An Analysis of Low Performing Courses in the Final Examination

Roziah Mohd Janor, Ahmad Mazli Muhammad, Muhammad Nadzri Mohd Yusoff, Rozzana Mohd Said and Sharipah Ruzaina Syed Aris

Curriculum Affairs Unit, Academic Affairs Division, Universiti Teknologi MARA (UiTM)

Corresponding author: a.mazli@salam.uitm.edu.my

Abstract

Every semester, the final examination results of several courses may record a high percentage of failure. Several contributing factors which may cause this phenomenon include student attributes, test item construction and teaching approach. The purpose of this study is to look at test item construction variables which may lead to low performance. As the Key Performance Indicator (KPI), the university (UiTM) has stipulated that all courses should attain below 25% failure rate. Based on this criterion, courses which registered 25% and above failure rate, were selected. Utilising a specially designed tool, the final examination papers were analysed based on two indicators: (1) mapping of examination papers with course learning outcomes and (2) construction of quality items which are valid, reliable, fair and consistent. The analysis revealed several disparities that showed non alignment between the final examination papers and the curriculum documents. In addition, there were several quality issues found in terms of language, information, relevance, context and presentation. This study suggests that Test Specification Tables must be referred to ensure and regulate quality of final examination papers.

Keywords:

Constructive alignment, Test item construction, Quality item, Validity, Reliability, Fairness, Consistency

Introduction

Assessment is a part of a curriculum that must comply with the standards described by the Malaysian Qualification Agency (MQA). It is specifically mentioned in Area 3 of the Code of Practice for Programme Accreditation (COPPA). Assessment is a tool of measurement to determine achievement of intended learning outcomes. In the design of a curriculum, intended learning outcomes are first to be determined followed by assessment method and later followed by method of delivery. Therefore, all these elements have to constructively align to each other (Biggs, 2003). Assessment itself is one of the most difficult aspects of teaching. The quality of an assessment depends on four factors: validity, reliability, fairness and consistency (Roziah Mohd Janor et. al, 2012).

Being the largest university in Malaysia, with more than 180 000 students and more than 300 programmes offered at multiple locations, Universiti Teknologi MARA (UiTM), face the problem of ensuring quality of assessment. UiTM handles approximately five thousand (5,000) in number of final examination papers every end of semester. About three thousand (3,000) papers are contributed by degree level programmes while about two thousand (2,000) courses come from diploma level programmes. This presents an immensely challenging task to the institution in ensuring that all final examination papers are validly, reliably, fairly and consistently constructed.

Course failure rate of below 25% is one of the Key Performance Indicators of UiTM academic performances. A report released by the University Board of Examination that shows for the past 3 consecutive semesters, several courses have recorded 25% or more failure rate. Thus, this warrants the university to embark on an investigation.

The primary objective of this study is to identify the root causes of the poor performance in several identified courses in UiTM. This study addresses the following research questions: (1) Do the examination questions map the course learning outcomes? and (2) Are the items constructed valid, reliable, fair and consistent? These research questions in turn address the constructive alignment in each of the courses analysed in this study.

Performance in summative assessments provides evidence that shows the attainment or mastery of skills as described in the course learning outcomes. Hence, to ensure quality assessment, it is imperative that when constructing test items, five (5) basic elements should be observed: Language, Information, Relevance, Context and Presentation (Berkel and Bax, 2006).

The Language criterion constitutes the following variables. Firstly, the scenarios, questions or items are written in correct grammar. This allows students to understand the constructed questions clearly. Secondly, the questions are constructed using simple sentences that are easy to comprehend. Thirdly, questions do not use double negatives. These will only confuse students. Fourthly, items must be concise. Hence, it does not allow multiple interpretations. Fifthly, questions should be as neutral as possible with no inclinations, bias or negative perceptions. Sixthly, words used to construct questions are not vague that may lead to misunderstanding. Finally, items must be focused so that the meaning is mutually understood by the students.

The Information criterion addresses the following variables. Firstly, questions should be formulated with sufficient information so that students can answer them correctly. Secondly, clear instructions on the length and form of answer should be explicitly indicated. Thirdly, all questions should clearly indicate whether the answer must be explained, described, or illustrated. Finally, the questions should clearly separate the scenario from the instruction so that students are guided clearly to the problem at hand.

The Relevance criterion focuses on the validity of the questions. Firstly, the question must clearly indicate the topic being assessed. This will direct the students to the respective knowledge taught earlier in class. Secondly, questions should not be answered by students

using their general prior knowledge. Constructed questions should only be answered using knowledge or skill assessed in the question. Thirdly, the questions should not give hints of the intended answer. Fourthly, items must be constructed according to the level of difficulty as stipulated in the Test Specification Table. Fifthly, the questions must not include additional information that may make the questions more difficult. Finally all constructed items fulfill the objective of the course. Hence the relevancy of all items is deemed valid.

The Context criterion encompasses three (3) variables. First, the context i.e., drawings, illustrations, texts and images are used effectively. Secondly, these contexts should be presented clearly and correctly. Finally, questions should not include information which is not relevant.

The Presentation criterion includes four (4) variables. Firstly, questions and sub questions can be well identified. Secondly, the numbering should be systematic, logical and clear, and according to the format stipulated by the Exam Department. Thirdly, correct use of writing conventions (such as spellings, punctuations and symbols) is observed. Finally, correct information regarding the sources used as references in the questions must be clearly written.

This study focuses solely on the analysis of open-ended questions. Hence the construction of quality items described above only considers the construction of open-ended test items in the final examination papers.

Method

This qualitative research is an exploratory study to look for possible reasons and explanations on what causes high percentage of failures (than a specified target) in course(s) at university level. Hence, it studies the constructive alignment of test items to course outcomes and also the quality of final examination papers.

The population for this study is the performance of all courses offered in both diploma and degree programmes in UiTM for the past three academic semesters i.e., from January 2012 to March 2013. Hence, this is a purposive sampling. For each level, the performance of 3 different cohorts was studied. For the degree level, the performance for three consecutive final examinations i.e., January 2012, June 2012 and January 2013 was observed while for the diploma level, the performance for the 3 consecutive semesters i.e., March 2012, October 2012 and March 2013 was analysed.

For all courses, from all programmes, their performances were screened and the top forty (40) courses with highest percentage of failures were identified. A course with 25% failure and recorded on all the 3 cohorts were then extracted. For this study, 14 courses were identified: 5 courses were from the degree level and 9 were from the diploma level. The courses identified come from different fields that include science and technology based, social science based and business management based.

For each of those identified courses, its syllabus and 3 consecutive final examination papers were analysed for alignment to its documented course outcomes. The alignment of each final examination question to its course outcome(s) was also evaluated using a specially developed tool. This tool comprehensively looked at the alignment of items from the Bloom's taxonomy level point of view, the time taken to answer the question, the validity of the item and the marks allocated to the item. In addition, this study also analysed the quality of the examination papers using a modified protocol originally developed by Berkel and Bax (2006). For confidentiality, the 14 courses were coded as alphabets from A to N.

Results

The analyses were reported in two parts: (1) constructive alignment of test items to the course outcomes, and (2) quality of the final examination papers. The analysis revealed several disparities that showed non alignment between the examination papers and the curriculum documents. In addition, there were several quality issues found in terms of language, information, relevance, context and presentation.

Mapping of examination questions to course learning outcomes (CLO)

In the analysis, an examination paper is considered aligned to the CLO if the highest Bloom's taxonomy level of the examination paper identified is the same as that of the CLO documented in the course syllabus.

Table 1.0 Highest Cognitive Level in CLO vs Highest Cognitive Level in Each Paper

Program level	Course	Highest Cognitive level in CLO	Highest cognitive level identified in each cohort of examination papers		
			1	2	3
Degree	A	4	3	5	5
	B	6	5	5	5
	C	2	5	5	5
	D	4	3	2	3
	E	4	4	4	4
Diploma	F	5	5	5	5
	G	3	4	4	4
	H	4	3	3	3
	I	3	3	3	3
	J	3	5	5	5
	K	3	3	3	3
	L	3	5	5	5
	M	4	5	5	4
	N	5	5	5	5

Table 1.0 shows the highest cognitive level recorded by each course learning outcome against the respective final examination paper. The total number of final examination papers analysed was 42. Out of this number, sixteen papers were found not aligned to the course learning outcomes. The items constructed in the 16 papers were found to be of higher cognitive level than the intended level stipulated in the course learning outcomes. Due to this non-alignment, those papers were deemed unfair and not valid in terms of level of difficulty.

The remaining examination papers were found to be aligned to the course outcomes. However, those papers were found to be inconsistent for the three consecutive semesters. In addition, for course F, offered in the first semester of year 1 (diploma level), carries the highest cognitive level of C5 in the course learning outcomes instead of an appropriate level of C4. Furthermore, some papers were too comprehensive i.e. most of the topics were assessed in the final examination. Due to this, too many items were put into the papers. Hence the students had limited time to answer all questions.

Analysis on quality of examination questions

In the analysis, the quality of most examination papers was found to have issues related to the five elements: language, information, relevance, context and presentation. The magnitude of the issues varies from one element to another. Table 2.0 shows the finding of the analysis.

Table 2.0 Frequency of five elements for quality examination questions based on three consecutive semesters

Program level	Course	Language	Information	Relevance	Context	Presentation
Degree	A	3	0	2	1	3
	B	0	0	3	0	0
	C	2	3	0	3	0
	D	0	0	0	0	0
	E	0	3	1	3	0
Diploma	F	0	0	3	0	0
	G	0	0	3	0	0
	H	0	0	0	0	0
	I	0	0	3	3	0
	J	0	0	0	0	0
	K	0	0	3	0	0
	L	0	0	3	3	0
	M	0	0	3	0	0
	N	0	0	3	0	0
Total papers	14	5	6	**27**	**13**	3

Based on Table 2.0, five (5) papers from 2 courses (A and C) had language issues, 6 papers from 2 courses (C and E) had information issues, 27 papers from 10 courses (A, B, E, F, G, I, K, L ,M, N) had issues in relevance, 13 papers from 5 courses had issues in context (A, C, E, I, L) and finally, 3 papers in course A had issues in presentation.

Language issues found include lack of or no instruction, unclear and ambiguous instruction, and the use of unsuitable verbs. Some questions were too wordy.

Information issues include too many information to be analyzed in one scenario and questions were too general.

Relevance issues include the level of difficulty is too high, non-alignment to the course learning outcomes, topic assessed is not in the syllabus, multiple level of cognitive levels for choice questions, in structured questions items were not arranged from easy to difficult, allocation of marks did not represent the nature of tasks and allocation of marks for sub-items were not indicated. All these issues contribute towards validity of the question papers.

Context issues include depiction of information (too wordy).

Presentation issues include unclear demarcation between the scenario and instruction, illogical arrangement of taxonomy level in structured questions and unrelated items in a structured question.

Discussion and Recommendation

Mapping of examination questions to course learning outcomes (CLO)

Sixteen papers were found not aligned to the course learning outcomes. The items constructed in the 16 papers were found to be of higher cognitive level than the intended level stipulated in the course learning outcomes.

For Course A, the cognitive level for some of the constructed items surpassed the cognitive level stipulated in the CLO by one level. Hence, students were not able to answer those items correctly simply because the items were assessing the level of ability which is beyond theirs. This caused the high percentage of failure in the final exam performance.

This disparity also happened to Course C. In this case, the cognitive level of several items was 2 times higher than that stipulated in the CLO. Likewise, Courses G, J, L and M recorded similar findings.

It was also found that in most of courses the final examination papers were constructed based on previous semester's final examination papers in terms of choice, question type and marks allocated. Hence, a previously badly constructed final examination paper would definitely make a bad reference to the subsequent semesters' final examination papers.

The validity, reliability, fairness and consistency of the final examination papers mentioned above are therefore questionable.

Issues were also found in final examination papers which were aligned to the respective CLO. Several final examination papers were found to be highly comprehensive. So many topics were assessed. As a result, there were just too many items constructed in one final examination paper. Hence, students had little time to answer all items correctly in the limited time given. This contributed to the overall poor performance in the final examination paper.

In addition, several papers were found to be inconsistent for the three consecutive semesters. The level of difficulty fluctuates throughout the 3 consecutive semesters. Hence, items were more difficult in one semester and easy in another semester. Hence, there was a lack of fairness and reliability in the final examination papers for 3 consecutive semesters.

Analysis on quality of examination questions

The quality of most examination papers was found to have issues related to the five elements: Language, Information, Relevance, Context and Presentation. Two elements showed striking disparity i) Relevance (27), and ii) Context (13).

Relevance issues include the incorrect pitching of the level of difficulty. Almost always, items constructed were of difficult level. At the diploma level, many courses were designed to include cognitive level C4 as the highest level of ability and the percentage for this highest level is small. But what transpired in several final examination papers showed otherwise. Many items constructed contributed more than half of the items. Therefore, the students did not perform well because they were solving excessive items which were deemed difficult. These final examination papers failed to make logical discrimination between good, average and poor students.

Another issue includes the non-alignment of test items to the course learning outcomes. In several cases, the topics assessed were not stipulated in the syllabi. How can students attempt to answer these items correctly when they had no knowledge of the subject matter?

The issue of relevance also includes multiple cognitive levels for choice questions. Many final examination papers offered choice questions. For instance, answer 5 out of 7 questions. The problem lies in the unfairness of each question. Most questions represented different levels of difficulty. For example, Question 1 is difficult, Question 2 is easy, Question 3 is moderate and so on. Hence, how can the test setter ensure that fairness to all students be achieved? For choice questions, all items must be of the same level of difficulty.

It was also found that in structured questions, items were not arranged from easy to difficult. In human psychology, especially when dealing with young adults' mental ability during final examination, it is imperative to consider their state of mind. It would help them tremendously when the items arranged could lead them in simple, easy and logical manner.

In some final examination papers, the allocation of marks did not represent the nature of tasks. Several easy items were given equivalent marks allocated for difficult items. How could

this ensure fairness? In addition, this confused students as to how much answer they were supposed to write. There was no mutual understanding between the intention and expectation of the test setter and the students. Furthermore, in some papers, the allocation of marks for sub-items was not indicated. How many marks would be awarded to students who answered the simple questions in a detailed manner as compared to students who answered difficult questions briefly? All these issues contributed towards the validity of the question papers.

In most of the final examination papers that recorded low performance, *Context* issues include depiction of the scenario. In several final examination papers, too much information was written in words. Many questions presented a scenario in more than one page. Students had to read, understand and analyse every single line. It is imperative to note that most students were not native speakers of English. Their use and understanding of English was fairly limited. Hence, it is suggested that lengthy scenarios be presented in interesting graphical illustrations that require minimal reading. Hence, students can spend more time answering the items than reading to understand the meaning embedded in the scenario.

Recommendation

Due to the inconsistencies and disparities in the quality of the final examination papers found above, it is highly recommended that Test Specification tables (TST) be constructed and adopted as a solution to overcome the problem.

A Test Specification Table (TST) is a blueprint that helps test setters to prepare a valid, reliable, fair and consistent final examination paper. In principle, *fairness* is addressed when consideration is deliberately made to ensure that the time spent teaching each topic is fairly represented by the marks allocated to it. This is done by converting both the teaching time and the allocation of marks to percentages. Both percentages should mirror approximately the same value.

In addition, *validity* is also ensured when each item is directly aligned to the respective course learning outcome(s). Hence, the relevancy of each item is clearly indicated. The *reliability* of the exam papers throughout the subsequent semesters is also addressed when the percentages for the level of difficulty for the respective cognitive levels (easy, moderate and difficult) is maintained. Hence, TST is a mechanism that ensures and regulates the quality of final examination papers.

The TST blueprint is designed to ensure that the four (4) principles of assessment i.e., validity, reliability, fairness and consistency are adhered to when final exam papers are constructed. As such the benefits of the blueprint include: (1) Fair distribution of topics for both final exam and continuous assessment. This also helps to avoid over testing certain topics or subject matter. (2) Test items are directly aligned to course outcomes. Hence, the assessment directly reflects students' attainment of skills at the end of the course. This enforces the validity of the assessment. (3) Test items are constructed according to level of ability/difficulty and the topics/sub topics stipulated in the syllabus. Thus, validity and fairness of test items are ensured at all times. Furthermore, the consistency of exam papers (in terms of difficulty level) in

subsequent semester is ensured. (4) The content of the TST can be justified to various committees at the department, faculty and university levels (Horst and Martens, 2010).

It is imperative to note that although these fourteen (14) analysed courses were found to have disparity in constructive alignment and quality, they only represent 0.28% of the total examination papers in UiTM. Hence, this percentage had an insignificant impact on the overall quality of final examination papers in UiTM. However, the findings of this study are deemed important to create awareness amongst test setters on the construction of quality final examination papers.

Limitation

This study focused on two aspects only i.e., the alignment of the examination papers to the course learning outcomes, and the quality of the final examination papers. It did not cover the aspects of teaching delivery and student ability. Hence, for the future, it is recommended that further research be carried out on the alignment of teaching delivery and student ability to the course learning outcomes.

References

Berkel, H.J.M. van & Bax, A.E. (2006). *Toetsen in Hoger Onderwijs*. Houten: Bohn Stafleu van Loghum.

Biggs, J. (2003). Teaching for Quality Learning at University (2nd. Edition). The Society for Research into Higher Education and Open University Press. Buckingham, UK: Open University Press.

Code of Practice for Programme Accreditation (COPPA). (2008). Agensi Kelayakan Malaysia (MQA).

Horst, H.T & Martens, R. (2010). The Specification Matrix. www.utwente.nl/mb/en/education/test blueprint.doc (Retrieved on Monday 14/10/2013. 09:47)

Horst, H.T & Martens, R. (2010). Open-ended Questions. www.utwente.nl/mb/en/education/test blueprint.doc (Retrieved on Monday 14/10/2013. 10:05)

Roziah Mohd Janor, Sharipah Ruzaina Syed Aris, Rozzana Said and Ahmad Mazli Muhammad (2012). Garis Panduan Kesetaraan Kualiti Penilaian Pelajar UiTM (Edisi 2). Unit Hal Ehwal Kurikulum (UHEK), Bahagian Hal Ehwal Akademik (BHEA): Unit Percetakan UiTM.

Roziah Mohd Janor, Zainab Mohd Noor, Salliza Sideni. (2010). Module 2: Constructive Alignment of Learning Outcomes, Instruction and Assessment. Training Module for IHLS. Bahagian Hal Ehwal Akademik (BHEA): Unit Percetakan UiTM.

Roziah Mohd Janor, Zainab Mohd Noor, Salliza Sideni. (2010). Module 3: Effective Course Delivery. Training Module for IHLS. Bahagian Hal Ehwal Akademik (BHEA): Unit Percetakan UiTM.