IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 10, Issue 12, December 2020, ||Series -II|| PP 06-11

Intelligent Methods for Objective Assessment of Learners in Online Testing

Uzunova-Dimitrova Boryana
Konstantin Preslavsky University of Shumen, the Faculty of Mathematics and Computer Science, Department Computer Systems And Technologies
Received 15 December 2020; Accepted 29 December 2020

ABSTRACT: A formalized approach to the assessment of test and written work of learners is proposed through the unconventional application of Rasch's dichotomous model using fuzzy evaluations of the criteria. The model allows the implementation of programs and can be embedded in e-learning platforms. Thus assessed the acquired knowledge, skills and competencies of the learners are objective, despite their remote assessment.
KEYWORDS: E-learning in higher education, assessment of students' written work, one-parameter Rasch model, Rasch measurement, formalization of the assessment process.

I. INTRODUCTION

Training at each level over the past year has been faced with serious problems caused by the global pandemic related to COVID-19, which has necessitated comprehensive distance learning and subsequent distance assessment.

In order to obtain an objective assessment of students' knowledge, various measuring tools are used: tests, essays, reviews, term papers, course projects, case studies, subject tasks, dissertations and others.

Item Response Theory (IRT) develops formal assessment models for test control of students' knowledge, some of which are implemented in software [1]: Winsteps (Dichotomous Rasch Model, Partial Credit Model, Rating Scale Model); 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), etc.

Test systems allow rapid and objective assessment of students' knowledge in a particular subject area, but also have major shortcomings, such as the ability to randomly select answers, the complexity of developing high quality tests, the lack of specialists in test theory, the inability to assess of abstract, creative knowledge, etc. In the conditions of electronic testing - the most serious problem is the verification with the Internet of the correct answer.

The desire to avoid the shortcomings of test systems increasingly draws the attention of experts to the use of written works as a tool for measuring the knowledge or skills that learners acquire in certain subject areas [2], [3]. The development of essays, course projects, etc. Develops creative thinking, ability to structure, summarize and analyze information, skills to use the acquired knowledge in practice and much more.

Creative works are more informative than formalized tests. Their evaluation is an intellectual process; its automation is associated with the implementation of high technology for text analysis. The models use artificial intelligence methods that allow adequate modeling of the decision-making process to assess students' knowledge. Semantic networks, neural networks, fuzzy sets, fuzzy logic, and other formalities for modeling in conditions of uncertainty are used to formalize knowledge graph models. These works are still theoretical and research in nature, but will undoubtedly lead to significant results and software methods for their implementation.

II. METHODOLOGY

In the current world situation, all teachers were faced with the need to find the most correct and correct way for the real assessment of the acquired knowledge, skills and competencies of the students, ie. the objective assessment of the learners' knowledge, which is present in a test conducted under normal conditions.

The problem of evaluation is informal and is solved in conditions of uncertainty of various types, determined by criteria and language rules for decision-making. Also, the other problem with objective assessment occurs if a test is used - with uncertain knowledge, the learner can play the lottery and accidentally give the correct answer or search for the correct answer on the Internet while solving the test, as well as another to solve the test on his behalf, etc ..

Avoiding these considerations requires the need to proceed to a comprehensive assessment of knowledge, namely to test the acquired knowledge, skills and competencies of learners through more than one approach. Therefore, in the forthcoming article the two approaches are considered, described and compared -

test testing and defense of a course project, and for this purpose the experiment was conducted with students from the University of Shumen studying the discipline "Informatics and Information Technology".

The modified model based on Rasch's one-parameter model was used for assessment in a test and for assessment of course projects created by the learners [4] in order to achieve an objective assessment corresponding to the real knowledge of the learners. The applied model allows software implementation and can be used in all forms of training, it is especially useful in e-learning.

Applicability of the Rasch model

When conducting a test, teachers follow informal linguistic models for decision-making to assess the preparation of learners, which depend on very important as well as non-essential parameters, which can be described as follows:

• If the learner <has a general idea of the subject, admits significant omissions in solving the test questions, has not mastered specific terms, etc.> - Medium (3). In the same way, through vague implications, the rules for judging with Good, Very Good and Excellent are described.

• In assessing the acquired knowledge, skills and competencies of students through test questions, it is possible to use a scale that serves both to measure students' knowledge through their numerical achievement for each of them and to measure test questions through the parameter difficulty for every one of them.

To objectify the test assessment, the one-parameter Rush model is used, which is one of the main formalization methods in probabilistic modeling for measuring student achievement (IRT). It establishes a correspondence between the observed results of the test and two sets related to the difficulty of the test (T) and the preparation of the learners (S) [5].

The possibility to use the one-parameter model of Rash [6], [7] to formalize the assessment of students' course projects stems from the following considerations, which do not contradict the actual practice, and the nature of the lecturer's thought process:

• The knowledge and difficulty of the developed Web site by the learners are assessment parameters that allow objective assessment, regardless of the assessment and measurement tool used.

• The subject of assessment is the hidden variable <functionality of the student's course project>, which does not allow direct measurement, but can be measured objectively.

• Dozens of different, equally competent teachers in the same job may differ slightly from each other due to unavoidable measurement errors, but not due to differences in competencies.

• Careful analysis of the teacher's thinking in the decision-making process for the formation of assessment shows that teachers have a subjective tendency to encourage knowledgeable students and negative expectations of ignorant, which affects the assessment. Similarly, due to its nonlinearity, Rasch's model favors knowledgeable students and is unfriendly to ignorant ones.

One-dimensional Rasch model

To paraphrase Rasch's model [6], [7], we can assume that the probability P for a student with experience S to develop a high-quality site with difficulty T is determined by the following formula, which is valid and correctly answers the test questions:

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

The function P (S, T) is called the success function, and the variables S and T are called latent variables. If we enter the following notations:

A = LN (S), S = EXP (A)

B = LN (T), T = EXP (B)

for P we get:

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

The resulting ratio is called the basic logistics model of Rasch. The last formula shows that the probability of success depends only on the difference B-A, which is why the Rasch model is one-parameter.

III. PERFORMANCE EVALUATION

Rasch's model is applied with the dichotomous scale {yes, no} ≡ {0, 1}, which is a bit informative. We choose the scale L = {bad, good, excellent} ≡ {1, 0.5, 1}. It is applicable with minor modifications to the Rasch model. This scale is convenient for teachers and gives a greater opportunity to form an unambiguous assessment compared to the scales in which the number of terms k> 3. This same scale L = {bad, good, excellent} ≡ {1, 0.5, 1 } is also convenient in the test test in terms of the fact that it allows to evaluate test questions that have more than one correct answer and those that are open, but with a short addition.

The evaluation is performed in the following sequence:

1. The teacher evaluates the work according to pre-selected criteria in scale L. As a result of the expert evaluation we get a matrix A with dimensions mxn, where m is the number of checked papers, n is the number of criteria for course projects / number of test questions.

2. We calculate the primary grade bi, i = 1, 2, ... m of the students as a sum of the grades on the lines:

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

3. Calculate the parameters pi, i = 1, 2, ... m by the formula:

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

We ignore the extreme estimates as follows: if bi = 0, we make pi = ε; if bi is equal to the maximum result, we make pi = 1 - ε, where ε is a very small number, for example ε = 0.001.

4. The initial approximation of the evaluation of the i-th work is calculated by the formula:

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

5. Calculate the primary score cj, j = 1, 2, ... m from the criteria obtained by adding the scores in the columns.

$$c_j = \sum_{i=1}^{m} a_{ij}$$

6. Calculate the parameters pj, j = 1, 2, ... n by the formula:

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

Like (3), if cj = 0, make pj = ε; if cj is equal to the maximum result, we make pj = 1 - ε.

7. Calculate the initial values of the difficulty criteria according to the formula:

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

8. The final grade of the course project or test is obtained by linear transformation Ai ∈ [min (Ai), max (Ai)], i = 1, 2, ... n] on the scale {2,3,4, 5.6} ≡ {bad, average, good, very good, excellent}.

To test the possibility of using the described method, an experiment was conducted with the developed course projects of students from Shumen University evaluated using the above-described modified Rasch model, as well as the online test on the same study material needed to develop their course project. Some of the 80 experiments performed are presented in Table 1 and Table 2.

Table 1: Experimental results from the evaluation of the test of students

test / tested person	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-e))) in logite	Evaluation
x_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	27	0,900	**2,187**	6
x_2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,5	1	1	1	1	1	1	1	1	1	0,5	1	1	1	29	0,967	**3,338**	6
x_3	1	1	1	1	1	1	1	1	1	1	0,5	1	1	1	1	1	1	1	1	1	1	1	1	0,5	1	1	1	1	1	1	29	0,967	**3,338**	6
x_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	25,5	0,850	**1,728**	6
x_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	22	0,733	**1,008**	5
x_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	27,5	0,917	**2,386**	6
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
x_{35}	1	1	1	1	1	1	0,5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,5	1	1	1	29	0,967	**3,338**	6
x_{36}	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	27,5	0,917	**2,386**	6
x_{37}	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	24	0,800	**1,381**	5
x_{38}	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	22	0,733	**1,008**	5
x_{39}	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	30	1,000	**6,908**	6
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
x_{74}	0	1	0	0	0	1	1	0	1	0	1	0	0	1	0	1	0	1	0	1	1	0	0	1	1	1	1	1	1	0	16	0,533	**0,131**	4
x_{75}	1	0	1	1	0,5	1	1	1	1	1	1	0	0	0	0	1	0	1	0	1	1	1	0	1	1	1	0	1	1	1	20,5	0,683	**0,766**	5
x_{76}	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	18	0,600	**0,403**	5
x_{77}	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	23	0,767	**1,185**	5
x_{78}	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	26,5	0,883	**2,016**	6
x_{79}	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	28	0,933	**2,624**	6
x_{80}	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	30	1,000	**6,908**	6
	12	9	12	11	9,5	12	12,5	13	13,5	10,5	12,5	9	9	12	10	13	10	13	9,5	10,5	12	10,5	11,5	11	11	13	10,5	12,5	13	12				
P	0,86	0,64	0,86	0,79	0,68	0,86	0,89	0,93	0,96	0,75	0,89	0,64	0,64	0,86	0,71	0,93	0,71	0,93	0,68	0,75	0,86	0,75	0,82	0,79	0,79	0,93	0,75	0,89	0,93	0,86				
Logit	-1,79	-0,59	-1,79	-1,30	-0,75	-1,79	-2,12	-2,56	-3,30	-1,10	-2,12	-0,59	-0,59	-1,79	-0,92	-2,56	-0,92	-2,57	-0,75	-1,10	-1,79	-1,10	-1,53	-1,30	-1,30	-2,56	-1,10	-2,12	-2,56	-1,79				

Table 2: Experimental results from the evaluation of course projects of students

Criteria / tested person	ESSENTIAL PART WEIGHT t=2												DESIGN CSS WEIGHT t=1		PROTECTION WEIGHT t=2		Primary ball b_i	P_i	$A_i=Ln(P_i/(1-(P_i-e)))$ in	Evaluation
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16				
x_1	1	0	0,5	0,5	0	1	1	0	1	1	0,5	1	1	1	1	0	19	0,6129032	0,4595323	5
x_2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	30	0,9677419	3,4011974	6
x_3	1	0,5	1	0,5	2	1	0,5	1	2	2	0,5	0,5	1	2	1	0,5	31	0,999	6,9067548	6
x_4	0,5	1	0	0	1	0	1	1	0	0,5	0,5	0	0,5	1	0,5	1	15,5	0,5	0	4
x_5	1	1	1	1	1	1	1	1	1	1	1	1	1	0,5	1	0,5	28,5	0,9193548	2,4336134	6
x_6	1	0,5	1	0,5	1	1	1	1	1	0,5	1	1	0,5	1	1	1	26,5	0,8548387	1,7730673	6
:																				
x_{35}	0	0,5	1	0	0,5	0	1	0	1	1	1	1	1	1	1	0,5	19	0,6129032	0,4595323	5
x_{36}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	30	0,9677419	3,4011974	6
x_{37}	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	15	0,483871	-0,064539	4
x_{38}	1	0	0	1	0	1	1	1	1	0	1	0	1	0	1	0	17	0,5483871	0,194156	4
x_{39}	1	1	1	1	1	0,5	1	0	1	1	1	1	0,5	1	1	1	26,5	0,8548387	1,7730673	6
:																				
x_{74}	0	0,5	0,5	0	0,5	0	1	0	0	0,5	0,5	0	0,5	0	1	1	11,5	0,3709677	-0,528067	3
x_{75}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,5	0,5	28	0,9032258	2,2335922	6
x_{76}	0,5	1	0,5	0,5	0,5	1	0,5	1	1	0,5	1	0,5	0,5	0	1	0	19,5	0,6290323	0,5280674	5
x_{77}	1	0	1	1	0	1	1	0,5	1	0	0	1	1	1	1	1	21	0,6774194	0,7419373	5
x_{78}	1	1	1	0,5	1	1	1	1	1	1	1	0,5	1	1	0,5	0,5	26	0,8387097	1,6486586	6
x_{79}	1	1	1	1	0,5	0,5	0,5	0,5	1	1	1	1	1	1	1	1	26	0,8387097	1,6486586	6
x_{80}	1	0,5	1	1	1	1	1	1	0,5	1	1	1	1	1	0	1	26	0,8387097	1,6486586	6
	3	1,5	2,5	2	3	3	2,5	2	4	4	2	2,5	3	4	3	1,5				
P	0,999	0,5	0,833	0,667	0,999	0,999	0,833	0,667	0,999	0,999	0,667	0,833	0,999	0,999	0,999	0,5				
Logit	-6,907	0	-1,609	-0,693	-6,907	-6,907	-1,609	-0,693	-6,907	-6,907	-0,693	-1,609	-6,907	-6,907	-6,907	0				

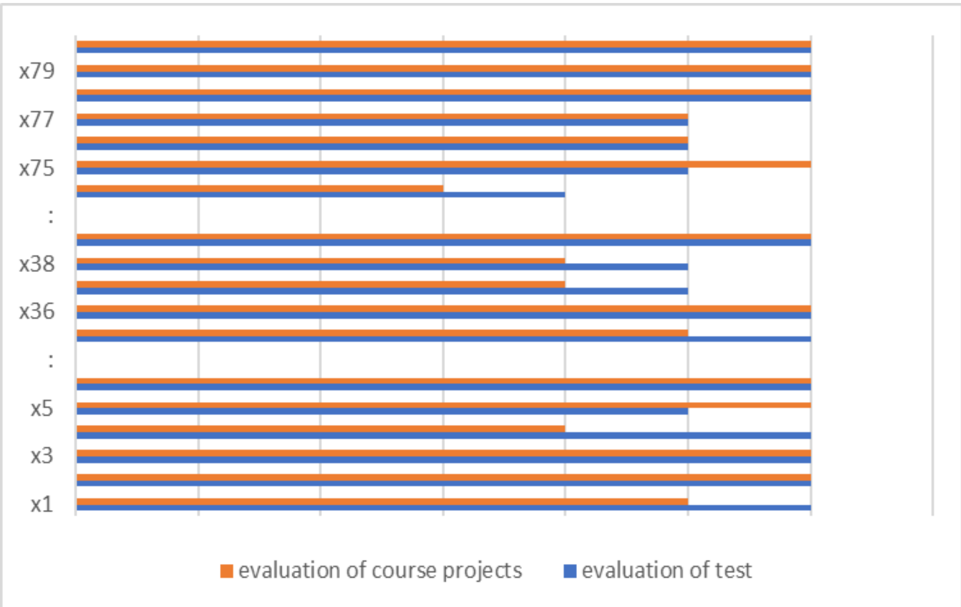

Figure 1: Comparing the results of the course projects and the test with the proposed model

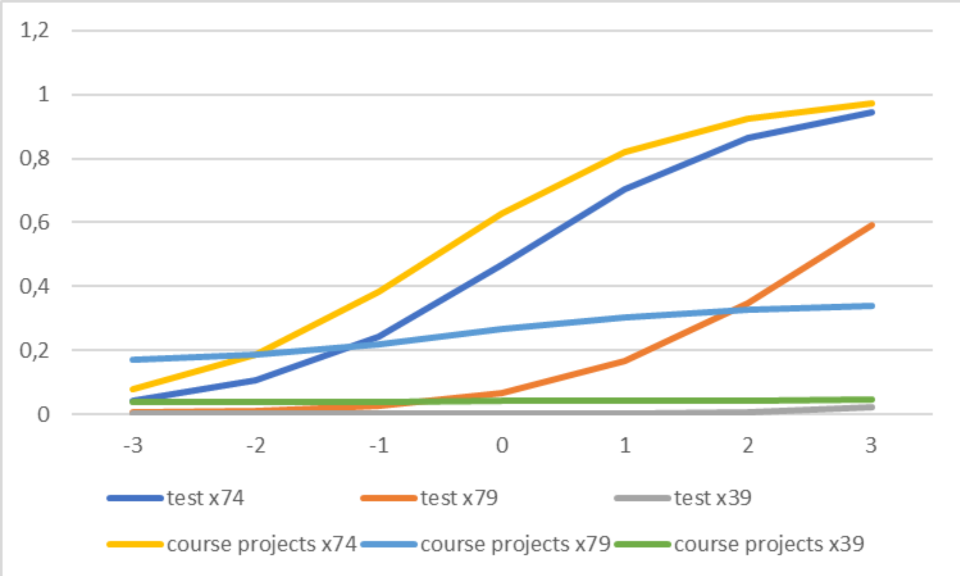

Figure 2: Comparison of the characteristic curves of three students x74, x79 and x39 compared to the results of the course projects and the test with the proposed model

Analyzing the results, we can draw the following conclusion:

• From the presented experimental results in fig. 1. it is evident that there is over 80% duplication of results. The differences in the results of the evaluation of the course projects compared to the test tested with the modified Rasch model are insignificant in favor of the test. This confirms the assumption that in order to obtain an objective assessment of learners it is necessary to select at least two different assessment methods, and that the selected two methods for assessment of acquired knowledge and skills by learners use the same assessment model, unifies the assessments.

• From the presented experimental results in fig. 2 it is observed that the characteristic curves of student x74 have a standard shape traditional for the one-parameter Rush model. Its interpretation would be the following that the learner has shown both acquired and not so well acquired knowledge, which speaks of a good assessment. The characteristic curves of x39 are completely concave and approximately parallel to the abscissa axis, which indicates a well-prepared student who has fully mastered the study material and his assessment should be excellent. The characteristic curves of x79 are close to those of x39, ie. it is a cold one that has done very well with the material being studied.

• In multiple real-world experiments, statistics can be used to "fit" the "input population-fit" of the input parameters (criteria) to the requirements of the models. Easy (which are satisfied by all) and difficult (which no job satisfies) criteria can be excluded.

IV. CONCLUSION

The results achieved from the use of the proposed model for assessment of knowledge, based on the assessment components in the respective discipline, so far are positive, both from the teaching and student point of view.

The proposed model for assessing students' knowledge is based on the development of the one-parameter Rush model using fuzzy set theory.

REFERENCES

[1]. Томашева Л. Ю. Система оценки качества знаний студентов дистанционного обучения. Retrieved 23.02.2016 from http://elib.altstu.ru/elib/books/Files/pa2010_2/pdf/196tomasheva.pdf

[2]. Black A. Tools for Education Assessment. Retrieved 23.02.2016 from http://www.ehow.com/info_7896270_tools-education-assessment.html

[3]. Methodology of Educational Measurement and Assessment. Series Editors: Veldkamp B., von Davier, M. Springer, Retrieved 19.03.2016 from http://www.springer.com/series/13206

[4]. Boryana, U.-D., Stanimir, Z., Hristo, P., Intelligent methods for evaluation of student written works, Journal of Engineering and Applied Sciences 12(Specialissue10), pp. 8780-8784 (sjr 2017: 0.163)

[5]. Челышкова М.Б., А.Г. Шмелев, Шкалирование результатов единнoго госэкзамена. Проблемы и перспективы. [09.2013] http://www.ege.ru/organisators/index.html

[6]. Bond T.G, James Cook Applying the Rasch Model. Fundamental Measurement in the Human Sciences, 2007. Retrieved 19.03.2016 from https://mmm1406.sanjose14-verio.com/rascho/books.htm

[7]. Maslak, A. A., G. Karabatsos, T. S. Anisimova, S. A. Osipov, Measuring and Comparing Higher Education Quality between Countries Worldwide, Journal of applied measurement, 6(4), 432–442, 2005. Retrieved 19.03.2016 from http://sgpi.ru/prepod/maslak/pdf/Measuring_and_Comparing_Higher_Education_Quality_between_Countries_Worldwide.pdf

Uzunova-Dimitrova Boryana. "Intelligent Methods for Objective Assessment of Learners in Online Testing." *IOSR Journal of Engineering (IOSRJEN),* 10(12), 2020, pp. 06-11.