RUNNING HEAD: Multiple-choice examinations in Psychology

Does correction for guessing reduce students' performance on multiple-choice examinations? Yes? No? Sometimes?

Lucy R. Betts*, Tracey J. Elder, James Hartley & Mark Trueman

Keele University, United Kingdom

Abstract

Multiple-choice (MC) examinations are becoming increasingly popular in higher education because they can be used effectively to assess breadth of knowledge in large cohorts of students. This present research investigated Psychology students' performance on, and experiences of, MC examinations with and without correction for guessing. In Study 1, data were collected from two cohorts of students across three Psychology MC examinations. The results revealed that students scored higher, and left fewer questions unanswered, when there was no correction for guessing. Furthermore, when the correction for guessing was removed from the theory MC examination, students who were told there was no correction for guessing did better than those told there was a correction. In addition, there was limited evidence of gender differences, with female students performing significantly better on one MC examination than males. In Study 2, a further set of first-year Psychology students reported their experiences of correction for guessing on open-book and closed-book MC examinations. Students reported feeling less anxious and more confident on the open-book MC examination. The findings of both of these studies have implications for instructors deciding whether or not correction for guessing is appropriate, and for the advice to be given to students preparing for MC examinations.

Keywords: Multiple-choice examinations, gender differences, correction for guessing, Psychology, open-book examinations, closed-book examinations, negative marking

Does correction for guessing reduce students' performance on multiple-choice examinations? Yes? No? Sometimes?

The use of multiple-choice (MC) examinations is becoming widespread within education and it is a practice that is viewed favourably by both instructors and students (Simkin & Kuechler, 2005). Instructors recognise the ability of MC examinations to: (1) examine efficiently large cohorts of students, (2) assess a broad array of topics in a single examination, and (3) give relatively quick feedback compared to other more traditional modes of assessment (Bush, 2001; Nicol, 2007; Simkin & Kuechler, 2005; Williams & Clark, 2004). Furthermore, through randomly selecting the order of question presentation for individual students, MC examinations can serve to reduce the growing concern about student cheating in examinations (see Simkin & Kuechler, 2005). Students regard MC examinations favourably because: (1) they are perceived as objective, (2) marks can be improved through making correct guesses, and (3) it is possible for students to gain marks on MC tests even if they do not reach the end of the paper in the allotted time (Simkin & Kuechler, 2005).

Despite the apparent benefits of MC examinations for both instructors and students, Burton (2005) urges that these examinations be used cautiously. MC examinations, it is argued, typically promote 'surface' rather than 'deep' learning (Nicol, 2007; Scouller, 1998; Williams & Clark, 2004) and fail to assess students' critical and communication skills (Paxton, 2000). MC examinations also fail to test students' ability to develop an argument – a skill actively encouraged in higher education (Paxton, 2005; Simkin & Kuechler, 2005). Additionally, there are concerns that students tend to perform better in MC examinations than in other 'traditional' types of assessment such as essay-based examinations (Downs, 2006; Simkin & Kuechler, 2005). However, this is not always the case (see Hartley, Betts & Murray, 2007).

One of the most widely voiced concerns about MC examinations is that students can answer correctly through guessing and thereby receive relatively higher marks, even when they know little of the area (Budescu & Hillel, 1993; Diamond & Evans, 1973; Frary, 1988). Paxton (2000) argues that if a student does select a correct answer the extent of their knowledge and understanding of the topic remains unclear – and this is even more problematic if the guess happens to be right. Indeed, some researchers argue that, when there is no correction for guessing, the best strategy is for students' to guess if they are unsure of the answer (Burton & Miller, 1999). Therefore, students' performance on MC examinations may not be an accurate reflection of their ability because students can achieve artificially inflated scores through guessing (Bush, 2001). Guessing also has implications for the reliability of MC examinations through the introduction of a random element (Burton, 2001, 2005). Further, test reliability is also influenced by length and the quality of items (Burton, 2004). According to Burton (2002, 2004, 2005), establishing the reliability of a MC examination is crucial to ensure that the test does what it claims to – that is to assign marks that accurately represent the students' performance.

To overcome these concerns, many MC examinations are now scored with a correction for guessing (where a percentage of a mark is deducted for each incorrect answer) or with negative marking (where whole marks are deducted for incorrect answers). Crucially, for both of these marking criteria, when a student chooses not to guess and a question is left unanswered, no proportion of a mark is deducted. The rationale behind these marking criteria is that students know that they will lose marks for incorrect answers and this reduces the likelihood of students guessing. Furthermore, correction for guessing or negative marking criteria also penalise students' misinformation whilst having little effect on the overall reliability of the examination

(Burton, 2004, 2005). Moreover, some authors argue that by implementing correction for guessing, MC examinations may be measuring students' answering strategies and risk taking behaviours rather than their subject specific knowledge (Budescu & Bar-Hillel, 1993; Hammond McIndie, Sansome & Spargo, 1998; Walker & Thompson, 2001). Also, from the students' perspective, another implication of such marking criteria is that partial knowledge is not rewarded in most MC examinations compared to essay-based examinations (Bush, 2001).

As the use of correction for guessing becomes more common, the effect of introducing such schemes on students' perceptions and performance should not be overlooked. Traditionally, in study guides, students have been told 'not to guess' when correction for guessing is implemented (e.g., Collins & Kneale, 2004; Freeman & Stone, 2006). However, Hammond et al. (1998) conclude that this advice does not facilitate students' performance and that students should be advised to guess if they have limited knowledge (e.g., if students know that two of four options may be correct) because such a strategy is rewarded through better performance. Further, Budescu and Bar-Hillel (1993) conclude, through theoretical modelling, that students should be *encouraged* to guess as this will serve to increase their scores. However, this effect has yet to be examined by measuring or manipulating students' performance on MC examinations.

Students' performance on MC examinations may also be influenced by a number of factors including gender, language ability, and examination format (Paxton, 2000; Simkin & Kuechler, 2005). Some studies report that males tend to outperform females on MC examinations (Anderson, 2002; Ben-Shakhar & Sinai, 1991; Bridgeman & Lewis, 1994), others report that females outperform males on MC examinations (Hartley et al., 2007), whereas others report no gender differences (Greene, 1997).

There are two possible explanations for these mixed findings. First, gender differences may emerge because, when there is correction for guessing, males typically leave fewer questions unanswered and this may serve to increase their scores (Anderson, 2002; Ben-Shakhar & Sinai, 1991; Von Schrader & Ansley, 2006). Second, it may be that these differences in performance result from the different topics assessed in MC examinations rather than from gender differences *per se*. In support of this argument, the MC examination reported by Hartley et al. (2007) assessed students' knowledge of neuropsychology, whereas the MC examination reported by Anderson (2002) assessed constructs aligned with mathematical ability.

MC examination performance may also vary as a function of the examination format. Specifically, some MC examinations are open-book, where students can take course materials into the examination, whereas other MC examinations are closed-book, where they cannot. Some argue that open-book examinations are useful to student learning because they reduce rote learning and foster higher-order thinking (Shine, Kiravu & Astley, 2004; Theophilides & Koutselini, 2000). Interestingly, most students perceive open-book and closed-book examinations to be of equal difficulty (Theophilides & Dionysiou, 1996). This may be because past experience has taught students that open-book examinations tend to foster an unrealistic sense of confidence and, consequently, they may prepare less for these examinations than for closed-book examinations, resulting in reduced performance (Ioannidou, 1997; Theophilides & Dionysiou, 1996; Theophilides & Koutselini, 2000).

The current research

This paper presents two studies on these issues. In Study 1, we examine the implications of correction for guessing[1], including the number of questions left unanswered and gender differences in students' performance, on MC examinations

designed to assess students' knowledge of Psychology theory and research methods. In Study 2, because previous research suggests differences in students' performance emerge according to whether examinations are open-book or closed-book, we examine students' perceptions of correction for guessing across open-book and closed-book MC examinations.

Study 1

Study one investigated the effects of different marking strategies on students' performance on MC examinations with two cohorts of first- and second- year Psychology undergraduates studying theory and research methods modules. We were able to do this for two reasons. First, the content of the modules assessed by the examinations was broadly consistent for the two cohorts and the questions for the two examinations were drawn from the same test bank. Second, with the first cohort of students there was no correction for guessing and the students were not told of any correction made for guessing. As such, this cohort could feel more relaxed about guessing as there was no penalty imposed. However, correction for guessing was introduced for these MC examinations the following academic year and, therefore, correction for guessing was used with the second cohort and the students were informed of this new procedure. This second cohort, therefore, had to consider the advantages and disadvantages of guessing depending on their confidence in their answers. To discourage guessing, a third of a mark was deducted from the students overall score for each incorrect answer.

Study 1 was designed to examine four main areas:

(1) how students performed on MC examinations before and after the correction for guessing policy was implemented (cohort one versus cohort two);

(2) the pattern of blank responses before and after the implementation of correction of guessing (cohort one versus cohort two);

(3) the performance of students when correction for guessing was *implemented* without the students' knowledge (cohort one);

(4) the performance of students when correction for guessing was *removed* without the students' knowledge (cohort two);

The purpose of the latter two points was to examine whether differences emerged when students were, and were not, aware of the correction for guessing procedure. Further, through examining the data, it was possible to examine whether or not the students' strategies for accommodating the correction for guessing were academically successful. Finally, because of the controversy surrounding gender differences on MC examination performance (Anderson, 2002; Ben-Shakhar & Sinai, 1991; Bridgeman & Lewis, 1994; Greene, 1997; Hartley et al., 2007), Study 1 also examined gender differences across the four areas outlined.

Method

Data were collected from two cohorts of undergraduate Psychology students for three MC examinations from three different modules. Table 1 shows the numbers of students who participated in each examination by cohort and gender. For each module, the MC examination represented one of a range of components used to assess students' overall performance. The modules were all part of a Dual Honours Psychology course at Keele University in the UK.

Insert Table 1 here

The first cohort of students were informed that there was no correction for guessing and the second cohort of students were informed that there was correction for guessing. The number of questions within the MC examinations ranged from 42 to 75. For both cohorts, the first-year research methods MC examination was open-book whereby students could refer to course textbooks and lecture notes during the examination. The other two MC examinations (first-year theory and second-year research methods) were closed-book. For all MC examinations there was a list of four possible answers for each question lettered from A to D. Those who were informed of the correction for guessing were told: "You should note that responses on this examination are weighted. You will be given one mark for each correct answer and will have a third of a mark deducted for each incorrect answer. Unanswered questions will receive a mark (weighting) of zero". When there was no correction for guessing no such instructions were given to the students.

Data coding

For each MC examination script, the percentage of correct answers with no correction for guessing, the percentage of correct answers with correction for guessing (i.e., with a third of a mark deducted for each incorrect answer) and the percentage of unanswered questions were calculated. Percentages were used because the number of questions varied across the MC examinations.

Results

The difference in students' scores on the MC examinations was investigated when there was no correction for guessing and when there was correction for guessing. In both cases, students were aware of this information and therefore it is safe to assume that the latter cohort may have guessed at their answers to a lesser extent.

Insert Table 2 here

Table 2 shows the means and standard deviations for performance according to MC examination and marking criteria implemented. For each MC examination, 2 (correction: correction versus no correction) x 2 (gender: male versus female) unrelated ANOVAs were used to examine the effect of correction for guessing and gender. The results showed that there was a significant main effect for correction for guessing for the first-year theory MC examination, $F(1, 520) = 167.37$, $p < .001$, $\eta^2 = .237$; the first-year research methods MC examination $F(1, 512) = 35.07$, $p < .001$, $\eta^2 = .064$; and the second-year research methods MC examination, $F(1, 415) = 46.92$, $p < .001$, $\eta^2 = .102$. In all cases students who were not subjected to the correction for guessing scored higher marks on the MC examinations than those students who were subjected to the correction for guessing.

The results for gender were mixed. For the first-year theory MC examination there was a main effect of gender, $F(1, 520) = 5.68$, $p = .018$, $\eta^2 = .011$, with females scoring higher than males ($M_{\text{Females}} = 47.09$, $SD = 15.28$; $M_{\text{Males}} = 43.87$, $SD = 14.23$). However, for the first-year research methods MC examination there was no significant difference in students' scores comparing males ($M = 59.69$, $SD = 14.92$) and females ($M = 62.04$, $SD = 14.80$), $F(1, 512) = 1.78$, $p = .18$, $\eta^2 = .003$. Similarly, for the second-year research methods MC examination, there was no significant difference in students' scores comparing males ($M = 54.96$, $SD = 14.70$) and females ($M = 52.99$, $SD = 15.28$), $F(1, 415) = .39$, $p = .53$, $\eta^2 = .001$.

Across all three MC examinations, it is evident that students scored higher when there was no correction for guessing than when there was correction for guessing.

However, the difference in overall response pattern remains unclear from these analyses. In other words, these data do not tell us whether different response strategies were used depending on the instructions given to students. To study this a series of 2 (correction: correction versus no correction) x 2 (gender: male versus female) unrelated ANOVAs were used to investigate whether significant differences emerged for the percentage of questions left unanswered according to marking criteria and gender[2]. Table 3 displays the means and standard deviations for the percentages of questions left unanswered for all three examinations.

Insert Table 3 here

The analyses showed that there was a main effect of correction for guessing on the percentage of questions left unanswered for the first-year theory MC examination, $F(1, 520) = 316.75$, $p < .001$, $\eta^2 = .379$; the first-year research methods MC examination $F(1, 512) = 225.08$, $p < .001$, $\eta^2 = .305$; and the second-year research methods MC examination $F(1, 415) = 189.85$, $p < .001$, $\eta^2 = .252$. Taken together these results suggest that when students are told that there is correction for guessing there are a significantly higher number of unanswered questions than when students are told there is no correction for guessing. In particular, 13% - 22% of the questions were left unanswered when there was correction for guessing. There were no gender differences in these results, suggesting that males and females leave an approximately equal percentage of questions unanswered.

The results presented so far indicate that students scored higher and left fewer questions unanswered when there was no correction for guessing. The next stage in the analysis was designed to examine the effects of changing the marking criteria once the

students had sat the examination. Specifically, in the first instance we examined the effect on students' scores of *introducing* correction for guessing after the students had been informed that there was *no* correction for guessing. In the second instance, we investigated the effect on students' scores of *removing* the correction for guessing when students had been told that there *was* correction for guessing. Gender differences in scores were also examined alongside these changes to the marking criteria.

The effect of introducing correction for guessing on those not aware

To examine whether performance changed when introducing correction for guessing when students were not aware of this procedure compared to when students were aware, 2 (correction: aware versus unaware) x 2 (gender: male versus female) unrelated ANOVA's were conducted for each MC examination. Table 4 displays the means and standard deviations for students' scores according to marking criteria and MC examination.

Insert Table 4 here

For the first-year theory MC examination there was no main effect of correction for guessing, $F(1, 520) = .41$, $p = .52$, $\eta^2 = .001$. Therefore, it seems that applying the correction for guessing to those students not aware that this marking strategy would be used did not significantly reduce their scores. This suggests that leaving several answers blank does not improve scores in the MC theory examination. However, there was a marginal main effect of correction for guessing for the first-year research methods MC examination, $F(1, 512) = 2.79$, $p = .096$, $\eta^2 = .005$. Contrary to the pattern of results for the theory MC examination, students scored marginally *higher* when they

were aware of the correction for guessing than when they were not. Therefore, students scored lower when the correction for guessing was implemented without their knowledge, although this was still not statistically significant. For the second-year research methods MC examination, there was no significant difference in students' scores between those aware of the correction and those not aware, $F(1, 415) = .38$, $p = .54$, $\eta^2 = .001$.

Looking at the effect of gender, females tended to score higher on the theory MC examination ($M = 39.50$, $SD = 14.47$) than males ($M = 35.88$, $SD = 13.49$) regardless of whether students were aware of the correction for guessing, $F(1, 520) = 6.22$, $p = .013$, $\eta^2 = .012$. However, for the first-year research methods examination there was no significant difference in scores when comparing males ($M = 54.23$, $SD = 15.82$) and females ($M = 56.14$, $SD = 16.16$), $F(1, 512) = 1.56$, $p = .21$, $\eta^2 = .003$. Similarly, for the second-year research methods examination there was no significant difference in scores when comparing males ($M = 48.51$, $SD = 14.40$) with females ($M = 47.10$, $SD = 15.86$), $F(1, 415) = .70$, $p = .41$, $\eta^2 = .002$.

The effect of removing the correction for guessing on those who believed correction would be applied

Students' scores were also examined when the correction for guessing was removed. So, for one cohort this would have meant that they were informed that correction for guessing was to be applied, when in fact it was not. The other cohort was informed that there was no correction for guessing and, in this case, none was applied. Table 5 displays the means and standard deviations for students' scores according to marking criteria and MC examination.

Insert Table 5 here

For the first-year theory examination, there was a significant main effect of the removal of the correction for guessing $F(1, 520) = 17.62$, $p < .001$, $\eta^2 = .033$. Students scored significantly higher when they were aware that there would be no correction for guessing than students who thought correction for guessing would be applied when in fact it was not. This suggests that because students think there will be correction for guessing they leave questions unanswered, which then has a detrimental effect on scores. For the first-year research methods examination there was no main effect of removing the correction for guessing, $F(1, 512) = .63$, $p = .43$, $\eta^2 = .001$, so students who were aware there was no correction for guessing scored equally as well as those students who thought correction for guessing would be applied when in fact it was not. Similarly, for the second-year research methods examination there was no main effect of removing the correction for guessing, $F(1, 415) = 2.02$, $p = .16$, $\eta^2 = .005$.

Looking at the effects of gender, for the first-year theory examination, females ($M = 52.51$, $SD = 11.85$) scored significantly higher than males ($M = 49.68$, $SD = 10.75$, $F(1, 520) = 5.85$, $p = .016$, $\eta^2 = .011$. For the first-year research methods examination there was no significant difference in scores when comparing males ($M = 63.82$, $SD = 12.80$) with females ($M = 65.53$, $SD = 12.76$), $F(1, 512) = 1.63$, $p = .20$, $\eta^2 = .003$. Again, for the second-year research methods examination there was no significant difference in scores when comparing males ($M = 59.77$, $SD = 11.85$) with females ($M = 58.24$, $SD = 12.64$), $F(1, 415) = .84$, $p = .36$, $\eta^2 = .002$.

Discussion

In summary, the results show that students scored higher and left fewer questions unanswered when there was no correction for guessing than when there was correction for guessing. In other words, the correction for guessing did reduce students' marks. However, implementing a correction for the first-year theory MC examination to those who were not aware of the correction did not lower students' scores compared with those who were aware of the correction for guessing. In fact, the implementation of correction for guessing for students not aware that one would be used meant that they still scored slightly higher than those who were aware of the correction. This suggests that for the theoretical MC examination the correction for guessing did not actually do the job it was supposed to. If it did, we would expect the marks for the not aware cohort to be significantly lower compared with the aware cohort. For the two research methods MC examinations there was little or no difference depending on whether students were aware or not aware of the correction for guessing.

When the correction for guessing was removed from the first-year theory MC examination, students who were told there was no correction for guessing scored higher than those who were told that there was correction. This is not surprising given that in this MC examination with correction for guessing students tended to leave a large percentage of questions unanswered. In the two research methods MC examinations there was no difference across the two cohorts when the correction for guessing was removed.

Study 2

The findings of Study 1 indicated that differences did occur in students' scores on MC examinations according to whether or not correction for guessing was implemented, but we still do not know how the correction for guessing affected

students' attitudes and strategies towards the examination. Therefore, Study 2 was designed to examine students' perceptions of the correction for guessing procedure with regard to both closed-book and open-book MC examinations. The distinction between open-book and closed- book examinations was made because research suggests that students approach these examinations differently (Ioannidou, 1997; Theophilides & Dionysiou, 1996; Theophilides & Koutselini, 2000).

Method

Participants. A further 116 (24 male and 88 female) first-year Psychology students evaluated both open-book and closed-book MC examinations. The mean age of the participants was 19.47 (*SD* = 4.20). Students completed the evaluations approximately two weeks after the examinations at the start of a lecture, before their marks were published. These students comprised a different cohort than those in Study 1, although the examination questions were drawn from the same question bank.

Materials. Five questions were constructed to investigate students' confidence and perceptions of performance in the MC examinations. Specifically, students were asked "Did the correction for guessing make you feel less confident about answering questions you were not sure about?", "Did the correction for guessing make you feel more anxious about the exam and your performance", "How confident did you feel about your performance after the exam?", "Do you think you prepared more for the exam knowing there was a correction for guessing?" and "Do you think having a correction for guessing is a good idea?" Students responded to the items using a 7-point Likert scale (from 1 = *not at all* to 7 = *very much*).

Additionally, students were asked "how many questions on the exam paper did you leave blank" and "how many questions would you have left blank if there had not been a correction for guessing". Students completed all questions twice: once with

regard to an open-book MC examination and once with regard to a closed-book MC examination. The open-book MC examination was designed to assess students' knowledge of research methods and the closed-book examination was designed to assess students' knowledge of Psychological theory. In the open-book examination students were able to use course materials and textbooks. Each of these MC examinations was one component of a wider range of assessment methods used to assess students' performance on two different first-year Psychology modules. The number of questions in the closed-book examination was 72 but in the open-book examination there were 45 questions. There were fewer questions in the open-book examination in order to allow there to be time for the students to carry out statistical calculations.

Results

Students' responses to the evaluation items were analysed separately through a series of 2 (examination condition: open-book versus close book) x 2 (gender: male versus female) mixed ANOVAs, with examination condition as a repeated measure. Table 6 displays the means and standard deviations for the students' evaluation of the MC examination.

Insert Table 6 here

There was a significant main effect of examination condition for whether the correction for guessing made students less confident about answering questions, $F(1,113) = 9.74$, $p = .002$, $\eta^2 = .079$; confidence in examination performance, $F(1,114) = 28.31$, $p < .001$, $\eta^2 = .199$; and anxiety surrounding the examination and their performance, $F(1,114) = 2.92$, $p = .09$, $\eta^2 = .025$. Students felt less confident about

answering questions, less confident in their examination performance and had higher anxiety about their performance when correction for guessing was used in conjunction with the closed-book MC examination than in the open-book examination. Similarly, when students were asked whether they thought the correction for guessing was a good idea, there was a significant main effect of examination condition, $F(1,114) = 15.61$, $p < .001$, $\eta^2 = .121$. Students thought it was better to have correction for guessing for the open-book examination than for the closed-book examination. There were no gender differences or interactions between gender and examination condition for these responses.

Interestingly, there was no significant main effect of examination condition on the amount of preparation students engaged in, $F(1,114) = .97$, $p = .33$, $\eta^2 = .008$. Students reported preparing for the open-book and closed-book examination to a similar extent. This is quite surprising given that students were aware that they were able to take their course material and textbooks into the open-book MC examination. However, there was a marginally significant interaction between examination condition and gender for preparation, $F(1,114) = 3.69$, $p = .057$, $\eta^2 = .031$. Repeated measures ANOVAs, preformed separately for each gender, revealed that females prepared more for the closed-book examination ($M = 4.49$, $SD = 1.43$) than for the open-book examination ($M = 4.12$, $SD = 1.34$), $F(1,90) = 9.58$, $p = .003$, $\eta^2 = .096$. There was no significant difference in the amount that males prepared for the open-book ($M = 3.88$, $SD = 1.54$) and closed- book examination ($M = 3.76$, $SD = 1.59$), $F(1,24) = .30$, $p = .59$, $\eta^2 = .012$. Further, there was no significant main effect of gender on preparation.

Questions left unanswered

Due to the variation in the number of questions for each examination, the students' estimates of the number of questions they left unanswered, and the number of

questions that they would have left unanswered if correction for guessing had not been used, were converted into percentages for each examination. As before, 2 (examination condition: open-book versus close book) x 2 (gender: male versus female) mixed ANOVAs were conducted, with examination condition as a repeated measure. There was a significant main effect of examination type on the estimated percentage of unanswered questions, $F(1,110) = 6.79$, $p = .010$, $\eta^2 = .058$. Students reported leaving a higher percentage of questions unanswered on the closed-book examination ($M = 9.82$, $SD = 10.12$) than the open-book examination ($M = 7.40$, $SD = 9.39$). Similarly, there was a marginal main effect of examination condition on the percentage of answers that students would have left unanswered if there was no correction for guessing, $F(1,114) = 3.72$, $p = .056$, $\eta^2 = .032$. Students estimated that if there would have been no correction for guessing they would have left more questions unanswered on the open-book examination ($M = .98$, $SD = 5.49$) than on the closed-book examination ($M = .50$, $SD = 5.49$). There was also a marginally significant interaction between examination condition and gender, $F(1,114) = 2.86$, $p = .093$, $\eta^2 = .024$.

Discussion

In summary, these results suggest that these students appeared to favour the use of correction for guessing for open-book MC examinations. Specifically, students reported that implementing correction for guessing was a good idea for open-book MC examinations but less so for closed-book examinations. Also, the students reported higher levels of confidence both in their performance and in selecting answers when the correction for guessing was used for the open-book MC examination than for the closed-book examination. It was also interesting to see that females prepared more for the closed-book examination than for the open-book examination, but that males reported equal levels of preparation for the two examinations. It is possible that this

may account for the results in Study 1 which showed that females tended to score higher on the closed-book examination than males. This result also suggests that females tend to be more strategic in their preparation for examinations than do males.

General Discussion

The aim of the research outlined in this paper was to investigate students' performance on, and attitudes towards, MC examinations. In particular, we were interested in the effects of correction for guessing because of its increasing popularity as a marking criterion in MC examinations. In our first study we examined how students' knowledge of correction for guessing affected their performance. This was possible because there was a change in marking criteria between the two cohorts: for one cohort the correction for guessing was not applied and therefore students were unaware of the correction, but for the second cohort the correction for guessing was introduced and therefore students were aware of the correction.

Students scored higher on the MC examinations when there was no correction for guessing than when one was implemented with their knowledge. However, implementing correction for the first-year theory MC examination to those who were not aware that a correction would be implemented did not lower scores compared to those who were aware of the correction for guessing. In fact, the implementation of correction for guessing for those not aware meant that they still scored slightly higher than those who were aware of the correction. For the research methods MC examinations there was little or no difference depending on whether students were aware or not aware of the correction for guessing.

When the correction for guessing was removed from the first-year theory MC examination, students who were told there was no correction for guessing scored higher than those who were told that there was correction. This is not surprising given that in

the latter examination students tended to leave a large percentage of the questions unanswered. In the research methods MC examinations there was no difference between the two cohorts when the correction for guessing was removed. Together, these results suggest that correction for guessing had only a minimal influence on students' performance in the research methods examinations when the examination was open-book. Furthermore, for the theory closed-book examination, the correction for guessing led students to leave more questions unanswered and as such their performance was negatively affected.

In addition to examining the effect of correction for guessing, we also examined gender differences in MC examination performance. In general, the present study added to the growing literature that suggests that there are only minimal gender differences in MC examination performance (Greene, 1997). However, there was evidence of gender differences on performance for the theory MC examination. Females tended to score higher than males on the theory MC examination, which is consistent with the findings of Hartley et al. (2007). Consequently, the pattern of gender differences may be a function of the differences in the MC examination content rather than gender *per se*. This difference in performance according to gender and MC examination topic warrants further investigation because any advantage for gender on a MC examination has implications for equality, especially given the increasing popularity of MC examinations within higher education.

Study Two revealed that students were more confident about answering questions in open-book MC examinations when there was correction for guessing compared to closed-book examinations. Similarly, students reported leaving more questions unanswered in closed-book MC examinations than in open-book examinations. This suggests that students are more likely to follow the advice 'not to guess' in closed-book

MC examinations than they do in open-book MC examinations. However, consistent with previous findings (Budescu & Bar-Hillel, 1993; Hammond et al., 1998), the results of this study suggest that this strategy may be detrimental for students' performance. One explanation for this may be that students are more confident in their guesses when the MC examination is open-book because their notes and textbooks provide some partial knowledge and in this situation it is best to guess (Budescu & Bar-Hillel, 1993; Hammond et al., 1998).

In addition to leaving fewer questions unanswered in the closed-book examination compared with the open-book examination, students also reported feeling less anxious and more confident in their performance in the open-book examination. This is surprising given that a previous study has reported that students perceive open-book and closed-book examinations to be of equal difficulty (Theophilides & Dionysiou, 1996). However, the pattern of results in the present study may be partly due to preparation levels. In previous research, some authors have argued that students' prepare less for open-book than they do for closed-book examinations (Ioannidou, 1997; Theophilides & Koutselini, 2000), but this pattern of results was not replicated in the present study. One explanation for the similarity in preparation for open-book and closed-book MC examinations is that some study guides encourage students to highlight key facts and information that they may need in the examination (e.g., Collins & Kneale, 2004). Interestingly, females reported that they prepared more for the closed-book MC examination than they did for the open-book examination; this finding is consistent with previous research (Ioannidou, 1997; Theophilides & Dionysiou, 1996; Theophilides & Koutselini, 2000).

Whilst there were differences in student evaluations of the open-book and closed-book MC examinations, future research is needed to explore this topic further. In

particular, we compared an open-book and closed-book MC examination that covered different topics and this could have influenced the findings. Future research could compare students' perceptions of an open-book and a closed-book MC examination assessing similar skills and topics. Such an investigation would lead to a greater understanding of whether the results are due to the examination topic or the examination condition.

There are implications from the present studies for both students and instructors. From the perspective of students, the findings of the present study seem to contradict the advice given to students in study guides, which is not to guess (Collins & Kneale, 2004; Freeman & Stone, 2006). The results of the present study support the advice given by Hammond et al. (1998) that students *should* guess if they have partial knowledge.

From the perspective of instructors, there are implications for deciding whether or not correction for guessing is an appropriate marking criterion. Our data suggest that implementing correction for guessing for open-book examinations does not result in significantly poorer performance. However, for the closed-book theory MC examinations the correction for guessing did have a detrimental effect on the number of questions students left unanswered which in turn negatively affected their performance. It could be argued that applying correction for guessing to theory MC examinations is not actually an effective means of correcting for guessing at all. Instead, our research suggests that students' lack of confidence in closed-book exams means that they are more likely to leave questions unanswered than to guess.

A second implication for instructors concerns the advice they give to students about whether or not to guess. Our data seem to suggest that contrary to what students are normally advised - not to guess - guessing does not significantly reduce

performance. This suggests that instructors should be more cautious about advising students not to guess. In fact, it may be better to advise students to guess, particularly when they are able to discount some of the available options and therefore arrive at a more educated ‘guess’.

References

Anderson, J. (2002) Gender-related differences on open and closed assessment tasks, *International Journal of Mathematical Education in Science and Technology, 33,* 495-503.

Ben-Shakhar, G. & Sinai, Y. (1991) Gender differences in multiple-choice tests: The role of differential guessing tendencies, *Journal of Educational Measurement, 28,* 23-35.

Bridgeman, B. & Lewis, C. (1994) The relationship of essay and multiple-choice scores with grades in college courses, *Journal of Educational Measurement, 31,* 37-50.

Budescu, D. & Bar-Hillel, M. (1993) To guess or not to guess: A decision-theoretic view of formula scoring, *Journal of Educational Measurement, 30,* 227-291.

Burton, R. F. (2001) Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answers, *Assessment and Evaluation in Higher Education, 26,* 41-50.

Burton, R. F. (2002) Misinformation, partial knowledge and guessing in true/false tests, *Medical Education, 36,* 805-811.

Burton, R. F. (2004) Multiple choice and true/false tests: reliability measures and some implications of negative marking, *Assessment & Evaluation in Higher Education, 29,* 587-595.

Burton, R. F. (2005) Multiple-choice and true/false tests: myths and misapprehensions, *Assessment & Evaluation in Higher Education, 30,* 65-72.

Burton, R. F. & Miller, D. J. (1999) Statistical modelling of multiple-choice and true/false tests: ways of considering, and of reducing, the uncertainties attributable to guessing, *Assessment & Evaluation in Higher Education, 24,* 399-411.

Bush, M. (2001) A multiple choice test that rewards partial knowledge, *Journal of Further and Higher Education, 25,* 157-163.

Collins, S. C. & Kneale, P. E. (2004) *Study skills for psychology students: A practical guide* (London: Arnold).

Diamond, J. & Evans, W. (1973). The correction for guessing, *Review of Educational Research, 43,* 181-191.

Downs, C. T. (2006) What should make up a final mark for a course? An investigation into the academic performance of first year bioscience students, *Assessment & Evaluation in Higher Education, 31,* 345-364.

Frary, R. B. (1988) NCME instructional module on formula scoring of multiple-choice tests (correction for guessing), *Educational Measurement: Issues and Practice, 7,* 33-38.

Freeman, R. P. J. & Stone, T. (2006) *Study skills for psychology: succeeding in your degree* (London: Sage Publications).

Greene, B. (1997) Verbal abilities, gender, and the introductory economics course: A new look at an old assumption, *Journal of Economic Education, 28,* 13-30.

Hammond, E. J., McIndoe, A. K., Sansome, A. J. & Spargo, P. M. (1998) Multiple-choice examinations: adopting an evidence-based approach to exam technique, *Anaesthesia, 53,* 1105-1108.

Hartley, J., Betts, L. & Murray, W. (2007) Gender and assessment: Differences, similarities and implications, *Psychology Teaching Review, 13,* 34-47.

Ioannidou, M. K. (1997) Testing and life-long learning: Open-book and closed-book examination in a university course, *Studies in Educational Evaluation, 23,* 131-139.

Nicol, D. (2007) E-assessment by design: using multiple-choice tests to good effect, *Journal of Further and Higher Education, 31,* 53-64.

Paxton, M. (2000) A linguistic perspective on multiple choice questioning, *Assessment & Evaluation in Higher Education, 25,* 109-119.

Simkin, M. G. & Kuechler, W. L. (2005) Multiple-choice tests and student understanding: what is the connection, *Decision Sciences Journal of Innovative Education, 3,* 73-97.

Scouller, K. (1998) The influence of assessment method on students' learning approaches: multiple choice question examination versus assignment essay, *Higher Education, 35,* 453-472.

Shine, S., Kiravu, C. & Astley, J. (2004) In defence of open-book engineering degree examinations, *International Journal of Mechanical Engineering Education, 32,* 197-211.

Tabachnick, B. G. & Fidell, L. S. (2001) *Using multivariate statistics* (4th ed.) (Boston: Allyn and Bacon).

Theophilides, C. & Dionysiou, O. (1996) The major functions of the open-book examination at the university level: A factor analytic study, *Studies in Educational Evaluation, 22,* 157-170.

Theophilides, C. & Koutselini, M. (2000) Study behavior in the closed-book and the open-book examination: A comparative analysis, *Educational Research and Evaluation, 6,* 379-393.

Von Schrader, S. & Ansley, T. (2006) Sex differences in the tendency to omit items on multiple-choice tests: 1980-2000, *Applied Measurement in Education, 19,* 41-65.

Walker, D. M. & Thompson, J. S. (2001) A note on multiple choice exams, with respect to students' risk preference and confidence, *Assessment & Evaluation in Higher Education, 26,* 261-267.

Williams, R. L. & Clark, L. (2004) College students' ratings of student effort, student ability and teacher input as correlates of student performance on multiple-choice exams, *Educational Research, 46,* 229-239.

Footnotes

[1]In the context of the two studies presented in this paper we use the term correction for guessing to denote that only a proportion of a mark was deducted and to reflect the fact that the marking criterion was used to discourage guessing (see Burton, 2005).

[2] We recognise that the descriptive statistics for these analyses may indicate that the data are not normally distributed. However, with such a large sample size, the effect of non-normal distributions is minimal (Tabachnick & Fidell, 2001).

Table 1.

Study 1. Number of students who participated in each MC examination according to cohort and gender

Examination	Number of participants according to cohort and gender					
	Cohort 1			Cohort 2		
	Male	Female	Total	Male	Female	Total
First- year theory	65	203	269*	64	192	256
First-year research methods	61	203	265*	67	185	252
Second-year research methods	44	146	202*	41	188	229

*In the first cohort, some students withdrew from the course and therefore their gender is unknown

Table 2.

Study 1. Means and standard deviations for performance on MC examinations according to different marking criteria (as percentages)

	Marking criteria	
Examination	No correction for guessing	Correction for guessing
First-year theory	54.33 *(10.65)* $_a$	37.90 *(14.46)* $_b$
First-year research methods	65.38 *(11.53)*$_a$	57.35 *(16.74)* $_b$
Second-year research methods	59.48 (*11.24)*$_a$	48.34 *(16.14)* $_b$

Note: Within rows means with different subscripts are significantly different at $p < .01$

Table 3.

Study 1. Means and standard deviations for percentage of questions left unanswered on MC examinations according to different marking criteria (as percentages)

	Marking criteria	
Examination	No correction for guessing	Correction for guessing
First-year theory	0.11 *(0.42)* $_a$	16.48 *(12.96)* $_b$
First-year research methods	0.40 *(0.40)*$_a$	12.52 *(11.44)* $_b$
Second-year research methods	0.39 (*2.42)*$_a$	21.74 *(18.88)* $_b$

Note: Within rows means with different subscripts are significantly different at $p < .001$.

Table 4.

Study 1. The effect of introducing correction for guessing on those not aware: Means and standard deviations for performance on MC examinations according to awareness of the correction for guessing (as percentages)

	Awareness of the correction for guessing	
Examination	Not aware of correction	Aware of correction
First-year theory	39.29 *(14.15)* $_a$	37.90 *(14.46)* $_a$
First-year research methods	54.06 *(15.29)*$_a$	57.35 *(16.74)* $_a$
Second-year research methods	46.24 (*14.81)*$_a$	48.34 *(16.14)* $_a$

Note: Within rows means with different subscripts are significantly different at $p < .05$

Table 5.

Study 1. The effect of removing the correction for guessing on those who believed correction would be applied: Means and standard deviations for performance on MC examinations according to awareness of the correction for guessing (as percentages)

	Effect of removing the correction for guessing	
Examination	Not aware correction would be removed	Aware of no correction
First-year theory	49.18 *(12.08)* $_a$	54.33 *(10.65)* $_b$
First-year research methods	64.82 *(13.98)*$_a$	65.38 *(11.53)*$_a$
Second-year research methods	57.78 (*13.41)*$_a$	59.48 *(11.24)* $_a$

Note: Within rows means with different subscripts are significantly different at $p < .01$

Table 6.

Study 2. Means and standard deviations for the students' evaluation of open-book and closed-book MC examinations

	Examination	
	Open-book	Closed-book
Did the correction for guessing make you feel less confident about answering questions you were not sure about?	5.48 *(1.34)* $_a$	5.82 *(1.13)* $_b$
Did the correction for guessing make you feel more anxious about the exam and your performance after the exam?	5.34 *(1.38)*$_a$	5.51 *(1.34)*$_b$
How confident did you feel about your performance after the exam?	4.24 (*1.23)*$_a$	3.59 *(1.24)*$_b$
Do you think that you prepared more for the exam knowing there was a correction for guessing?	4.07 *(1.38)*$_a$	4.34 *(1.49)*$_a$
Do you think having a correction for guessing is a good idea?	3.52 (*1.80)*$_a$	2.83 *(1.59)* $_b$

Note: Within rows means with different subscripts are significantly different at $p < .01$, except for anxiety surrounding performance where $p = .09$