Proposal Models for Personalization of e-Learning based on Flow Theory and Artificial Intelligence

Anibal Flores[1]
E.P. Ingeniería de Sistemas e Informática
Universidad Nacional de Moquegua, Moquegua, Perú

Luis Alfaro[2], José Herrera[3], Edward Hinojosa[4]
Doctorado en Ciencia de la Computación
Universidad Nacional de San Agustin, Arequipa, Perú

Abstract—**This paper presents the comparison of the results of two models for the personalization of learning resources sequences in a Massive Online Open Course (MOOC). The compared models are very similar and differ just in the way how they recommend the learning resource sequences to each participant of the MOOC. In the first model, Case Based Reasoning (CBR) and Euclidean distance is used to recommend learning resource sequences that were successful in the past, while in the second model, the Q-Learning algorithm of Reinforcement Learning is used to recommend optimal learning resource sequences. The design of the learning resources is based on the flow theory considering dimensions as knowledge level of the student versus complexity level of the learning resource with the aim of avoiding the problems of anxiety or boredom during the learning process of the MOOC.**

Keywords—*Massive Online Open Course; MOOC; e-learning; flow-theory; learning resource sequence; case based reasoning; reinforcement learning; q-learning*

I. INTRODUCTION

The e-learning is a teaching-learning process, oriented to the acquisition of a series of competences and skills by the student, characterized by the use of web-based technologies, the sequencing of content and structured activities [1], that since its beginning has had and still has aspects to improve.

In recent years, a huge number of sites have appeared and offer online training and education services, such as: coursera, udacity, udemy, etc. "in press" [2]; the offered courses in these sites, generally MOOCs, which are not far from a traditional classroom course, in the sense that these courses have been planned or prepared considering that all students learn equally, regardless of the level of knowledge or skills that they can have.

Some of the current deficiencies detected in MOOCs are: Instructors and designers of the MOOCs may lack knowledge of the contemporary instructional design principles or learning theories [3]; the access to contents of the learning sessions is linear [4], that is, there is a single sequence of learning resources for all students; the content of the learning resources is structured for all students equally [4], without considering the level of knowledge on the subject that each student has; there is no adequate feedback, it is necessary to clarify misunderstandings or misconceptions [5]. In MOOCs, commonly there is no online tutor or teacher, the student can stagnate because they may not understand properly any resource or learning activity; teaching strategies and the use of

learning resources often do not take advantage of the benefits offered by information and communication technologies.

As described above, it generates an inadequate management of learning resources which contributes to an inadequate learning process, generating dissatisfaction in the participants or students, which can culminate in dropout.

Fig. 1 shows what was described above in a problem tree, in such a way that the causes, the problem and the corresponding effects are appreciated for a better understanding.

The solution proposed in the present study focuses on solving the causes mentioned as linear access to contents and non-flexible resources according student's knowledge level, and the personalization of the sequencing of learning resources is considered, for which, in this work, two personalization models are evaluated under similar conditions in order to determine which is the most suitable as a solution according to the results of experimentation.

The first model implements Case Based Reasoning and recommends learning resource sequences that have been successful in the past, Euclidean distance is used to determine similarity between past cases and a new case and recommend a personalized learning resource sequence. In the second model, Q-Learning algorithm of Reinforcement Learning is used to generate an optimal sequence base which is used to recommend a learning resource sequence.

The learning resources were designed considering the flow theory [6], considering the dimensions of the student's level of knowledge and level of complexity of the learning resource. For the present study, resources of two levels of complexity were designed: basic and advanced.

The content of this paper is organized into ten sections, the first of them summarizes the content of this paper, the problem and some causes; the second section shows a review of the state of the art in relation to the problem and the solutions for it; then, the third section, describes the theoretical background necessary for the adequate understanding of the paper content; likewise, in the fourth section, the proposal models in the present work are described; next, in the fifth section, the step by step process for implementing the models is described; in the sixth section, the main features of the MOOC design and the case study are described; in the seventh section, the experimental design used in the study is described; in the eighth section, the results

achieved in the study are shown, also, the results are discussed with similar works; then, in the ninth section, the conclusions reached at the end of the study are shown, and finally the future work section is shown with the improvements that can be made in subsequent works.
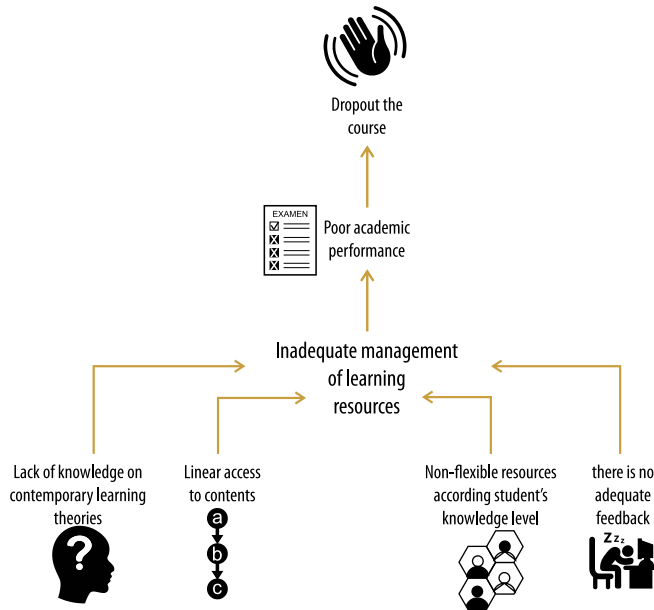

Fig. 1. Dropout in MOOCs.

II. RELATED WORK

A. Proposal Model for e-Learning based on Case based Reasoning and Reinforcement Learning

In this work "in press" [2] the authors proposed a personalized learning management model based on flow theory and Case Based Reasoning and Reinforcement Learning, they used these techniques in a complementary manner. A case of study was implemented working with an experimental group and a control group, the results obtained show that the experimental group achieved a better academic performance with respect to the control group. Authors concluded and highlighted the importance of the personalization of the learning resources sequences in e-learning.

B. Intelligent Model for Personalized Learning Management in a Virtual Simulation Environment based on Instances of Learning Objects

In this work [7], the author presents the results of applying Case Based Reasoning (CBR) to solve the problem of personalization of content in virtual environments. The proposed intelligent learning management system considers Case Based Reasoning for the identification of learning styles and selection of teaching-learning strategies. In the process of identifying learning styles, Case Based Reasoning reached the best classification rate (99.50) compared with other techniques such as Simple Logistic (98.99), Naive Bayes (97.98), Tree J48 (96.98), RN Multilayer Perceptron (94.97). Likewise, in the experimentation process comparing the experimental group with the control group, on a rating scale from 0 to 100, the first reached a general average of 60.5, while the second

reached only 39.5, demonstrating the importance of the personalization of contents.

C. Optimization of Personalized Learning Pathways based on Competencies and Outcome

In this work [8], the author formulated the selection of learning routes as an optimization problem based on competencies and evaluation of student learning. The goal was to find the optimal personalized learning path that allows the student to achieve the best possible learning outcome. The author's proposal consists of a course with a set of competences, each competence has associated a set of learning objects, and likewise the competences are associated with evaluation modules that allow measuring the mastery of the competence on the part of the student. The problem was modeled as a Markov Decision Process and the technique or algorithm that was used for its solution were the Temporal Differences included in the set of reinforcement learning techniques, since it is a work in progress, it does not show final results.

D. A Reinforcement Learning-Based Framework for the Generation and Evolution of Adaptation Rules

In this work [9], the authors propose a framework based on Reinforcement Learning (RL) for the generation and evolution of adaptation rules. This framework has two key capabilities for self-adaptation through a two-phase process:

1) The automatic learning capability of adaptation rules from different configurations of objectives in the offline phase (as a result a case base is obtained that includes a set of different configurations of objectives and the corresponding optimal rule sets).

2) The ability of automatic evolution of adaptation rules from real-time information about the environment and user goals in the online phase (As a result, a continually updated case base is obtained that reflects the dynamics of the environment of more precise way and includes a set of possible configurations of objectives with greater degree of coverage). Based on the two capacities, in the online phase, the case that best fits will be recovered from the base of cases to carry out the adaptation, and will be continuously evolved from the actual feedback information.

As it is a work in progress paper, it does not show final results.

E. CBR based Approach for Adaptive Learning in e-Learning System

In this work [10], the authors presented a C Programming based adaptive learning system that removes static learning delivery and accommodates individual student needs and differences to improve their programming learning aspects. This proposal adopts an adaptive approach of four phases based on case base reasoning (CBR) to develop adaptive learning in programming system. On basis of different programming aspects like syntax error, logical error and application usage feasibility, student performance being predicted and impact on their characteristics at different levels are identified.

They concluded that results verify the feasibility and performance of programming learning system using a control and experimental group, where the experimental group had better learning performance than the control group in terms of syntax, logical and application feasibility findings. Individual student needs and differences can be accommodated easily with such personalized adaptive C Programming based e-learning system.

III. THEORETICAL BACKGROUND

A. MOOCs

Massive Open Online Courses (MOOCs) are open online courses that generally allow anyone to register and follow the course without a fee (at least for the basic course) [11]. MOOCs, like most online courses, offer learners the flexibility of self-paced learning without the constraints of time and place [12].

To enable self-paced learning, many activities in MOOCs are asynchronous in nature, whereby learners watch a series of videos, take quizzes, or participate in discussion forums. Yet, unlike online courses that offer credits, MOOCs have no enrolment restrictions and can be taken by any interested individual at little or no cost [13]. Therefore, MOOCs have a much larger and more diverse learner population than other online learning environments. In that respect, designing instructions to support the highly diverse learners in MOOCs is important but challenging [13]. The phenomenon of MOOCs has recently attracted considerable attention in the fields of higher education, lifelong learning, and distance education [14].

B. Flow Theory

The flow is a state in which an individual is completely immersed in an activity without reflective self-awareness, but with a deep sense of control. Someone in the flow condition is so focused that he has no room for other thoughts or distractions. The flow theory [6] was initially presented by Mihaly Csikszentmihalyi who used the flow term to represent optimal individual experience focusing on her participation in an activity.

Although the flow was constructed from several complex variables, skill and challenge are the two most important [15]. In general, the theory of flow poses three conditions: the optimal conditions (flow state), the condition of restlessness (anxiety) and the condition of boredom [16]. The optimal condition is achieved when a person's ability is in balance with the given challenge. When the skill required to complete an action or task is lower than the challenging action, learners become anxious or frustrated [17]. When learner skill is higher than the challenging action, learners become bored.

Fig. 2 shows the effect of the relationship between skills and challenge on states of anxiety and boredom.

C. Personalized Learning

In the area of e-learning, "personalization" has a wide range of new meanings. One of the best explanations could be that "personalized learning is the adaptation of pedagogy, curriculum and learning environments to meet the needs and learning styles of individual students" [18].

The subject of personalization is strictly related to the shift from a teacher-centered to a student-centered and competition-oriented perspective. Unlike conventional e-learning, which tends to treat students as a homogeneous entity, personalized e-learning recognizes students as a heterogeneous mix of people [18].

Essentially, personalized e-learning offers students the customization of a variety of elements of the online education process:

- The learning environment: the content and its appearance for the student (such as backgrounds, themes, font sizes, etc.)

- The content of learning itself: multimedia representations (such as text, graphics, audio, video, etc.)

- Interaction: includes facilitator, student and learning content (for example, mouse, keyboard, touch / slide: through questionnaires, online discussions, "games", tutorials, adaptive learning approaches).

D. Case Based Reasoning (CBR)

Given a large set of problems and their individual solutions case based reasoning seeks to solve a new problem by referring to the solution of that problem which is "most similar" to the new problem. Crucial in case based reasoning is the decision which problem "most closely" matches a given new problem [19].

Case-based reasoning (CBR), is a problem to solving paradigm and that utilizes the knowledge of past cases to solve new cases. A past case denotes a previously experienced situation that has been captured and learned, and based from it a new case denotes an unexperienced situation to be resolved [9].
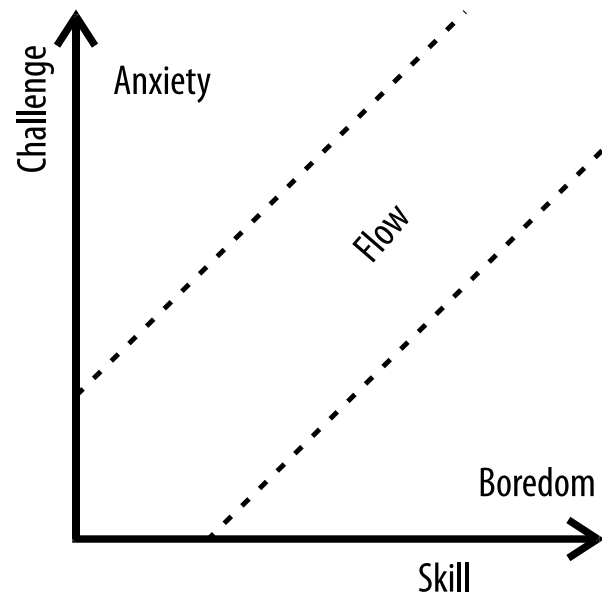

Fig. 2. Flow States.

A past case is stored in a case base, and is characterized from three aspects: 1) Problem description, which depicts the state of the world when the case occurred; 2) Problem solution, which states the derived solution to that problem; and 3) Results, which describes the state of the world after the case occurred [9].

Based in the past cases, a new case is resolved with the following four steps:

1) Retrieve the most similar.

2) Propose a solution to the new case by reusing the information and knowledge in the most similar past case.

3) Revise proposed solution.

4) Retain the information and knowledge of the solution for the new case.

Fig. 3 shows a graphic view of CBR process.

E. Reinforcement Learning

The reinforcement Learning refers to the problem of an agent that aims to learn optimal behavior through trial-and-error interactions with a dynamic environment [20]. The algorithms for reinforcement learning share the property that the feedback of the agent is restricted to a reward signal that indicates how well the agent is behaving.

In Reinforcement Learning, the decision-maker, i.e. the agent, interacts with an environment over a sequence of observations and seeks a reward to be maximized over time.

Formally, the model consists of a finite set of environment states S, a finite set of agent actions A, and a set of scalar reinforcement signals (i.e. rewards) R. At each iteration i, the agent observes some representation of the environment's state $s_i \in S$. On that basis, the agent selects an action $a_i \in A(s_i)$, where $A(s_i) \subseteq A$, denotes the set of actions available in state s_i. After each iteration, the agent receives a numerical

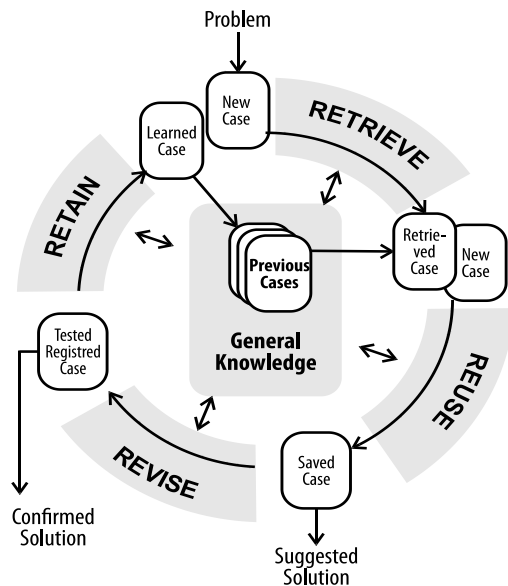reward $r_{i+1} \in R$, and observes a new state s_{i+1} [20].

Fig. 3.   CBR Process.

In order to store current knowledge, the reinforcement learning method introduces a so-called state-action function $Q(s_i,a_i)$, that defines the expected value of each possible action a_i in each state s_i. If $Q(s_i,a_i)$ is known, then the optimal policy $\pi*(s_i,a_i)$ is given by the action a_i, which maximizes $Q(s_i,a_i)$ given the state s_i [20]. Consequently, the learning problem of the agent is to maximize the expected reward by learning an optimal policy function $\pi*(s_i,a_i)$.

F. Q-Learning

Q-Learning [21] is an off-policy method proposed by Watkins to solve Markov Decision Processes (MDP's) with incomplete information. From the point of view of control theory, it is an adaptive direct method, and it is based on the learning of Q according equation (1).

As the agent moves forward from an old state to a new one, Q-Learning propagates the estimates of Q backwards from the new state to the old one.

Although the Q-Learning cycle takes place infinitely in theory, in practice learning is done by episodes (or trials), where each episode begins in a certain initial state until reaching a condition defined by the designer of the learning system (such as: reaching the target state, reaching an absorbing state, exceeding a maximum number of iterations, etc.).

$$Q(s_i,a_i) \leftarrow Q(s_i,a_i) + \alpha \left[r_{i+1} + \gamma \max_{a_{i+1} \in A_{s_i}} Q(s_{i+1},a_{i+1}) - Q(s_i,a_i) \right] \quad (1)$$

Fig. 4 shows Q-Learning algorithm [22].


```
Initialize Q(s, a), ∀s ∈ S, a ∈ A(s), arbitrarily, and Q(terminal-state, ·) = 0
Repeat (for each episode):
    Initialize S
    Repeat (for each step of episode):
        Choose A from S using policy derived from Q (e.g., ε-greedy)
        Take action A, observe R, S'
        Q(S, A) ← Q(S, A) + α[R + γ maxₐ Q(S', a) − Q(S, A)]
        S ← S';
    until S is terminal
```

Fig. 4.   Q-Learning Algorithm.

IV. PROPOSAL MODELS

The proposal models in the present work are very similar, both have four modules, the module COURSE, KNOWLEDGE and E-TUTOR is the same in the two models. The fourth module is different for the two models; the first model (CBR), it is formed by the base of successful sequences; and the second model (RL), it is formed by the base of optimal sequences. In both models, the fourth module contains a sequence retrieval sub-module.

Next, each of the modules of the proposal models is briefly described.

A. Course

This module contains general information of the course, likewise, it contains the base of questions of the pretest and postest, in addition, of the tests of each learning session with the respective solutions.

B. Knowledge

The Knowledge module contains the results of the application of various tests, including the pretest, the postest and the tests of each learning session. The results of these tests are used to implement the base of successful sequences that will be used by CBR and the base of optimal sequences that will be obtained using the Q-Learning algorithm of Reinforcement Learning (RL).

C. E-Tutor

The E-Tutor module contains various learning resources, including educational games (puzzles, crosswords and alphabet soups), videos and PDF documents at basic and advanced levels of complexity. It has a sub-module that implements a learning resource selecting process according proposal models.

D. CBR or RL

This is the main module of our proposals, which contains a base of success cases or a base of optimal sequences, which are required by the E-Tutor module in the learning resource selecting process. For first model, the algorithm of success case retrieving is based on Euclidean distance, and for second model, a random process is used. For the second model, the optimal sequence retrieving is random, considering a higher retrieval probability for the optimal sequences with greater reward. A detailed process is described on next section. Fig. 5 shows proposal model 1 and Fig. 6 shows proposal model 2.
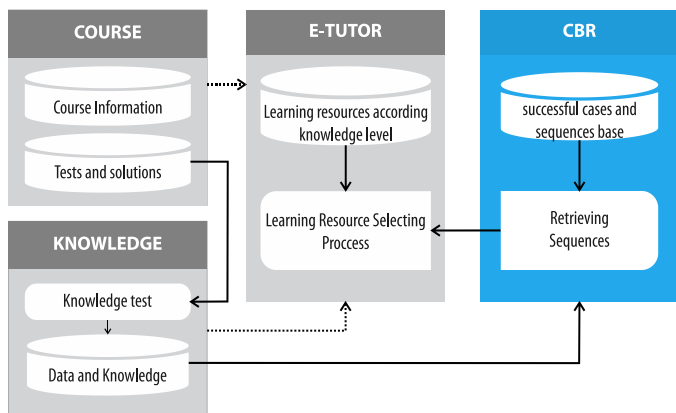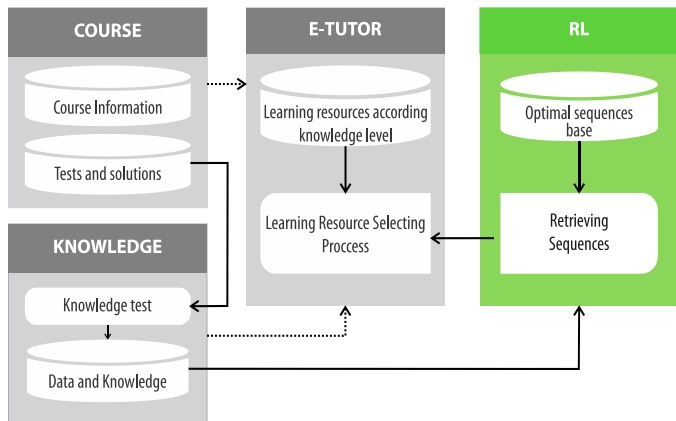

Fig. 5.   Proposal Model 1.

Fig. 6.   Proposal Model 2.

V.   PROCESS

A topic of general interest was chosen as a case of study to evaluate our proposals, in such a way that a large number of participants could be enrolled in the course; this topic is about Chiribaya Culture and the MOOC was titled "Conociendo la Cultura Chiribaya", having as a secondary objective of improving cultural identity in the Moquegua region, in the south of Peru.

Fig. 7 shows a summary of process.

1) Application of Pretest: In this stage, a pretest of twenty (20) questions was applied to fifty five (55) students enrolled in the MOOC titled "Conociendo la Cultura Chiribaya".

2) Assignment of sequences randomly: For each learning session, based on pretest results and designed learning resources, all possible sequences were determined and these are assigned randomly to the students according criteria shown in Table I.

In Fig. 8, we can see a graph and possible sequences for first session. Sequence starts at a resource and ends at e resource. For this case; b1, c1 and d1 are basic resources; b2, c2 and d2 are advanced resources; *a* is the session starting resource and e is a quiz about what was learned in this session.

Sessions 2, 3, 5, 6 and 7 have a similar graph, this is shown in Fig. 9 with the difference that the resources were different in the nodes of the graph, as shown in Table II.

Fig. 7.   Process of CBR and RL.

TABLE I.   SEQUENCE ASSIGNMENT CRITERIA

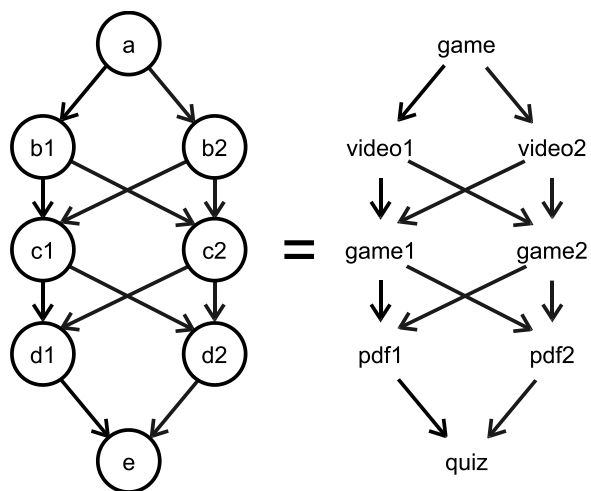| Pretest Score	Sequences	Resource Level	
		Basic	*Advanced*
[0-5]	a,b1,c1,d1,e	3	0
[6-7]	a,b1,c1,d2,e a,b1,c2,d1,e a,b2,c1,d1,e	2	1
[8-9]	a,b1,c2,d2,e a,b2,c2,d1,e a,b2,c1,d2,e	1	2
[10-20]	a,b2,c2,d2,e	0	3

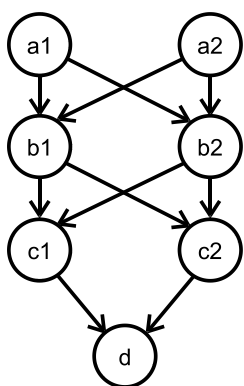Fig. 8. First Session–Learning Resources Sequence Graph.

Fig. 9. Graph for Sessions: 2, 3, 4, 5, 6 y 7.

TABLE II. TYPE OF RESOURCES PER SESSION (2,3,4,5,6 AND 7)

Session	a1	a2	b1	b2	c1	c2	d
2	Video1	Video2	Game1	Game2	PDF1	PDF2	Quiz
3	Video1	Video2	Game1	Game2	PDF1	PDF2	Quiz
4	PDF1	PDF2	Game1	Game2	PDF1	PDF2	Quiz
5	PDF1	PDF2	Game1	Game2	PDF1	PDF2	Quiz
6	Video1	Video2	Game1	Game2	PDF1	PDF2	Quiz
7	PDF1	PDF2	Game1	Game2	PDF1	PDF2	Quiz

3) Content development and learning evaluation: Once learning resources sequences were assigned, sessions were developed. Table III shows number of participants who completed course sessions.

TABLE III. STUDENTS PER SESSION

Learning Session	Number of students
1	55
2	51
3	46
4	43
5	40
6	39
7	38

4) Generation of success case base for CBR: The success cases base is generated for each session.

Learning in each session was evaluated with 4 questions, 5 points per question, thus, then possible scores were: 0, 5, 10, 15 or 20. In Table IV, first records of case base are shown.

According to Table IV, for success cases selection, score session was considered. For example, case Id *1* may be eligible as a success case for most sessions, except for session *5*. Similarly, case Id *2* may be eligible as a success case for the first three sessions (1, 2 and 3), but not for the last 4 (4, 5, 6 and 7). Same selection criterion was used for the rest of the cases, besides Euclidean distance.

5) Generation of optimal sequences base: For getting optimal sequences, Q-learning was used, this algorithm generates Q tables and from these tables optimal learning resource sequences for every session were obtained. Table V, shows Q Table for first session. Table VI shows optimal sequences for every session.

TABLE IV. CASE BASE–FIRST CASES

Id	Pretest questions					Score	Session scores						
	1	*2*	*3*	*…*	*20*		*1*	*2*	*3*	*4*	*5*	*6*	*7*
1	1	0	1		0	8	15	15	15	15	10	15	15
2	1	1	0		0	6	15	15	15	0	0	0	0
3	0	1	0		0	3	10	10	10	15	15	15	10
4	1	0	0		0	2	15	15	10	15	15	15	15
5	1	1	0		1	6	15	15	15	15	10	10	15
6	1	1	0		0	5	15	15	10	15	10	15	10
…													

TABLE V. Q TABLE FOR FIRST SESSION

	c1	c2	a	e	b1	b2	d1	d2
.ab1c2d1	0.000000	0.000000	0.000000	0	0.000000	0.000000	0.000000	0.00000
.ab1c2d2	0.000000	0.000000	0.000000	0	0.000000	0.000000	0.000000	0.00000
.ab2c1d1	0.000000	0.000000	0.000000	0	0.000000	0.000000	0.000000	0.00000
.ab1	5.254556	4.857422	0.000000	0	0.000000	0.000000	0.000000	0.00000
.ab2c1d2	0.000000	0.000000	0.000000	0	0.000000	0.000000	0.000000	0.00000
.ab2	5.119506	3.096000	0.000000	0	0.000000	0.000000	0.000000	0.00000
.ab1c1d1	0.000000	0.000000	0.000000	0	0.000000	0.000000	0.000000	0.00000
.ab1c1d2	0.000000	0.000000	0.000000	0	0.000000	0.000000	0.000000	0.00000
.ab2c1	0.000000	0.000000	0.000000	0	0.000000	0.000000	2.561231	0.64000
.ab2c2	0.000000	0.000000	0.000000	0	0.000000	0.000000	0.000000	1.80000
.	0.000000	0.000000	5.668461	0	0.000000	0.000000	0.000000	0.00000
.ab1c1	0.000000	0.000000	0.000000	0	0.000000	0.000000	2.750853	1.00000
.ab1c2	0.000000	0.000000	0.000000	0	0.000000	0.000000	2.721600	3.68928
.a	0.000000	0.000000	0.000000	0	5.517102	6.367814	0.000000	0.00000
.ab2c2d1	0.000000	0.000000	0.000000	0	0.000000	0.000000	0.000000	0.00000
.ab2c2d2	0.000000	0.000000	0.000000	0	0.000000	0.000000	0.000000	0.00000

TABLE VI.   OPTIMAL SEQUENCES

Learning Session	Optimal Sequences
1	a,b2,c1,d1,e a,b1,c1,d1,e
2	a2,b1,c1,d a1,b2,c1,d
3	a2,b2,c2,d a1,b1,c1,d
4	a1,b2,c1,d
5	a2,b1,c2,d
6	a1,b2,c1,d a2,b1,c2,d
7	a1,b1,c1,d a2,b1,c2,d

6) Preparation of the algorithm for the assignment of personalized sequences: At this stage, once success cases base and optimal sequences base were built, two algorithms were implemented, one for each proposal model.

For the proposal model 1:

a) First, the algorithm receives as input (the problem), a vector of correct and incorrect answers from student's pretest.

b) Second, the input vector is compared with each case, according similarity determined by Euclidean distance (2).

$$d(p,q) = d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + ... + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \qquad (2)$$

Table VII shows implementation of Euclidean distance algorithm in PHP language.

Thus, when the student logs in or enters the MOOC, the sequences of learning resources of each session or topic are loaded into a matrix, that is accessed according to the student's interaction with the MOOC interface. Learning resource sequence matrix can be seen in Fig. 10.

TABLE VII.   EUCLIDEAN DISTANCE ALGORITHM IN PHP

```
function euclidean_distance($v1,$v2)
{
    $s=0.0;
    $nv=count($v1);
    for($i=0;$i<$nv;$i++)
    {    $d=(int)$v1[$i]-(int)$v2[$i];
        $s+=pow($d,2);
    }
    return sqrt($s);
}
```
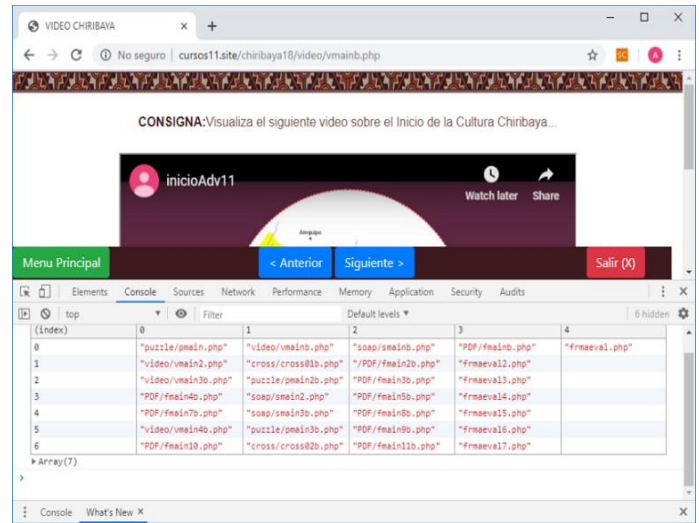

Fig. 10. Learning Resource Sequence Matrix for a Student.

For the proposal model 2:

a) First, the algorithm receives as input (the problem), a vector of correct and incorrect answers from student's pretest.

b) Second, a sequence from optimal sequence base is randomly assigned; ensuring that sequences with highest reward have a higher occurrence.

VI. MOOC DESIGN

The MOOC of the Chiribaya Culture was designed with the contents shown in the Table VIII.

Fig. 11 shows main screen of Chiribaya Culture MOOC.

Fig. 12 shows some educational games designed for the MOOC. Also, Table IX shows the fundamental differences between basic and advanced resources in the designed MOOC.

The learning resources are accessed by the students according to the learning resource sequence determined by CBR or RL module corresponding to proposal model 1 or proposal model 2, respectively.

TABLE VIII.   MOOC CONTENT

Content	Learning Resources
Pretest	Pretest
1. Inicio	Game, Video, Game, PDF, Quiz
2. Agricultura	Video, Game, PDF, Quiz
3. Ganadería	Video, Game, PDF, Quiz
4. Textilería	PDF, Game, PDF, Quiz
5. Cerámica	PDF, Game, PDF, Quiz
6. Pesca y Recolección	Video, Game, PDF, Quiz
7. Momias	PDF, Game, PDF, Quiz
Postest	Postest

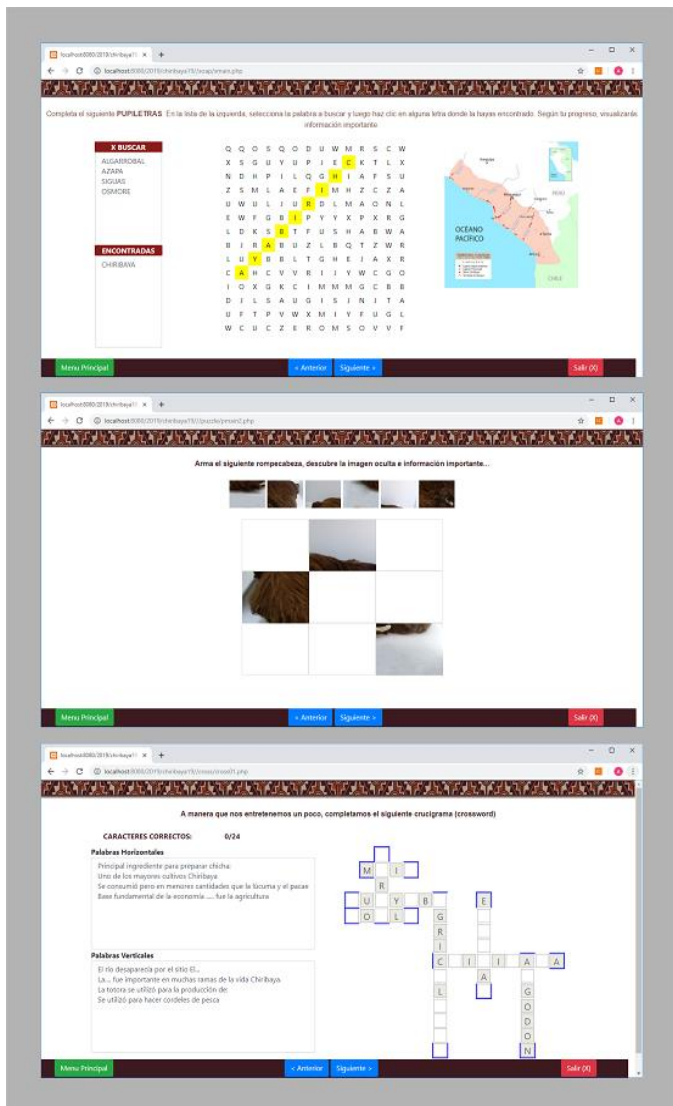Fig. 11. Chiribaya Culture MOOC–Main Screen.

Fig. 12. MOOC's Educational Games: Alphabet Soup, Puzzle and Crossword.

Table X shows some learning resources with the exact difference between basic and advanced level in different MOOC learning sessions.

TABLE IX. DIFFERENCES BETWEEN BASIC AND ADVANCED RESOURCES

Learning resource		Basic Level	Advanced Level
Video		Less time Less content	More time More content
PDF		Less pages Less content	More pages More content
Game	Crossword	Less words	More words
	Alphabet Soup	Less words Smaller matrix	More words Bigger matrix
	Puzzle	Less pieces	More pieces

TABLE X. BASIC AND ADVANCED RESOURCES EXAMPLES

Learning session	Learning resource		Basic Level	Advanced Level
I	Video		1.35	1.58
III	PDF		2 pages	4 pages
II		Crossword	6 words	8 words
II	Game	Alphabet Soup	4 words Matrix 13x13	6 words Matrix 14x14
IV		Puzzle	6 pieces	9 pieces

VII. EXPERIMENTAL DESIGN

Once the process was completed, an experiment was carried out with the following experimental design:

EG1 O1 X1 O2
EG2 O3 X2 O4

CG   O5   O6

Where:

EG1 : Experimental Group 1
O1   : Experimental Group 1 Pretest
O2   : Experimental Group 1 Postest
X1   : Proposal Model 1
EG2 : Experimental Group 2
O3   : Experimental Group 2 Pretest
O4   : Experimental Group 2 Postest
X2   : Proposal Model 2
CG   : Control Group
O5   : Experimental Group 3 Pretest
O6   : Experimental Group 3 Postest
X3   : Proposal Model 3
CG   : Control Group
O5   : Control Group Pretest
O6   : Control Group Postest

Likewise, the number of students that participated in each group is shown in Table XI.

Experimental results are described on section Results and Discussion.

TABLE XI. NUMBER OF STUDENTS PER GROUP

Group	Number of students
Experimental Group 1 (EG1)	11
Experimental Group 2 (EG2)	10
Control Group 1 (CG)	15

VIII. RESULTS AND DISCUSSION

Results were organized in two dimensions: Academic and Technical. For the academic dimension we will analyze some statistics of academic performance of the students under proposal models, the results of the pretest and postest were evaluated on the scale from 0 to 20; and in the technical dimension we will analyze a common metric to evaluate effectiveness of recommended sequences by proposal models, such as precision.

A. Academic Results

According to Table XII, comparing the mean of the pretest with the mean of the postest, it is appreciated that the proposal model 1 (CBR), reached the highest performance, a total of 7.81 points of increase. Second, we have the proposal model 2 (RL) with an increase of 7.5 points.

Likewise, the two models for personalization of learning resources sequences obtained superior performances regarding not using personalization in the MOOC that obtained an increase of just 5.93 points.

Fig. 13 shows graphically the mean differences between the described proposal models, and Fig. 14 shows a mean comparison of Experimental Groups with Control Group.

TABLE XII. COMPARISON OF PROPOSAL MODELS

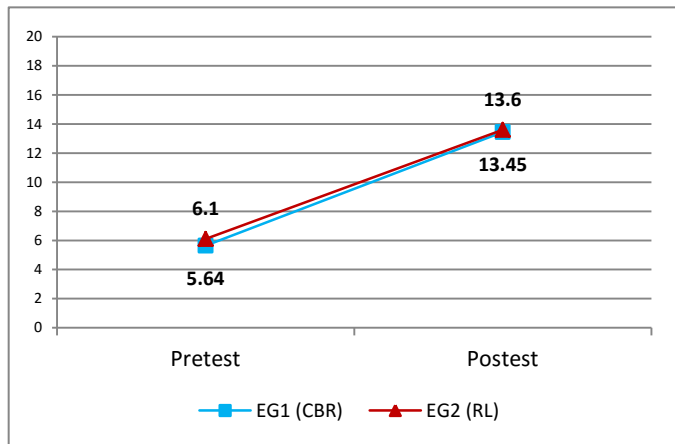| Statistics	Mean		Deviation	
Technique	Pretest	Postest	Pretest	Postest
CBR	5.64	13.45	3.17	3.39
RL	6.1	13.6	2.6	2.95
No personalized	5.40	11.33	1.96	1.95

Fig. 13. Mean Comparison of Pretest versus Postest.

Comparing the standard deviation of the pretest versus the postest, in the CBR model an increase in dispersion of 0.22 points is observed, while in the RL Model the increase is greater 0.35 points.

Results of proposal models in this work were compared with proposal model of paper "in press" [2], Fig. 15 shows a better academic performance of hybrid CBR+RL model proposed "in press" [2], we attribute this difference with respect to proposal models to the use of optimal sequences in a complementary manner when the CBR success case base does not contain enough cases.
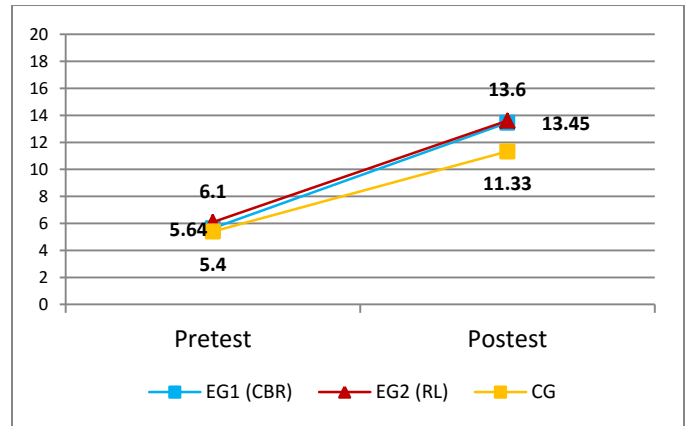

Fig. 14. Mean Comparison between Experimental Groups and Control Group.
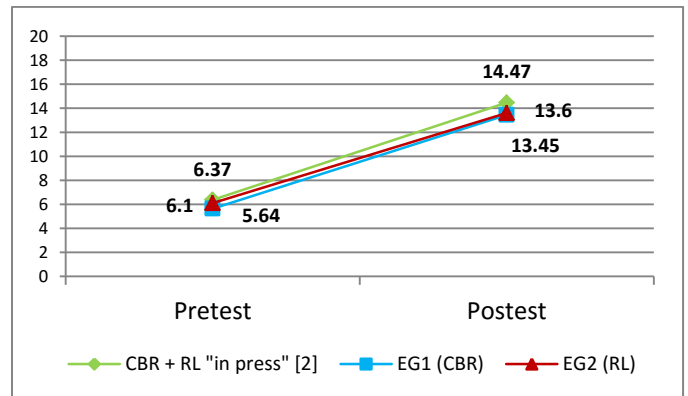

Fig. 15. Mean Comparison between Proposals Models Versus CBR+RL Model "in press" [2].

However, despite not having managed to overcome the results of the CBR + RL model "in press" [2], the results of the proposal model 1 and proposal model 2 are very promising, since despite the small number of cases (55) the academic results were very close to the CBR + RL model, we consider that, with a greater number of cases, results of the proposal model 1 could equal or exceed the hybrid model. In addition, it is necessary to work with larger samples, 10 or 11 students are not enough.

Also, we compared results of proposal models with results of work [7], which was based on the personalization of content based on student learning styles and Case Based Reasoning. In a Kinematics course, 100 students were evaluated on a scale of 0 to 100, where the postest average

reached 60.5 points, which converted to a scale of 0 to 20 equals 12.1 points, which is lower than those obtained by our proposal models 13.45 and 13.6 respectively, although the greater complexity of the Kinematics course with respect to our case study should be highlighted.

B. Hypothesis Contrast

For the hypothesis contrast, the normality tests of Shapiro-Wilk [23] and Anderson-Darling [24] were first applied to the experimental groups, in both cases the samples passed the tests, so the hypothesis contrast was performed with the Welch Two Sample T – Test [25].

Our hypothesis are:

H_0: True difference in means of proposal model 1 and proposal model 2 is equal to 0

H_1: True difference in means of proposal model 1 and proposal model 2 is not equal to 0

Considering a significance level: $\alpha = 0.05$; and running the t.test function from R software we obtain the results shown in Fig. 16.

According to Fig. 16, since p-value > α (0.9174>0.05), H_0 is accepted.

The mean of the Experimental Group 1's population is considered to be equal to the mean of the Experimental Group 2's population or the difference between the mean of the Experimental Group 1 and Experimental Group 2 populations is not big enough to be statistically significant.

C. Technical Results

In this part, it is very important to analyze effectiveness of retrieved sequences by our proposals, for it, the summary in Table XIII was elaborated. Equation (3) was used to calculate precision metric.

$$Precision = \frac{\#true\ positives}{\#true\ positives + \#false\ positives} \qquad (3)$$

The Precision indicates how well retrieved sequences match student's interest. It is the ratio of the number of relevant sequences retrieved to the total number retrieved.

```
          Welch Two Sample t-test

data:  muestraX and muestraY
t = -0.10513, df = 18.974, p-value = 0.9174
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.041435  2.750526
sample estimates:
mean of x mean of y
 13.45455  13.60000
```

Fig. 16. Results of Welch Two Sample T-Test in R.

TABLE XIII.   EFFECTIVENESS OF PROPOSAL MODELS

Proposal Model	Precision
CBR	0.909
RL	0.900

According Table XIII, a precision comparison of proposal models was made, observing that the proposal model 1 (CBR) reached the highest precision 0.909, with respect to the proposal model 2 (CBR) that reached a precision of 0.90. In this aspect, the proposal model 1 outperforms the proposal model 2.

IX. CONCLUSIONS

We conclude that personalization based on flow theory considering knowledge level of the students and complexity level of resources is very important to improve the academic performance in MOOCs. In the two proposal models, a higher academic performance was achieved respect the traditional linear access strategy (non-personalized) offered from most MOOC sites of the two proposal models, for the study case of the Chiribaya Culture teaching, the proposal model that achieved the best academic performance was the second one, the based on Reinforcement Learning and the proposal model who achieved the best precision of the recommended learning resource sequences was the first one, the based on Case Based Reasoning. We must emphasize that there is no significant statistical difference between the means of both proposal models.

X. FUTURE WORK

The proposal models can be improved by working with a bigger number of students (55+) used for the training phase and during the development of the course it was decreasing until get 38 students. Dropout is a big problem in most MOOCs [26]. So, it was no possible to have a good size for the successful sequence base.

It would be important to analyze the personalization proposals in other more complex teaching areas such as Mathematics, Physics, etc.

Also, models presented on this work can be improved, updating structure variables of cases, containing not only pretest questions, but also considering other aspects of flow theory as challenge, control, focused attention, presence, flow and positive affect [27].

REFERENCES

[1] J. Garcia-Peñalvo, A. Seoane Pardo, "Una revisión actualizada del concepto de e-learning," Education in the knowledge society, 2015.

[2] A. Flores, L. Alfaro, J. Herrera, "Proposal model for e-learning based on Case Based Reasoning and Reinforcement Learning," in III IEEE World Engineering Education Conference - EDUNINE, Lima, Perú, 2019.

[3] A. Margaryan, M. Bianco, A. Littlejohn, "Instructional quality of Massive Open Online Courses (MOOCs)," Computers & Education, no. 80, pp. 77-83, 2015.

[4] R. Zatarain Cabada, M. Barrón Estrada, "Authoring tool for identifying learning styles, using self-organizing maps on mobile devices," Revista Electrónica de Investigación Educativa, no. 13, pp. 43-55, 2011.

[5] B. Groves, Teaching in a digital age: strengths and weaknesses of MOOCs, opentextbc.ca, 2015.

[6] M. Csikszentmihalyi, The Psicology of Optimal Experience, 1990.

[7] B. Maraza, "Modelo Inteligente de Gestión de Aprendizaje Personalizado en un Ambiente de Simulación Virtual Basado en Instancias de Objetos de Aprendizaje," Universidad Nacional de San Agustin, Arequipa, Perú, 2015.

[8] J. Lin, "Optimization of personalized learning pathways based on competencies and outcome," in 16th International Conference on Advanced Learning Technologies, 2016.

[9] T. Zhao, W. Zhang, H. Zhao, Z. Jin, "A reinforcement learning based framework for the generation and evolution of adaptation rules," in IEEE International Conference on Autonomic Computing, 2017.

[10] B. Pandey, R. Mishra, A. Khamparia, "CBR based approach for adaptative learning in e-learning system," in Asia Pacific World Congress on Computer Science and Engineering, 2014.

[11] T. R. Liyanagunawardena, "Massive Open Online Courses," Humanities, vol. 4, no. 35, pp. 35-41, 2015.

[12] E. Jung, D. Kim, M. Yoon, S. Hoon Park, B. Oakley, "The Influence of instructional design on learner control, sense of achievement, and perceived effectiveness in a supersize MOOC course," Computers & Education, 2018.

[13] J. Wong, M. Khalil, M. Baars, B. De Koning, F. Paas, "Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course," Computers & Education, 2019.

[14] A. Bozkurt, E. Akgun-Ozbek, O. Zawacki-Richter, "Trends and patterns in Massive Open Online Courses: Review and content analysis of research on MOOCs (2008-2015)," International Review of Research in Open and Distributed Learning, vol. 18, no. 5, 2017.

[15] G. Moneta, On the Measurement and Conceptualization of Flow, London, UK: Springer, 2012.

[16] D. Sihabudin, L. Nugroho, P. Santosa, "Modeling the flow experience for personalized context aware e-learning," in 8th International Conference on Information Technology and Electrical Engineering, Yogyakarta, Indonesia, 2016.

[17] C. Yonghui, "Study of flow theory and experiential learning," in Second International Conference on MultiMedia and Information Technology, Kaifeng, China, 2010.

[18] A. Klasnja-Milicevic, B. Vesing, M. Ivanovic, Z. Budimac, E-learning systems, intelligent techniques for personalization, Springer, 2017, pp. 22-23.

[19] J. Dippon, P. Fritz, M. Kholer, "A statistical approach to case based reasoning with application to breast cancer data," Computational Statistics & Data Analysis, vol. 40, pp. 579-602, 2002.

[20] N. Proellochs, "Reinforcement Learning in R," cran.r-project.org, 2018.

[21] C. Watkins, P. Dayan, "Technical note Q-Learning," Machine Learning, vol. 8, pp. 279-292, 1992.

[22] R. Sutton, A. Barto, Reinforcement Learning: An Introduction, Cambridge, Massachusetts, 2014.

[23] S. Shapiro, M. Wilk, "An analysis of variance test for normality (complete samples)," Biometrika, vol. 52, no. 3-4, pp. 591-611, 1965.

[24] M. Stephens, "EDF statistics for goodness of fit and some comparisons," Journal of the American Statistical Association, vol. 69, no. 347, pp. 730-737, 1974.

[25] B. Welch, "The generalization of student's problem when several different population variances are involved," Biometrika, vol. 34, no. 1-2, pp. 28-35, 1947.

[26] D. Onah, J. Edwin Sinclair, R. Boyatt, "Dropout rates of massive open online course: Behavioural patterns," Semantic Scholar, 2014.

[27] I. Rodríguez-Ardura, A. Meseguer-Artola, "Flow in e-learning: what drives it and why it matters," British Journal of Educational Technology, vol. 48, pp. 899-915, 2016.