Correlation Between Test and Examination for Different Types of Tests – A Hong Kong Study

TO, W. T. D.,
WONG, L. T.,
MUI, K. W.
Department of Building Services Engineering,
The Hong Kong Polytechnic University

Abstract: Assessment is an important part of the teaching and learning process. Assessment allows the instructor to assess the understanding of the students on subject matters taught. Essay type questions with calculation components are reckoned to be an accurate tool of assessment and have become the typical examination and test format in many universities worldwide.

Multiple choice questions has the advantage that marking can be done very quickly and rapid feedback can be given to students. In the case of online multiple choice questions tests, the test result can be obtained immediately after the test. If it was found that the correlation between the multiple choice questions to that of examination is as good as the conventional essay type of questions, we can use it to assess the students with good confidence that it will accurately indicate the students' performance.

In a first year engineering subject at the department of Building Services Engineering, The Hong Kong Polytechnic University, the author has attempted different types of tests at different years over a 4 year's period. The test results are compared with the examination results for the respective years. The purpose of the comparison is to find out whether there is good correlation between the different types of tests to that of examinations.
Keywords: Examination, Test, Multiple choice, Students' performance

Introduction

Assessment is a major contributor to raise standards in universities in terms of teaching, learning and students' achievement. The quality of assessment has a significant impact on attitudes to learn, challenge students to work hard and encourage teachers to focus on how to improve the learning attitude of individual students (Office for Standards in Education, 2003). Assessments serve a series of functions, being used for certification, diagnosis, improvement of learning and teaching, accountability, evaluation and motivating students and teachers (Louis, Lawrence & Zeith, 2004). Taught programmes in universities are rationalized on an ongoing basis that subject examiners are responsible to review of subject syllabus, learning and teaching objectives on a regular basis. Externally, programme development is also facilitated through comments from the professional accreditation bodies. Active collaboration with the local profession, industry, oversea and local universities also ensure the programmes/subjects are cognizant with developments in education.

Department of Building Services Engineering (BSE) established at 1981 at The Hong Kong Polytechnic University is a major educational unit in Hong Kong and offers a number of academic programmes at sub-degree, degree and postgraduate levels to educate students in the discipline of building services engineering (BSE, 2004). Engineering subjects involve engineering design and there is much computations involved. The conventional assessment approach is essay type of questions with computations, result analysis and discussions. In the department of BSE, we have taken this approach for many years.

Although multiple choice (MC) or multiple selection (MS) type of assessment has been in place for many years, they were regarded as inappropriate for assessing engineering subjects because they only ask for possible answer(s) among a list of 4 or 5. It is impossible to assess an engineering design task which involves many steps and a myriad of design considerations. Of course it is possible to breakdown the steps in an engineering design and set MC questions for each part. However by nature the MC questions should be all independent of each other and should not have linkage between them, i.e. one would not expect that getting the right answer in one question is a pre-requisite for answering the next question. Therefore there is a limitation to use MC questions to assess engineering problems. However for engineering subjects at fundamental level, the subject matter involves simple calculations as a means of grasping the engineering concepts and theoretical principles. There are also a lot of engineering terminologies that students have to master. MC questions could be an appropriate approach to test the students' understanding.

One of the authors has adopted this line of thinking and has used the MC type of questions to assess the students for several years for a first year engineering subject. The test was administered to the students during a mid-term test. MC type of questions was introduced starting in year 2001. In year 2001, the test consists of 70% essay type of questions and 30% MC type of questions (sample A1). In year 2002, the test was administered online and there are altogether 44 MC/MS questions. MC questions have only one correct answer among a list of 4-5. MS questions have more than one correct answer among a list of 4-5 (sample B1). In year 2003, it is again 70% essay type and 30% MC type (sample A2). In year 2004, 44 MC type of questions were provided some of which are similar to those online MC/MS questions given in 2002 (sample B2).

For final examinations, the conventional approach using essay type of questions is taken. We have so far accumulated 4 years of results. It is rather interesting to see whether there are any correlations between the two sets of assessments for the same group of students. Although it involves 4 different cohorts of students, their entry standards indicate that there should not be marked differences among the 4 sample groups.

Association of assessment methods

The degree of association between the test and examination results can be presented by a contingency table of i by j cells and evaluated by Chi-square test with the null hypothesis of no association (Kokoska & Zwillinger, 2000). Assumptions related to sample size with contingency tables are based on the expected frequency. In addition to the requirement that each cell contain an expected frequency of at least 1, no more than 20% of the cells should contain expected frequency less than 5. A statistic Chi-square value χ_D^2 is given by where, n_{ij} and $\hat{\varepsilon}_{ij}$ are the observed count and the expected count at the ij-th cell; $n_{i.}$ and $n_{.j}$ are the i-th row total and j-th column total; n is the grand total of the table; I and J are the number of row and column, and D is the degree-of-freedom respectively,

$$\chi_D^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij}-\hat{\varepsilon}_{ij}\right)^2}{\hat{\varepsilon}_{ij}} \quad \text{… (1)}$$

$$\hat{\varepsilon}_{ij} = \frac{n_{i.}n_{.j}}{n} \quad \text{… (2)}$$

$$D = (I-1)(J-1) \quad \text{… (3)}$$

A significant association would be indicated by a p-value, which is arbitrarily set at $p \leq 0.05$ for this study. In order to account for small expected frequencies, a Yates' correction for continuity would be used to adjust the Chi-square values and it was reported that the application of the Yates' correction would increase the type II error (Kokoska & Zwillinger, 2000).

$$\chi_D^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(n_{ij}-\hat{\varepsilon}_{ij}-0.5\right)^2}{\hat{\varepsilon}_{ij}} \quad \text{…(4)}$$

The correlation between the test and examination results would be studied by rank correlation by replacing the results with their ranks. Ties were assigned the average for the ranks associated with the tied observations. The correlation between the ranks would be indicated by the Spearman rank correlation r_s, where d_i is the difference between test and examination ranks.

$$r_s = 1-\frac{6\sum d_i^2}{n^3-n} \quad \text{… (5)}$$

A significant correlation, t_{n-2} is indicated by a test statistic at a degree of freedom n–2,

$$t_{n-2} = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \quad \text{… (6)}$$

The correlation is good when the number of ties is small in comparison to grant total of the table, n. The rank correlation in each sample group with i, i+1, … n students can also be indicated by the Kendall's rank correlation coefficient, where, κ is the sum of $n(n-1)/2$ counts for; ζ is a dummy variable; ϕ_1 and ϕ_2 are the test and examination results respectively.

$$r_K = \frac{\kappa}{n(n-1)/2} \quad \text{… (7)}$$

$$\kappa = \sum_i \zeta_i \qquad \dots (8)$$

$$\zeta_i = \begin{cases} 1 & ;\lambda > 0 \\ 0 & ;\lambda = 0 \\ -1 & ;\lambda < 0 \end{cases} \qquad \dots (9)$$

$$\lambda = \left(\phi_{1,i+1} - \phi_{1,i}\right)\left(\phi_{2,i+1} - \phi_{2,i}\right) \qquad \dots (10)$$

For moderate to large n and few ties, an approximate standard normal test statistic, t is given by,

$$t = \frac{\kappa}{\sqrt{n(n-1)(2n+5)/18}} \qquad \dots (11)$$

When there are ties in the results ϕ_1 and ϕ_2, the variability is reduced. An adjusted asymptotic formula would be used (Hollander & Wolfe, 1999), where, g and h are the distinct values in ϕ_1 and ϕ_1 with ϕ_{1j} and ϕ_{2k} observations at the j-th and k-th tied value.

$$t = \frac{\kappa}{\sqrt{\sigma_\kappa^2}} \qquad \dots (12)$$

$$\sigma_\kappa^2 = \frac{n(n-1)(2n+5)}{18} - \sum_{j=1}^{g} \frac{\varphi_{1j}(\varphi_{1j}-1)(2\varphi_{1j}+5)}{18} - \sum_{k=1}^{h} \frac{\varphi_{2k}(\varphi_{2k}-1)(2\varphi_{2k}+5)}{18}$$
$$+ \frac{\left(\sum_{j=1}^{g} \varphi_{1j}(\varphi_{1j}-1)(\varphi_{1j}-2)\right)\left(\sum_{k=1}^{h} \varphi_{2k}(\varphi_{2k}-1)(\varphi_{2k}-2)\right)}{9n(n-1)(n-2)} + \frac{\left(\sum_{j=1}^{g} \varphi_{1j}(\varphi_{1j}-1)\right)\left(\sum_{k=1}^{h} \varphi_{2k}(\varphi_{2k}-1)\right)}{2n(n-1)} \qquad \dots (13)$$

For the investigated subject, the assessment is made up of 40% continuous assessment and 60% examinations. For the continuous assessment, there are two presentations each accounts for 10% of the total subject grade and a mid-term test which accounts for 20% of the total subject grade. The presentations, test and examination marks are all converted to grades before they are combined to obtain the overall subject grade. In the present study, we used only the test grade and examination grade for correlation study; and assumed the results of these 2 components were independent. The number of students in the class for different years and the nature of the mid-term tests are shown in Table 1 and a total of 429 data pairs were collected for analysis.

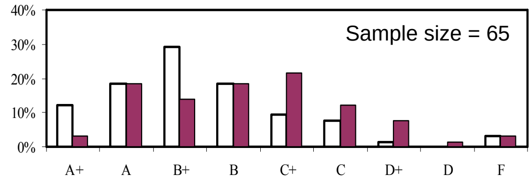
Table 1: *The detail of data pairs in this study*

Year	2001 sample A1	2002 sample B1	2003 sample A2	2004 sample B2
No. of student	65	65	72	68
Contribution in the mid-term test (Essay; MC)	(0.7; 0.3)	(0.0; 1.0)	(0.7; 0.3)	(0.0; 1.0)

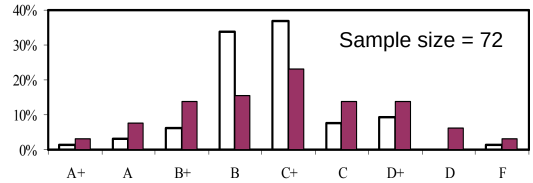
Results, analysis and discussions

Figure 1 shows the student results of the subject for four consecutive academic years (2001/02, 2002/03, 2003/04 and 2004/05) categorized into 9 ranks according to the University Grade-Point system, where, (A+) and (A) interpreted as 'Excellent' for Grade Point (GP) ranges 4.15-4.5 and 3.75-4.14; (B+) and (B) are 'Good' for GP ranges 3.25-3.74 and 2.75-3.24; (C+) and (C) are 'Satisfactory' for GP ranges 2.25-2.74 and 1.75-2.24; (D+) and (D) are 'Marginal' for GP ranges 1.25-1.74 and 0.60-1.24; and (F) is failure grade with

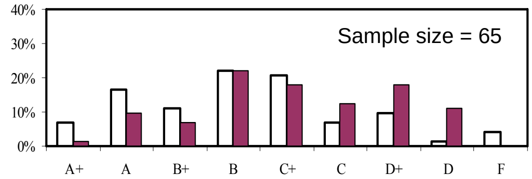
GP range from 0-0.59, respectively. A total of 65-72 students per year took the subjects and all student results were used for the correlation analysis.



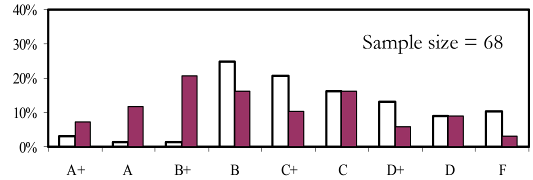
(a) Sample A1: Written test and examination



(b) Sample A2: Written test and examination



(c) Sample B1: Multiple choice test and written examination



(d) Sample B2: Multiple choice test and written examination

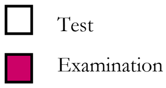


Figure 1: *Subject test and examination results*

Contingency table for subject results of sample groups A1 and A2 is shown in Table 2(A), and for sample groups B1 and B2 is shown in Table 2(B), with combining the data using the same assessment method of test. In order to satisfy the assumptions of the Chi-square test, the student grades were grouped and a 3 by 3 contingency table was presented for each assessment method of test. The classification of student according to the results are, (1) good students, who awarded B+ or above; (2) average students, who awarded C+ or B; and (3) less good students, who award C or below.

A test for association between the row and column categories in Table 2 was conducted. The results showed that a significant evidence that an association between test and examination results for (A) that $\chi^2 = 20.2$, D = 4, p = 0.0005 and (B) that $\chi^2 = 57$, D = 4, $p = 1\times10^{-11}$, respectively, whatever the testing method was adopted. It is noted that the results were pooled into 3 categories and the result accuracy would be improved by increasing the sample size good for a 9 by 9 contingency table.

The association between the test and examination results was further evaluated by Spearman's rank correlation coefficients and Kendall's rank correlation coefficient. Table 3 summarized the results of the rank correlation for sample groups (A) and (B) in year basis. It is reported that the results for both assessments were significantly correlated with the examination results (P<0.05).

Table 2: *Assessment results*

(A) Written test and examination (Sample size = 137)

Test	Examination			Row Total
	B+ or above	C+ or B	C or below	
B+ or above	26	23	15	64
C+ or B	8	25	16	49

C or below	2	7	15	24
Column total	36	55	46	137

(B) Multiple-choice test and examination (Sample size = 133)

Test	Examination			Row Total
	B+ or above	C+ or B	C or below	
B+ or above	32	8	10	50
C+ or B	17	5	16	38
C or below	15	9	21	45
Column total	64	22	47	133

Table 3: *Rank correlation*

Sample size, n	Spearman's rank correlation		Kendall's rank correlation	
	r_s	P-value	r_K	P-value
Written test and examination				
65 (A1)	0.3306	0.0072	0.2637	0.0019
72 (A2)	0.3583	0.0020	0.2782	0.0005
Multiple-choice test and examination				
65 (B1)	0.2841	0.0218	0.2201	0.0095
68 (B2)	0.3900	0.0010	0.3124	0.0002

Conclusions

Assessment is taking place all the time because making judgment is something that we all do about ourselves and the others. The assessment methods of an engineering subject as an important part of the teaching and learning process are evaluated in the perspective of the correlation to a student's examination result. An indicative assessment allows the instructor to assess the understanding of the students on subject matters taught, before the final examination. This study discussed the rationale for students studying Building Services Engineering at The Hong Kong Polytechnic University to use 2 different forms, namely essay type written test and multiple-choice/multi-selection test as a mean of assessment. The assessed student results of these components were compared with those obtained in the final examination. In particular, subject assessment results in academic years 2001/02 to 2004/05 were used to study the correlations between the test result and the examination result for 2 groups of students, assessed by a multiple-choice/multiple-selection assessment or assessed by an essay-type written test. The results showed that the examination results correlated with the test results for all groups of students. Similar correlations were found for both assessment methods for test. Multiple choice questions has the advantage that marking can be done very quickly and rapid feedback can be given to students. With the increasingly application of online assessment adopted in university education, proper use of multiple-choice question would be an effective tool for assessment in terms of marking time and instant feedback to student, in supplement to the essay type written assessment.

Acknowledgement

This work is undertaken in the Department of Building Services Engineering, The Hong Kong Polytechnic University.

References

Office for Standards in Education. (2003). *Good assessment in secondary schools*. London: OFSTED.

Louis, C., Lawrence, M., & Keith, M. (2004). *A Guide to Teaching Practice 5th Edition.* Routledge Falmer, ISBN 0-415-30675-2.

BSE, HKPU., (2004). *Programme documents, BEng(Hons) in Building Services Engineering (Full-time or part-time; government funded or self-financed),* Higher diploma in Building Services Engineering, Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong.

Kokoska, S. & Zwillinger, D. (2000). *Standard probability and statistics tables and formulae.* New York: Chapman and Hall/CRC.

Hollander, M., & Wolfe, D.A. (1999). *Nonparametric statistical methods.* 2nd Ed. New York: Wiley.