

# EVALUATION OF A NEW SCHEME FOR CONVERTING MCQ TEST SCORES TO PERCENTAGE MARKS

Yuyuan Zhao

The University of Liverpool, UK

## ABSTRACT

Multiple choice question (MCQ) tests are a widely used assessment methodology. In summative assessments, MCQ raw scores must be converted into more meaningful marks or grades. This paper introduces a recently developed conversion scheme and evaluates the applications of the scheme to an engineering computing module over two academic years. The assessment of the module consisted of 5 MCQ tests. The conversion scheme was applied to each test and then the marks of the 5 tests were averaged to give the final module marks. There was a reasonable correlation between the students' marks of this module and their overall average marks of all the other modules taught in the same semester. The changes in subject difficulties and improvement in teaching method were evidenced in the changes in the average marks of the individual tests. The MCQ assessment and the application of the conversion scheme are proved to be successful.

---

## INTRODUCTION

Multiple choice question (MCQ) tests are a widely used assessment methodology. They are objective, easy to mark and quick to obtain results. Properly designed MCQ tests are an efficient tool for assessing a large number of candidates, especially in knowledge or fact based subjects.

Designing MCQ tests and interpreting the test results are not as straightforward as setting and marking conventional examination papers. A conventional examination paper consists of a relatively small number of questions, which are explicitly or inexplicitly divided further into smaller elements. Each element is assigned a fixed mark. The mark that a student obtains in the examination is simply the sum of the marks of the correctly answered elements. The mark is normally expressed in percentage and is considered to be a true measure of the students' performance. Any MCQ test, however, invites some guesswork and therefore involves some uncertainty in the test results. This uncertainty issue must be addressed in summative assessments using MCQ tests.

The reliability of a MCQ test depends on the suitability of the subject content for MCQs, the format of the questions, and the method adopted for converting the marks. For a MCQ test to serve as a reliable and effective assessment method, three criteria must be met.

**Criterion 1:** each question is properly designed as a suitable measure of a learning outcome. This criterion is not always easy to be met. The main reason is that MCQ assessments are not suitable for some subjects. Engineering education involves acquiring knowledge, describing processes and phenomena, solving problems, developing experimental methodologies, creating designs, deriving mathematical formulas, analysing data, performing calculations and developing communications skills etc. It is difficult, if not impossible, to formulate MCQs for some of the above elements. For example, it is extremely difficult to split a design project into small and independent assessable problems. Even if MCQ assessments are suitable for the subject, badly devised MCQs can impair the effectiveness of the assessment. A good MCQ should have equally feasible choices of answers and the correct answer cannot

be identified by a layman. The quality of the design of MCQs is largely dependent upon the skills of the question setter, as subject relevance is concerned.

**Criterion 2:** the format of the MCQ test is designed to ensure that the fluctuation in the test scores as a result of random guesses is minimised. Zhao (1) analysed the effect of the format of a MCQ test on the part played by guesswork on the test scores. The probabilistic analysis confirms that the optimum number of choices of answers for each MCQ question is four (1). In a test composed of questions with two- or three-choices, there is a high chance of obtaining a pass score by pure guesswork and the scores often fall within a narrow range. It is often difficult to differentiate the students' performance. Introducing five or more choices of answers does not offer significant benefits in reducing the effect of guessing. From a subject point of view, questions with five or more choices of answers are also difficult to construct. The probabilistic analysis shows that for any given type of MCQs, the number of questions in a test has a determining effect on the reliability of test scores and thus test marks (1). For a four-choice question test, the minimum number of questions needed to reduce the probability of obtaining a mark above 40 by pure guesswork to below 5%, 1% or 0.01% is 8, 18 or 48, respectively. While the more questions the better, 20 four-choice questions are sufficient for ensuring a high level of reliability. MCQ tests containing 10 or more four-choice questions are often good enough.

**Criterion 3:** raw test scores are converted to marks that are a true measure of the students' performance. McLachlan and Whiten (2) differentiated marks, scores and grades and pointed out that raw scores of MCQ tests should not be used directly. Instead, scaling should be applied before the marks of the individual assessment units are aggregated. In well established tests involving a large number of participants, such as TOEFL, complex scaling schemes are often used. These schemes are developed on the basis of extensive research on the statistics of the

past tests, as demonstrated in Wainer and Wang (3). While these schemes are good for gauging the relative competence of the candidates, the converted test scores are not compatible with the percentage marks normally used in engineering assessments. Recently, Zhao (4) developed a scheme for converting MCQ raw scores to conventional percentage marks based on probability theory. The conversion scheme is independent of class size and historical data. It removes the guesswork element so that the converted marks become a true measure of the students' knowledge and competence. The converted marks are compatible with the standard marking scheme. MCQ tests can therefore be used standalone or as units of an assessment including conventional assessment units.

The author has used a series of computer based MCQ tests as the sole summative assessment method for a 7.5 credit module, *Introduction to Computing*, for the past two academic years. This module is compulsory for all first year students in the Department of Engineering, the University of Liverpool. The aim of the module is to equip the students with key computing skills for engineering applications. Traditionally, the module had been assessed by a series of coursework plus computer based tests conducted on a one-assessor-to-one-student basis. Because of the large number of skills to be assessed and the large number of students taking the module, the assessment was extremely time-consuming. It also had some shortcomings, such as difficulty in detecting copying and cheating, and long delay in giving feedback to students. The MCQ tests were introduced with an aim to achieve efficient and responsive assessments without sacrificing effectiveness and reliability.

This paper demonstrates the procedure of applying the conversion scheme developed by Zhao (4) and evaluates the outcomes of the applications of the scheme to the *Introduction to Computing* module.

## IMPLEMENTATION OF CONVERSION SCHEME

The simplest representation of the conversion scheme is in the form of conversion tables. Table 1 is for converting raw scores to standard percentage marks for MCQ tests with questions of two, three, four or five choices of answers (4). A raw score represents the percentage points a student scores in a MCQ test before any conversion is applied, and is simply termed 'score' in this paper. A standard percentage mark represents the percentage points a student deserves in a MCQ test, and is simply termed 'mark' in this paper. Take four-choice question tests as an example. A student obtaining a score equal to or below 25 is awarded a zero mark. A score of 60 corresponds to a mark of 53.

The conversions of scores of an MCQ test can be carried out by using a spreadsheet programme such as Microsoft Excel. For demonstration purposes, let us assume that a class of 20 students have taken a MCQ test composed of 20 four-choice questions, each of which is worth 5 points. The possible scores the students can obtain vary from 0 to 100 in steps of 5. Figure 1 demonstrates how a list of scores of the class is converted into a list of marks by using Excel. Firstly, we enter the conversion table in columns A and B from row 3 to row 23, with the scores in column A and their corresponding marks in column B. Secondly, we enter the student names in column D and their scores in column E, starting from row 2. Thirdly, we use the VLOOKUP function to perform the conversion. This is realised by entering the following formula in cell F2: =VLOOKUP(E2,\$A\$3:\$B\$23,2). The formula searches the value of E2 in column A and returns the corresponding value in the same row from column B. Finally, we select cell F2 and pull down the filler handle so that all the scores in column E are converted and the marks are displayed in column F.

In practice, the arrangement of the data in Excel can be improved to give better

presentation and clarity. For example, the conversion table may be entered on a separate worksheet. The same conversion table can be applied to the scores of a number of tests entered in different columns.

## EVALUATION OF ASSESSMENT

The *Introduction to Computing* module was composed of 5 units: MS Word, Excel Basics, Excel Optimisation, MATLAB Basics and MATLAB Programming. Each unit was assessed by a one-hour MCQ test with four-choice questions. Each test required the students to independently perform a series of operations using the software package being tested. In the test, the students were allowed to consult training materials. The module mark for each student was obtained by converting the scores of the 5 tests into standard percentage marks followed by averaging these marks.

In the current academic year (2005/06), the marks of three modules taken concurrently by all the first year students in the first semester have been made available. The average score before conversion and average mark after conversion of the *Computing* module have been compared with the average mark of the other two modules for 231 students. Figure 2 shows the relationship between the scores of *Computing* and the average marks of the other two modules. Figure 3 shows the relationship between the marks of *Computing* and the average marks of the other two modules. While the class average of the average marks of the other two modules is 54.3, the class averages of the scores and marks of *Computing* are 73.3 and 67.7, respectively. Fitting the data to a straight line also shows that the scores and marks of *Computing* are on average 28% and 19% higher than the average marks of the other two modules. As the other two modules are examination based, it is understandable that the marks of *Computing* can be considerably higher than the marks of these two modules. However, using the scores of MCQ tests directly would result in unacceptably high marks. The effect is especially

Table 1 Conversion table for MCQ tests with questions of two, three, four or five choices of answers, corresponding to columns indicated by [2], [3], [4] and [5]

Score	Mark				Score	Mark				Score	Mark			
	[2]	[3]	[4]	[5]		[2]	[3]	[4]	[5]		[2]	[3]	[4]	[5]
≤20	0	0	0	0	47	0	22	35	43	74	38	60	69	75
21	0	0	0	2	48	0	23	36	44	75	40	61	71	76
22	0	0	0	4	49	0	25	38	46	76	41	62	72	77
23	0	0	0	5	50	0	26	39	47	77	43	64	73	78
24	0	0	0	7	51	2	28	41	48	78	45	65	74	79
25	0	0	0	9	52	3	29	42	49	79	47	66	75	80
26	0	0	2	11	53	5	31	43	51	80	48	68	76	81
27	0	0	3	12	54	6	32	45	52	81	50	69	77	82
28	0	0	5	14	55	8	33	46	53	82	52	70	78	83
29	0	0	7	16	56	9	35	47	55	83	54	72	79	84
30	0	0	9	17	57	11	36	49	56	84	56	73	80	84
31	0	0	10	19	58	12	38	50	57	85	58	74	81	85
32	0	0	12	20	59	14	39	51	58	86	60	76	83	86
33	0	0	14	22	60	15	41	53	59	87	62	77	84	87
34	0	1	15	24	61	17	42	54	61	88	64	79	85	88
35	0	3	17	25	62	18	43	55	62	89	66	80	86	89
36	0	4	18	27	63	20	45	56	63	90	68	81	87	90
37	0	6	20	28	64	22	46	58	64	91	70	83	88	91
38	0	8	21	30	65	23	48	59	65	92	72	84	89	92
39	0	9	23	31	66	25	49	60	66	93	74	86	90	92
40	0	11	25	33	67	26	50	61	67	94	77	87	91	93
41	0	12	26	34	68	28	52	62	69	95	79	89	92	94
42	0	14	28	36	69	30	53	64	70	96	82	90	93	95
43	0	16	29	37	70	31	54	65	71	97	85	92	95	96
44	0	17	31	39	71	33	56	66	72	98	88	94	96	97
45	0	19	32	40	72	35	57	67	73	99	92	96	97	98
46	0	20	34	41	73	36	58	68	74	100	100	100	100	100

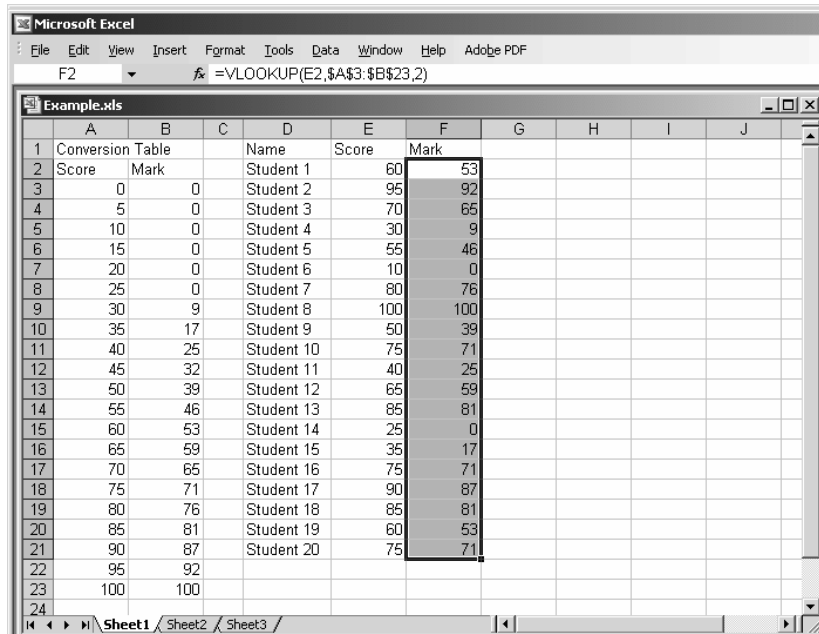


Figure 1 A screen snapshot showing the conversions of scores to marks by using Excel

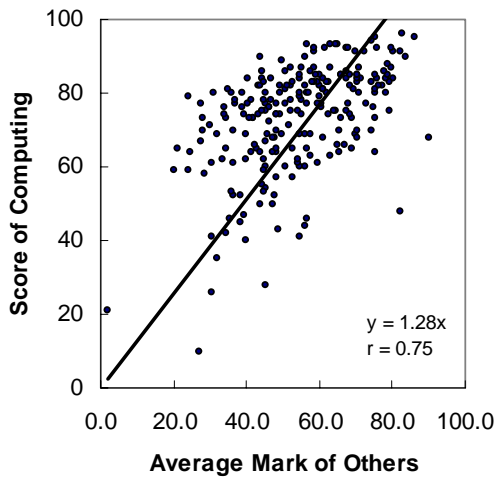


Figure 2 Relationship between the scores of *Computing* and the average marks of the other two modules.

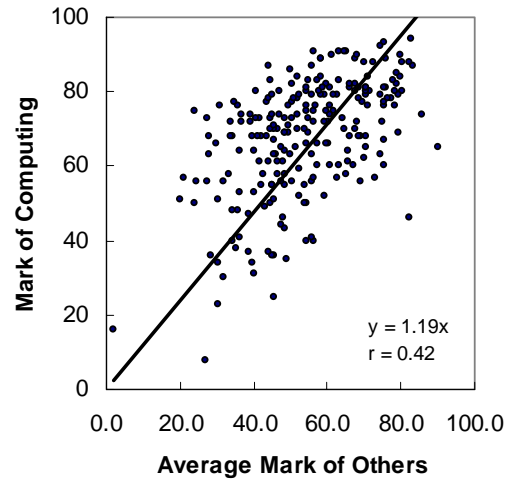


Figure 3 Relationship between the marks of *Computing* and the average marks of the other two modules.

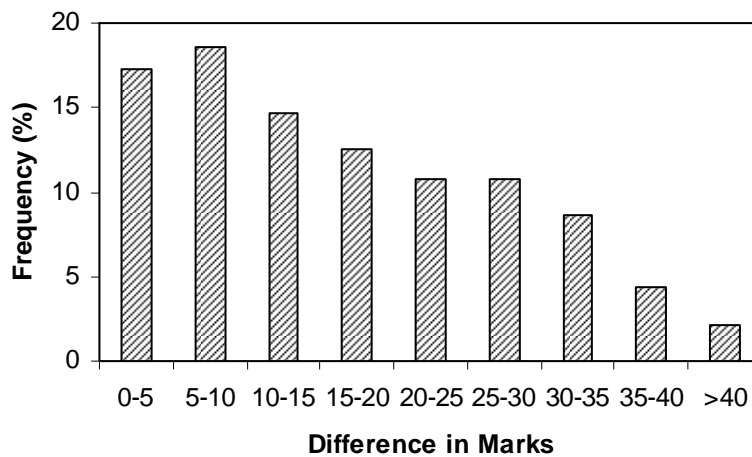


Figure 4 Histogram of the differences between the marks of *Computing* and the average marks of the other two modules taken in the same semester

pronounced for weak students. As a consequence, using scores directly would result in an increase in pass rate from 93.5% to 97.8%. The data reinforces the point that MCQ test scores must be converted to standard marks.

Figure 4 shows the histogram of the differences between the marks of *Computing* and the average marks of the other two modules. 36% of the students have a difference within 10 marks, 63% within 20 marks and 85% within 30 marks. It should be pointed out that the

correlation between *Computing* and the other modules in academic year 2005/06 is not as strong as that in academic year 2004/05 (1). The higher degree of scatter is mainly because of the fewer modules available for comparison in 2005/06.

Table 2 compares the average marks of the 5 individual MCQ tests in *Computing* between academic years 2004/05 and 2005/06. In 2004/05, there were big variations in the average marks among the 5 tests, which correctly reflected the relative difficulties of the 5 topics. In

2005/06, lectures and demonstrations were introduced to improve the students' learning of the difficult topics. The difficulties of the test questions were also adjusted. As a consequence, the overall module average was increased and the variations in the average marks among the 5 tests were significantly reduced.

Table 2 Average marks of individual MCQ tests in 2004/05 and 2005/06

Test	2004/05	2005/06
Word	82.1	79.3
Excel I	72.1	66.6
Excel II	60.8	73.5
MATLAB I	62.1	73.0
MATLAB II	49.9	54.5

## SUMMARY

Computer based MCQ tests were used as a summative assessment method for an engineering module over two academic years. The conversion scheme developed by Zhao (4) was applied to each of the 5

MCQ tests. The final module marks were obtained by averaging the marks of the 5 tests. There was a reasonable correlation between the students' marks of this module and their overall average marks of all the other concurrent modules. The changes in difficulties of topics and the improvements in teaching and assessment methods were evidenced in the changes in the average marks of the individual tests. The MCQ assessment and the conversion scheme were proved to be as reliable as the traditional assessment methods.

## REFERENCES

1. Zhao Y., 2006, *Int. J. Eng. Edu.*, 22, (in press)
2. McLachlan J.C. and Whiten S.C., 2000, *Med. Edu.*, 34, 788-797
3. Wainer H. and Wang X., 2001, *TOEFL Technical Report TR-16: Using a new statistical model for testlets to score TOEFL*, Education Testing Services, Princeton, USA
4. Zhao Y., 2005, *Int. J. Eng. Edu.*, 21, 1189-1194.