Serious Game Evaluation System based on Fuzzy AHP

Kamal Omari
Polydisciplinary Faculty University Ibn Zohr
Ouarzazate
Morocco

Yassine Chajri
Sultan Moulay Slimane University
Beni Mellal
Morocco

Ayoub Oukhatar
Higher School of Technology
University Ibn Zohr
Ouarzazate
Morocco

ABSTRACT
The rapid advancement of technology has led to a surge in popularity for digital games, prompting training and educational institutions to incorporate them into their curriculums as educational games, commonly referred to as "serious games". These games provide learners with the opportunity to develop knowledge and skills in a fun and interactive manner by virtually recreating learning situations. However, in order to effectively adopt serious games, it is essential to obtain solid evidence of their quality. To address this issue, an evaluation system for serious games is proposed, based on four essential dimensions that a game must satisfy to effectively fulfill its intended purpose. These dimensions have been validated and weighted according to the context of use using the fuzzy AHP method. The system was tested on a serious game called "ROBOCODE," which was validated by an educational commission and used by first-year students at Hassan II University. The results showed the usefulness and relevance of the proposed evaluation system in assessing the adaptability of Serious Games.

General Terms
Serious Game

Keywords
Serious Game- Evaluation system- Fuzzy AHP

1. INTRODUCTION
Over the years, educators have often overlooked the potential of games as an educational tool. However, with the rapid growth of the gaming industry and the unique characteristics of the current generation, there has been a growing interest in the use of Serious Games for educational purposes. According to [1] and [2], Serious Games offer several advantages, including engaging and motivating learners, providing immediate feedback, promoting the development of critical thinking and problem-solving skills, and facilitating active learning.

These Serious games vary widely in content and quality, and evaluating their pedagogical, technical, and ludic aspects is essential to ensure their effectiveness and appropriateness for specific educational contexts [3]. Unfortunately, there is currently no consensus on a model or tool for evaluating Serious Games specifically for learning purposes [4]. This lack of consensus makes it difficult to determine the effectiveness and appropriateness of Serious Games in different learning contexts.

Therefore, the objective of this article is to propose a system for evaluating Serious Games based on four necessary dimensions that a Serious Game must meet to fulfill its intended purpose.

These dimensions include pedagogical, technical, ludic, and behavioural dimensions, are weighted and validated using the Fuzzy Analytic Hierarchy Process (Fuzzy AHP), a multi-criteria decision-making method that ensures internal consistency when evaluating the Serious Game based on its intended context, providing a reliable evaluation system.

2. STATE OF THE ART
Since the creation of Serious Games, several researchers have attempted to address the various challenges associated with them. Among these challenges is the issue of evaluation, as without proper evaluation, it would not be possible to integrate Serious Games into learning situations effectively. To this end, numerous researchers have focused their efforts on this subject and have developed frameworks and approaches aimed at providing practical solutions. These approaches can be classified as follows:

Approaches based on player performance: These approaches measure player performance based on the game's objectives (skills, knowledge acquisition, etc.) [5] [6].

User experience-centered approaches: These approaches evaluate the player's interaction with the game, including smoothness, usability, and design appeal [7] [8].

Engagement and motivation evaluation: Player engagement and motivation are essential to determine whether the game has successfully captured attention and maintained interest (Engagement surveys and behavioral studies) [9] [10].

Educational and learning approaches: This approach evaluates the effects of a Serious Game on players' knowledge, skills, or attitudes (Pre-test/post-test, case studies, questionnaires on learning outcomes, etc.) [11] [12].

Evaluation of social and community impact: This approach evaluates the impact of the game on behavioral or social changes within the target community [13] [14].

Approaches based on qualitative feedback (qualitative analysis): Qualitative evaluation explores players' perceptions and gathers their opinions, emotions, and reflections on the game [15] [16].

Experimental and quasi-experimental approaches: These approaches compare the effects of a Serious Game with a control group to evaluate its effectiveness. [17] [18]

Approaches based on the validation of theoretical models: This approach aims to validate a theoretical model of learning, engagement, or motivation through Serious Games

(SDT – Self-Determination Theory) [19].

Approaches based on technology and data analysis: Use of advanced analytical tools to study how players interact with the game, often within a Big Data framework (Analysis of game data logs, player progression and choices, analysis of player behaviors in virtual worlds) [20] [21].

Approaches based on heuristic evaluation: Experts evaluate the game based on predefined criteria, such as playability, interface clarity, and pedagogical consistency [22] [23].

In this paper, the focus is on works and studies that have concentrated on evaluating the quality and effectiveness of Serious Games, as well as those that have presented a holistic evaluation.

2.1 Quality evaluation of a Serious Game

Since the emergence of Serious Games, evaluation has been a critical issue that needs to be addressed to ensure their effective adoption in learning situations. A variety of evaluation frameworks and methodologies have been proposed by researchers to address this challenge. One critical aspect of evaluating Serious Games is their quality, and several researchers have proposed evaluation approaches that provide solid evidence of their quality from different perspectives.

An evaluation and analysis grid was proposed to assess the quality of Serious Games for educational purposes [24]. The grid consists of four sections: Identification of the Serious Game, Pedagogical specifications, Fun specifications, and Technical specifications. The evaluation results revealed that each Serious Game possesses unique strengths and weaknesses, yet they all exhibit unmet requirements within the educational sphere, especially concerning the monitoring of learners' progress.

In addition, [25] proposed a methodological evaluation tool named MEDGE for integrating Serious Games in education. This tool is a simplified version of an evaluation framework that uses different data collection techniques, such as surveys, examinations, observation and reflection, to identify the different parameters influencing the qualitative integration of Serious Games in the classroom. More than 62 teachers in different countries have tested the methodology, but the results were inconclusive about its ability to lead to appropriate evaluations between the Serious Game and its educational objectives.

Similarly, [26] proposed quality indicators to analyze the quality of Serious Games under design from different angles, including pedagogical quality, playful potential and usefulness in a formative context. These indicators are inspired by previous work [27] and [28] but are also homogenized and granularized to make them easier for designers to quantify. However, the author did not specify how to validate these indicators, which could create confusion when using this tool for evaluating Serious Games.

An evaluation framework was identified for evaluating the various quality characteristics of Serious Games [29]. This framework categorizes quality characteristics into primary and secondary dimensions based on their literature usage and assesses multiple aspects of Serious Games, including usability, comprehensibility, motivation, commitment, and user experience. The findings of the study highlighted an issue with comprehensibility, and it was suggested that the use of tutors or the addition of tutorials could assist users in understanding the objectives, concepts, and procedures of Serious Games.

Finally, [30] developed a model named MEEGA+ that provides comprehensive support for quality evaluation of Serious Games. The model was validated following a reliability and validity analysis based on the data collected from case studies involving a population of 1048 computer science students, as well as on the perspectives of 19 experts in educational games. The study indicates that MEEGA+ is a valid and reliable method that provides systematic support for quality evaluations of Serious Games.

2.2 Efficiency evaluation of a Serious Game

There is a consensus that Serious Games have significant potential as a teaching tool, but their effectiveness in terms of learning outcomes is still being studied due to the complexity of evaluating intangible measures. However, some researchers have proposed reliable and automated solutions to measure their efficiency [31].

One such solution is a methodology that uses integrated non-disruptive monitoring to evaluate the effectiveness of Serious Games. This methodology facilitates the measurement of learning outcomes by using automated evaluations instead of traditional questionnaires outside the Serious Games. Additionally, it provides a systematic means of evaluation analysis, which can complement formal experiments to measure the effectiveness of Serious Games [32].

For her part, an evaluation-based design framework was presented [33], which provides guidance for planning and conducting a Serious Game evaluation. This framework highlights the role of assessing the effectiveness of Serious Games in terms of their objectives, rather than just simple usability tests. While this framework does not measure whether the Serious Game is effective in relation to its designated objective, it provides a basis for constructive discussion and draws attention to the fact that the intended impact should be at the center of the design process of a Serious Game.

In addition, a framework called the serious game stakeholder experience assessment method (SGSEAM) was described [34], which uses the strategy of simultaneous triangulation of mixed methods of collection and analysis of quantitative and qualitative data [35]. SGSEAM does not judge a Serious Game according to one standard or compare it against another, but instead identifies the main strengths and weaknesses of a Serious Game from the perspective of the main stakeholders, including players, system administrators, game designers, game managers, and developers. For each stakeholder, a set of questions and approaches are proposed.

The SGSEAM evaluation case study on the Serious Game dedicated to sustainability (Makahiki) provides concrete indications on how to improve this Serious Game for stakeholders. However, the study's small sample size limits the confidence in the data obtained, which is a disadvantage.

2.3 Holistic evaluation of a Serious Game

Recognizing the complexity of evaluating Serious Games, researchers have developed initial frameworks for their holistic evaluation, many of which provide observations on the generic way to evaluate them. One such framework was proposed by [36], who developed a toolkit for evaluating Serious Games by defining generic criteria for evaluation and analysis. The criteria are presented under a list of features that group together (Scripting, Game Theory, Pedagogy), which are essential for a good methodological evaluation of Serious Games. The methodology was tested on several Serious Games available on the web, and the results were satisfactory in terms of the information needed for the adoption of a Serious Game in a learning process. However, the methodology needs to be refined to rank Serious Games according to the skills and knowledge offered and the priority in a given learning situation. Another framework was developed by [37], which includes a generic evaluation methodology for Serious Games that

resolves the dilemma between generality and standardization necessary for comparative research based on theory, specificity, and flexibility necessary for the evaluation of singular cases. The framework consists of eight essential steps and provides the most comprehensive evaluation recommendations for Serious Games. However, it is limited to the extent that learning outcomes are given the highest priority in assessment, with relatively less attention to Serious Game systems considerations. A Serious Games evaluation framework was proposed [38], emphasizing baseline evaluation in each of the four key conceptual areas: theoretical, technical, empirical, and external. This framework provides guidelines that highlight the importance of anchoring an evaluation strategy in each of the four conceptual areas, making it valuable for future researchers and practitioners interested in leveraging Serious Games as an effective training tool. Although there is a range of evaluation criteria addressed in different frameworks and methodologies dedicated to the evaluation of Serious Games, there is no consensus on a very specific evaluation approach because each evaluation approach has strengths and weaknesses. While the methodology proposed by [37] is positioned as the most comprehensive approach [39], no information on its applicability and validity has been found. On the other hand, the Serious Game evaluation model named MEEGA+ proposed by [30] is the most applied and evaluated Serious Game evaluation model in the scientific literature, but it cannot generally evaluate all Serious Games, requiring other specific criteria [40]. Thus, a more consistent and uniform approach is needed to systematically evaluate Serious Games not as a training tool but as the culmination of a project to develop a tool dedicated to use in a formative context.

3. SERIOUS GAMES EVALUATION SYSTEM DIMENSIONS

Building on the state of the art presented earlier, it is concluded that a more coherent and unified approach is needed to systematically evaluate Serious Games, not only as a training tool but also as a project focused on developing tools specifically designed for use in training environments. To address this need, the proposed Serious Game evaluation system is based on four main dimensions: pedagogical, technological, ludic, and behavioural. Each dimension will be measured according to well-defined criteria.

3.1 Pedagogical dimension (PD)

The objective of this dimension is to evaluate the pedagogical content provided by a Serious Game. It is essential that the pedagogical content of a Serious Game is well balanced to create harmony between gameplay and learning objectives. The following table (Table 1) describes the criteria used to evaluate this dimension.

Table 1. Pedagogical Dimension (PD)

Criteria	Description
Targeted skills (Ts)	Players need to realize that their skills are at a level where it is possible to overcome the challenges of the game. As the difficulty increases, challenges should require the player to develop skills to progress in the game and have fun simultaneously [41-42].
Pedagogical consideration (Pc)	Which refers to the educational content provided by a Serious Game. It is essential that the educational content of a Serious Game is well adjusted and balanced. According to [43], the Serious Game must be designed appropriately to create harmony between the game and the learning objectives.
Learning result (Lr)	Learning outcomes are the knowledge and skills that users acquire through their interaction with a Serious Game. Many studies first evaluate the learning outcomes of a Serious Game before examining other dimensions or characteristics. Analysis of these learning outcomes is often done using both qualitative and quantitative methods [44-45]. In some studies, learning outcomes have been categorized into different types such as cognitive, motor, affective, and communicative. For example [46] divided learning outcomes into these four categories.
Error management (Em)	Serious Games must allow learners to make mistakes so that they become aware of the consequences of their actions [47]. The game should then suggest the appropriate solution to guide learners back to the right path of learning. In Serious Games, errors are considered a source of learning and should not be eliminated or denied, but rather used to enhance the learning experience [48].

3.2 Technological dimension (TD)

This dimension is important, because a Serious Game must be very attractive and benefit in its design from the technological advances of the video game industry. These criteria are described in the following table (Table 2):

Table 2. Technological Dimension (PD)

Criteria	Description
Game design (Gd)	The design of a Serious Game refers to its visually appealing and artistic appearance. It is an important factor that can influence user satisfaction and engagement with the game [47-49]. The design of the game has been measured and evaluated in various studies.
Performance (P)	What can be described as the functioning and efficiency of a Serious Game is referred to as its performance characteristic. This characteristic has been measured in studies evaluating the functionality of the Serious Game [50]. It has also been used to evaluate a Serious Game during its development process [51].
User Interface (UI)	Referring to the interaction that occurs between a user and a Serious Game, it is essential to design user interfaces for maximum accessibility and usability [52]. Researchers have indicated that Serious Games are unlikely to succeed without an appropriate and well-designed user interface [53].

Usability (U)	Usability has been defined by the international standard "ISO 9241-11" as "the extent to which a product can be used by specific users to achieve specific objectives effectively, efficiently, and satisfactorily in a given use context [54]." According to this definition, usability is a construct that has three dimensions: effectiveness, efficiency, and satisfaction. Similarly, [55] stated that ease of use is associated with five attributes: learning ability, efficiency, memorization, errors, and satisfaction. Numerous studies have measured the usability characteristic of Serious Games, such as [56-57].

3.3 Ludic dimension (LD)

This dimension encompasses all interactive components of a Serious Game, aiming to ensure and sustain the learner/player's motivation throughout the learning process. Its main measurement criteria are presented in the following table 3:

Table 3. Ludic Dimension (PD)

Criteria	Description
Challenge (C)	The challenge of a Serious Game refers to the balance between the difficulty level and the player's competence. It is essential that the challenges are stimulating enough to maintain the player's interest, but not so difficult as to cause frustration. The increase in difficulty must be gradual and occur at appropriate times to ensure a smooth learning curve. Introducing new obstacles and situations throughout the game can also help minimize fatigue and maintain student engagement [41].
Fun (F)	To improving motivation and engagement, incorporating elements of fun in a Serious Game can also promote learning by inducing a positive emotional state that can enhance memory retention and knowledge acquisition. Emotions are known to be a significant factor in the learning process, and research has demonstrated that positive emotions can lead to improved learning outcomes [58]. Fun, in this context, can be defined as a sensation of pleasure, happiness, relaxation, and diversion [59].
Gameplay (G)	This dimension encompasses all the experiences and interactions that a player has while engaging with a Serious Game [60].
Immersion (I)	The concept of immersion in a Serious Game refers to the player's ability to become deeply engaged and involved in the game, creating a sense of challenge that is centered on the game world and causing the player to forget about the outside world while playing [61].

3.4 Behavioural dimension (BD)

Behaviour is a dimension that encompasses the involvement and attitude of players/learners towards the Serious Game. The description of these criteria is provided in Table 4.

Table 4. Behavioural Dimension (PD)

Criteria	Description
Motivation (M)	This refers to the ability of a Serious Game to impact the user's motivation and encourage its usage. According to [62], Serious Games stimulate students' intrinsic motivation by encouraging curiosity and a desire for challenge. Additionally [63], in their analysis of previous studies on the effectiveness of Serious Games, found that motivational factors were key to the successful use of Serious Games. The characteristic of motivation has been measured and studied in several research works such as [62-45].
Engagement (E)	When it comes to involving students in the use of a Serious Game, engagement sis a complex issue, as indicated by [64], where the study considers motivation as the root of commitment.
User Experience (Ue)	This refers to the emotions, perceptions, and attitudes of learners towards the use of a particular Serious Game [65].

The numerous criteria associated with the selected evaluation dimensions make it challenging for evaluators to determine their appropriate weighting. This task is further complicated by the requirement for the weighting to be both objective and contextually aligned with the use of Serious Games. Consequently, it was determined that a multi-criteria decision-making approach would be appropriate for weighting the dimensions and criteria involved in the evaluation process.

4. CRITERIA WEIGHTING METHOD

Due to the complexity and multitude of criteria involved in the evaluation of Serious Games, it is essential to weight these criteria appropriately. To achieve this, a multi-attribute decision support method (MADM) [66] was considered necessary. MADM methods have been successfully applied in various domains, including logistics, education, and technology, to facilitate informed decision-making by structuring complex problems and explicitly considering multiple criteria. However, incomplete and vague information often requires the inclusion of the theory of fuzzy sets. Fuzzy sets were first introduced by [67] to mathematically represent imprecision relative to certain classes of objects. A fuzzy set consists of values that belong to a particular class with a degree of certainty. Fuzzy sets are useful in modeling uncertain or ambiguous data commonly encountered in real-life scenarios and allows for considerable flexibility in reasoning, making it possible to consider inaccuracies and uncertainties [67-68]. The combination of fuzzy sets and MADM methods has been well-suited to manage ambiguity encountered in multi-attribute decision-making problems. This combination allows decision-makers to describe the problem environment and its properties more realistically, thereby building a rational decision-making model [68]. Since the evaluation space is discrete, the Fuzzy Analytic Hierarchy Process (Fuzzy AHP) described by [69] has been chosen for implementation in the proposed Serious Game evaluation system. The Fuzzy AHP method utilizes language variables represented by triangular fuzzy numbers to calculate relative weights for each measurement of the evaluation dimensions. A triangular fuzzy number is represented by a triplet (a, b, c) that respectively gives the lower bound, the modal value, and the

upper bound of the fuzzy set, as shown in Figure 1.

The height of the set M, denoted as h(M), corresponds to the upper limit of the membership function of M, which is defined as $h(M) = \sup\{U\_M(x) \mid x \in X\}$.
The support of M is the set of elements in X that have a nonzero degree of membership in M. In other words, it is the set $supp(M) = \{x \in X \mid U\_M(x) > 0\}$.
The core of M is the set of elements in X that have a maximum degree of membership in M. In other words, it is the set $noy(M) = \{x \in X \mid U\_M(x) = 1\}$.

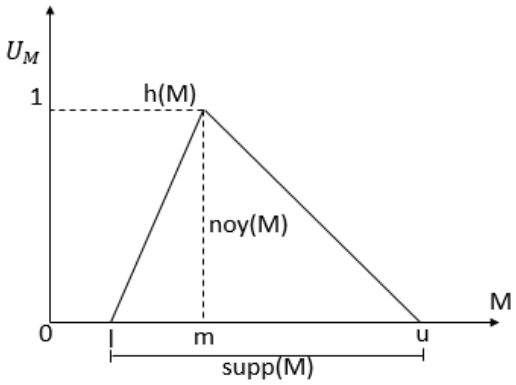

Fig 1: Triangular fuzzy number [68]

Using the fuzzy AHP method, evaluation criteria can be grouped into different levels and categories based on similar characteristics by constructing a hierarchical structure. This ensures logical consistency in the judgments used to determine the priorities of our evaluation criteria while taking into account the vagueness and uncertainty in the evaluator's judgments.

5. FUNCTIONAL PRINCIPLE

To use the proposed evaluation tool, the evaluator first defines the context in which the Serious Game will be used. This involves assigning weights to the relevant evaluation dimensions. The validity of these weights is then assessed using the fuzzy AHP method, which ensures logical consistency and accounts for any vagueness or uncertainty in the evaluator's judgments. The evaluation process employs four dimensions to assess the Serious Game, and the resulting feedback provides the evaluator with insight into the extent to which the Serious Game aligns with the intended context of use.

5.1 Evaluation process

The evaluation process for the proposed Serious Game begins with the evaluator comparing the criteria using linguistic values. The comparison is then processed to obtain an appropriate weighting for each criterion in the evaluator's decision, while maintaining a consistency ratio that does not exceed 0.10, as required by the Fuzzy AHP method.

Afterwards, students are invited to use the selected Serious Game and complete a predefined questionnaire provided by the evaluator. The results are then generated from the analysis of the questionnaire.

The assessment process is illustrated in Figure 2, which includes three main blocks:

- SG: represents the Serious Game to be evaluated.
- Serious Game Evaluation Tool: consists of several sub-blocks including:
  - ✓ Linguistic evaluation scale: where the evaluator expresses pairwise comparisons between all criteria using linguistic variables.
  - ✓ Triangular fuzzy number: where the expressed language variables are translated into triangular fuzzy numbers.
  - ✓ FMADM: where the fuzzy AHP method is used.
  - ✓ MCQ extraction: where a questionnaire is extracted to evaluate the selected Serious Game.
  - ✓ Presentation of the MCQ: where obtained questionnaire is presented to students to evaluate the selected Serious Game.
- Results: includes the evaluation results obtained.

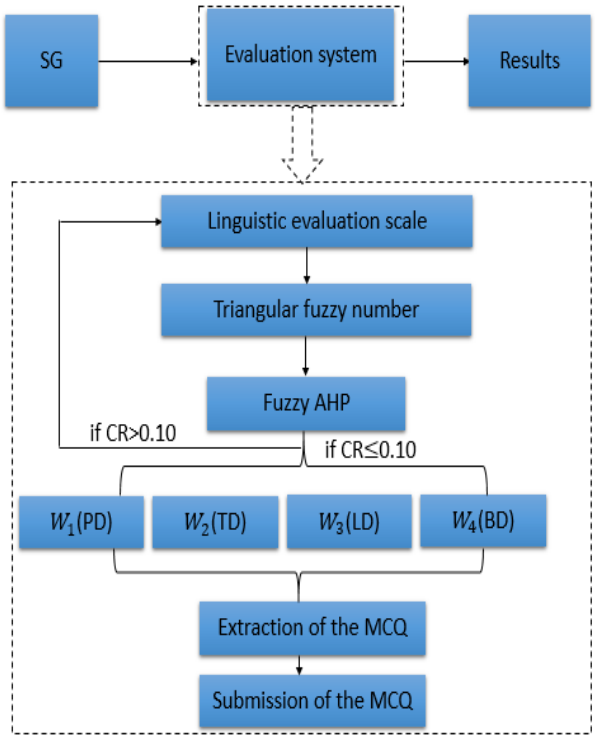

Fig 2: Serious Games Evaluation Process

6. APPLICATION OF THE EVALUATION SYSTEM

To test our proposed evaluation system, the Pedagogical Commission of the Department of Mathematics and Informatics at the University Hassan II of Casablanca approved an experiment to evaluate the Serious Game "ROBOCODE." This experiment aims to assess the effectiveness and suitability of our evaluation system for Serious Games.

6.1 Criteria weighting process

The seven steps involved in the weighting process for Serious Game evaluation criteria using the fuzzy AHP method are as follows:

Step 1: Pair Comparison between Criteria

As an evaluator, the first step is to compare the evaluation dimensions by pairs, assigning a linguistic value to each comparison. As shown in Table 5, PD was prioritized over all other dimensions, given that the context is purely formative. Additionally, the target population is academic and scientific, and therefore, familiar with new information technologies, leading us to favor TD over BD and LD. Furthermore, the use of Serious Game in an educational activity has led us to prioritize BD over LD. These linguistic values will be then translated into triangular fuzzy numbers $\widetilde{P}_{ij}$.

Table 5. Language Value Mapping – Triangular Fuzzy Number

Pairwise comparison	Linguistic value	Relative importance	Triangular fuzzy number $\tilde{P}_{ij}$
(PD, TD)	high importance	$\tilde{5}$	[4,5,6]
(PD, BD)	high importance	$\tilde{5}$	[4,5,6]
(PD, LD)	Very high importance	$\tilde{9}$	[8,9,9]
(TD, BD)	Moderate importance	$\tilde{3}$	[2,3,4]
(TD, LD)	high importance	$\tilde{5}$	[4,5,6]
(BD, LD)	Equal importance	$\tilde{1}$	[1,1,1]

Step 2: Construction of the fuzzy judgment matrix $\widetilde{M}$

The fuzzy judgment matrix, composed of triangular fuzzy numbers, is defined by aggregating all the triangular fuzzy numbers $\tilde{P}_{ij}$, represented as $\{l_{ij}, m_{ij}, u_{ij}\}$.

$$\widetilde{M} = \begin{pmatrix} (1,1,1) & \cdots & \tilde{P}_{ij} \\ \vdots & \ddots & \vdots \\ \tilde{P}_{ij} & \cdots & (1,1,1) \end{pmatrix}$$

With: $\tilde{P}_{ij}=\tilde{P}_{ij}^{-1}$ and $\tilde{P}_{ij}^{-1} = \left(\frac{1}{u_{ij}}, \frac{1}{m_{ij}}, \frac{1}{l_{ij}}\right)$

Thus, our fuzzy judgment matrix $\widetilde{M}$:

$$\widetilde{M} = \begin{pmatrix} (1,1,1) & (4,5,6) & (8,9,9) & (4,5,6) \\ \left(\frac{1}{6},\frac{1}{5},\frac{1}{4}\right) & (1,1,1) & (4,5,6) & (2,3,4) \\ \left(\frac{1}{9},\frac{1}{9},\frac{1}{8}\right) & \left(\frac{1}{6},\frac{1}{5},\frac{1}{4}\right) & (1,1,1) & (1,1,1) \\ \left(\frac{1}{6},\frac{1}{5},\frac{1}{4}\right) & \left(\frac{1}{4},\frac{1}{3},\frac{1}{2}\right) & (1,1,1) & (1,1,1) \end{pmatrix}$$

Step 3: Calculate the consistency ratio

At this stage, the evaluation system may suggest revisiting the initial comparisons to ensure a consistency ratio of no more than 0.10, as recommended by [70], to validate the selections made. The consistency ratio is defined as the ratio between the consistency index (CI) of the evaluation matrix and the consistency index of a randomly generated matrix (RI).

$$CI = \frac{\lambda_{max} - n}{n-1} \text{ and } CR = \frac{CI}{RI} \leq 0.10$$

With λmax is the eigenvalue and n is the number of criteria. The value of RI is 0.90.

In our case, we obtained λmax= 4.185, n = 4 and CI = 0.061, resulting in a CR of 0.0685<0.10. This indicates that the pair comparisons made is consistent and valid.

Step 4: Calculate the geometric mean

The system computes the fuzzy geometric mean for each criterion using the following equation:

$$\tilde{r}_i = \left(\prod_{j=1}^{n} \tilde{P}_{ij}\right)^{\frac{1}{n}} ; \text{ i = 1,2, ……., n}$$

Step 5: Fuzzy weight calculation

The fuzzy weight of each criterion ($\widetilde{w}_i$) is calculated by multiplying each fuzzy geometric mean by the inverse of its vector, using the equation below.

$$\widetilde{w}_i = \tilde{r}_i \otimes \left(\sum_{i=1}^{n} \tilde{r}_i\right)^{-1} ; \text{ i = 1,2, ……., n}$$

Step 6: Defuzzification of the fuzzy weights

The defuzzification of the fuzzy weights is done by applying the equation below:

$$M_i = \frac{lw_i + mw_i + uw_i}{3} ; \text{ i = 1,2,….., n}$$

Step 7: Normalization of fuzzy weights

The normalized fuzzy weights are obtained using the equation below:

$$N_i = \frac{lw_i + mw_i + uw_i}{\sum_{i=1}^{n} M_i} \quad \text{i = 1,2, ……., n}$$

After applying the steps mentioned on all the criteria, the following results are obtained:

Table 6. Language weightings of the criteria

Dimension	Criterion	r_i	W_i	M_i	N_i	priority
DP (0,631)	**Ts**	**(3.364 ,3.873 ,4.243)**	**(0.493 ,0.637 ,0.807)**	**0.645**	**0.631**	**0.398**
	Pc	**(1.075 ,1.316 ,1.565)**	**(0.157 ,0.216 ,0.298)**	**0.224**	**0.219**	**0.138**
	Lr	**(0.452 ,0.508 ,0.595)**	**(0.066 ,0.084 ,0.113)**	**0.088**	**0.086**	**0.054**
	Em	**(0.369 ,0.386 ,0.420)**	**(0.054 ,0.063 ,0.080)**	**0.066**	**0.064**	**0.040**
TD (0,219)	**Gd**	**(1.075 ,1.316 ,1.565)**	**(0.157 ,0.216 ,0.298)**	**0.224**	**0.219**	**0.048**
	P	**(0.452 ,0.508 ,0.595)**	**(0.066 ,0.084 ,0.113)**	**0.088**	**0.086**	**0.019**

	Ui	(0.369 ,0.386 ,0.420)	(0.054 ,0.063 ,0.080)	0.066	0.064	0.014
	U	(3.364 ,3.873 ,4.243)	(0.493 ,0.637 ,0.807)	0.645	0.631	0.138
LD (0,064)	C	(1.075 ,1.316 ,1.565)	(0.157 ,0.216 ,0.298)	0.224	0.219	0.014
	F	(0.452 ,0.508 ,0.595)	(0.066 ,0.084 ,0.113)	0.088	0.086	0.006
	G	(3.364 ,3.873 ,4.243)	(0.493 ,0.637 ,0.807)	0.645	0.631	0.041
	I	(0.369 ,0.386 ,0.420)	(0.054 ,0.063 ,0.080)	0.066	0.064	0.004
BD (0,086)	M	(0.369 ,0.386 ,0.420)	(0.059 ,0.069 ,0.087)	0.072	0.070	0.006
	E	(3.364 ,3.873 ,4.243)	(0.540 ,0.695 ,0.883)	0.706	0.690	0.059
	Ue	(1.075 ,1.316 ,1.565)	(0.173 ,0.236 ,0.326)	0.245	0.239	0.020

After completing the priority calculation for each criterion, the following hierarchy is obtained (Figure 3).

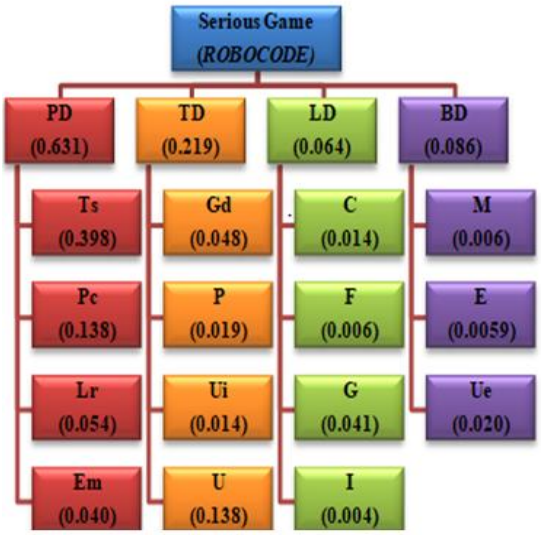

Fig 3: Serious Game Rating System (ROBOCODE) weightings

The figure represents a hierarchical structure of the evaluator's decisions regarding the selected criteria.

6.2 The Serious Game chosen

A pedagogical commission from the Department of Mathematics and Computer Science at the University Hassan II of Casablanca confirmed the selection of the Serious Game "ROBOCODE", for an experimental study. The objective of this Serious Game is to teach the basics of object-oriented programming (OOP). This Serious Game, "ROBOCODE," is a free educational programming video game created and distributed by IBM. It is specifically designed to teach the Java programming language (as shown in Table 7).

Table 7. Identification of the Serious Game (ROBOCODE)

Name of the game	ROBOCODE
Game developer	Flemming N. Larsen (IBM)
Game genre	Action (combat simulation)
Game type	programming game
Gaming experience required	beginner
Description	Programming game, where the goal is to develop a robotic battle tank to fight against other tanks in Java or .NET.
language	English

The player programs the robot by giving it commands to respond to events on the battlefield. Therefore, ROBOCODE provides a platform for students and learners to learn and apply their knowledge of OOP, which includes tasks such as writing, reading, analyzing and using existing code, event management and message transmission for robots to engage in automated battles on a rectangular battlefield. The robots are capable of movement, searching, and shooting at each other.

6.3 Data collection instrument

To gather information about the test of the evaluation system conducted with students from the University Hassan II of Casablanca, a questionnaire based on the Likert scale was created. The Likert scale is a psychometric scale commonly used in research to measure respondents' attitudes by asking them to what extent they agree or disagree with a particular statement or question [71]. Due to its simplicity and popularity, the traditional Likert scale has been extended to many variants, including the 4-point Likert scale, which is a forced Likert scale [72]. This means that users are required to form an opinion, with no neutral option available in the choices [73]. The questionnaire was designed based on the adopted Serious Game evaluation criteria to measure learners' attitudes toward the Serious Game "ROBOCODE" after completing its "hot evaluation" sequence.

7. RESULTS AND DISCUSSION

After playing the Serious Game "ROBOCODE", the students were invited to evaluate it by answering our questionnaire. The consistency of the internal reliability was measured through this evaluation, based on the correlations between the different elements of the questionnaire. This measurement was made using an index called "Cronbach's Alpha," which is a popular method for assessing the reliability of a questionnaire [74]. Generally, Cronbach Alpha values between 0.70 and 0.95 are considered acceptable [75]. In our study, the total Cronbach Alpha value was calculated to be 0.938, indicating acceptable internal coherence not only in the students' reactions to the Serious Game but also inter-coherence between the items related to each of the aspects (Pedagogical, Ludic, Technological, and Behavioural). This implies that the answers between the items are consistent and accurate, indicating the reliability of the standardized elements of the questionnaire.

Based on the results obtained (Figure 4), the Serious Game "ROBOCODE" appears to be more suitable for use in a purely educational context, as confirmed by the strong value achieved in the pedagogical dimension.. Note also that this Serious Game has excellent operability and also it facilitates the learning of programming including the notion of POO. Thus, it has clear and relevant rules for the interests of students and offers excellent focused attention so as not to have monotonous activities. For the technological dimension, this Serious Game presents a very attractive design that allows easy and fluid interaction, avatars of excellent quality, as well as performances that allow students fluidity and flexibility during the game. Regarding the playful dimension, this Serious Game offers a strong interaction and produces playful moments for players, challenges that arouse the interest of players to interact with the Serious Game. Concerning the behavioural dimension, this Serious Game presents a medium attention to the players, even if the students do not forget their environment. Thus, this Serious Game also provides feelings of trust and satisfaction between players and moments of conviviality and pleasure between players.

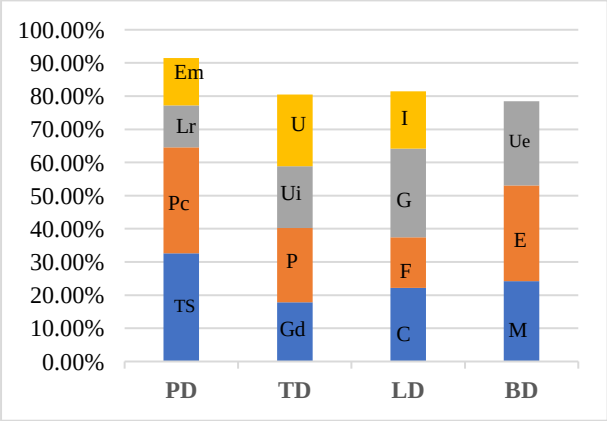

Fig 4: Serious Game "ROBOCODE" analysis results

Overall, this Serious Game was beneficial for the students, as they learned many concepts about programming and in particular about the object-oriented programming concept. Thus, they enjoyed the introduction of this Serious Game during the class session by discovering a new and playful way of learning.

8. CONCLUSION

This article presents the implementation of a Serious Games evaluation system designed to consider four dimensions: pedagogical, technological, behavioral, and ludic. The multi-criteria fuzzy decision support method "Fuzzy AHP" was utilized and integrated into a program developed using Matlab to weight and to validate these dimensions. The results of the experimentation have shown that the Fuzzy AHP method can be effectively employed to weight the criteria, leading to an evaluation system that is flexible and well-suited to the specific context of Serious Games. Furthermore, the results obtained have demonstrated the usefulness and effectiveness of the evaluation system in practice.

In the future works, the application of the proposed evaluation system will be extended to various Serious Games to gain a deeper understanding of its effectiveness across different contexts and game types.

9. ACKNOWLEDGEMENTS

The authors would like to express their gratitude to all those who contributed to this research project.

10. REFERENCES

[1] D. Michael and S. Chen, "Serious Games: Games That Educate, Train, and Inform," Boston, Mass, Thomson Course Technology PTR, 2006.

[2] M. Papastergiou, "Digital game-based learning in high school Computer Science education: Impact on educational effectiveness and student motivation," Computers & Education, vol. 52, pp. 1-12, 2009, doi : 10.1016/j.compedu.2008.06.004.

[3] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey and J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," Computers & Education, vol. 59, pp. 661-686, 2012, doi : 10.1016/j.compedu.2012.03.004.

[4] S. De Freitas and S. Jarvis, "Serious games: a new paradigm for education?," Serious games, Springer, Dordrecht, pp. 9-23, 2006, doi: 10.1007/978-1-4471-2161-9_2.

[5] Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless... EDUCAUSE Review, 41(2), 16-30.

[6] Aldrich, C. (2009). Learning by doing: A comprehensive guide to simulations, computer games, and pedagogy in e-learning and education. Jossey-Bass.

[7] Koenig, R., and Roth, D. (2015). The effectiveness of serious games for training and education. Journal of Educational Technology Systems, 43(3), 259-279. https://doi.org/10.1177/0047239515582715

[8] McMahan, A. (2003). Immersion, engagement, and presence: A method for analyzing 3D video games. In Proceedings of the International Conference on Human-Computer Interaction, 3, 459-463.

[9] Deci, E. L., and Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01

[10] Sweetser, P., and Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. In Proceedings of the 2005 Conference on Future Play, 1-8. https://doi.org/10.1145/1100000.1100002

[11] Gee, J. P. (2003). What video games have to teach us about learning and literacy. Computers in the Schools, 19(3-4), 19-30. https://doi.org/10.1300/J025v19n03_03

[12] Anderson, C. A., and Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. Journal of Personality and Social Psychology, 78(4), 772-790. https://doi.org/10.1037/0022-3514.78.4.772

[13] Mackey, T. K., and Nayyar, G. (2016). A review of the role of games in public health education. Journal of Health Communication, 21(3), 267-274. https://doi.org/10.1080/10810730.2015.1069771

[14] Susi, T., Johannesson, M., and Backlund, P. (2007). Serious games – An overview. University of Skövde, Sweden.

[15] Rugelj, M., and Novak, M. (2013). Serious games: The impact on learning and skill development. Journal of Educational Multimedia and Hypermedia, 22(1), 5-24.

[16] Schell, J. (2008). The art of game design: A book of lenses. Morgan Kaufmann.

[17] Clark, D. B., and Estes, A. (2008). The influence of video games on learning. International Journal of Gaming and Computer-Mediated Simulations, 1(1), 46-68. https://doi.org/10.4018/jgcms.2008010104

[18] Agarwal, R., and Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. MIS Quarterly, 24(4), 665-694. https://doi.org/10.2307/3250951

[19] Liu, M., and Chen, M. (2012). The effects of serious games on student learning and engagement. Computers and Education, 58(2), 220-229. https://doi.org/10.1016/j.compedu.2011.08.007

[20] Siang, C., and Lee, W. (2015). Educational games: Data analytics for assessment and improvement. International Journal of Game-Based Learning, 5(3), 51-64. https://doi.org/10.4018/IJGBL.2015070104

[21] Hainey, T., and Connolly, T. M. (2011). Serious games and their educational potential. In Proceedings of the European Conference on Games Based Learning, 243-251.

[22] Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen and R. L. Mack (Eds.), Usability inspection methods (pp. 25-62). John Wiley and Sons.

[23] Togelius, J., and Yannakakis, G. (2013). The art of game design. CRC Press.

[24] Y. El Borji, "Comparative Study to Develop a Tool for the Quality Assessment of Serious Games Intended to be Used in Education," International Journal of Emerging Technologies in Learning (iJET), vol. 9, pp. 50-55, 2014, doi: 10.3991/ijet.v9i9.4150.

[25] M. Videnovik, A. Madevska Bogdanova and V. Trajkovik, " SERIOUS GAMES EVALUATION METHODOLOGY", 2018, doi: 10.21125/iceri.2018.1308.

[26] I. Marfisi-Schottman, "Methodology, Models and Tools for Designing Learning Games", 2012.

[27] M. R. Lepper and T. W. Malone, "Intrinsic motivation and instructional effectiveness in computer-based education," Aptitude, learning, and instruction. Hillsdale, NJ : Erlbaum, pp. 255-286, 1987.

[28] E. Sanchez, "Key criteria for Game Design. A Framework", Rapport du MEET Project, European Commission, 2011.

[29] A. J. Abdellatif, B. McCollum and P. McMullan, "Serious Games: Quality Characteristics Evaluation Framework and Case Study," in Proc. IEEE Integrated STEM Education Conference (ISEC), 2018, pp. 112-119, doi: 10.1109/ISECon.2018.8340460.

[30] G. Petri, C.Gresse von Wangenheim and Borgatto, "MEEGA+: A Method for the Evaluation of the Quality of Games for Computing Education", in Brazilian Symposium on Games and Digital Entertainment, 2019, doi: 10.5753/cbie.wcbie.2019.951.

[31] F. Bellotti, B. Kapralos, K. Lee, P. Moreno-Ger and R.Berta, "Assessment in and of Serious Games: An Overview", Advances in Human-Computer Interaction, 2013, doi: 10.1155/2013/136864.

[32] A. Serrano-Laguna, B. Manero, M. Freire and B. Fernández-Manjón, "A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes," Multimedia Tools and Applications, vol. 77, 2018, doi: 10.1007/s11042-017-4467-6.

[33] K. Emmerich and M.Bockholt, "Serious Games Evaluation: Processes, Models, and Concepts", Lecture notes in computer science, 2016, doi: 10.1007/978-3-319-46152-6_11.

[34] Y. Xu, P. M. Johnson, C. A. Moore, R. S. Brewer and J. Takayama, "SGSEAM: Assessing serious game frameworks from a Stakeholder Experience perspective", in ACM International Conference Proceeding Series, 2013, doi: 10.1145/2583008.2583018.

[35] J. W. Creswell, "Research design: qualitative, quantitative, and mixed methods approaches," 2nd ed. ed. Sage Publications, Thousand Oaks, California, 2003.

[36] A. Ghannem, "Characterization of serious games guided by the educational objectives", in TEEM '14: Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality, 2014, pp.227–233, doi: 10.1145/2669711.2669904.

[37] I. Mayer, G. Bekebrede, C. Harteveld, H. Warmelink, Q. Zhou, T. van Ruijven, J. Lo, R. Kortmann and I. Wenzler, "The research and evaluation of serious games: Toward a comprehensive methodology,".British Journal of Educational Technology. Vol. 45, pp. 502-507, 2014, doi: 10.1111/bjet.12067.

[38] D. W. Wilson et al., "Serious games: an evaluation framework and case study," System Sciences (HICSS), 49th Hawaii International Conference, 2016, pp. 638-647, doi: 10.1109/HICSS.2016.85.

[39] I. Mayer, "Towards a comprehensive methodology for the research and evaluation of serious games," Procedia Comput. Sci. vol. 15, pp. 233–247, 2012, doi: 10.1016/j.procs.2012.10.075.

[40] G. Petri, "A METHOD FOR THE EVALUATION OF THE QUALITY OF GAMES FOR COMPUTING EDUCATION", 2018.

[41] P. Sweetser and P. Wyeth, " GameFlow: a model for evaluating player enjoyment in games," Computers in Entertainment, vol. 3, pp. 1-24, 2005, doi: 10.1145/1077246.1077253.

[42] J. Takatalo, J. Häkkinen, J. Kaistinen, G. Nyman, "Presence, Involvement, and Flow in Digital Games," In Bernhaupt, R. (Ed.). Evaluating User Experience in Games: Concepts and Methods, pp. 23-46, Springer, 2010, doi: 10.1007/978-1-84882-963-3_3.

[43] S. D. Freitas and M. Oliver, "How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?," Computers & Education, vol. 46, pp. 249-264, 2006, doi:10.1016/j.compedu.2005.11.007.

[44] A. Baker, J. Zhang and R. E. Caldwell, "Reinforcing Array and Loop Concepts Through a Game-Like Module," The 17th International Conference on Computer Games, 2012, pp. 175-179.

[45] G. C. Wangenheim, R. Savi and F. A. Borgatto, "DELIVER! – An educational game for teaching Earned Value Management in computing courses," Information and Software Technology, vol. 54, pp. 286-298, 2011, doi: 10.1016/j.infsof.2011.10.005.

[46] P. Wouters, E. van der Spek and H. van Oostendorp, "Current practices in serious game research: A review from a learning outcomes perspective," Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices, pp. 232-255, 2009, doi: 10.4018/978-1-60566-360-9.ch014.

[47] K. Amokrane and D. Lourdeaux, "Virtual Reality Contribution to training and risk prevention," the proceeding of the International Conference on Artificial Intelligence, 2009, pp. 466-472, http://hal.archives-ouvertes.fr/hal-00462120.

[48] J. P. Astolfi, "L'erreur, un outil pour enseigner," ESF, 1997.

[49] N. El-Kechaï and C. Després, "A plan recognition process, based on a task model, for detecting learner's erroneous actions," Lecture Notes in Computer Science, vol. 4053, pp. 329-338, 2006, doi:10.1007/11774303_33.

[50] J. Zhang, R. E. Caldwell and E. Smith, "Learning the Concept of Java Inheritance in a Game," The 18th International Conference on Computer Games (CGAMES), 2013, pp. 212-216, doi:10.1109/CGames.2013.6632635.

[51] B. Manero, C. Fernandez-Vara and B. Fernandez-Manjon, "E-Learning Takes the Stage: From La Dama Boba to a Serious Game," IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, Vol. 8, pp. 179-204, 2013, doi :10.1109/RITA.2013.2285023.

[52] S. Lanyi, J. Brown, P. Standen, et al., "Results of user interface evaluation of serious games for students with intellectual disability," Acta Polytechnica Hungarica, Vol. 9, pp. 225-245, 2012.

[53] Z. Mikovec, P. Salvik and J. Zara, "Cultural Heritage, User Interfaces and Serious Games at CTU Prague,"15th International Conference on Virtual Systems and Multimedia, 2009, pp. 211-116.

[54] ISO/IEC 9126:1998 Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on usability, 1998.

[55] J. Nielsen, "Usability Engineering," Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 1993.

[56] M. Muratet, P. Torguet, F. Viallet, et al., "Experimental feedback on Prog&Play: a serious game for programming practice," Computer Graphics Forum, vol. 30, pp. 61-73, 2011.

[57] D. Rodríguez-Cerezo, A. Sarasa-Cabezuelo, M. Gomez-Albarran and J. Sierra, "Serious games in tertiary education: a case study concerning the comprehension of basic concepts in computer language implementation courses," Computers in Human Behavior, vol. 31, pp. 558-570, 2014, doi: 10.1016/j.chb.2013.06.009.

[58] B. L. Fredrickson, "The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions," American psychologist, vol. 56, pp. 218-226, 2001, doi: 10.1037/0003-066X.56.3.218.

[59] K. Poels, Y. D. Kort, W. Ijsselsteijn, "It is always a lot of fun!: exploring imensions of digital game experience using focus group methodology," Proc. of Conf. on Future Play, Toronto, Canada, 2007, pp. 83-89.

[60] C. V. Carvalho, "Is game-based learning suitable for engineering education?," Proc. of the Global Engineering Education Conf., IEEE, 2012, pp.1-8, doi: 10.1109/EDUCON.2012.6201140.

[61] F. Fu, R. Su and S. Yu, "EGameFlow: A scale to measure learners' enjoyment of elearning games", Computers & Education, vol. 52, pp. 101-112, 2009, doi: 10.1016/j.compedu.2008.07.004.

[62] M. Wrzesien and M. Alcaniz, "Learning in serious virtual worlds: evaluation of learning effectiveness and appeal to students in the E-Junior project," Computers & Education, vol. 55, pp. 178–187, 2010, doi: 10.1016/j.compedu.2010.01.003.

[63] C. Girard, J. Ecalle and A. Magnant, "Serious games as new educational tools: how effective are they? A meta-analysis of recent studies," Journal of Computer Assisted Learning, vol. 29, 2013, doi: 10.1111/j.1365-2729.2012.00489.x.

[64] O. Dele-Ajayi, R. Strachan, J. Sanderson and A. Pickard, "Learning Mathematics Through Serious Games: An engagement framework," Frontiers in Education Conference (FIE), 2016, doi: 10.1109/FIE.2016.7757401.

[65] L. E. Law and X. Sun, "Evaluating user experience of adaptive digital educational games with Activity Theory," International Journal of Human-Computer Studies, vol. 70, pp. 478-497, 2012, doi:10.1016/j.ijhcs.2012.01.007.

[66] E. Zavadskas, Z. Turskis and S. Kildiene, "State of art surveys of overviews on MCDM/MADM methods," Technological and Economic Development of Economy, vol. 20, pp. 165-179, 2014, doi: 10.3846/20294913.2014.892037.

[67] A. Zadeh, "Fuzzy sets," Information and Control, Vol. 8, pp. 338-353, 1965, doi: 10.1016/S0019-9958(65)90241-X.

[68] A. Lazim, "Fuzzy Multi Criteria Decision Making and its Applications: A Brief Review of Category", Procedia - Social and Behavioral Sciences, 2013, vol. 97, pp. 131-136, doi: 97. 10.1016/j.sbspro.2013.10.213.

[69] J. J. Buckley, "Fuzzy hierarchical analysis," Fuzzy Sets and Systems, vol. 17, pp. 233-247, 1985, doi:10.1016/0165-0114(85)90090-9.

[70] T. L. Saaty, “Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World”, RWS Publications, Pittsburgh, 1995.

[71] R. Likert, “A technique for the measurement of attitudes,” Archives of psychology, vol. 140, pp. 5–55, 1932.

[72] S. Y. Chyung, K. Roberts, I. Swanson and A. Hankinson, “Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale,” Performance Improvement. vol. 56, pp. 15-23, 2017, doi: 10.1002/pfi.21727.

[73] Z. Awang, A. Afthanorhan and M. Mamat, “The Likert scale analysis using parametric based Structural Equation Modeling (SEM),” Computational Methods in Social Sciences (CMSS). "Nicolae Titulescu" University of Bucharest, Faculty of Economic Sciences, vol. 4, pp. 13-21, 2016.

[74] E. G. Carmines and R. A. Zeller, “Reliability and Validity Assessment”, vol. 17, Thousand Oaks, CA, SAGE Publications, 1979, doi: 10.4135/9781412985642.

[75] R. F. DeVellis, “Scale development: theory and applications”, vol. 26, 2nd ed. Thousand Oaks, CA, SAGE Publications, 2003.