Journal on Efficiency and Responsibility in Education and Science
ISSN: 1803-1617

EVALUATION OF TEST QUESTIONS USING THE ITEM ANALYSIS FOR THE CREDIT TEST OF THE SUBJECT OF MATHEMATICAL METHODS IN ECONOMICS IN THE MOODLE LMS

Kučera, P., Kvasnička, R., Vydrová, H.

Czech University of Life Sciences Prague
kucera@pef.czu.cz

Abstract

The Moodle system has been used within the CULS in teaching for almost two years. This paper deals with its utilization in the subject of Mathematical Methods in Economics II. The Moodle course of this subject provides students with basic and organization information, figures and tasks for exercises, software for solving them if necessary, video-records of methods and, of course, tests. The aim of this contribution is to analyze one of these tests and to summarize recommendations for improving it. The item analysis is used for this purpose. First the definitions of the item analysis indicators are given. Then they are computed using the Moodle system. All of them were very different at particular questions and it means that single questions had a different quality. The quality even depended neither on the topic nor on the form of the question (it often happened that two similar questions were evaluated very differently). The proper analysis is shown on several particular examples of questions. The item analysis has shown to be a very efficient tool for evaluation of test questions and there is no other way how to get some of the information obtainable by this analysis.

Key Words

Test, test question, test item, item analysis, LMS Moodle, mathematical methods in economics

Kucera, P., Kvasnicka, R., Vydrová, H. (2009), "Evaluation of Test Questions Using the Item Analysis for the Credit Test of the Subject of Mathematical Methods in the Moodle LMS", Journal on Efficiency and Responsibility in Education and Science, Vol. 1, No. 2, ISSN: 1803-1617, pp 41-50, [on-line] www.eriesjournal.com/_papers/article_69.pdf [2009-06-30]

Introduction

Since the last academic year the learning management system (LMS) Moodle has been used in teaching at the CULS in Prague. The LMSs are the narrowest part of electronic education support systems, which integrates in itself tools for controlling the teaching, i.e. offers a variable environment for creating and administrating Internet courses. Furthermore, it provides communication tools for all the participants of the course (in addition, with ability to differ single roles of users in a given course).

Currently the Moodle system is utilized in the most of subjects taught within the CULS. The first two of the authors of this paper use it in the subject of Mathematical Methods in Economics II which is concerned in this paper. It is the second semester of a two-semester course of the operations research methods. The topics contained in the subject are various methods and approaches which often occur in practice, e.g. larger models of linear programming, different types of transportation tasks, game theory, decision models, multiple-criteria decision, Markov chains and some of other stochastic systems etc.

The introduction of the Moodle course of this subject contains the syllabus, information about the conditions for obtaining the credit, a forum for sharing news within teachers and students and a software tools necessary for calculations at the exercises (macros in the Visual Basic for Applications specially designed by the members of the department for this purpose). Next the course is divided into 14 thematic units according to the single teaching weeks (lectures and exercises). The typical content of a unit is as follows:

First, figures and tasks for the exercises are available for students. Their amount is sufficiently large to cover all the time of the exercise and some tasks remain for students for their home individual training. They are in the MS-Word document form.

Another regular part is a page of checking questions for students. These questions are made up so that they guide students to the repetition of important parts of the taught problem. In the first year of the utilization of the Moodle system they were originally in the form of a text page, for the second (i.e. last) year the form was changed to the small tests (so called home preparation) which are evaluated and students get points necessary for obtaining the credit.

Some of the units provide also video-records of methods for exercises. They are applied with such topics which are problematic for students.

The structure of testing students consists of two parts. The former one is the home preparation mentioned above. The latter and the main part of testing is a (main) test divided into two parts. The first part takes place in the middle of the semester and the second part in the end of the semester. Both the home preparation and the main test are designed in the test interface provided by the Moodle system and consist of various types of questions: multiple-choice (with both one and more right answers), numerical and true/false.

The minimum number of points necessary for obtaining the credit is approximately one third of the whole possible amount for the whole testing (both the main test and the home preparation) and at the same time approximately one third of the maximum possible amount for the home preparation.

The aim of this paper is to analyze the first part of the main test and to suggest changes, i.e. deletion or modification of unsuitable questions etc. This analysis will be realized using the item analysis.

More information about the Moodle courses on the Mathematical Methods in Economics is in Kučera, Kvasnička (2008a), Kučera, Kvasnička (2008b), Kučera, Kvasnička and Burdych (2008).

Material and Methods

The item analysis is used for exploring the characteristics of single items of the test. It is suitable for judging the quality of each of its questions. It involves two types of analysis: quantitative (e.g. correlation, burdensomeness, time demand) and qualitative (lucidity, content and format adequacy, etc.). The item is the smallest unit of the test according to which it is possible to judge a given test using it.

More information about the theoretical respect of the test results evaluation and the item analysis is e.g. in Vyškovský (1982), Hniličková, Josífko and Tuček (1972), Komenda (2003), Smékal, Švec and Zajac (1973). The Moodle system provides computing the following statistical parameters.

The Ease

The ease indicates how the question is easy for a student. In the case of the dichotomic form of the question (right/wrong answer) this parameter is equal to proportional (or percentage) expression of the right answers of students. This parameter can be clearly described using the formula:

$$ease = \frac{X_{mean}}{X_{max}},$$

where:

X_{mean} is the average number of points obtained by all the students for a given item and

X_{max} is the maximum possible number of points.

The Standard Deviation

The standard deviation (SD) is a commonly and widely used statistical parameter. It determines the dispersion of the students' answers in the whole sample. It shows how much students differ among each other in their knowledge of the given problems. Thus the standard deviation is an indicator of the discrimination ability of the test but it does not identify good and bad students. If all the answers are identical, the standard deviation is equal to zero.

The Discrimination Index

The discrimination index (DI) shows how much (to what degree) an item makes difference between successful and unsuccessful students. Using this index it is possible to compare the result of e.g. this item, or the whole test, and results of all the other items, or other tests, respectively. In general, one can state that a student with good results will write the test well and, on the contrary, a bad student will not succeed. The discrimination index is a rough indicator of the efficiency of each item at a given students group.

For determining the value of this index one third of all the students (no matter whether the evaluated question has been randomly chosen for a student and thus he/she has actually solved it) with the best results and one third of students with the worst results are taken and then it is found out how these groups of students have solved a given particular question. In the ideal case the best students should succeed and the worst ones should fail. For each of these two groups the mean value of the proportional expressions of the results of all its members is computed and the mean value of the latter group is subtracted from the mean value of the former group.

The value of this index ranges in the interval from –1 to +1. Negative values show that the right answer has been given more often by worse students than by the best students at the single question. Such negative values of this index should signalize to the creator of the test that it is necessary to remove or reformulate such a question.

The Discrimination Coefficient

From the statistical point of view the discrimination coefficient (DC) is a correlation coefficient between the score for a given item and for the whole test. It shows how much (to what extent) the results would be different if we differed between clever and less clever students. As in the case of the discrimination index, this index gets values from –1 to +1. Positive values show the difference of the clever students. Negative values indicate items which have been answered wrong by the best students. For the creator of the test it signalizes the necessity of removing the given question.

The advantage of this coefficient is that it uses the data of all the results for computing, and not only the results of one third of the best and the worst students.

Results

The test analyzed here is based on a file of 205 questions divided into 17 categories. As mentioned in the introduction, all the questions are simple (unstructured) and so they form items for the analysis at the same time. Each category contains from 6 to 14 (usually 10) questions. A single test is created by the Moodle system using a random choice of questions, from 1 to 4 questions representing each category. Number of students who solved each question varied from 24 to 142, but at each category this number was similar for each question.

The ease of questions was very different: the lowest one was 5 p.c. and one question had an ease of 100 p.c., i.e. it was right answered by all the students who got it.

There was no item with a negative discrimination index. In the worst case this index was just 0.000. On the other hand, 20 items had the discrimination index equal to 1.

The discrimination coefficient varied from –0.197 to 0.658. There were only 10 questions with a negative discrimination coefficient. Let us remind that for the one question with the 100 p.c. ease the discrimination coefficient is not defined.

Discussion

For demonstration of applying the item analysis 19 questions were selected, all of them being of the multiple-choice type with an offer of four answers, just one of them being right. The following data were taken from the form of evaluation provided by the Moodle system. Because the original table containing these data had too many columns and so it was too large and nontransparent, the data are divided into two parts. In the following part (Text of Problematic Questions) there are given for each question its number from the Table 1, the text (input) of the question and the possibilities given students for choice. In the Table 1 there is the following information about the test and its item analysis results:

In the first column there is its serial number. As all the questions have four answers to choice, the information about each question is contained in four columns. In the column No. 2 the offered answers are labeled by letters from a) to d) and the column No. 3 contains the partial credit for each answer, i.e. 1 for the right one and 0 for the wrong ones. The following column shows how many students have chosen this answer and how many

students have solved this question, so these two numbers form the ratio of students who have chosen this option. The next column contains the same ratios in the percentage form. In the column No. 6 the number of students solving the task is repeated again and the last four columns contain the statistical indices described in the chapter of Material and Methods.

1	2	3	4	5	6	7	8	9	10
Question number	**Answers**	**Partial credit**	**Number/ answers**	**P.c. of answers**	**Num. of students**	**Ease**	**SD**	**DI**	**DC**
1	a)	0,00)	2/90	(2%)	90	6	0,2303	0,071	0,118
	b)	0,00)	1/90	(1%)					
	c)	1,00)	5/90	(6%)					
	d)	0,00)	82/90	(91%)					
2	a)	1,00)	2/37	(5%)	37	5	0,2292	0,111	0,117
	b)	0,00)	21/37	(57%)					
	c)	0,00)	12/37	(32%)					
	d)	0,00)	2/37	(5%)					
3	a)	(0,00)	20/46	(43%)	46	22	0,4170	0,111	-0,110
	b)	(1,00)	10/46	(22%)					
	c)	(0,00)	13/46	(28%)					
	d)	(0,00)	3/46	(7%)					
4	a)	(0,00)	5/44	(11%)	44	25	0,4380	0,182	0,000
	b)	(0,00)	3/44	(7%)					
	c)	(0,00)	24/44	(55%)					
	d)	(1,00)	11/44	(25%)					
5	a)	(0,00)	5/50	(10%)	50	18	0,3881	0,111	0,118
	b)	(0,00)	5/50	(10%)					
	c)	(0,00)	31/50	(62%)					
	d)	(1,00)	9/50	(18%)					
6	a)	(1,00)	6/50	(12%)	50	12	0,3283	0,375	0,388
	b)	(0,00)	28/50	(56%)					
	c)	(0,00)	13/50	(26%)					
	d)	(0,00)	3/50	(6%)					
7	a)	(0,00)	6/56	(11%)	56	29	0,4558	0,500	0,247
	b)	(0,00)	27/56	(48%)					
	c)	(0,00)	6/56	(11%)					
	d)	(1,00)	16/56	(29%)					
8	a)	(0,00)	2/45	(4%)	45	29	0,4584	0,500	0,250
	b)	(0,00)	2/45	(4%)					
	c)	(0,00)	27/45	(60%)					
	d)	(1,00)	13/45	(29%)					
9	a)	(1,00)	15/42	(36%)	42	36	0,4850	0,625	0,258
	b)	(0,00)	15/42	(36%)					
	c)	(0,00)	7/42	(17%)					
	d)	(0,00)	4/42	(10%)					
10	a)	(1,00)	8/44	(18%)	44	18	0,3902	0,375	0,340
	b)	(0,00)	24/44	(55%)					
	c)	(0,00)	10/44	(23%)					
	d)	(0,00)	2/44	(5%)					
11	a)	(1,00)	14/43	(33%)	43	33	0,4741	0,429	0,283
	b)	(0,00)	22/43	(51%)					
	c)	(0,00)	5/43	(12%)					
	d)	(0,00)	2/43	(5%)					
12	a)	(0,00)	9/28	(32%)	28	25	0,4410	0,000	-0,197
	b)	(0,00)	6/28	(21%)					
	c)	(1,00)	7/28	(25%)					
	d)	(0,00)	6/28	(21%)					

13	a)	(0,00)	13/35	(37%)	35	26	0,4434	0,200	0,075
	b)	(1,00)	9/35	(26%)					
	c)	(0,00)	3/35	(9%)					
	d)	(0,00)	10/35	(29%)					
14	a)	(0,00)	5/36	(14%)	36	78	0,4216	0,429	-0,131
	b)	(1,00)	28/36	(78%)					
	c)	(0,00)	0/36	(0%)					
	d)	(0,00)	3/36	(8%)					
15	a)	(0,00)	4/39	(10%)	39	51	0,5064	0,100	-0,097
	b)	(0,00)	3/39	(8%)					
	c)	(1,00)	20/39	(51%)					
	d)	(0,00)	12/39	(31%)					
16	a)	(1,00)	16/36	(44%)	36	44	0,5040	0,625	0,212
	b)	(0,00)	14/36	(39%)					
	c)	(0,00)	4/36	(11%)					
	d)	(0,00)	2/36	(6%)					
17	a)	(0,00)	8/25	(32%)	25	36	0,4899	0,600	0,100
	b)	(1,00)	9/25	(36%)					
	c)	(0,00)	5/25	(20%)					
	d)	(0,00)	3/25	(12%)					
18	a)	(1,00)	13/32	(41%)	32	41	0,4990	0,429	0,009
	b)	(0,00)	2/32	(6%)					
	c)	(0,00)	2/32	(6%)					
	d)	(0,00)	15/32	(47%)					
19	a)	(1,00)	51/53	(96%)	53	96	0,1924	0,818	-0,118
	b)	(0,00)	0/53	(0%)					
	c)	(0,00)	1/53	(2%)					
	d)	(0,00)	1/53	(2%)					

Table 1: Evaluation of Questions by the Moodle System

The first two questions are the most difficult among all. These questions are the only ones with the less than 10 p.c. ease. Their discrimination index and discrimination coefficient are relatively small, too. But there is a significant difference between the evaluations of these two questions. If a question of 5 or 6 p.c. ease solved the same number of both the best and the worst students, its discrimination index might be at most 0.15 or 0.18, respectively. The question No. 1 has a discrimination index equal to 0.071, while the one No. 2 has 0.111 which is relatively close to this boundary. The only problem of the question No. 2 is its burdensomeness. If there were other difficult questions, there would be no reason to remove it from the test. This matter is analyzed thereinafter in comparison to other questions.

On the contrary, the question No. 1 has a small discrimination index also in comparison with its small ease. But first of all, as one can easily see, the most of students have chosen the wrong answer d) (while at the previous question different wrong answers occur, especially b) and c)). Thus the first idea is that this question should be removed or the answer d) should be replaced by more suitable and less confusing one. Let us concentrate on the formulation of this question. This question is on the analysis of the solution of a transportation task and it is the only question where the costs are expressed in monetary units. In all the other such questions tons and kilometers are used. The problem probably is that students are not used to operate with the monetary units and the suggestion how to repair this problem is to concentrate on it more during the exercises and to add other similar questions into the questions database.

The questions with serial numbers from 2 to 11 are the all the questions of the category of the interchange of criterion function and limiting condition. They are selected as an example how much the similar questions on the same topic may differ in their

results of the item analysis. Here is obviously confirmed that the question No. 2 is all right because there are four of all ten questions of this category among ten questions with the smallest ease (less than 20 p.c.). The only question of this category which should be removed according to the item analysis is the question No. 3 because it has a negative discrimination coefficient (and at the same time a relatively small discrimination index) but it is not clear why its parameters are so bad and thus whether it is really necessary to remove it.

The question No. 12 seems to be good if we do not notice its discrimination index and discrimination coefficient. It has ease of 25 p.c. and all four answers were chosen approximately by the same number of students. But both discrimination index and discrimination coefficient are the worst from all the questions and it shows that good students answer this question wrong and vice versa. Thus this question must be removed from testing. Herewith, there exists a similar question No. 13 with a good evaluation again and it is hardly to say what the problem is at the question No. 12.

The question No. 14 is the worst evaluated one of the category of the theoretical questions. Its negative discrimination coefficient indicates that it is confusing for clever students and so it should be modified or removed.

Another question with the negative discrimination coefficient has no. 15. In this case, three other similar questions (on judging which variants in the multiple-criteria analysis are feasible subject to the aspiration levels) follow (from no. 16 to no. 18) with approximately the same ease, slightly higher discrimination coefficient and much higher discrimination index in comparison to the question no. 15 and there is an easily notable difference between the bad evaluated question and the others. The question no. 15 is the only one of these questions where there is offered an answer that there is no variant feasible subject to the aspiration levels and at the same time the right answer is that there are two such variants (in all other questions there is just one such a variant). It seems to be surprising that these facts which make the question more difficult are troubling just clever students. Anyway, this question should be removed.

The last question (No. 19) has also a bad discrimination coefficient. Herewith, it has 96 p.c. ease and its bad discrimination coefficient is caused by only two good students who have answered wrong. It may be only a bad luck of these students. Moreover, the high value of the discrimination index of a question with such a high ease indicates that almost nobody of the students who have got it is a bad student in the whole testing (this fact may even cause the high ease). Therefore, it may be kept in testing.

Text of Problematic Questions

Question No. 1

The cost for the transportation of the optimum amount of the potatoes from the plot of field H2 to the potato store B1 (see model "potatoes - variant B, 1 km cost is 8 CZK) is:

a) 400 CZK
b) 2600 CZK
c) 16000 CZK
d) 32000 CZK

Question No. 2

If the interchange of criterion function and limiting condition to the optimization model with two criteria z1 = 2x1 - 2x2 + 4x3 --> min with the optimum value 33 and z2 = 4x1 - x2 + 4x3 --> min with the optimum value 72 is applied, the following is obtained: e.g.

a) a new constraint 2x1 - 2x2 + 4x3 <= 33 and the criterion z2 = 4x1 - x2 + 4x3 --> min80; 260; 50; 70; 8.9
b) a new constraint 2x1 - 2x2 + 4x3 <= 39 and the criterion z2 = 4x1 - x2 + 4x3 --> min
c) a new constraint 2x1 - 2x2 + 4x3 <= 39 and the criterion z1 = 2x1 - 2x2 + 4x3 --> min
d) a new constraint 2x1 - 2x2 + 4x3 <= 33 and the criterion z1 = 2x1 - 2x2 + 4x3 --> min

Question No. 12

	Vitamin A	Vitamin B1	Vitamin B2	Vitamin C	Price
Cabbage	80	260	50	70	3,70
Cauliflower	90	200	100	70	8,90
Cucumber	1	120	30	6	13,70

In the table single sorts of vegetables are evaluated according to the content of vitamins in mg per kg and to the price in CZK per kg. To which aspiration levels is no variant a feasible subject?

a) 1; 120; 30; 6; 13.8
b) 80; 260; 50; 70; 8.9
c) 90; 200; 50; 70; 3.7
d) 90; 200; 100; 70; 3.7

Question No. 14

Systems analysis is

a) an applied cybernetics
b) an applied systems science
c) the only systems science
d) a closely specialized team science

Question No. 15

	Vitamin A	Vitamin B1	Vitamin B2	Vitamin C	Price
Cabbage	80	260	50	70	3,70
Cauliflower	90	200	100	70	8,90
Cucumber	1	120	30	6	13,70

In the table single sorts of vegetables are evaluated according to the content of vitamins in mg per kg and to the price in CZK per kg. If the aspiration levels are set to the values of 50; 90; 35; 10; 10, then the following variants are feasible subject to the aspiration levels:

a) cauliflower
b) cucumber
c) cabbage and cauliflower
d) no variant

Question No.19

Transportation tasks belongs to the group of tasks called:

a) distribution tasks
b) quadratic tasks
c) balance tasks
d) NP-complete tasks

Conclusion

The item analysis has shown to be a very efficient tool for evaluating the test questions. Applying on the first part of the main test of the subject of Mathematical Methods in Economics II, it determined the burdensomeness (or ease) of each single question which reaches a wide scale of different values. It shows that one question was solved right by all the students and few others by almost all of them.

About 10 p.c. of the whole amount of questions had discrimination index equal to one which means they are answered right by good students and wrong by bad students. On the other hand, about 5 p.c. had a negative discrimination coefficient, i.e. they differ badly between good and bad students. Such questions should be modified or removed from the file of questions. But it is necessary to approach to single questions individually, as shown in the discussion. Thus one cannot simply say that the unsuitable questions are just those with the small discrimination coefficient.

The questions with a bad evaluation were on different topics, i.e. the single topics did not differ in the burdensomeness of creating the questions on them. Two similar questions often differed very much in their evaluation and it is hardly to say why. It means that only the item analysis can check them in such details.

If questions of such a category where more than one is chosen into a single test differ too much in their ease, it is suitable to divide this category into several smaller ones according to the ease of the questions.

The only drawback of the Moodle system item analysis is that it does not inform about the number of good and bad students who answers particular questions. Such information would enable a better analysis of the computed parameters of the item analysis, especially the ease and the discrimination index.

The item analysis was very helpful to analyze the test and according to its results it will be improved for the next year.

Acknowledgements

The paper is supported by the grant No. 2735 of the Found of the Development of the Universities (FRVŠ) of the Czech Ministry of Education, Youth and Sports.

References

Hniličková, J., Josífko, M., Tuček, A.: Didaktické testy a jejich statistické zpracování, SPN, Praha, 1972.

Komenda, S.: Měření a metaměření znalostí, Univerzita Palackého, Olomouc, 2003.

Kučera P., Kvasnička R.: Possibilities of the LMS Moodle in Teaching the Subject of the Mathematical Methods in Economics I, In: Agrarian Perspectives, CULS, Prague, 2008a, p. 761-764.

Kučera P., Kvasnička R.: Tvorba kurzů v LMS Moodle, In: Kvantitativní metody v ekonomii, JČU, České Budějovice, 2008b, p. 64-69.

Kvasnička R., Kučera P., Burdych P.: Application of the Moodle System in the Teaching of Quantitative Methods, In: International Scientific Days, SUA, Nitra, 2008.

Smékal, V., Švec, V., Zajac, J.: Didaktické testy a jejich vyhodnocování, Středisko pro výzkum učebních metod a prostředků, Brno, 1973.

Vyškovský, P.: Základy měření výsledků výuky. Tvorba didaktického testu, ČVUT, Praha 1982.