Does the multiple choice question structure in examinations have an effect on student performance?

Dragi Klimovski[a] and Antonio Cricenti[a].
Swinburne University of Technology[a]
Corresponding Author Email: dklimovski@swin.edu.au

Structured abstract

BACKGROUND
The existence of multiple-choice questions (MCQs) as an examination method has been around for more than a century. The attraction of MCQ tests/examinations is twofold. They allow a large amount of material to be tested over a short period of time and many cognitive abilities can be assessed. These tests can be marked relatively quickly and accurately. The majority of MCQs are structured with a stem (the question) with options (containing a single key and multiple distractors) (Cohen & Wollack, 2000). In addition the test can allocate negative marks for incorrect choices to avoid guessing or it can have no penalty for guesses. There are no hard and fast rules to determine the optimal number of distractors (Nitko, 2004).

PURPOSE
The focus of this project is to investigate the effect on student results of changing the structure of MCQs in examinations. In particular the effect of replacing 4 optioned MCQs with negative marking by 6 optioned MCQs with no penalties will be evaluated.

DESIGN/METHOD
The method used in this research was to change the format of the MCQs. Originally the format was 4 alternatives with a mark of +3 for the key, 0 for no answer and -1 for any of the distractors. The new format was to have 6 alternatives with +3 for the key and 0 for no answer or any of the distractors. The data is collected through end of semester final written examinations. The data collected has equal samples of years of average examination results of both negative marking MCQs and non-negative marking MCQs. The data is analysed on the performance of the various year's students and then compared and contrasted. The comparison is carried on both the MCQ section results as well as the total examination result. A student survey is conducted to attain the students' views.

RESULTS
The results show that there was an improvement in the performance of the students in the MCQ section, but overall across the whole exam, which includes the MCQ section as well as a short answer section, there was only a marginal improvement.
A student survey was conducted to determine the student's views on the different MCQ formats in the examination. The survey shows that 60% prefer to have the 6 alternative MCQs with no penalties. When negative marking could not be avoided 80% of the students chose to have the largest differential possible ie (+4,-1), thereby reducing the impact of negative marking.

CONCLUSIONS
The effect of changing the structure of MCQs in examinations has shown that the overall results in the examination have not changed greatly statistically. In fact the overall marks improved slightly as you would expect with no negative marking, but of no great significance and yet the students believe that the new format is better for them. The students feel less stressed in doing the examination.

KEYWORDS
Multiple choice question, examination, negative marking.

Introduction

Student assessment has been a challenge for as long as mankind has been learning/teaching. Over the centuries since Socrates invented the teaching practice of pedagogy (Boghossian, 2003 Vol 23 Book 2), Western learning has developed and tried many approaches to assess the level of knowledge attained. The evolution of the transference of knowledge and the assessment of the level of understanding has led to modern day assessment styles which need to deal with the great abundance of information courtesy of the Internet. One well known and used method is the multiple choice question (MCQ).

The existence of multiple-choice questions as an examination method has been around for more than a century (McArthur, 1983; Fuchs & Trewin, 2007). The attraction of MCQ tests/examinations is twofold. They allow a large amount of material to be tested over a short period of time and many cognitive abilities can be assessed. These tests can be marked relatively quickly and accurately. Well-structured MCQ tests are considered to be objective because no judgment has to be made as to the correctness of an answer at the time of marking, as the correct answer is already pre-determined. The majority of MCQs are structured with a stem (the question) with options (containing a single key and multiple distractors) (Cohen & Wollack, 2000). In addition the test can allocate negative marks for incorrect choices to discourage guessing or it can have no penalty for guesses. There are no hard and fast rules to determine the optimal number of distractors (Nitko, 2004).

In this paper the background for using MCQs as a form of assessment is presented, followed by an outline of the approach used in gathering the data needed to undertake the evaluation of the research investigation. A discussion of the results shown in graphs is presented as well as the rationale and need to conduct a student survey. Lastly the conclusions and recommendations are presented.

Background

At Swinburne University of Technology in the Faculty of ICT in the Telecommunications Academic Group, many examinations were composed of short answer questions. This type of questions required the students to read and comprehend the information provided, followed by some analysis and then a discussion, calculation or synthesis of the data to provide their answer. The migration to using MCQs as part of the end of semester written examination came about as a result of large numbers of students choosing the academic group's courses. In general, MCQs have been used in subject examinations to try and assess the breadth of a student's knowledge and short answer questions to assess the depth of their knowledge for any one particular area. The authors have found this combination of assessment works very well in determining the student's level of knowledge and understanding.

Over a number of years a fundamental Telecommunications Engineering communications principles subject has been taught to deliver the required knowledge to students enrolled in a number of engineering degrees. This subject has taken advantage of the use of MCQs to assess large areas of the syllabus as found by Nicholls (2002) up to the 4th level (Lister, 2000) of Bloom's Taxonomy (Bloom, 1956). There are many different forms and styles of MCQs. One main difference is the number of options available to choose the key from. Commonly there are three, four and five optioned MCQs. These multi-optioned MCQs can be with or without negative marking. The MCQs without negative marking have champions for any number of options. Both Landrum (1993) and Rodriguez (2005) provide evidence and data as to why three optioned MCQs are best to use, whilst Nitko (2004) strongly argues there are no clear rules to determine the best number of options. Van Blerkom (2009) argues that purely from a psychometric perspective the more options the better, as it reduces guessing and therefore removing the unfair advantage. However, he also states that the greater the number of options the more work required deciphering the information and

determining the optimal choice. It is also argued that the four option MCQ provides the best option discrimination (Van Blerkom, 2009).

MCQs ideally use negative marking to remove the possibility of rewarding candidates who guess and do not hold actual knowledge (Holt, 2006). Others state that in MCQs without negative marking the effect of lucky guesses is that someone with only 20% of the knowledge can attain, on average, a mark of close to 40% (Bush, 2001). Hence there is a justification to have penalties in MCQs and not reward guesswork. However, MCQs with penalties can have a negative impact on the students in terms of fairness and not rewarding partial knowledge (Gardener-Medwin & Gahan, 2003). A possible way to address this shortcoming is to provide confidence based MCQ assessment (Gardener-Medwin & Gahan, 2003; Farrell & Leung, 2006). In confidence based MCQ assessment the student may allocate a percentage of the available mark to different options in the question. For example, a student may allocate 70% of the mark to option (a) and 30% of the mark to option (d). One shortcoming of the confidence based MCQ is that this method could be very costly in terms of time, as students may spend a significant portion of their available time trying to optimize the percentage allocation of their marks to the various options.

Over many years of running this subject anecdotal evidence was gathered as to the detrimental impact the negative marking of the MCQ section of the examination had on students' perception and confidence. Many students had commented on the negative mark aspect of the MCQs:

- *"they're not fair…"*
- *"don't truly reward my level of knowledge…"*
- *"make me scared to answer the questions…"*

The staff involved with this subject had a number of thorough discussions as to how best to address these concerns and about how to change these student perceptions. The outcome of the discussions was to restructure the MCQ section so that there were no negative marks for the MCQs and that the staff felt comfortable about not giving too large of an unfair reward to those students who guessed the answer. The approach undertaken in this research to overcome the negative impact on students is to increase the number of options in the MCQs so that anyone guessing is only slightly rewarded.

Purpose

The focus of this research is to investigate the effect on student results of changing the structure of MCQs in examinations from 4 optioned MCQs with negative marking to 6 optioned MCQs with no penalties. Part of the research is to show that the large body of anecdotal evidence collected by the authors was truly present in the majority of student beliefs through the use of an anonymous survey.

Approach

The method used in this research was to change the format of the MCQs from 4 alternatives with a mark of +3 for the key, 0 for no answer and -1 for any of the distractors to having 6 alternatives with +3 for the key and 0 for no answer or any of the distractors. Whilst recognising that partial marks are being awarded for guesses, it is significantly less generous than the 4 alternatives scenario without penalties. The academic staff involved with this subject were comfortable with this compromise. The data for the different MCQ scenarios is collected through end of semester final written examinations. The data collected has equal samples of years, of average examination results, for both negative marking and greater number of alternatives with no negative marking to choose from. The data is analysed on the performance of the various year's students and then compared and contrasted. The comparison is carried out on both the year average of the MCQ section results as well as the average of the short answer section of the examination result. The students who undertook the subject between the years 1999 and 2005 had the 4 optioned MCQs with the negative marking. In 2006 and 2007 the MCQ structured was changed to 5 alternatives without

penalties. The students who undertook the subject between the years 2008 and 2012 had the 6 optioned MCQs without penalties.

To ensure that the results data from different years can be compared appropriately, the design of the examination questions were of a similar level of difficulty and complexity as well as ensuring they covered the same topics. This was achieved by selecting questions from a bank of panel moderated questions covering the various topics delivered in the subject. The fact that this was a fundamental communications principles subject ensured the topics did not change greatly over the period over which the results were collected.

The method used to obtain the student views on the MCQ structure was through an anonymous survey, so that the students felt they could answer honestly, conducted on the current batch of students. The survey questions were structured so that the students could choose from a number of options as well as provide written input/feedback. A small sample of the questions and answers from the survey are shown in figures 1 and 2.

2) When answering Multiple Choice Questions (MCQs), which format do you prefer?

| Option A | 4 alternatives with penalties (+3 for correct, 0 no answer, -1 incorrect) | |
| Option B | 6 alternatives with no penalties (+3 for correct, 0 no answer, 0 incorrect) | ✓ |

Why?

The Incorrect answer doesn't mean always that we don't know the Question, but sometimes they are tricky so, we dont desireve to get (-1) for the incorrect Q.

Which is fairer and Why?

Option B

3) When answering MCQs as per Option A, when you **do not know the correct answer**, how likely are you to guess for the following situations? Choose one

| Always | Only if I can narrow my choice down to 1 out of 3 | Only if I can narrow my choice down to 1 out of 2 | Never |
| | | | ✓ |

4) When answering MCQs as per Option A, when you are **not sure of the correct answer**, how likely are you to guess for the following situations? Choose one

| Always | Only if I can narrow my choice down to 1 out of 3 | Only if I can narrow my choice down to 1 out of 2 | Never |
| | | | ✓ |

Figure 1: Sample of multiple questions on student survey

2) When answering Multiple Choice Questions (MCQs), which format do you prefer?

Option A	4 alternatives with penalties (+3 for correct, 0 no answer, -1 incorrect)	
Option B	6 alternatives with no penalties (+3 for correct, 0 no answer, 0 incorrect)	✓

Why?

Less guessable while not encouraging gambling.

Which is fairer and Why?

B. less likely to be overcome by exam-induced self doubt and not answer something you probably know.

Figure 2: Sample of different answer to Q2 on student survey

Results

The results of the students' performance over the years from 1999 to 2012 are shown in figure 3. Note these results have been normalised so that they are relative to the 1999 results. This has been achieved by dividing the current year result by the corresponding 1999 result. The results are broken down into the yearly students' average of the MCQ section (MCQ), the short answer section (SECB) and the overall combined mark (TOTAL). Note that only the TOTAL student marks were available for the year 2009.

Examining the graphs in figure 3 it can be seen that there was an improvement in the performance of the students in the MCQ section over time. A significant improvement in the results of the MCQ section is shown in 2006 and 2007. The reason for this is that these were the years when the first change to the MCQ format was made, and in these years the MCQ section comprised of 5 options and no penalties. In 2008 the MCQ format was changed to 6 options with no penalties and it can be seen that the student performance dropped to a similar level as in the 4 options with penalties period. The 2009 results show that there was an increase in the TOTAL mark, however we cannot speculate on this result of the effect of the MCQ section as the breakdown is not available.

The results in Figure 3 also reveal that once the penalties were removed from the MCQ section, the students' performance in that section of the examination was better than in SECB. The reason for this was twofold. Firstly, the students no longer felt restricted by the prospect of possibly losing marks by making a calculated guess and secondly the extra time taken to understand the extra alternatives meant less time was spent on answering the questions in SECB.

The most telling result is that the overall examination result (TOTAL), which includes both the MCQ and SECB sections, exhibits only a marginal improvement. The gradient for the trend line of the TOTAL result is only +0.5, which is very close to flat. Whilst the expectation might have been to see a greater improvement in the TOTAL mark, a possible reason for this is based on the fact that the length of the examination has remained the same over the period this research. Since the examination length is constant it can be expected that the average student would take more time to complete the MCQ section and hence have less time to do the short answer section, SECB, where they are awarded partial marks for any correct working shown. In general students are always advised to do the examination questions they think they know first to build their confidence, then the short answer questions and lastly the

questions they find most challenging. However each student has their own style for completing written examinations.
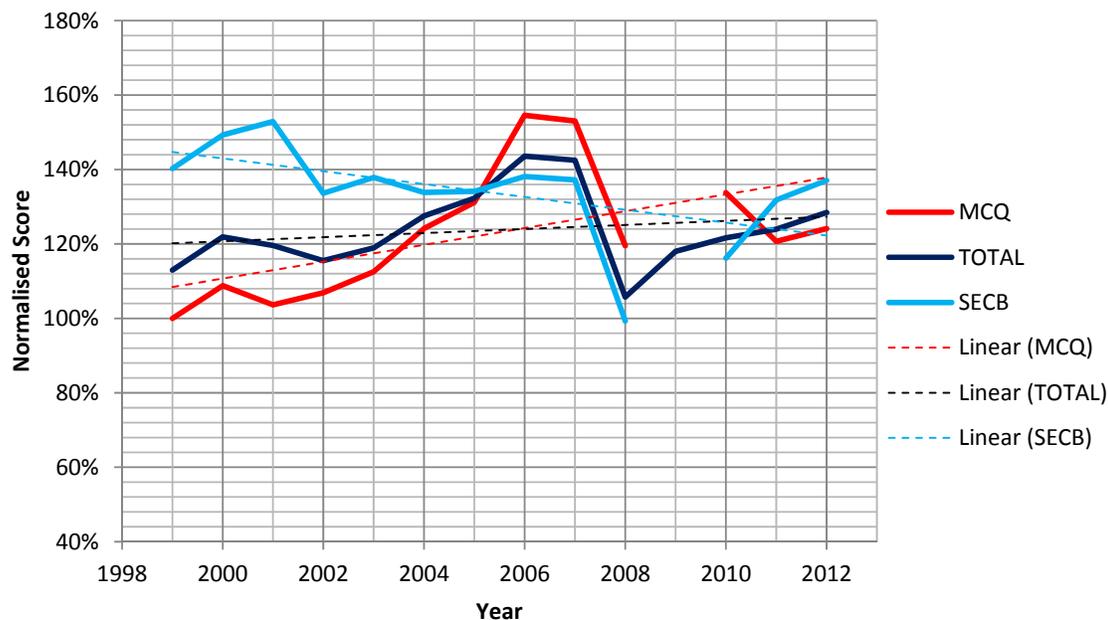

Figure 3: Yearly average of student results for the components of the examination

A student survey, of sample size 60, was conducted to determine the students' views on the different MCQ formats in examinations. The survey shows that 60% prefer to have the 6 optioned MCQs with no penalties. Their justification for this choice is that they do not want to lose marks for a partial lack of knowledge. The students perceive the 6 optioned MCQs with no penalties examination structure to be fairer and are happier to sit for the examination. Of the surveyed student population, 46% found that negative marking increased their stress levels and had an impact on their examination performance. Whilst 16% of the students stated that they would not answer a MCQ with penalties unless they were absolutely certain. When negative marking could not be avoided 80% of the students chose to have the largest differential possible i.e. (+4,-1), thereby reducing the impact of negative marking. The survey results confirm the anecdotal evidence gathered in the past about the students' views and preferences on the examination structure.

Conclusions

The effect of changing the structure of MCQs in examinations from 4 options with penalties to 6 options without penalties has yielded overall examination results which do not vary significantly.

The students who sit an examination with no penalties suffer less stress during the examination and are happier to take the examination. As one might expect the overall examination marks improved slightly due to the removal of penalties as statistically the students who guess are better off. This outcome is a win-win scenario as the students do not feel disadvantaged and are happier to sit for the examination and the staff are happy because this new format does not give unfair advantage since the overall results are similar.

The conclusions drawn in this research were based on results compared over many years and different cohorts. A recommendation from this work is to redesign the examination structure of a number of subjects using the different MCQ assessment methods. Thereby ensuring that each cohort of students is exposed to the various formats and a comparison can be made on a per student basis.

References

Bloom, B. (1956). Taxonomy of educational objectives. In B. Bloom, *Handbook 1. The Cognitive Domain.* New York: David McKay Co Inc. .

Boghossian, P. (2003 Vol 23 Book 2). How Socratic Pedagogy Works. *Informal Logic - Reasong and Argumentation in Theory and Practice (Teaching Supplement)*, 17-25.

Bush, M. (2001). A Multiple Choice Test that Rewards Partial Knowledge. *Journal of Further and Higher Education, 25*(2), 157-163.

Cohen, A., & Wollack, J. (2000). *Handbook on Test Development: Helpful Tips for Creating Reliable and Valid Classroom Tests.* Retrieved June 6, 2013, from http://testing.wisc.edu/Handbook%20on%20Test%20Construction.pdf

Farrell, G., & Leung, Y. (2006). A comparison of an innovative web-based assessment tool utilizing confidence measurement to the traditional multiple choice, short answer and problem solving questions. *10th CAA International Computer Assisted Assessment Conference* (pp. 195-204). Loughborough: © Loughborough University.

Fuchs, A. H., & Trewin, S. A. (2007). History of Psychology: Robert Yerkes' Multiple-Choice Apparatus, 1913-1939. *The American Journal of Psychology, 120*(4), 645-660.

Gardener-Medwin, M., & Gahan, A. (2003). Formative and Summative Confidence-Based Assessment. *Proc.7th International Computer-Aided Assessment Conference*, (pp. 147-155).

Holt, A. (2006). An Analysis of Negative Marking in Multiple-Choice Assessment. *The 19th Annual Conference of Computing & Information Technology Research & Education New Zealand.*

Landrum, R. E., Jeffrey, R., Cashin, A., Kristina, D., & Theis, S. (1993). More evidence in favor of three option multiple choice tests. *Educational And Psychodogical Measurement*, 771-778.

Lister, R. (2000). On blooming first year programming, and its blooming assessment. *Proceedings of the Australasian conference on Computing education*, (pp. 158-162). Melbourne, Australia.

McArthur, D. L. (1983). *Educational Testing and Measurment: A Brief History.* University of California, Los Angeles, Centre For The Study of Evaluation. Graduate School of Education. Los Angeles: University of California, Los Angeles.

Nicholls, G. (2002). *Developing Teaching and Learning in Higher Education.* Routledge Farmer.

Nitko, A. J. (2004). *Educational assessments of students.* Englewood Cliffs, NJ: Prentice Hall.

Rodriguez, M. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24(2)*, 3-13.

Van Blerkom, M. (2009). *Measurement and Statistics for Teachers.* Hoboken: Routledge.

Copyright statement