Models and Tools for Information Support of Test Development Process in Learning Management Systems

Olena Kuzminska[1], Mariia Mazorchuk[2]

[1]National University of Life and Environmental Sciences of Ukraine,
16a Heroev Oborony St., Kyiv, Ukraine

o.kuzminska@nubip.edu.ua

[2] National Aerospace University "KhaI",
Chkalova str. 17, Kharkiv, Ukraine

mazorchuk.mary@gmail.com

Abstract. Many current educational trends are based on the digital technologies support for content modeling and delivering as well as for control services in the teaching and learning processes. E-learning methods allow us significantly to enhance the teaching and learning effectiveness. These methods also show the rapid development of software of a new generation intended for the teaching and studying process support. Nevertheless there still exist problems to deal with in order to make further progress. One of these problems lies in the creating effective systems for student educational progress assessment. This article proposes a new original item bank information system creating method based on the Open Source tools: LMS Moodle and package R. This method was tested in the National University of Life and Environmental Sciences of Ukraine and the National Aerospace University "KhaI". Obtained results are presented and practically illustrated.

Keywords. Educational Technologies, Test Items, Testing, Information and Communication Technologies, Experience

Key Terms. Academia, Teaching Process, Development, Information Communication Technologies, Model

1 Introduction

The modern stage of educational system development is characterized by appearance new educational technologies, which are come together with high rate of informatization. Influence analysis of macro, meso and micro trends; designing of educational spaces and models are the subjects of the research [1]. In its annual reports New Media Consortium (NMC) are describing technologies, which will be having significant impact on the educational processes, including higher educational establishments, as in: flipped classroom, learning analytics, blending learning (b-learning), personalizing

ICTERI 2016, Kyiv, Ukraine, June 21-24, 2016

learning, bring your own device (BYOD), maker spaces, the Internet of things, adaptive learning technologies, open educational resources and massive open online courses (MOOC) [2, 3]. Concerning e-learning implementation [4] the publications widely represent the experience b-learning application, as well as learning management systems (LMS) for e-learning organization [5, 6]. Simultaneously stay relevant the issues of supporting the objective testing of knowledge quality of students using e-learning systems, which are applied both in MOOC, and in knowledge control systems at universities. Especially burning is the problem of using the qualitative tests in the systems of e-learning, because the users of on-line courses may have different levels of preparation and need individual approach for education, which cannot be ensured by the current e-learning systems. For example, in LMS Moodle, which is used in most Ukrainian universities, provides the functional part of test quality rating [4]. However, the data rates not always are reliable, as they are often determined on the basis of small population of participants and primary test results. The existing methods and models of evaluation of the test quality rating have its own application field; nevertheless they still don't allow solving the tasks connected with performance of effective on-line rating in complex. Virtually, there are no tools that can provide complex support for test designers in the process of testing.

The objective of this research is the development of models and tools of information support of tests forming in systems of distance learning, providing the adequate level of quality depending on the ability level of students.

2 Methods and Quality Analysis Models of Education Tests

For tests quality analysis mainly are used methods of classic theory, based on the calculation of such main parameters as, complexity, ability to differentiate, correlation of tests with total test grade and so[7, 8]. For more detailed analysis the method of threshold group is used, which allows to build the curve and frequency tables of distractor choice for threshold group of test takers, that sufficiently represents the information about the quality of developed test items [9]. In LMS Moodle system, which possesses wide functional, the block of quality tests results interpretation is not established. In case of small populations, the evaluation parameters methods can have serious errors; there is no possibility to compare the results of learners groups.

Today there are a lot of supporters of Item Response Theory, which allows to get tests results in metric scale. The literature [8], [10, 11] gives exhaustive information about methods and models of IRT. The important feature of models of modern tests theory is the limited conditions of its use, which are as follows:

- tests modeling by Rush function (not always possible);
- compatibility of participants response with Gutmann condition (unipath continuum);
- unidimensionality of test (the test must be measured by only one construct);
- items test are independent.

The main disadvantages of this theory are the calculation and results interpretation complexity, and also high demand to study population volume (at least 500 individuals), which is difficult to assure in conditions of functioning of modern educational institutions (tests participants groups may not exceed 15 persons).

User experience of LMS Moodle [12] has shown, that it is possible to guarantee the accumulation of statistical data by tests results in small groups (10-15 persons) during long time period, so it is effectively to use this platform for experiments.

Using LMS Moodle it is possible to save various tests results of students: test parameters (duration of the test, quantity of the used attempts to answer the questions); total points; parameters of test questions (correct and incorrect answers). All results can be presented in convenient format for following analysis.

To analyze the quality of tests it is necessary to evaluate tests parameters, which allow estimating their reliability and validity. Input data for analysis are the matrix of tests results. Matrix of tests results is matrix of NxM dimension, where N – quantity of tests participants, M – numbers of test items:

$$A = \left(a_{ij}\right)_{i,j=1}^{N,M} . \qquad (1)$$

Evaluation schemes for the majority of tests can be classified into dichotomous and polytomous data [8]. Dichotomous scale is used in this paper.

To evaluate psychometric tests characteristics it has been decided to use Item Response Theory, since it allows to get results in metric scale and to compare groups of participants. For research of psychometric characteristics of tests various IRT models are applied: classical model Rasch, 1PL, 2PL and 3PL. More detailed information about these models may be found in [10, 11].

Let us consider the three-parameter models 3PL, which allows to receive the most fitting results (in fact all other models are the particular case of 3PL models). In this model the conditional probability of correct performing of j-test for examinee with ability level Θ depends on three parameters: difficulty parameter δ_j, discrimination parameter d_j and guessing parameter c_j:

$$P_j(\Theta) = P\{ x_{ij} = 1 \mid \delta_j, d_j, c_j \} = c_j + (1 - c_j)\frac{e^{Dd_j(\Theta - \delta_j)}}{1 + e^{Dd_j(\Theta - \delta_j)}} . \qquad (2)$$

Here the constant multiplier D=1.7 for better model fitting with the model of normal ogive [10, 11]. On the ground of this dependence are formed characteristic curves for every j task, the position of which in Cartesian plane is determined by the quality of tests. More detailed interpretation of received values and analysis of characteristic curves are also presented in [10, 11].

This way it is possible with the use of 3PL model to calculate probabilistic characteristics of test items, on the ground of which the items, which meet the demands of reliability, may be chosen.

3 Tests Development Model in Learning Management Systems

For supporting the process of decision making for tests formation in the process of distance learning the generalized structural model is suggested, which can be realized in LMS Moodle (Fig. 1). The main test development steps with account of ability level students are the following:

- qualitative learners ability level analysis by primary testing (initial check);
- selection of relevant category training courses; material learning analysis by intermediate tests (tests 1, 2, … i, … N-1);
- qualitative analysis of intermediate tests (using IRT models);
- accumulation of tests base (data of tests are corrected after each entering the test);
- forming the final tests (test N) with appropriate psychometric characteristics by all courses categories;
- the final control and estimation of ability level learners.

The courses may be presented by the separate modules of one discipline or a number of courses, which need to be learnt to acquire specific knowledge.

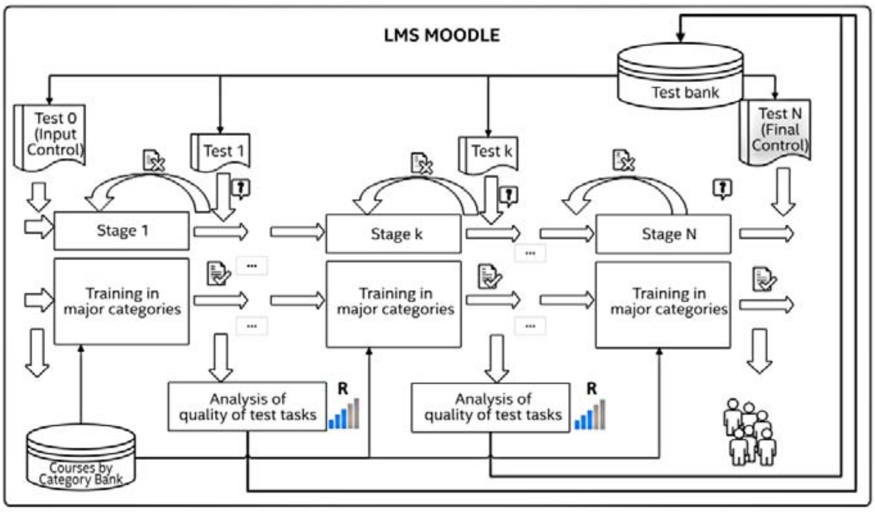


Fig. 1. Structural model of test development in the process of e-learning using LMS Moodle (Source: Own work)

The collection of empirical results of testing is the preparation step. In fact, the receiving of solid results, reflecting the reliable estimates regarding the preparation level of learners, is possible only after some iteration of system operation with current bank of tests. This is necessary while the information accumulation will be carried out during certain time and psychometric indexes of tests will be recalculated.

Total points, received in login scale (according to the IRT models), allows to estimate the tested by the ability level. The initial difficulty level of test (for the initial check test) is determined by the teacher.

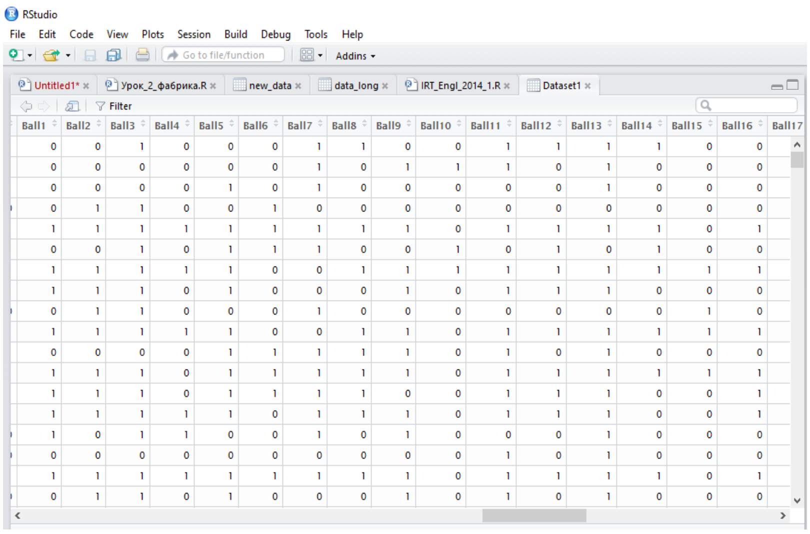
The selection of tests is carried out in random manner, which provides proper evaluation in the process of testing (the students cannot know the answers, cannot pass the information to each other about right answers, all are taking the test in equal conditions). In this research have not yet been considered the issues of automatic selection of courses in accordance with primary training of learners and the time test characteristics have not been monitored.

It is suggested to carry out all calculations of psychometric test characteristics using software functions with open source code R-Studio. Special package *ltm* exists for the determining of the main characteristics and evaluation of received models quality. Description of the package functions can be found both in R-Studio reference material, and in scientific journal Journal of Statistical Software [13].

4 Quantitative Experiment

Let us consider some calculation results of tests parameters, which were received based on the testing results of intermediate stage of discipline "Information technologies" study cycle in National University of Life and Environmental Science (NULES) of Ukraine (http://elearn.nubip.edu.ua/enrol/index.php?id=230) and in National Aerospace University "KhAI" (http://stm.khai.edu/course/index.php?categoryid=4). The experimental research was carried out during 2013-2015. The volume of sampling population amounted to 520 students – masters of 1 year of study from the two universities. Students were offered the same sets of testing for the input testing of ICT essentials. The level of initial preparation of students can be considered the same, because the level of IC-competence doesn't influence entering the masters course.

Testing results are represented in form of the rectangular matrix with zeros and, ones, since there were only dichotomous tasks in the tests. The results were received in LMS Moodle and exported to R-Studio (Fig. 2).



Ball1	Ball2	Ball3	Ball4	Ball5	Ball6	Ball7	Ball8	Ball9	Ball10	Ball11	Ball12	Ball13	Ball14	Ball15	Ball16	Ball17
0	0	1	0	0	0	1	1	0	0	1	1	1	1	0	0	
0	0	0	0	0	0	1	0	1	1	1	0	1	0	0	0	
0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	
0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	
1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	
0	0	1	0	1	1	1	0	0	1	0	1	0	1	0	0	
1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	
1	1	1	0	1	0	0	0	1	0	1	1	1	0	0	0	
0	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	
1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	
0	0	0	0	1	1	1	1	1	0	1	0	1	0	0	0	
1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	
1	1	1	0	1	1	1	1	0	0	1	1	1	0	0	1	
1	1	1	1	1	0	1	1	1	0	1	1	1	0	0	1	
1	0	1	1	0	0	1	0	1	0	0	0	1	0	0	0	
0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	
1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	
0	1	1	0	1	0	0	0	1	0	1	0	1	0	0	0	

Fig. 2. Matrix of testing results (Source: Own work)

There were analyzed 48 test items on the basis of different IRT models. In the process of research were evaluated the parameters of test items by classical theory and using 3PL model. For each tasks were calculated the parameters of difficulty, discrimination and guessing. Table 1 shows some results of calculation for three-parameter model for the easiest and difficult items. In the last column the probability to give correct answer for each test items (P(x=1|z=0)) is given.

Table 1. Psychometric characteristic according to IRT three-parameter models

Test item (number)	Model 3PL			
	Guessing	Difficulty	Discrimination	P(x=1|z=0)
7	0,660	1,462	3,236	0,663
11	0,091	-0,921	2,351	0,906
3	0,084	-0,859	2,372	0,894
…	…	…	…	…
34	0,205	1,168	4,064	0,213
18	0,033	1,551	2,587	0,050
21	0,132	1,760	1,940	0,161
26	0,227	2,872	1,065	0,261
25	0,186	1,907	3,511	0,188

The analysis of the given results has shown that test items 18, 34, 21, 26, 25 and 26 do not satisfy the demands of coherence and reliability of tests by different parameters: they are difficult, have high level of guessing or low discrimination, which, in its turn, causes the low probability to receive the correct answer for given tasks. These tasks were removed from bank of test items and were not included into the final test.

According to the results of calculation also were built characteristic and information curves, which reflect the quality of test items. Fig. 3 and Fig. 4 shows the curves for 3PL model.

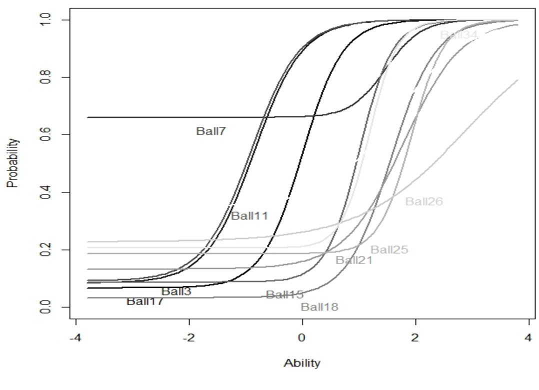


Fig. 3. Item Characteristic Curves for some test items data-set under model 3PL (Source: Own work)

It can be seen on the graph, that the curves have accumulated to the left of zero, which confirms the received results – the most of test items are difficult. Also, it is apparent that, some items have low parameters of discriminatory power or high parameters of guessing (for example 7, 11).

It can also be seen on the figure, which tasks have the low probability of receiving the correct answer according to model 3PL: 15, 18, 25.

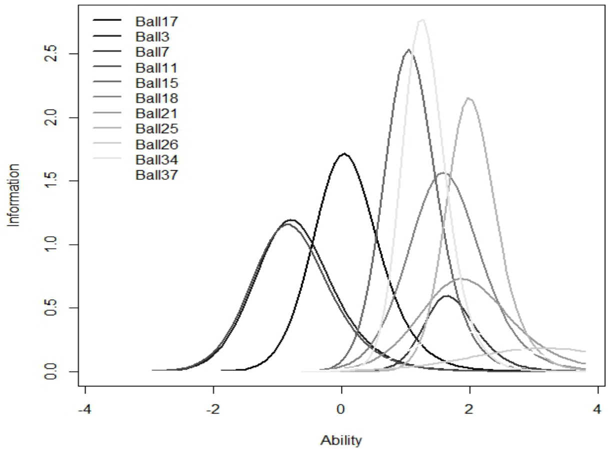


Fig. 4. Item Information Curves for some test items data-set under model 3PL
(Source: Own work)

Thus, the functional R allows us to fully estimate the quality of test tasks and the test in general. In the suggested system of e-learning (Fig.1) the analysis of psychometric characteristics is proposed to be made automatically, i.e. user is offered the list with tasks numbers, which do not meet the demands.

5 Conclusions

This paper describes a method as a technical tool for the quality test items bank forming. Classic test theory and dichotomous logistic item response models are used to estimate the psychometric test items characteristics.

The main assets of this approach use the Moodle system to accumulate and to keep the intermediate results of testing during the process of studying, which allows us to create the calibrated test items bank. That is the reason why using the R package allows us quickly and accurately to calculate the main characteristics of test items and tests overall. Due to the R package, various items analyzing methods become available.

Therefore, due to the logistic IRT model performed in the R package, this method is flexible and easy to administrate, as far as it requires no special knowledge in the field of statistical data processing.

The results of the test received during the experiment held in two Ukrainian universities show us the effectiveness of the method.

We believe this paper is only a small step towards this direction, both on the methodological and the practical aspects for creation quality tests for e-Learning. In perspective it is suggested to consider other working parameters with distance courses, which will allow to evaluate the quality of training in whole, on the basis of analyzing different activities of students, not only performance at tests.

References

1. Miller, R., Shapiro, H., Hilding-Hamann, K. E.: School's Over: Learning Spaces in Europe in 2020: An Imagining Exercise on the Future of Learning. Office for Official Publications of the European Communities, online: http://ftp.jrc.es/EURdoc/JRC47412.pdf (2008)
2. Johnson, L., Adams Becker, S., Estrada, V., Freeman, A.: NMC Horizon Report: 2014 Higher Education Edition, New Media Consortium, online: http://cdn.nmc.org/media/2014-nmc-horizon-report-he-EN-SC.pdf (2014)
3. Johnson, L., Adams Becker, S., Estrada, V., Freeman, A.: NMC Horizon Report: 2015 Higher Education Edition, New Media Consortium, online: http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf (2015)
4. Moore, J. L., Dickson-Deane, C., Galyen, K.: E-Learning, online learning, and distance learning environments: Are they the same? Internet and Higher Education, 14(2), 129—135 (2011)
5. Hongjiang, X., Mahenthiran, S., Smith, K.: Effective Use of a Learning Management System to Influence On-Line Learning. In: Proc. 11th Int. Conf. on Cognition and Exploratory Learning in Digital Age (CELDA), Porto, Portugal, Oct 25-27, 2014. online: http://files.eric.ed.gov/fulltext/ED557395.pdf (2014)
6. Dias, S. B., Diniz, J. A.: Towards an Enhanced Learning Management System for Blended Learning in Higher Education Incorporating Distinct Learners' Profiles. Educational Technology & Society, 17, 307--319 (2014)
7. Kim, V. S.: Testirovanie uchebnyih dostizheniy. Monografiya. Izd-vo UGPI, Ussuriysk, Russian Federation (2007). (in Russian)
8. Crocker, L., Algina, J.: Introduction to Classical and Modern Test Theory. Cengage Learning Pub., Ohio, USA: (2006)
9. Mazorchuk, M., Dobryak, S. S., Bondarenko, E. O.: Otsenka kachestva testov na osnove analiza distraktorov po metodu porogovyih grupp. Radioelektronni i komp'yuterni sistemi, 62 (3). 39--44 (2013) (in Russian)
10. Lisova, T.V.: Modeli ta metody suchasnoyi teoriyi testiv: Navchal'no-metodychnyy posibnyk Vydavets' PP Lysenko M.M., Nizhyn, Ukraine (2012) (in Ukrainian)
11. Baker, F.B.: The Basics of Item Response Theory. ERIC Clearing house on Assessment and Evaluation, USA (2001)
12. Anisimov, A. M.: Rabota v sisteme distantsionnogo obucheniya Moodle. HNAGH, Harkov, Ukraine (2008) (in Russian)
13. Rizopoulos, D.: An R package for latent variable modelling and item response theory analyses. Journal of Statistical Software, 17(5). 1—25 (2006)