Can learning outcomes in cognitive domain be assessed effectively using multiple choice questions? A study in an undergraduate pharmacy curriculum

Hui Meng Er, Srinivasan Ramamurthy, Peter CK Pook

Abstract

Background: The widespread use of multiple choice questions (MCQ) in examinations is attributed to its logistical advantage and broad coverage of content within a short duration. The end-of-semester examinations for several modules in the pharmacy programme previously employed a combination of written examination tools including MCQ, short answer questions (SAQ) or essays for assessing learning outcomes in the cognitive domain. Concerns regarding assessment fatigue and subjectivity in marking have led to a review of the assessment formats in the examinations. Various types of MCQ were consequently introduced as the only assessment tool. This study was conducted to evaluate the performance of students in the examinations as a result of the change.

Methodology: Analyses were carried out on the end-of-semester examination results of two cohorts of students for each module, one based on a combination of MCQ, SAQ or essay and the other based on MCQ alone. The class means were compared, and t-test was used to determine the difference between the performances.

Results: Although the difference in the mean scores of the two groups is statistically significant in 13 of the 20 modules, the difference is less than 5% in 10 modules.

Conclusion: The findings provide evidence that well-constructed MCQ can effectively assess cognitive skills.

IeJSME 2014 8(3): 9-18

Keywords: Cognitive; learning outcomes; multiple choice questions; pharmacy; written examination.

Introduction

Educational objectives are divided into three domains according to the Bloom's taxonomy: cognitive (knowledge), affective (attitude) and psychomotor (skills) [1]. The cognitive domain can be measured at different levels, which are knowledge, comprehension, application, analysis, synthesis and evaluation, using common written assessment formats such as multiple choice questions (MCQ), short answer questions (SAQ) and essay. While short answer questions and essay questions are useful for testing higher order thinking including the ability of students to organise their ideas, the amount of materials assessed using these formats can be restrictive due to time constraints. In addition, marking can be subjective despite the use of marking rubrics. On the contrary, MCQ testing is a reliable and valid written assessment for testing of knowledge and it has been shown to correlate well with overall competence and performance [2]. Well-constructed MCQs can be used to assess higher-order cognitive skills such as interpretation, synthesis and application of knowledge [3]. Widespread use of MCQs in examinations is attributed to its logistical advantage and testing efficiency such as broad coverage of content within a short time. However, the drawbacks include unfairness, promotion of regurgitation and lack of professional authenticity [4,5].

The Bachelor of Pharmacy programme in International Medical University (IMU) is a four-year (8 semesters) programme. The curriculum is outcome-based, focusing on producing graduates who are knowledgeable, competent, professional, ethical, empathic, able to work effectively with other healthcare professionals, life-long learners, committed to continuing professional development as well as evidence-based practitioners. The learning outcomes of the modules and assessment activities are aligned to the outcomes of the three phases, i.e. Phase 1 (Semesters 1 and 2), Phase 2 (Semesters 3 to 5) and Phase 3 (Semesters 6 to 8). The outcomes of the phases are shown in Table I. The Phase 1 and Phase 2 outcomes define the desired student attributes and levels of competence, while the Phase 3 outcomes define the desired graduate attributes and competence. The cognitive skills in Phase 1 and Phase 2 are mainly assessed in written examinations. In Phase 3, the students are increasingly assessed on their psychomotor and affective skills in addition to cognitive skills,

School of Pharmacy, Faculty of Medicine and Health, International Medical University, MALAYSIA

Address for Correspondence:

Assoc. Prof. Er Hui Meng, International Medical University, No 126 Jalan Jalil Perkasa 19, Bukit Jalil, 57000 Kuala Lumpur, MALAYSIA

Email: huimeng_er@imu.edu.my Tel: 603-2731 7223 Fax: 603-8656 7228

using a combination of written examinations, clinical examinations and work-based assessments.

Subsequent to an assessment review in the programme, the formats of the end-of-semester written examinations for selected modules in Phase 1 and Phase 2 of the programme have been changed from a combination of MCQ, MEQ, SAQ and/or essay to solely MCQ. In this study, the students' performances in the written examinations before and after the assessment review were compared. The objective was to determine whether MCQ alone is an adequate and appropriate tool for assessing knowledge in Phase 1 and Phase 2 of the programme. The findings of the study would also contribute to the continuous quality improvement exercise in enhancing and strengthening the outcome-based curriculum.

Methods

The modules in Semesters 1 to 5 in which the learning outcomes are primarily cognitive were selected for analysis in this retrospective study. The cohorts undertaking the examinations comprising of a combination of MCQ, MEQ, SAQ and/or essay were assigned as Group A, whereas the cohorts undertaking the examination comprising of MCQ as the only tool were assigned as Group B.

In the examinations using solely MCQ, various types of MCQs were used to test the knowledge recall, application, reasoning, critical thinking and problem solving skills using A-type, B-type, E-type, K-type [3]. Negative marking was not practised. Prior to the implementation, faculty training was conducted on constructing the various types of MCQs. Each MCQ contained a question stem followed by five options, one of which being the correct answer. The A-type items were the one-best-option items, which were further categorised into the direct factual recall items (identified as Type A1) and problem solving items (identified as Type A2). The B-type items were the matching items consisting of a list of options followed by several numbered items, whereby each option could be selected once, more than once or not at all. The E-type items were the two-part true/false items that involved analysis of relationships, with the second part being the reason for the first part. It enabled evaluation of reasoning skills and understanding of the basic principles. The K-type items consisted of a stem followed by four options, one or more could be correct.

The difficulty index of each MCQ item in the examinations undertaken by Group B was calculated using the following formula:

Difficulty index = Number of students who answer the item correctly / Total number of students

The percentage of questions in each examination paper having difficulty indices (DI) of 0.70 or less, between 0.3 and 0.70, and less than 0.3 are determined. The correlations of these percentages with the class means of the modules are established. Besides, the mean difficulty index for each type of MCQ was computed for both Phase 1 and Phase 2 modules.

The class mean scores for Groups A and B were compared, and independent-samples T-test was used to determine the significant difference between the performances of the two groups of students. All analyses were carried out using PASW version 18.0 (SPSS, USA).

Results

The p-value for the T-test and the class mean scores of Group A and Group B for each module are presented in Tables IIa and IIb. The difference is considered significant when the p-value is less than 0.05 (indicated with an asterisk in the table).

Of the twenty modules selected in this study, thirteen modules show statistically significant difference in the scores of the two groups. Among these, the scores of the Group A are lower than those of Group B by more than 5%

in ten modules (Human Biology, Cardiovascular System, General Pharmacology, Haematology, Natural Products, Pharmaceutical Analysis I, Renal System, Respiratory System, Complementary Medicine and Community Pharmacy). On the other hand, the scores of the two groups differ by less than 5% in three other modules (Biological Science, Pharmaceutics I and Immunology). While the Pharmaceutics I and Immunology modules show slightly lower score for Group A, the Biological Science module shows slightly higher score for Group A. The scores of the two groups are not significantly different in the remaining 7 modules and the differences are less than 5%.

The percentage of MCQ items having difficulty indices of 0.70 or less and that between 0.30 and 0.70 for each module in Group B are presented in Table III. The graph of the mean scores of the modules versus the percentage of MCQs having difficulty indices of 0.70 or less is shown in Figure I. The Pearson correlation value is determined to be -0.941, which is significant at the 0.01 level. Meanwhile, the Pearson correlation value between the mean scores of the modules and the percentage of MCQs having difficulty indices between 0.30 and 0.70 is -0.821, which is also significant at the 0.01 level (Figure II).

The mean difficulty indices for the various types of MCQ are presented in Figure III. The mean difficulty indices for A1-type, A2-type, B-type, E-type and K-type are 0.76, 0.76, 0.87, 0.60 and 0.72 for Phase 1 respectively; and 0.78, 0.72, 0.72, 0.65 and 0.67 for Phase 2 respectively.

Discussion

The study has shown that the mean scores of the two groups are significantly different in thirteen modules. Among them, the mean scores of three of the modules differ by less than 5%. Meanwhile, the mean scores of the two groups are not significantly different (less than 2.5%) in seven other modules. The findings provide evidence to support that MCQ alone can be effective for assessing cognitive skills in summative assessment instituted in a formal setting. Nevertheless, the effectiveness depends on the construction of the MCQs that ensures higher order cognitive skills are assessed.

Further analysis indicates that there is a significant negative correlation between the mean scores and the percentage of MCQ items having difficulty indices of 0.70 or less in the modules. The difficulty of a question increases with decreasing difficulty index. The ideal range of difficulty index is between 0.30 and 0.70, and a difficulty index of less than 0.30 indicates a difficult question [6]. It was observed that the percentage of MCQ items having difficulty indices of 0.70 or less for those modules that did not show significant difference in the mean scores of Groups A and B ranged between 25 to 53%, whereas that for the modules that showed significant difference in mean scores of more than 5% between the two groups ranged between 12.5 to 55%. This indicates that although the scores are affected by the difficulty of the questions, other factors may also play a role. In the modules that show significant differences in the mean scores of the two groups, examinations using MCQ as the only tool have higher mean scores. It is likely that MCQs help students to avoid losing marks due to grammatical errors and poor writing skills [7]. Students could have guessed the correct answers in some instances [8]. Furthermore, the lack of preparedness of the students in answering questions of SAQ, MEQ and essay formats could have also contributed to this finding, as most of the formative assessments in these modules are of MCQ format only. As a principle of good assessment practice, students must be given guidance on the formats used in summative assessments, in particular the assessment criteria and marking schemes. This should also be accompanied by good feedback to improve learning [9].

Among the various types of MCQ, the mean difficulty indices for the E-type items were the lowest (i.e. most challenging to the students) in both Phase 1 and Phase 2. This was not unexpected as the E-type items required

good reasoning skills and understanding of the basic principles. The B-type items were commonly used in integrated body system modules. In Phase 1, the B-type items were relatively easy for the students (highest difficulty index among the different types of MCQ), but they became more challenging in Phase 2. This could probably be attributed to the increase in the complexity of the body systems studied in Phase 2. The A1-type (factual recall) items increased slightly in difficulty index progressing from Phase 1 to Phase 2. However, the A2-type (problem solving) and K-type items decreased in difficulty index from Phase 1 to Phase 2. This indicated that in-depth knowledge and understanding were required in answering A2 and K-types MCQ in Phase 2.

The decision on which assessment method to use is determined by its validity, reliability, educational effect, feasibility and acceptability to learners and faculty [5]. Besides the benefit to teaching and learning, the cost to the individual trainee, institution and society at large should be taken into account when considering the feasibility of an assessment [10]. Each assessment method has its own strengths and weaknesses. Multiple choice questions are useful for a large scale assessment (ie. a large number of students). Evidence has shown that MCQs can be used to assess higher cognitive learning successfully, as effective as essay questions [11]. They also allow evaluation of a large domain of knowledge in a relatively short time and can be graded objectively [12]. This is an important consideration for addressing the issue of assessment fatigue often faced by students, which in turn will have an impact on the effectiveness of summative assessment as a genuine measure of students' ability and achievement of learning outcomes.

Conclusion

This study showed that MCQs alone can be an effective tool for assessing knowledge in an undergraduate pharmacy curriculum. However, appropriately constructed MCQs are crucial for an examination that is designed for testing higher order thinking in the cognitive knowledge domain. Continuous faculty training is important to ensure that high quality MCQs are generated to appropriately assess the desired learning outcomes. Item analysis of MCQs is also essential to ensure the reliability and validity of the test items. On the other hand, a variety of assessment tools are still necessary for a comprehensive assessment of the other learning outcomes in an undergraduate pharmacy curriculum. These can be evaluated using work-based, clinical or practical skill assessment as well as other formats for assessing writing skills during the course. In any case, guidance must be given to students on the assessment criteria and marking schemes in formative assessments to ensure the effectiveness and fairness of examinations.

REFERENCES

1. Bloom BS. Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc, 1956.
2. McCoubrie P. Improving the fairness of multiple-choice question: a literature review. Medical Teacher 200; 26(8): 709-712.
3. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. 3rd ed. Philadelphia: National Board of Medical Examiners, 2001.
4. Kaufman DM. Assessing medical students: hit or miss. Student BMJ 2001; 9: 87-88.
5. Van der Vleuten C. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ 1996; 1: 41-67.
6. Miller MD, Linn RL. Gronlund, N.E. Measurement and assessment in teaching. 10th ed. New Jersey: Prentice Hall, 2009.
7. Zeidner M. Essay versus multiple-choice type classroom exams: The student's perspective. Journal of Educational Research 1987; 80(6): 352-58.
8. Kennedy PE, Walstad WB. Combining multiple-choice and constructed-response test scores: An economist's view. Applied Measurement in Education 1997; 10: 359-75.
9. Meiers M. Feedback improves learning. EQ Australia 2005; Jan 2005: 20-22.
10. Shumway JM, Harden RM. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. Medical Teacher 2003; 25(6): 569-84.
11. Palmer E, Devitt P. Constructing multiple choice questions as a method for learning. Ann Acad Med Singap 2006; 35: 604-8.
12. Tiemeier AM, Stacy ZA, Burke JM. Using multiple choice questions written at various Bloom;s taxonomy levels to evaluate student performance across a therapeutics sequence. Innovations in Pharmacy 2011; 2(2): 1-11.

Table I. Phase outcomes in the BPharm curriculum of the International Medical University.

Phase	Outcomes
1 (Semesters 1 & 2)	The student demonstrates • an essential grasp of core knowledge in the sciences underpinning pharmacy; • competence in basic areas of practice; • appropriate attitudes towards ethics and professional responsibilities.
2 (Semesters 3, 4 & 5)	The student demonstrates • essential understanding and the ability to apply core knowledge in specified areas of pharmaceutical science; • he ability to think critically and to solve problems; • the ability to reflect upon the learning process; • the ability to work effectively, compassionately and ethically within a team.
3 (Semesters 6, 7 & 8)	The graduate demonstrates • the ethos of scientific research; • basic proficiency in the provision of pharmaceutical services; • a sense of empathy with patients, clients and healthcare professionals; • sound ethical values; • the skills required for life-long learning, evidence-based practice and continuing professional development: reflective learning, self appraisal and audit.

Table IIa. The mean scores and T-test results of the written examinations for Group A and Group B of each module (semesters 1 and 2).

Semester	Modules	Group	Cohort	Number of students (N)	Format of written examinations (number of questions and duration are shown)	Mean score (%)	p-values
1	Biological Science	A	A-S12	145	40 MCQ, 3 SAQ (2 hrs)	75.5	0.002*
		B	B-S12	165	40 MCQ (1 hr)	72.1	
	General Chemistry	A	A-S12	145	40 MCQ, 3 SAQ (3 hrs)	74.9	0.33
		B	B-S12	166	40 MCQ (2 hrs)	76.4	
	Genetics	A	A-S12	145	40 MCQ, 3 SAQ (2 hrs)	75.3	0.77
		B	B-S12	166	40 MCQ (1 hr)	75.7	
	Human Biology	A	A-S12	146	20 MCQ, 4 SAQ (2 hrs)	62.4	0.000*
		B	B-S12	166	40 MCQ (1 hr)	74.7	
2	Cardiovascular System	A	A-S12	138	40 MCQ, 4 SAQ (3 hrs)	71.0	0.000*
		B	B-S12	164	40 MCQ (1 hr)	78.4	
	General Pharmacology	A	A-S12	137	40 MCQ, 2 Essay (2 hrs)	74.6	0.000*
		B	B-S12	164	40 MCQ (1 hr)	85.2	
	Haematology	A	A-S12	137	40 MCQ, 3 SAQ (2 hrs)	63.5	0.000*
		B	B-S12	164	40 MCQ (1 hr)	75.1	
	Pharmaceutics I	A	A-S12	138	40 MCQ, 3 SAQ (2 hrs)	73.3	0.01*
		B	B-S12	164	40 MCQ (1 hr)	75.4	
	Physical Pharmacy	A	A-S12	138	20 MCQ, 4 SAQ (3 hrs)	68.1	0.71
		B	B-S12	165	50 MCQ (2 hrs)	68.6	

* The difference in the mean scores between Groups A and B is statistically significant.

Table IIb. The mean scores and T-test results of the written examinations for Group A and Group B of each module (semesters 3 to 5).

Semester	Modules	Group	Cohort	Number of students (N)	Format of written examinations (number of questions and duration are shown)	Mean score (%)	p-values
3	Natural Products	A	A-S34	125	10 MCQ, 3 SAQ (2 hrs)	55.1	0.000*
		B	B-S34	138	30 MCQ (1 hr)	75.4	
	Pharm Analysis I	A	A-S34	125	20 MCQ, 2 SAQ (2 hrs)	59.7	0.000*
		B	B-S34	138	30 MCQ (1 hr)	73.1	
	Pharmaceutics II	A	A-S34	125	30 MCQ, 2 SAQ (2 hrs)	70.0	0.06
		B	B-S34	138	40 MCQ (1 hr)	71.3	
	Renal System	A	A-S34	125	2 MEQ, 1 Essay (2 hrs)	61.7	0.000*
		B	B-S34	138	40 MCQ (1 hr)	68.2	
	Respiratory System	A	A-S34	125	2 MEQ, 1 Essay (2 hrs)	56.6	0.000*
		B	B-S34	138	40 MCQ (1 hr)	64.3	
4	Immunology	A	A-S34	123	2 MEQ, 1 Essay (2 hrs)	65.2	0.002*
		B	B-S34	138	40 MCQ (1 hr)	69.0	
	Medicinal Chemistry	A	A-S34	123	20 MCQ, 2 SAQ (2 hrs)	66.5	0.24
		B	B-S34	139	30 MCQ (1 hr)	64.7	
	Pharmaceutical Biotechnology	A	A-S34	123	4 SAQ (2 hrs)	65.9	0.11
		B	B-S34	138	40 MCQ (1 hr)	68.4	
	Pharm Analysis II	A	A-S34	123	20 MCQ, 2 SAQ (2 hrs)	67.0	0.91
		B	B-S34	139	30 MCQ (1 hr)	66.8	
5	Community Pharmacy	A	A-S5	123	4 SAQ (2 hrs)	67.6	0.000*
		B	B-S5	119	40 MCQ (1 hr)	80.1	
	Complementary Medicine	A	A-S5	123	10 MCQ, 3 SAQ (2 hrs)	75.5	0.000*
		B	B-S5	119	40 MCQ (1 hr)	81.0	

* The difference in the mean scores between Groups A and B is statistically significant.

Table III. The percentage of MCQ items with difficulty indices of 0.70 or less, between 0.30 and 0.70 and less than 0.30 in the examinations undertaken by Group B.

Semester	Modules	Percentage of MCQ items with difficulty indices of 0.70 or less	Percentage of MCQ items with difficulty indices of 0.30-0.70	Percentage of MCQ items with difficulty indices of less than 0.30
1	Biological Science	40%	35%	5%
	General Chemistry	37.5%	37.5%	0%
	Genetics	25%	20%	5%
	Human Biology	27.5%	22.5%	5%
2	Cardiovascular System	27.5%	25%	2.5%
	General Pharmacology	12.5%	7.5%	5%
	Haematology	33.3%	28.2%	5.1%
	Pharmaceutics I	35%	30%	5%
	Physical Pharmacy	49%	38.8%	10.2%
3	Natural Products	40%	40%	0%
	Pharm Analysis I	30%	23.3%	6.7%
	Pharmaceutics II	42.5%	35%	7.5%
	Renal System	45%	40%	5%
	Respiratory System	55%	45%	10%
4	Immunology	42.5%	32.5%	10%
	Medicinal Chemistry	53.3%	43.3%	10%
	Pharmaceutical Biotechnology	42.5%	27.5%	15%
	Pharm Analysis II	50%	36.7%	13.3%
5	Community Pharmacy	22.5%	17.5%	5%
	Complementary Medicine	17.5%	12.5%	5%

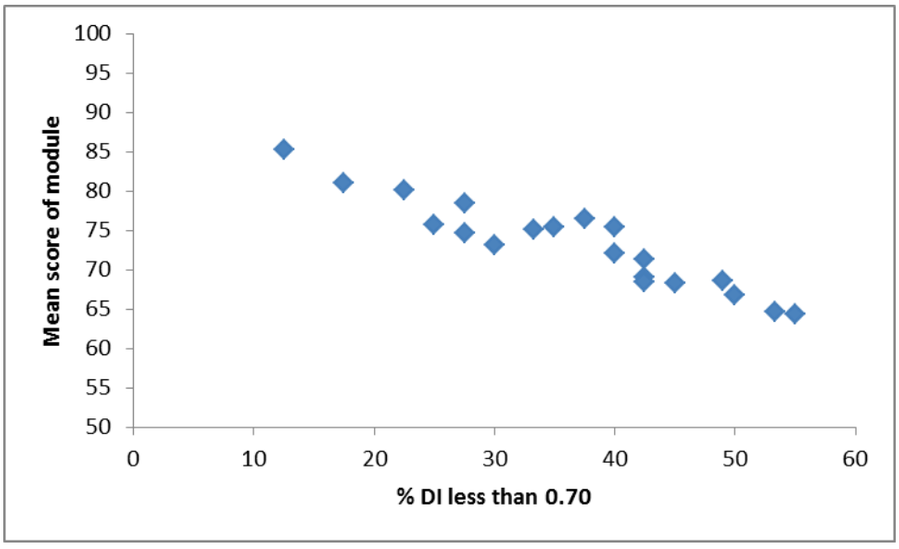

Figure I. Mean score of module versus percentage of MCQs with difficulty indices (DI) of 0.70 or less.

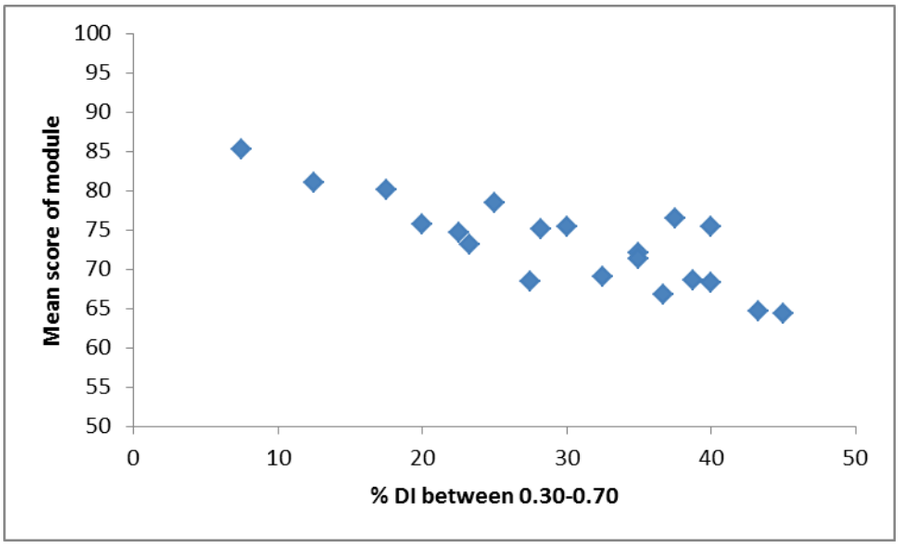

Figure II. Mean score of module versus percentage of MCQs with difficulty indices (DI) between 0.30 and 0.70.

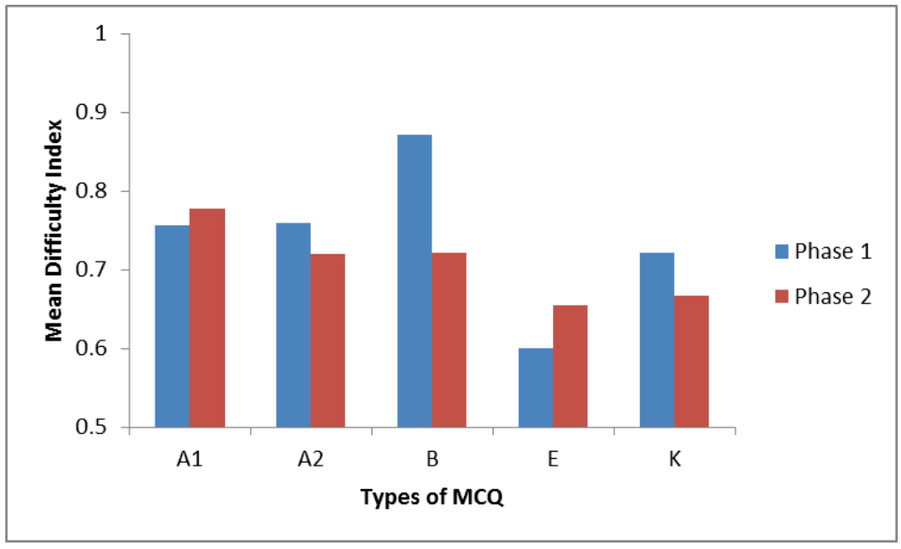

Figure III. Mean difficulty indices for various types of MCQ in Phase 1 and Phase 2 modules.