TEM Journal. Volume 12, Issue 1, pages 303-308, ISSN 2217-8309, DOI: 10.18421/TEM121-38, February 2023.

Objective Assessment During a Pandemic

Boryana Uzunova-Dimitrova [1], Hristo Paraskevov [1], Stanimir Zhelezov [1]

[1] *Konstantin Preslavsky University of Shumen, Faculty of Mathematics and Computer Science, 115 Universitetska str, Shumen, Bulgaria*

Abstract **– Life during the pandemic tested the objective assessment of learners' knowledge. By using criteria and grammatical rules for decision-making, the informal evaluation problem is resolved in a variety of unstructured situations.**

The current study's objective is to evaluate and compare the knowledge that has been acquired and competencies by the learners on the basis of online presented learning material and subsequent online assessment. The achieved results obtained during the evaluation of a test and exclude a constructed course project on the teaching material in the discipline "Web technologies" are compared.

A formalized approach based on non-standard application of the one-parameter Rush model with fuzzy evaluations of the criteria was used for the two applied evaluation methods. The model enables program implementation and can be embedded in e-learning platforms.

Keywords **– online evaluation, formalization of the evaluation process, one-parameter Rasch model, Rasch measurement.**

1. Introduction

During the past year, training has posed significant challenges to the global COVID-19 pandemic, necessitating extensive distance learning and subsequent distance assessment.

DOI: 10.18421/TEM121-38
https://doi.org/10.18421/TEM121-38

Corresponding author: Boryana Uzunova-Dimitrova, *Konstantin Preslavsky University of Shumen, Faculty of Mathematics and Computer Science, 115 Universitetska str, Shumen, Bulgaria*
Email: *b.uzunova@shu.bg*

Received: 10 October 2022.
Revised: 09 December 2022.
Accepted: 14 January 2023.
Published: 27 February 2023.

In practice, the classical methods of teaching and assessment have been replaced by the use of information and communication technologies, which transform traditional learning into electronic or distance learning.

The purpose of this study is to analyze the results obtained, as the material is transmitted entirely online, in the case of videoconferencing via Google Meet. The experiment was conducted with 52 students studying the discipline "Web technologies". The examination of the acquired knowledge, skills and competencies was carried out by means of an electronic test and a developed course project. The test consists of 30 multiple-choice questions and answers, which are divided according to difficulty into two groups, with the test questions falling at random from the test bank, which contains twice as many questions. The Theory of Position Response (IRT) was used to evaluate students' test control, and some of the models were implemented in software [8]: Bilog, Multilog, Parscale (Birnbaum models and their extensions for polytomy tasks); Veneers (Rasch's multi-faceted model); RUMM (one-dimensional Rasch measuring models); Conquest (multidimensional models), and so on.

The course project is more informative [2], [10] than the formalized tests, through it the shortcomings of the test systems are avoided and the creative thinking is developed, the acquired knowledge is realized in practice. Its evaluation is an intellectual process [6]. A model is employed for the evaluation of the course materials utilizing artificial intelligence techniques that enable adequate modeling of the decision-making process for the evaluation of students' knowledge. Knowledge is formalized using fuzzy set models, fuzzy logic, and other formalities for modeling in uncertain situations.

2. Objectives, methodology and research design

Purpose of the study

In situations of various types of uncertainty, the informal evaluation problem is resolved using standards and linguistic conventions for making judgments.

A formalized approach to evaluation by unconventional application of the one-parameter Rasch model is proposed. The paradigm enables software implementation and is applicable to all training formats, but it is particularly helpful for distance e-learning and appropriate assessment.

Formalization of the problem for evaluation of a course project

The formalization and adequate modeling of the teacher's thought process in the process of making a decision for the formation of the assessment are prerequisites for the automation of the control procedures. The task for assessment of students knowledge is linguistically described, multi-criteria and informal, which makes it difficult to model. It does not lend itself to strict formalization and the process of interaction between the evaluator and the evaluated. At the same time, the development of software for control of students' knowledge requires, if not complete, then at least partial formalization of the educational subject area, as only objective and formal methods can generate unambiguous and reproducible assessments.

We use criteria representing the linguistic variables that make it easier to judge, with each criterion assuming a value from a predetermined term set {very bad, bad, not very good, good, very good, excellent, perfect}.

For formalization of the task and interpretation of the results of the criterion-oriented testing and evaluation of the quality of the course project the following solution is proposed, and in case of different interpretation of the basic parameters another model could be presented. To do this, let's introduce the following notations:

$S=\{s_1, s_2, \dots, s_m\}$ – a finite, discrete set of individuals to be evaluated;

D – finite, discrete set of diagnoses, final estimates {2, 3, 4, 5, 6};

$T=\{t_1, t_2, \dots, t_n\}$ – a finite, discrete set of expertly set test with test questions t_n, n=1, 2, …, k or criteria for assessing the value of the course project;

A = ||aij||, i=1,2,…m, j=1,2,…n – a matrix containing the results of the evaluation, where $a_{ij} \in L$ is the evaluation of the s_i-th work on the t_j-th test question / criterion;

L={0, 1} – discrete scale of values of the scores of each individual test question or criterion {1-correct answer, 0-wrong answer }.

Therefore, the task of interpreting the results could be formulated as follows: "For each evaluated person $s_i \in S$ to determine a final grade $d \in D$, based on the results obtained in A from the evaluations of the test questions / criteria $t_j \in T$ set on the rock L".

Assuming that each diagnosis from the term set of the studied variable is given linguistically and the test subject chooses the answer li, to test question ti from the scale L, representing the interval [0; 1] or a linear ordinance $(l_1, l_2, \dots, l_i)$.

Consider the Cartesian product $\overline{T}$ of the fuzzy test questions / criteria t_i ():

$$\overline{T} = t_1 \times t_2 \times \dots \times t_n \tag{1}$$

$\overline{T}$ represents a multidimensional fuzzy set on the set of those evaluated with an affiliation function μ ():

$$\mu: X \rightarrow L_1 \times L_2 \times \dots \times L_n, \tag{2}$$

if provided that $L_1 = L_2 = \dots = L_n$, the fuzzy set is homogeneous otherwise it is heterogeneous.

We assume that as a result of processing an expert opinion a number $\xi_i \in [0,1]$ can be determined, representing a quantitative assessment of the question / criterion t_i on the diagnosis $d \in D$. The vector $(\xi_1, \xi_2, \dots, \xi_n)$ is built for each diagnosis and can be thought of as a vector-valued degree of membership in the multidimensional fuzzy set $\overline{T}$.

On the other hand, the vector $(l_1, l_2, \dots, l_n)$ of the received answers $l_i \in L$ of the evaluated person $si \in S$ to the questions / criteria $t_i \in \overline{T}$ is a subjective measure of how much si corresponds to the studied variable, the meaning of which is formalized by $\overline{T}$.

In the general case, the vector $(l_1, l_2, \dots, l_n)$ contains numerical values 0 and 1. For each scale L_i we set the function γ() expertly:

$$\gamma_i: L_i \in [0; 1], \tag{3}$$

by which we establish an unambiguous correspondence between the sets $L_1 \times L_2 \times \dots \times L_n$ and $[0; 1]^n$. Thus, each answer $l_i \in L$ can be transformed into $\varphi_i \in [0,1]$. Formally, the transformed vector of answers $(\varphi_1, \varphi_2, \dots, \varphi_n)$ can be considered as a vector-valued degree of belonging of x to $\overline{T}$.

As a result, for which for each diagnosis and for each evaluated person we obtain an expertly set or subjectively determined degree of belonging to $\overline{T}$.

It is clear that the predominant value of the vector is determined by the answers of the tested or the realized criteria in its work, then the value can be used to determine its diagnosis.

Thus, the task of determining the diagnosis of the evaluated is reduced to estimating the closeness between two elements s and d, defined by their degrees of belonging to the same fuzzy set $\overline{T}$.

As a weight for the proximity between the two elements s and d in T it is possible to use the following formula:

$$\lambda(x, d) = \frac{1}{n}\sum_{i=1}^{n}\left|\xi_i - \varphi_i\right|. \tag{4}$$

Then, the diagnosis of the evaluated s can be determined by:

$$\min(\lambda(x, d_1),\ \lambda(x, d_2),\ \dots\ \lambda(x, d_n)), \qquad d_i \in D.$$

The proposed method for modeling a class of tasks allows for appropriate interpretation of test results or fulfillment of certain criteria.

The use of scales other than the binary one increases the adequacy of the model and allows the test takers to more easily assess their compatibility with the qualitative features subject to research and evaluation.

Despite the complex mathematical justification, the method can be easily applied in practice. Allows software implementation, as a result of which the work on the interpretation of the results can be automated.

The one-parameter Rush model, one of the primary formalization techniques in probabilistic modeling for quantifying student achievement, is used to objectify the assessment (IRT). It develops a relationship between the test's observed outcomes and two sets pertaining to the test's difficulty (T) and the learners' level of preparation (S) [4].

At the heart of Rush's model is the assumption that the higher the learner's preparation, the more likely he or she is to answer the test questions correctly, which does not contradict real practice. In addition, the model assumes that the preparation of learners and the difficulty of the tests used are objective variables and can be measured objectively. The level of preparation of the test takers, whose outcomes serve as the basis for determining the difficulty, determines the test's level of difficulty. Trainee training, measured repeatedly by different tests / criteria with different difficulty, will differ only from the inevitable measurement errors, but not as a result of differences in the difficulty of the tests / course projects [1], [3], [5].

To apply the Rush model, we assume that the learner's preparation and the complexity of the test / course project are assessment parameters that allow for objective assessment.

According to Rush's model [3], [7], we can assume that the following formula (5) determines the probability P for a student with experience S to answer a test question correctly / produce a high-quality project with difficulty T:

$$P(S,T) = \frac{S}{S+T}, \tag{5}$$

If we input the notations

LN(S)=A, $\Rightarrow$ S = EXP(A),

LN(T)=B, $\Rightarrow$ T = EXP(B),

for P we get the basic logistic model of Rush:

$$P(S,T) = \frac{EXP(A)}{EXP(A) + EXP(B)} = \frac{1}{1 + EXP(B-A)} \tag{6}$$

The model only has one parameter because the likelihood of success is determined by the difference between B and A. Logs are used to measure the parameters of models A and B, which describe the learner's readiness and the test's difficulty.

The probability that a standard learner will correctly answer a standard test is 0.5 if A = B = 1 logit. The likelihood of getting the right answer is P $\rightarrow$ 1 if the learner's preparation exceeds the test's difficulty (B - A $\rightarrow$ - ∞). If the learner's preparation is significantly less than the test's level of difficulty (B - A $\rightarrow \infty$), the likelihood that they will provide a correct response is zero (P $\rightarrow$ 0).

The dichotomous scale of {yes, no} ≡ {0, 1} is used to apply Rasch's model; with a few minor adjustments, the Rasch model can be used with the scale L = {bad, good, excellent} ≡ {0, 0.5, 1} [9]. Compared to scales where the number of terms is k> 3, this scale is convenient for teachers and provides a greater possibility to create an unambiguous assessment.

The evaluation is carried out in the order listed below:

1. On a scale of L = {0, 0.5, 1}, the teacher grades the project or test questions. The outcome of the expert evaluation is a matrix A with the dimensions m × n, where m represents the quantity of tested exams or course projects and n represents the quantity of test questions or criteria.

2. We total the grades by rows to determine the students' primary score b_i, i = 1, 2,… m:

$$b_i = \sum_{j=1}^{n} a_{ij} \tag{7}$$

3. Using the formula, we can calculate the p_i, i = 1, 2,… m:

$$p_i = \frac{b_i}{n} \tag{8}$$

The extreme scores are disregarded in the following manner: if b_i = 0, we set p_i = ε; if b_i is equal to the highest score, we set p_i = 1 - ε, where is a sufficiently small number, for instance ε = 0.001.

4. The following formula determines the initial approximation of the evaluation of the i-th work:

$$A_i = LN\left(\frac{p_i}{1-p_i}\right), i = 1,2, \ldots m \tag{9}$$

5. Calculate the primary score of the criteria, which is the sum of the scores for each of the pillars, j = 1, 2,..., m.

6. Use the formula to calculate the values p_j, j = 1, 2,..., m.

$$p_j = \frac{c_j}{m} \tag{10}$$

Similar to item (3), if c_j = 0, we presumptively assume p_j = ε; if c_j is equal to the maximum score, we presumptively assume p_j = 1 - ε.

7. In accordance with the formula, we determine the initial meanings of the test questions' and criteria's degree of difficulty:

$$B_j = LN\left(\frac{1-p_j}{p_j}\right), j = 1,2,\dots m \qquad (11)$$

8. A linear transformation of $A_i \in [\min(A_i), \max(A_i)]$, i = 1, 2, … n in the scale [2,3,4,5,6] ≡ {weak, medium, good, very good, excellent} yields the final evaluation of the completed test or course project.

Criteria for evaluation of a course project

The choice of criteria for evaluating the quality of the project is essential for the objectivity of the evaluation. The criteria are described linguistically. The lecturer follows the informal rules, which lead to an ambiguous assessment due to the subjectivity of the experts and the linguistic uncertainty of the terms used.

The criteria used to assess the quality of course projects are obtained on the basis of analysis of Internet sources and consultations with teachers using this method of assessing the knowledge of students in our faculty, which are precisely specified for the creation of a website.

For the purposes of the experiment, 30 criteria for evaluation of the quality of the created course project were selected, which are divided into four categories, having different weights for the formation of the final evaluation.

***Criteria*:**

Functional and logical completeness - completeness 1

1. The site has a clearly defined thematic goal and implements it through content.
2. Completeness and reliability of the information.
3. Functionality.

Convenient to use - weight 2

4. Well-named program elements in HTML and CSS.
5. Arranged structure of working directories.
6. Code order and readability without compression.
7. Intuitive navigation with a clear hierarchy of content and design of elements.
8. Intuitive interface. Easy orientation in the navigation structure of the site.
9. Availability of a working, convenient and intuitive navigation system (menu).
10. Site loading speed.

Graphical interface and design - weight 2

11. Aesthetically sound and pleasant graphical interface.
12. Conformity of the graphical interface in terms of content and navigation.
13. Respect for copyright (for images, icons, buttons, etc.) when presenting information on the site.
14. Use of author's (personal) resources (pictures, icons, buttons).
15. Evenly distributed pictures and text.
16. Proportionally sized and quality images.
17. Effects or processing of graphic objects and menus.
18. Using a gallery.
19. Screen readability (appropriate font families, contrast and size).
20. Appropriately selected color combinations on the site.
21. Shaped style for page titles.
22. Use text and paragraph formatting other than the default formatting.
23. Well-chosen background of the pages.
24. Using different types of hyperlinks.
25. Video, audio or light animation to improve the quality of the site.
26. The site contains data for feedback - contacts, form and / or forum.
27. Individuality in the implementation of the practical task.

Bonus elements - weight 1

28. Well-formed documentation and disk media.
29. Creativity and aesthetics.
30. Contains a minimum of 9 pages.

3. Experimental results

To see if the suggested procedure might be used, an experiment was run. The generated course projects and completed exams by University of Shumen students majoring in "Web technologies" were assessed using the above-described approach that makes reference to Rush's model. Tables 1 and 2 represent some of the 52 experiments which were conducted.

Table 1. Results of experiments for assessing pupils' test scores

	test questions: rank 1																	test questions: rank 2																
Student	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}	t_{29}	t_{30}	Primary ball b_i	P_i	$A_i=Ln(P_i/(1-P_i))$ in logits	Evaluation
s_1	1	1	1	1	1	1	0,5	1	1	1	1	1	0	1	1	1	0	1	1	0,5	1	1	1	1	1	1	1	1	1	1	48	0,889	2,070	6
s_2	1	1	1	0	0	1	1	0,5	1	0	1	1	0	1	1	1	1	0	0	0	1	1	1	1	0	1	1	1	1	1	37	0,685	0,776	5
s_3	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	44	0,815	1,476	6
s_4	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0,5	1	1	1	1	1	1	1	1	46	0,852	1,742	6
s_5	1	1	1	0	0	1	1	1	1	0	1	1	0	1	1	1	1	0	0	0	1	1	1	1	0	1	1	1	1	1	38	0,704	0,862	5
s_6	1	1	0,5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	49,5	0,917	2,386	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
[illegible]	1	1	1	0	0	1	0,5	1	1	0	1	1	1	1	0,5	1	1	0,5	0	0,5	1	1	0	1	1	0	1	0	1	1	37	0,685	0,776	5
[illegible]	1	1	0,5	1	1	1	1	1	1	0,5	1	1	1	1	0,5	1	1	1	1	0,5	1	1	1	1	1	1	1	0,5	1	1	50	0,926	2,512	6
[illegible]	1	1	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	42	0,778	1,248	5
[illegible]	1	0	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	1	0,5	1	1	0,5	1	1	1	0	1	0	41	0,759	1,144	5
[illegible]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	54	0,999	6,214	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
[illegible]	1	0	1	1	0,5	1	1	0,5	1	0,5	1	0	1	0,5	0,5	1	0	1	0	1	1	0,5	0	0,5	0	1	0	1	1	0	33	0,611	0,449	5
[illegible]	1	0	1	0	1	0	0	1	0,5	1	1	0	1	1	0,5	1	1	1	0,5	1	0	0	0,5	0	0	1	0	1	1	1	31	0,574	0,296	3
[illegible]	0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	42	0,778	1,248	5
[illegible]	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0,5	1	1	1	1	1	1	1	0	1	48	0,889	2,070	6
[illegible]	1	1	1	0	0	1	0,5	1	1	0	1	1	0	1	1	0,5	1	0	0	0	1	1	0	1	0	1	1	1	1	1	34	0,630	0,528	5
[illegible]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	54	0,999	6,214	6
	12	8	12	10	8,5	11	11,5	12,5	12,5	8	12	9	9	11,5	10	12	10	11,5	7,5	8	11	10	10,5	10	8	11	10	10,5	12	11				
P	0,92	0,62	0,92	0,77	0,65	0,85	0,88	0,96	0,96	0,62	0,92	0,69	0,69	0,88	0,77	0,92	0,77	0,88	0,58	0,62	0,85	0,77	0,81	0,77	0,62	0,85	0,77	0,81	0,92	0,85				
Logit	-2,48	-0,47	-2,48	-1,20	-0,66	-1,70	-2,04	-3,22	-3,22	-0,47	-2,48	-0,81	-0,81	-2,04	-1,20	-2,48	-1,20	-2,04	-0,31	-0,47	-1,70	-1,20	-1,44	-1,20	-0,47	-1,70	-1,20	-1,44	-2,48	-1,70				

Table 2. Results of experiments for assessing pupils' course projects

criterion	Functional and rank 1			Convenient to use rank 2							Graphical interface and design rank 2																	Bonus items rank 1						
Student	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}	t_{29}	t_{30}	Primary ball b_i	P_i	$A_i=Ln(P_i/(1-P_i))$ in logits	Evaluation
s_1	1	1	1	1	1	1	0,5	1	1	1	0	1	0	1	1	1	0	1	1	0,5	1	0,5	0,5	1	1	1	1	1	1	1	44	0,815	1,476	6
s_2	1	1	1	0	0	1	1	0,5	1	0	1	1	0	1	1	1	1	0	0	0	1	1	1	1	0	1	1	1	1	1	37	0,685	0,775	5
s_3	1	1	1	0	0	1	1	1	1	0	0	1	0	1	1	0	1	0	0	0	1	1	1	1	0	1	1	1	1	1	34	0,630	0,528	5
s_4	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0,5	1	1	1	1	1	1	1	1	46	0,852	1,742	6
s_5	1	1	1	0	0	1	1	1	1	0	1	1	0	1	1	1	1	0	0	0	1	1	1	1	0	1	1	1	1	1	38	0,704	0,862	5
s_6	1	1	0,5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	49,5	0,917	2,386	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
[illegible]	1	1	1	0	0	1	1	1	1	0	0,5	1	0	1	0,5	1	1	0,5	0	0	1	1	0,5	1	0	0	1	0	1	1	33	0,611	0,449	5
[illegible]	1	1	0,5	1	1	1	1	1	1	0,5	1	1	1	1	0,5	1	1	1	1	0,5	1	1	1	1	1	1	1	0,5	1	1	50	0,926	2,512	6
[illegible]	1	1	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	42	0,778	1,248	5
[illegible]	1	0	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	1	0,5	1	1	0,5	1	1	1	0	1	0	41	0,759	1,144	5
[illegible]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	54	0,999	6,214	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
[illegible]	1	0	1	1	0,5	0	0,5	0,5	0,5	0	0,5	1	1	0	0	1	0	1	0	0	0,5	0,5	0	0,5	0	0	0	1	0	0	21	0,389	-0,454	3
[illegible]	1	0	1	0	1	0	0	1	0,5	1	1	0	1	1	0,5	1	1	1	0,5	1	0	0	0,5	0	0	1	0	1	1	1	31	0,574	0,296	4
[illegible]	0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	42	0,778	1,248	5
[illegible]	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0,5	0,5	0,5	0,5	0,5	1	0,5	1	1	1	1	0	1	43	0,796	1,358	5
[illegible]	1	1	1	0	0	1	0,5	1	1	0	1	1	0	1	1	0,5	1	0	0	0	1	1	0	1	0	1	1	1	1	1	34	0,630	0,528	5
[illegible]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	54	0,999	6,214	6
	13	9	13	9	7,5	11	12	13,5	13	7,5	11	11	8	12	10,5	11,5	11	10	7	6	11,5	10,5	11	10,5	7	11	11	11,5	12	12				
P	0,93	0,64	0,93	0,64	0,54	0,79	0,86	0,96	0,93	0,54	0,79	0,79	0,57	0,86	0,75	0,82	0,79	0,71	0,50	0,43	0,82	0,75	0,79	0,75	0,50	0,79	0,79	0,82	0,86	0,86				
Logit	-2,56	-0,59	-2,56	-0,59	-0,14	-1,30	-1,79	-3,30	-2,56	-0,14	-1,30	-1,30	-0,29	-1,79	-1,10	-1,53	-1,30	-0,92	0,00	0,29	-1,53	-1,10	-1,30	-1,10	0,00	-1,30	-1,30	-1,53	-1,79	-1,79				

In fig. 1. the characteristic curves of the students in relation to their test score are shown, and in fig. 2. - according to the course project developed by them.

It is observed that the curves differ only in their location, ie. the curves are parallel and non-intersecting as is characteristic of the characteristic curve of the one-parameter Rush model in IRT theory. To construct a characteristic curve, the student must be prepared (θ) and have used the relevant criteria in his or her course project.

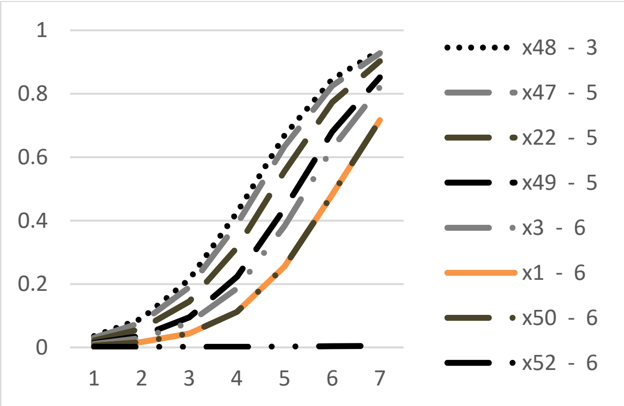

Figure 1. Achieved results of a test against the Rush assessment

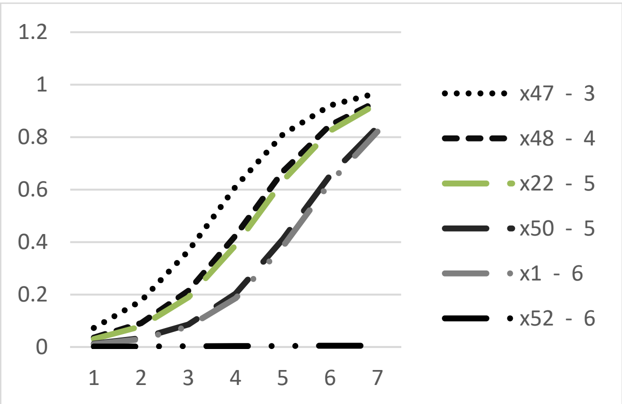

Figure 2. Achieved results of a course project compared to the Rush assessment

The following conclusions can be reached after analyzing the findings:

• The proposed model implements a multicriteria approach for measuring the quality of the test / course project of students.

• Test questions / criteria must allow for an objective assessment. Therefore, they must be accurate, clear, unambiguous, without the use of logical connections (and, or, not).

• The described algorithm allows obtaining an approximate assessment of the test / course project. Assessment according to the Rush model is favorable to knowledgeable students, answering test questions independently or developing their own course projects, their characteristic curve is completely concave (x_{47}, x_{48}) and, accordingly, unfavorable to ignorant ones, whose characteristic curve is convex, and students with standard characteristic curve for the Rush model or parallel to the abscissa (x_1, x_{22}, x_{50}, x_{52}) correspond to well-prepared students.

• The results of the thus assessed tests / course projects of the students shown in Figures 1 and 2 show that the assessment method used is reliable and objective.

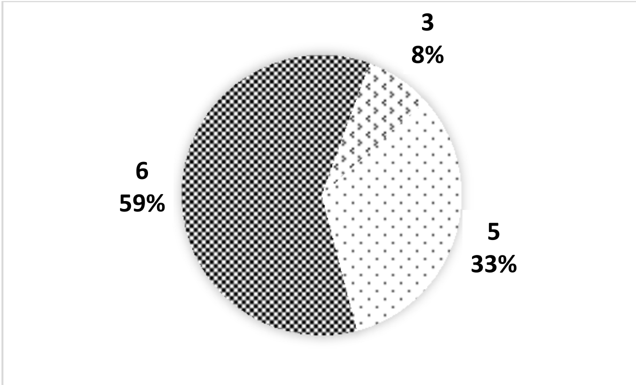

Figure 3. Ratio of achieved test results and number of students

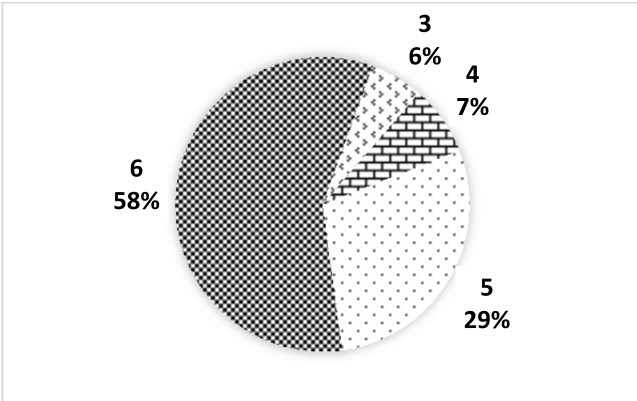

Figure 4. Ratio of achieved results of the course project and number of students

Figure 3 and Figure 4 show the ratio of the achieved results of the test and the presented course project and the number of students. It is observed that the obtained results overlap or approach each other, which in turn shows that the assessments of students are objective to their acquired knowledge. The evaluation of the test is commensurate and realistic with the evaluation of the course project.

The expectations for overlapping the grades from the two types of tests are justified, which is shown in Figure 5 where the results achieved by the learners are presented.

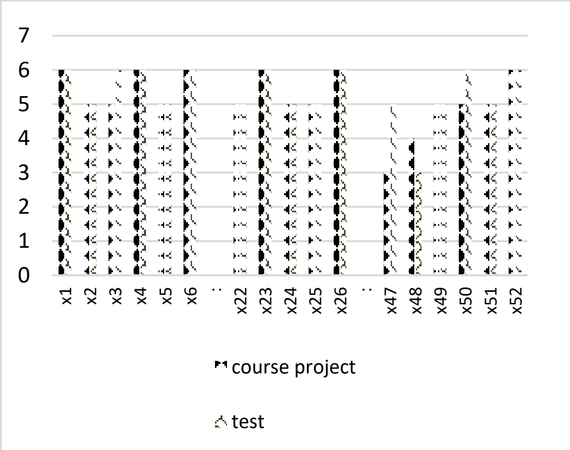

Figure 5. Achieved results when testing a practical task and through a test

4. Conclusion and future work

In an online learning, the learner needs at least twice as much time to prepare and master the learning material. When teaching new material online, the teacher must also distribute his attention: to follow the sequence of his presentation; to present short and clear examples; to receive an immediate response from the learners, whether the material is understood or there is a need for re-explanation, following their faces, etc.

The learning process in a pandemic environment requiring only online learning, teachers must provide their learners not only through the teaching material taught in the traditional way, but also, if possible, to record lessons and upload to an e-learning platform from where there will be an opportunity for further viewing, each individual lecture should have a well-structured content and be lined with many examples, to upload materials on the platform for the taught material, to give independent work to students through which the teacher can see the level of assimilated material. The relationship between teacher and learner in an online learning environment should be constant, which would contribute to an effective and quality learning process. Assessment in an online learning should be based on at least two different assessment methods, not just one, because in this way the teacher can be sure that the assessment is real and commensurate with the learner's knowledge.

The proposed method for assessment of knowledge could be implemented in software as part of a developed "Web-based intelligent system for students knowledge management", which will facilitate teachers for objective assessment electronically.

Acknowledgements

The article is partially funded by the project RD-08-147/02.03.2022.

References

[1]. Baker, F. B. (2001). *The basics of item response theory.* United States of America: Heinemann.
[2]. Black, P., & Dylan Wiliam. (2006). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102
[3]. Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (4th ed.). New York: Routledge.
[4]. Chelyshkova M.B., & Shmelev A.G. (2004). Scaling results Unified exam: problems and prospects. Educational Issues, no. 4, 168-186.
[5]. Fischer, G. H., & Molenaar, I. W. (Eds.). (2012). *Rasch Models: Foundations, Recent Developments, and Applications.* Springer Science & Business Media.
[6]. Garvanov Ivan, & Magdalena Garvanova. (2021). New Approach for Smart Cities Transport Development Based on the Internet of Things Concept. In *17th Conference on Electrical Machines, Drives and Power Systems (ELMA).* IEEE. doi:10.1109/ELMA52514.2021.9503084
[7]. Maslak, A. A., G. Karabatsos, T. S. Anisimova, S. A. & Osipov, S. A. (2005). Measuring and Comparing Higher Education Quality between Countries Worldwide. *Journal of applied measurement, 6*(4), 432–442.
[8]. Tomasheva, L. Y. (2010). System for assessing the quality of knowledge of distance learning students. *Polzunovskiy al'manakh*, 196-198. Retrieved from: http://elib.altstu.ru/elib/books/Files/pa2010_2/pdf/196tomasheva.pdf [accepted: 05 September 2022].
[9]. Uzunova-Dimitrova, B. (2016). Methods and algorithms of artificial intelligence for objectification in assessment in higher education. Shumen: Konstantin Preslavsky University of Shumen.
[10]. von Davier, M., & Lee, Y. S. (2019). Handbook of diagnostic classification models. *Cham: Springer International Publishing.*