Answering multiple-choice questions in high-stakes medical examinations

MARTIN R FISCHER, SYBIL HERRMANN & VERONIKA KOPP

OBJECTIVES To examine whether changing initial answers during a multiple-choice question (MCQ) test in medicine brings about better overall test results, as has been shown in other academic fields.

METHODS A total of 36 answer books from the German Second National Medical Board Examination, with 580 MCQs (where 1 answer out of 5 must be selected), were used for analysis.

RESULTS We confirmed that high-stakes MCQ test scores in medicine did indeed improve when students changed their answers once. Further changes of answers did not improve the scores.

CONCLUSIONS In written, high-stakes medical examinations, we recommend that students be encouraged, after further reflection, to change their answers in MCQ tests for questions for which they had previously had doubts about the answers.

KEYWORDS humans; male; female; *educational measurement; *choice behaviour; students, medical/*psychology; education, medical, undergraduate/*methods.

Medical Education 2005; 39: 890–894

INTRODUCTION

Students continue to believe, irrationally, that the first answer that comes to mind in a multiple-choice question (MCQ) test format is the best. Consequently, despite coming up with an alternative answer that might, upon reflection, be better, they rarely change their initial answers. Candidates for the National Medical Board Examination (NBE) thus expect to gain no advantage from changing their answers. Facts, however, speak against this belief. In a meta-study of 33 studies carried out between 1928 and 1983, Benjamin et al. showed that the majority of answer changes are from incorrect to correct and that most students who change their answers improve their test scores. On average, between 2% and 9% of the test answers were changed and of these 57.8% of the changes went from incorrect to correct, whereas only 20.2% went from correct to incorrect and 22.8% of the changes went from incorrect to incorrect. More recent studies have corroborated these findings. A study by Friedman-Erickson reported that 56% of changed answers went from incorrect to correct, 24% of changes from correct to incorrect and 20% of changes from incorrect to incorrect. McMorris et al. and Wagner et al. repeatedly noted the positive relationship between the number of changed answers to MCQs and an increased test score.

The studies referenced above come from various academic domains, yet scarcely any equivalent studies have been carried out in medicine. Some evidence comes, indirectly, from nursing studies, which also show that students profit from changing answers on questions when they are unsure of their initial answers. This raises the question of whether a study of medical students might not show similar results. Would the majority of changed test answers go from incorrect to correct in our NBE, as shown in other academic spheres? With reference to a complete test score, what is the relationship between test answers that are changed only once to those that are changed more than once?
METHODS

Participants

A total of 36 students (21 men and 15 women) from the universities of Freiburg, Heidelberg, Ulm and Munich volunteered for this study. They took their second NBE in the autumn of 2003.

Instruments

Instruction

The test group was asked to describe their attitudes toward their answers on the second NBE by denoting in their original examination booklets every concern or doubt about their original answers to each examination question. If they regarded an answer as inadequate (out of 5 possible answers for each examination question), they were asked to draw a line through the initial answer and write their new – and hopefully better – answer next to the old one. When they were finished with this exercise, they were to send the reworked examination booklet to us, the authors of this study.

National Medical Board Examination

The 580 MCQs (developed by the IMPP, the national Institut für Medizinische und Pharmazeutische Prüfungsfragen [Institute for Medical and Pharmaceutical Examination Inspection]) were divided into the following groups: non-invasive medical specialties (190 questions); neurology and psycho-sociological specialties (100 questions); general surgical or invasive specialties (190 questions), and occupational and general medical specialties (100 questions). Each question contained 5 possible answers, only 1 of which was right. This second NBE carried a reliability score of 0.97 (Cronbach’s alpha).

Analysis

The data from the reworked 36 Second State Medical Examination booklets were extracted and analysed. Standard descriptive statistical analysis with SPSS Version 11.5 was applied. Cronbach’s alpha was calculated to check for inner consistency.

RESULTS

The average score for the study group as a whole for their second NBE was 460.7 points (SD = 39.1), which showed that they had performed better than the overall student population taking the NBE (n = 4009; mean = 435.0 points, SD = 48.7). Each of the examination booklets we received – after the group had examined their previous answers – contained at least 1 changed answer. The aggregate random sample was 20 880 (36 participants and 580 questions per examination). Altogether the answers to 819 MCQs had been changed. As some items had been changed twice or more often, we observed 846 answer change processes in total (4.3%, SD = 2.7). Out of the entire set of changed answers, 55% (SD = 15.8) of the changes went from incorrect to correct, 25% (SD = 14.4) went from correct to incorrect, and 20% (SD = 13.0) went from incorrect to incorrect. Altogether the students enjoyed, on average, a 6.5-point (SD = 5.9) increase on their previous test scores or a percent increase of 1.1% (SD = 1.0). The success rate was calculated as the total number of answers changed from incorrect to correct minus the number of answers changed from correct to incorrect (Table 1).
The first answer change per examination question led to a clear-cut improvement from incorrect to correct (mean $\bar{X} = 55\%$, SD = 15.8). The second answer change – as well as further changes – produced more incorrect than correct answers (mean $\bar{X} = 41\%$, SD = 39.0). However, the total examination results clearly showed the advantage of the first answer substitution per question (mean $\bar{X} = 6.5$, SD = 5.9), over the second and following answer changes (mean $\bar{X} = 0.2$, SD = 1.3).

As expected, the number of answer changes are highly correlated with the net gain of points from these changes ($r = 0.66$; $P < 0.01$) (Fig. 1). In contrast, the overall examination score is negatively correlated with the number of changed answers per student ($r = -0.27$; not significant) (Fig. 2), indicating that high-performing students correct fewer answers.

**DISCUSSION**

Guessing is an immanent problem in all MCQ test formats. When a ‘don’t know’ option is provided and the correct minus the incorrect answers are used as test scores (formula scoring), it has been shown that medical students perform better when they are instructed to answer all questions.$^{13,14}$ Consequently, correct answers may be the result of partial knowledge or lucky guessing. Despite these concerns, most high-stakes MCQ tests have abandoned the ‘don’t know’ option and have moved to the 1 best response.
format with a number-right scoring. This predominant format was also used in our study, with 1 out of 5 possible answers being correct, giving examinees a 1 in 5 chance of guessing correctly. As our students improved their overall test scores by changing their initial answers once, we assume that this was related to careful reconsideration. With second and further answer changes, guessing may have been the predominant factor as the overall test scores decreased.

As our study group outperformed the overall student population taking the NBE, our sampling is not representative. However, as a high total examination result is negatively correlated with the number of answer changes per student, the observed effects of answer changes are rather underestimated. It is also unclear how many answer changes were not noted correctly in the examination booklets. This too could result in an underestimate of the percentage of answers changed. On the other hand, there is another potential sampling bias: All the students in our study who turned in their examination booklet to us changed at least 1 answer to an examination question. There is the possibility that the students who made no changes to their examination questions did not return their booklets to us. If this is so, then our change rate is overestimated.

Overall, our results are in accordance with those of other studies. In our case 4.3% of answers were changed. Benjamin et al. reported a range of 2–9% of changes in various domains. The most common answer change, in our study, was from incorrect to correct (55%). This finding also agrees with those of similar studies in other academic fields, which additionally found that in second and further answer changes, the majority of changes were from correct to incorrect. An improvement in test scores from the initial answer change is, then, to be expected. Additional answer changes (2 or more changes) do not improve overall examination scores, a result confirmed by our study as well.

To what extent these findings hold true for less experienced learners should be established in a further study. We should also consider asking examinees to indicate their confidence level for each of their responses, both the initial and the changed response, in order to learn more about the role of guessing.

**Practical implications**

The general consonance between the results of our study and those in other fields makes it clear that medical students need to be educated that the received wisdom, which claims that the initial answer to an MCQ is the best answer, is inappropriate.

Therefore, students need to be counselled to follow their doubts and rethink their answers. The chance that a substitute answer will be correct and will increase an overall examination score is statistically worth taking, providing the examinee does not change horses more than once.

---

**Contributors:** MRF was responsible for the conceptual design of this study and its accomplishment, was co-writer of the paper and conducted the literature search. SH and VK contributed to data analysis, the literature search and the writing of the paper.

**Acknowledgements:** We are indebted to D Neumann and R Krebs for methodological advice and helpful criticism.

**Funding:** The project was funded by the Bundesministerium für Bildung und Forschung (BMBF) as part of the CASEPORT-Project (FRZ: 01 NM 111 A).

**Conflicts of interest:** None.

**Ethical approval:** Authors state that ethical approval was not required.

---

**REFERENCES**

4. Morse DT, Morse LW. Change that answer! *Teach Prof* 1997;4:5.


Received 12 August 2004; editorial comments to authors 1 October 2004; accepted for publication 10 November 2004
Copyright of Medical Education is the property of Blackwell Publishing Limited. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.