate crashing” by Karandikar<sup>11</sup> and rewards partial knowledge of topics.<sup>26</sup>

This study revealed no significant association between risk-taking behavior and mean test scores or performance. This suggests that this personality trait places no undue advantage or disadvantage on any particular group and further strengthens the role of MCQs as a fair and balanced tool for assessing the critical-thinking skills of students.

Interestingly, most of the students preferred the True or False format to the Best option and the Matching type MCQ formats. This may be related to the ease of making a correct guess in the True or False compared to other MCQ types. In the True or False MCQ format, the probability of making a correct guess is always 50%, while in other MCQ formats, this probability gets much lower depending on the number of options one is to choose from.

**Table 3.** Association between risk taking behavior and performance in the second test.

RISK TAKING BEHAVIOUR	WITH PENALTY		$\chi^2$	P-VALUE
	PASS	FAIL		
Risk taker (answer all questions both the ones I know and those am not sure about)	0 (0.0)	1 (4.5)	3.818	0.148
Cautious (answer those I know and make an educated guess for those am not sure about)	3 (50.0)	18 (81.8)		
Conservative (only answer questions am absolutely sure about their answers)	3 (50.0)	3 (13.6)		



A significant number of the students (61.3%) did not agree with the assertion that negative marking reduces guess work. This interesting observation is probably a reflection of the “cautious” approach adopted by most of them. This study found no significant differences between male and females in their MCQ answering patterns. This is in keeping with other studies that show evidence of gender neutrality in the tendency to take risks.<sup>26,29</sup>

The degree of aversion for negative marking was so high that most students would like it to be removed from the MCQ examination format, and these views are consistent with reports from other studies.<sup>26,28</sup> However, it is more desirable to adopt fair penalties for discouraging guessing in negative marking schemes than excessive penalties for incorrect answers, which could intimidate students.<sup>30,31</sup> This view is highlighted in the pediatrics test administered to the students, where the median score without negative marking was significantly higher compared to the median score with negative marking. Objective assessment of students should therefore strike a balance between the interests of students who would like to attain success at the expense of accurate and detailed knowledge of subject content. Though negative marking is less preferred by students, it is the view of the authors that it should not be completely discarded due to its benefits. Rather, a modification with the lesser penalty of one quarter (1/4) as recommended by Hammond et al<sup>16</sup> may be a fairer alternative. In view of this, there is a need for continued research for a more effective and objective assessment tool that will ensure that the students’ final score in a test is a true representation of their knowledge level.

**Limitations.** The data strongly suggest a high influence of the grading scheme on student performance; however, a potential limitation is how representative the sample size was of the population. In addition, despite the precautions taken to minimize bias in the administration of the tests, we are aware that the students improving on the questions (by reading them up) may have introduced some bias. However, by doing the test first as number right scoring method and then negative marking order, we hoped to reduce but not completely eliminate this bias.

## Conclusions

In the assessment of students, it is more desirable to adopt fair penalties for discouraging guessing rather than excessive penalties for incorrect answers, which could intimidate students in negative marking schemes. There is no consensus on the penalty for an incorrect answer. The authors recommend a fair penalty based on the  $1/(n - 1)$  formula. This may ensure that the students’ final score in a test truly represents their level of knowledge.

## Acknowledgments

The authors wish to express their gratitude to all the medical students who participated in this study. We also wish to thank Ikenna Uche who assisted in the data entry and analysis.

## Author Contributions

The study was conceived by IKN. The study design, methodology, and questionnaire design were done by IKN, UE, DICO, SNU, and INA. Data collection was done by IKN and SNU. The data analysis and preparation of the results were done by IKN and DICO. There was equal contribution to the discussion by IKN, UE, DICO, SNU, INA, MON, OFA, IBO, JMC, and CJGO. All authors reviewed the final draft of the manuscript. IKN, UE, DICO, and INA are equal authors.

## REFERENCES

1. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian*. 2005;12:19–24.
2. Multiple-choice Question Formats. 2006. Available at: [http://www.uwo.ca/tsc/resources/selected\\_teaching\\_topics/assessment/multiple\\_choice\\_workshop\\_2006/mcq\\_format.html](http://www.uwo.ca/tsc/resources/selected_teaching_topics/assessment/multiple_choice_workshop_2006/mcq_format.html). Accessed July 26, 2015.
3. Veloshki JJ, Rabinowitz HK, Robeson MR. A solution to the cueing effects of multiple choice questions: the Un-Q format. *Med Educ*. 1993;27:371–375.
4. Bloom B. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain*. New York, NY: Longman; 1956.
5. Kim M, Patel RA, Uchizono JA, Beck L. Incorporation of bloom’s taxonomy into multiple-choice examination questions for a pharmacotherapeutics course. *Am J Pharm Educ*. 2012;76(6):114.
6. Lord FM. Formula scoring and number right scoring. *J Educ Measure*. 1975;12:7–12.
7. Lesage E, Valcke M, Sabbe E. Scoring methods for multiple choice assessment in higher education—is it still a matter of number right scoring or negative marking? *Stud Educ Eval*. 2013;39:188–193.
8. Espinosa MP, Gardeazabal J. Optimal correction for guessing in multiple-choice tests. *J Math Psychol*. 2010;4:415–425.
9. Kurz TB. A review of scoring algorithms for multiple-choice tests. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX; 1999.
10. Betts LR, Elder TJ, Hartley J, Trueman M. Does correction for guessing reduce students’ performance on multiple-choice examinations? Yes? No? Sometimes? *Assess Eval Higher Educ*. 2009;34(1):1–15.
11. Karandikar RL. On multiple choice tests and negative marking. *Curr Sci*. 2010;99(8):1042–1045.
12. Frary RB. Formula scoring of multiple-choice tests (correction for guessing). *Educ Measure*. 1988;7:33–38.
13. Choppin BH. Correction for guessing. In: Keeves JP, ed. *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford: Pergamon Press; 1988:384–386.
14. Bar-Hillel M, Budescu D, Attali Y. Scoring and keying multiple choice test: a case study in irrationality. *Mind Soc*. 2005;4:3–12.
15. Tweed MJ, Wilkinson T. A randomized controlled trial comparing instructions regarding unsafe response options in a MCQ examination. *Med Teach*. 2009;31:51–54.
16. Hammond EJ, McIndoe AK, Sansome AJ, Spargo PM. Multiple-choice examinations: adopting an evidence-based approach to exam technique. *Anaesthesia*. 1998;53:1105–1108.
17. Ahmed ME, Michail M. Evaluation of different MCQ-scoring-system in a medical school. *East Afr Med J*. 1993;70(12):787–788.
18. Budescu D, Bar-Hillel M. To guess or not to guess: a decision-theoretic view of formula scoring. *J Educ Meas*. 1993;30(4):277–291.
19. Duffield KE, Spencer JA. A survey of medical students’ views about the purposes and fairness of assessment. *Med Educ*. 2002;36:879–886.
20. De Laet T, Vanderoost J, Callens R, Vandewalle J. How to remove the gender bias in multiple choice assessments in engineering education? Experimental Validation and theoretical analysis using prospect theory. In: 43rd Annual SEFI Conference; June 29–July 2, 2015; Orléans, France.
21. Kansup W, Hakstian AR. A comparison of several methods of assessing partial knowledge in multiple choice tests: testing procedures. *J Educ Meas*. 1975;12(5):231–239.
22. Berbery-Meyer Y, Meyer J, Flasher OM. Prospect theory analysis of guessing in multiple choice tests. *J Behav Decis Mak*. 2002;15:313–327.
23. Pirie M. How exams are fixed in favour of girls. *The Spectator*. 2001. Available at: <http://archive.spectator.co.uk/article/20th-january-2001/12/how-exams-are-fixed-in-favour-of-girls>. Accessed July 26, 2015.



24. Kelly S, Dennick R. Evidence of gender bias in True-False-Abstain medical examinations. *BMC Med Educ.* 2009;9:32.
25. Ng AWY, Chan AHS. The testing methods and gender differences in multiple-choice assessment. In: Ao SL, Chan AHS, Katagiri H, Xu L, eds. *Transactions on Engineering Technologies.* New York: AIP Publishing LLC. 2009:236–243.
26. Bond AE, Bodger O, Skibinski DOF, et al. Negatively-marked MCQ assessments that reward partial knowledge do not introduce gender bias yet increase student performance and satisfaction and reduce anxiety. *PLoS One.* 2013;8(2): e55956.
27. Longstaffe JA, Bradfield JWB. A review of factors influencing the dissemination of the London Agreed Protocol for Teaching (LAPT)—a confidence based marking system; 2005. Available at: [www.ucl.ac.uk/lapt/CBM\\_review.doc](http://www.ucl.ac.uk/lapt/CBM_review.doc). Accessed July 26, 2015.
28. Goonewardene IP. A study of student perceptions towards the use of true/false type MCQs at undergraduate medical examinations with time course. *J Ceylon Coll Physicians.* 1999;32:35–41.
29. Maxfield S, Shapiro M, Gupta V, Hass S. Gender and risk: women, risk taking and risk aversion. *Gender Manage.* 2010;25:586–604.
30. Woodford K, Bancroft P. Multiple choice questions not considered harmful. ACM International Conference Proceeding Series; Vol. 106 Proceedings of the 7th Australasian conference on computing education-Volume 42. Newcastle, NSW, Australia; 2005:109–116.
31. Holt A. An analysis of negative marking in multiple-choice assessment. Available at: [www.citrenz.ac.nz/conferences/2006/papers/115.pdf](http://www.citrenz.ac.nz/conferences/2006/papers/115.pdf). Accessed September 4, 2015.