Journal of Quality Measurement and Analysis
Jurnal Pengukuran Kualiti dan Analisis

JQMA 11(1) 2015, 69-74

EVALUATING UKM'S LINEAR ALGEBRA FINAL EXAMINATION QUESTIONS IN LINE WITH OUTCOME BASED EDUCATION USING BLOOM'S TAXONOMY AND ITEM ANALYSIS

(Menilai Soalan Peperiksaan Akhir Aljabar Linear UKM Selaras dengan Pendidikan Berasaskan Hasil Menggunakan Taksonomi Bloom dan Analisis Item)

FATIMAH ABDUL RAZAK, SYAHIDA CHE DZUL-KIFLI & ABDUL GHAFUR AHMAD

ABSTRACT

Crafting exam questions is an art in itself. Even though there is no strict formula for producing perfect exam questions, the current line of thinking points in the direction of Outcome Based Education (OBE). In this study, we evaluate the final exam questions of the 2012 Linear Algebra course using OBE standards in two separate phases. Firstly, the questions are categorised according to the Course Learning Objectives (CLO) and evaluated in light of the six cognitive domains of Bloom's taxonomy. This first phase aims to assess whether the examination paper has the right balance of questions in each domain. The second phase involves analysing the results of the students in the same examination using item analysis techniques. The analysis includes finding the discrimination index and difficulty index derived from the answers of the students. This second phase focuses on determining the effectiveness of the questions in discriminating students according to their grasp of the CLO. Based on the results, some improvements are suggested. We conclude that except for the analysis level, the cognitive levels of all the questions are generally acceptable.

Keywords: Bloom's taxonomy; item analysis; outcome based education

ABSTRAK

Penyediaan soalan peperiksaan mempunyai seni yang tersendiri. Walaupun tiada sebarang rumus yang boleh digunakan untuk menghasilkan suatu soalan peperiksaan yang sempurna, namun ahli akademik pada masa ini lebih bersandarkan kepada suatu konsep yang dinamai Pendidikan Berasaskan Hasil. Dalam kajian ini, soalan peperiksaan akhir Aljabar Linear 2012 dinilai dengan menggunakan piawaian yang digariskan oleh Pendidikan Berasaskan Hasil dalam dua fasa berasingan. Dalam fasa yang pertama, soalan dikategorikan mengikut Objektif Pembelajaran Kursus dan dinilai berdasarkan enam domain kognitif taksonomi Bloom. Fasa ini dijalankan dengan tujuan untuk menilai tahap keseimbangan soalan dalam setiap domain. Fasa kedua melibatkan analisis keputusan pelajar dalam peperiksaan tersebut menggunakan teknik analisis item. Teknik ini melibatkan penghitungan nilai indeks pembeza layan dan indeks kesukaran yang diperoleh daripada jawapan pelajar. Fasa kedua ini adalah untuk menentukan keberkesanan soalan dalam membezakan kemampuan pelajar berdasarkan Objektif Pembelajaran Kursus. Berdasarkan keputusan yang diperoleh daripada kedua-dua fasa, beberapa penambahbaikan dicadangkan. Kesimpulannya melainkan untuk tahap analisis, tahap kognitif bagi semua soalan tersebut secara umumnya boleh diterima.

Kata kunci: taksonomi Bloom; analisis item; pendidikan berasaskan hasil

1. Introduction

Outcome Based Education (OBE) is a student focused approach of education where the main concern is the performances of the students. In order to measure these performances, Program Learning Objectives and Course Learning Objectives in line with the Bloom's taxonomy are now made mandatory for all courses in Universiti Kebangsaan Malaysia (UKM). In this paper we focus on what may be viewed as the most important measure of student performance (usually carrying the highest percentage of marks), the final examination questions. Therefore, we have selected a recent examination paper (Semester II 2012/2013) and results of the Linear Algebra course to be examined. In the School of Mathematical Sciences, UKM, the Linear Algebra course is a core mathematical course compulsory for all students from all three programmes (mathematics, statistics and actuarial science). There are five questions in the exam paper and each question has five sub-questions that are interrelated. In the first section we shall look at the questions as 25 separate questions because each sub-question maybe on different cognitive levels. The classification of these questions is according to the six cognitive domains of Bloom's taxonomy. The results obtained are compared to the expected percentage of question from each level. In the second part we use a method called item analysis to evaluate the quality of questions in the exam paper. Here, we analyse the marks for each question obtained by each student.

2. Bloom's Taxonomy

In the 1950s Benjamin Bloom and his colleagues undertook the task of classifying education goals and objectives (Bloom 1984; Noraini 2001). Work on the cognitive domain is now commonly referred to as Bloom's taxonomy of the Cognitive Domain. The idea of the taxonomy is that statements of educational objectives can be arranged in a hierarchy from less to more complex (higher) level. The categories or levels for the cognitive domain and illustrative verbs for each level are as follows (Bloom 1984; Noraini 2001):

a) Knowledge (K): recalling information, ideas, and principles in the approximate form in which they were learned. Sample verbs: define, write, list and label.
b) Comprehension (C): demonstrating understanding of terms and concepts. Sample verbs: describe, interpret and give examples.
c) Application (A1): applying learned concepts or skills to solve a problem. Sample verbs: calculate, solve and use.
d) Analysis (A2): breaking things down into their elements, formulating theoretical explanations for observed phenomena. Sample verbs: derive, explain and what is the difference.
e) Synthesis(S): creating something, combining elements in novel ways. Sample Verbs: formulate, make up and conclude.
f) Evaluation (E): making and justifying value judgments or selections. Sample verbs: determine, select and critique.

Levels (4)–(6) are known as higher-level (or higher-order) thinking skills. A good question paper is expected to be normally distributed throughout levels 1 to 6 where the peak coincides with level (3) and (4) so that questions are mostly on the intermediate level.

3. Table of Specification (TOS)

Due to Ahmed *et al.* (2013), Table of Specification is an instrument that teachers use to formulate an examination. The table is aimed to establish a comparison and organise the number of questions dedicated to each tier of Bloom's taxonomy.

We consider all the 25 sub-questions according to the Bloom Cognitive Domain. Table 1 displays number of questions allocated in each domain. Distribution of content is focused more on the three last topics since the first and second topics were already covered in the mid semester exam. Distribution of cognitive domain is not as expected since there are many questions on the application (A1) level and not nearly as much question on the analysis (A2) level. However, this could be due to the nature of Linear Algebra itself where many questions will require application of theorems and not as much explaining or deriving equations. Nevertheless, this also points out the need to diversify the questions and not depending too much on theory application to problems.

Table 1: Table of specification for Linear Algebra 2012 examination questions

Content	Bloom's Cognitive Domain						Total	Percentage
	K	C	A1	A2	S	E		
Linear Equation & Matrices			2				2	8
Determinant			3				3	12
Vector Space	1	1	2	1			5	20
Eigenvalues & Eigenvectors			2		2	1	5	20
Linear Transformation		3			1	1	5	20
Inner product space	2		3				5	20
Total	3	4	12	1	3	2	25	100
Percentage	12	16	48	4	12	8	100	

4. Item Analysis

Item analysis is a process which examines students' responses to individual test items (questions) in order to assess the quality of those items and of the test as a whole, typically for the purpose of test construction and revision (Adedoyin & Mokobi 2013). The process in this method examines students (examinees) responses to individual test items in order to assess the quality of those items and of the test as a whole. In this section we evaluate the quality of each question in the Linear Algebra 2012 final examination by using Item Analysis method. The purpose of this test is to improve test questions and identify unfair or biased items. On top of that, we are hoping to identify some specific areas of the course content that needs greater emphasis or clarity. The item analysis includes finding the discrimination index and difficulty index derived from the answers of the examinees. The guideline for item analysis procedures and the concepts of item difficulty and discrimination for an essay examination is given by Evaluation and Examination Service of The University of Iowa.

4.1 *Item Difficulty*

Item difficulty may be interpreted as "how hard is this item?" That is, how does the performance of a group of examinees compared with the highest possible level of performance? (Whitney 1970). The proportion of accumulated marks indicates the difficulty level of the item. The more marks accumulated by a group of examinees, the less difficult the item is. Therefore the difficulty index, P is the ratio;

$$P = \frac{\overline{X} - X_{\min}}{X_{\max} - X_{\min}},$$

where $\overline{X}$ is the average item score, $X_{\min}$ is the smallest item score and $X_{\max}$ is the highest item score. The index is represented as a fraction and varies between 0 and 1, where a small value of P indicates that the item is hard while a high value of P indicates that the item is too easy for a group of examinees. Therefore our desirable standard value of P is 0.5, which indicates a moderate difficulty.

4.2 *Item Discrimination*

According to Wilson (2005), item difficulty is the most essential component of item analysis. However, it is not the only way to evaluate test items. Discrimination goes beyond determining the proportion of people who answer correctly and looks more specifically at who answers correctly. In other words, item discrimination determines whether those who did well on the entire test did well on a particular item (Eaves & Bradley 2009). The index of discrimination is the difference in item difficulties between groups of examinees with high and low tests scores. Therefore we have to select two extreme groups, compute index P for each group and find the difference between P-values (Whitney 1970). Kelly (1939), Downing and Yudkowsky (2009) suggested that, in most conditions, the best method of grouping students for this computation is to take the highest and lowest 27% of the examinees group. However, the method of grouping is not practical for a small group of examinees. Evaluation and Examination Service of The University of Iowa suggested that to retain enough examinees in the two groups (to avoid distortion due to a single examinee's response) would be to use the highest and lowest 10 test scores for classes with 20 to 40 students. For larger classes, groups containing 25% or 27% of the class may be used.

To determine the discriminative index, D the difficulty index of the item will be calculated for each extreme group, D_U for an upper group and D_L for a lower group. The discrimination index is defined as

$$D = D_U - D_L .$$

From the formula, positive discrimination for an item results is when a high scoring group obtains a higher average score on the item than does a low scoring group. A test composed of items with high positive discrimination indices will more likely yield reliable scores (hence reliable grades) than a test whose items have low and negative discrimination indices. Item discriminations of $D = 0.50$ or higher are considered excellent, where $D = 0$ means that the item has no discrimination ability. On the other hand, $D = 1.00$ means the item has a perfect discrimination ability.

4.3 *Result and Analysis*

The method of item analysis described in the previous section is then used to evaluate the quality of each item (question) in the Linear Algebra 2012 final examination. To determine the discriminative index, we followed Kelly's suggestion of grouping method. Table 2 presents the difficulty and discriminative indices for each question.

Table 2: The value of difficulty and discrimination indices

	Question 1	Question 2	Question 3	Question 4	Question 5
P	0.64	0.37	0.68	0.29	0.34
D	0.08	0.01	0.30	-0.08	0.08

Questions 1, 2 and 5: The D-values near zero may indicate that the question had little relationship to the other questions or that it was a very easy (or very difficult) question so that nearly all students got a large number of points (or very few). Therefore, estimating difficulty index P will help determine whether the latter is an appropriate explanation. In this case, Question 1 was of moderate difficulty, so the D-value near zero means that the item had a very little relationship to the others. Discrimination index can be improved by rewriting the command in the questions clearly. However, the difficulty index of Questions 2 and 5 indicated that the items were slightly easier than the other questions, thus causing the near zero D-values. These items could be selected for inclusion in the question bank subject to some modifications.

Question 3: P-value near 0.5 which indicated that the item was of moderate difficulty. D-value was 0.3, so the question has a very acceptable discrimination. This question should be selected and kept in the question bank.

Question 4: A very poor item where P-value near 0 indicated that the item was difficult and negative D-value indicated either, the question may not logically "fit" with the other items contributing to the total test score, or the question may not have indicated clearly enough what kind of responses were desired. Another explanation for a negative D-value was that better performance students somehow read the question differently than was intended. This question should be rejected or drastically altered.

5. Conclusion

We have categorised the questions into the six cognitive domains of the Bloom's taxonomy and displayed the results in a Table of Specification. We found that questions need to be diversified even more especially to increase the number of questions on the analysis level. The second part involved analysing the results of the student in the exam using item analysis techniques. Based on the outcomes, one of the questions should be kept in the question bank without any correction made, three of the questions may be kept in the question bank with minor modification and one of the questions should be rejected or drastically altered. In a nutshell, except for the analysis level, the cognitive levels of the questions are generally acceptable.

Acknowledgement

The authors gratefully acknowledge the financial supports received in the form of research grants from the Universiti Kebangsaan Malaysia (GGPM-13-067 and GGPM-2013-031).

References

Adedoyin O.O. & Mokobi T. 2013. Using IRT psychometric analysis in examining the quality of junior certificate mathematis multiple choice examination test items. *International Journal of Asian Social Science* **3**(4): 992-1011.

Ahmed A., Nisa A. & Zarif T. 2013. Table of specification development and usage: an overview. *Interdisciplinary Journal of Contemporary Research in Business* **4**(12): 354-361.

Bloom B.S. 1984. *Taxonomy of Educational Objectives: Cognitive Domain*. New York: Longman.

Downing S.M. & Yudkowsky R. 2009. *Assessment in Health Professions Education.* New York: Routledge.

Eaves S. & Erford B. 2009. Item Analysis. http://www.education.com/reference/article/item-analysis/ (12 Nov 2014).

Kelley T.L. 1939. The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology* **30**(1): 17-24.

Noraini I. 2001. *Pedagogi dalam Pendidikan Matematik*. Kuala Lumpur: Utusan Publications and Distributors Sdn. Bhd.

Whitney D.R. 1970. Preparing and Evaluating Essay Test Questions. http://www.uiowa.edu/~examserv/ resources_fees/Technical_Bulletins/Tech%20Bulletin%2036.pdf (12 Nov 2014).

Wilson M. 2005. *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum.

School of Mathematical Sciences
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor DE, MALAYSIA
E-mail: fatima84@ukm.edu.my, syahida@ukm.edu.my, ghafur@ukm.edu.my*

*Corresponding author