Quality model to evaluate the usability and portability of Serious Games

Moisés Santiago Arévalo
Faculty of Engineering
Universidad de Cuenca
Cuenca, Ecuador
moises.arevalo@ucuenca.edu.ec

Vanessa Romero
Faculty of Engineering
Universidad de Cuenca
Cuenca, Ecuador
vanessa.romero@ucuenca.edu.ec

Cristina Sanchez-Zhunio
Faculty of Engineering
Universidad de Cuenca
Cuenca, Ecuador
cristina.sanchezz@ucuenca.edu.ec

Daniela Prado-Cabrera
Department of Computer Science
Universidad de Cuenca
Cuenca, Ecuador
daniela.pradoc@ucuenca.edu.ec

Priscila Cedillo
Department of Computer Science
Universidad de Cuenca
Cuenca, Ecuador
priscila.cedillo@ucuenca.edu.ec

***Abstract*—Serious games are designed to educate, train, or change behavior while entertaining players. These are increasingly popular due to the fact that they provide great health benefits, for example, improvement of cognitive functions, many of these games do not apply methods and / or standards to improve the quality of usability and portability. Therefore, this work proposes a method based on the ISO / IEC 25040 standard that contains the activities and artifacts that will be used to perform the quality evaluation of serious games, this method is called "Serious Game Quality Assessment Method" (SG- QUAM), furthermore this work describes a quality model based on the ISO / IEC 25010 standard for the evaluation of serious games. The model is based on two approaches, the quality in use and the quality of the product where usability and portability are evaluated. Finally, 16 software engineers participated in the evaluation of the method through a controlled experiment, the results obtained allowed to establish that SG-QUAM is perceived as easy to use, useful and that it could be used in the future.**

Keywords— Serious games, quality model, evaluation method, usability, portability

I. INTRODUCTION

Cognitive functions are brain actions that involve understanding and functioning of the environment [1]. These deteriorate with the aging process and the degree of deterioration depends on the type of function, the individual characteristics of the subject and the context in which it operates, family history, previous psychiatric conditions, and addictions [2]. Stimulation of the brain has a positive impact on cognitive functions, helping to reduce their deterioration [3]. There are tools for cognitive stimulation, such as exercise books, neurotechnology, games, among others.

Specifically, a game is a type of activity where players try to achieve a goal according to imposed rules [4]. Nowadays Video games have a great presence, it is estimated that more than 2.5 billion people use them [5]. Taking advantage of this situation, serious games have been developed, these are understood as "games that are designed to educate, train, or change behavior while entertaining the players" [4]. These games can be used in different areas such as mental wellness, education, and defense, providing a wide variety of benefits, for example, motivation for learning, greater ability to solve problems, and improvement in cognitive abilities [6] [4]. Especially in cognitive training, these games are highly effective since they involve people in demanding cognitive tasks, as well as in the simultaneous training of multiple cognitive processes [7].

Several studies show that not all games are developed correctly, so they do not reach the objectives for which they were designed [5] [8]. Research in this area commonly assumes that video games tend to appeal to audiences and, therefore, when applied in serious contexts the experience should be more memorable and potentially more effective [9]. As these objectives are not fully achieved, other research suggests the analysis of metrics applicable to serious games, Cevallos and Santorum [5] establish that the following characteristics should be taken into account: usability, motivation, commitment, user experience, and understandability.

Hence, this work proposes a quality model for the evaluation of the usability and portability of serious games developed to improve cognitive health. This model is based on the ISO/IEC 25010 standard, which provides a series of standards whose main objective is “to guide the development of software products by specifying requirements and evaluating quality characteristics” [10]. Additionally, the evaluation of two serious games is presented and is based on the guidelines of the ISO/IEC 25040 standard, which proposes a reference model to evaluate software where the inputs, restrictions, and resources needed to obtain the outputs are considered [11].

The proposed model assesses the quality of the software product and the quality in use. Understanding software quality as the degree to which a software product satisfies explicit and implicit needs when used under specific conditions, and in-use quality as the degree to which specific users can use a product or system to satisfy their needs [12]. Specifically, the model is designed to evaluate usability, portability, and satisfaction. ISO/IEC 25010 defines these three characteristics as follows: i) usability refers to the ability of the product to be understood, learned, used and be attractive to the user, when used under certain conditions; ii) portability is the ability of a product or component to be transferred effectively and efficiently; and iii) satisfaction is the degree to which user needs are met when a system is used in a given context [10].

This work is structured as follows: in section II the discussion of related works is presented, where it has been determined that there is no quality model to assess the usability and portability of serious games, in section III the methodology for evaluating serious games is described, in section IV the quality model is detailed, in section V the viability of the quality model is tested by evaluating two games, finally in section VI they have presented conclusions and future work.

II. Related Works

There are several contributions related to the quality assessment in serious games [8] [6] [13] [14] [15] [4] [16].

On the one hand, the authors in [8] propose three basic principles for usability evaluation in serious games in cognitive rehabilitation: movement control in 3D virtual environments, instructional modality and feedback time. For the evaluation of these characteristics, they carried out a study with 30 patients with brain injury and obtained the results that the interaction controlled by the mouse is more effective than the keyboard and that the patients prefer instructions based on video instead of text. This work does not use any standard to define the characteristics to be evaluated.

On the other hand, in [6] a quality model based on the ISO / IEC 25010 standard is presented. In this study the authors have not considered software evolution metrics that are related to the portability characteristic. Furthermore, within the aforementioned study, the model for the evaluation of serious games has not been applied.

In [13] a quality model for serious games is proposed based on the motivational measurement model, where it groups the characteristics into two categories: related to the player and related to the software. In the first category is balance, usability, attention, relevance, satisfaction and trust, and in the second group is portability, maintenance, compatible services and performance.

Finally, the authors of [14] propose a five-dimensional evaluation framework and apply it to a serious game called Robocode. Within the five characteristics that were investigated, three sub-characteristics were considered, these are the following: usability (learning ability, errors, ease of use), motivation (challenge, enjoyment, curiosity), commitment (purpose, interest, control), experience user (competence, social interaction, immersion), and understandability (clarity, simplicity, independence). They obtained as a result that compressibility obtained the lowest rating (62.2%) while usability obtained the highest rating (71.3%). They did not take any standard as the basis for the selection of these characteristics.

In addition, there are studies that carry out an evaluation of quality in digital games, as in [15], which presents a systematic review to study the playability and experience of older adults in digital games. For this study, he made a review of 34 articles, and they determined the following characteristics: perception, cognition, dexterity, navigation, content, visual design and interaction, where the sub-characteristics were considered: commitment, motivation and satisfaction.

In [16], they evaluate the gameplay as an extension of quality in video game use, this offers different metrics, such as gameplay characteristics, which can be of great interest in a serious game context. Similarly, in this research they make use of the ISO / IEC 25010 and ISO / IEC 9126-4 standards.

In [8] [13] [14] quality models are proposed to evaluate usability, but a model based on the ISO / IEC 25010 standard has not been found to be applied in a serious game focused on cognitive health. Neither models proposed to evaluate portability, in the same way these models have not been applied to evaluate serious games.

Serious games must have the ability to motivate the player, from the beginning to the end of the game. All this while understanding the learning material provided by the game [4]. When it comes to measuring the quality of serious games, the aforementioned standards may not cover metrics for the evaluation of content that improves the motivation of the player. Therefore, the use of four metrics for motivational evaluation is proposed, which are: Attention, relevance, trust and satisfaction (ARCS) [4].

III. Evaluation Method

To carry out the evaluation of serious games, a method based on ISO 25040 has been carried out. In the proposed method, the following roles were taken into account (Table I):

TABLE I. Roles present in the evaluation method.

Role	Table Column Head
Evaluator	People in charge of the entire quality assessment process.
Developer	Serious game developers
Serious Game Expert	People with extensive knowledge of serious games
Assessment Designer	People in charge of developing the evaluation.
Project Manager	Responsible for serious game

In the Fig. 1, each phase of the proposed method is observed:

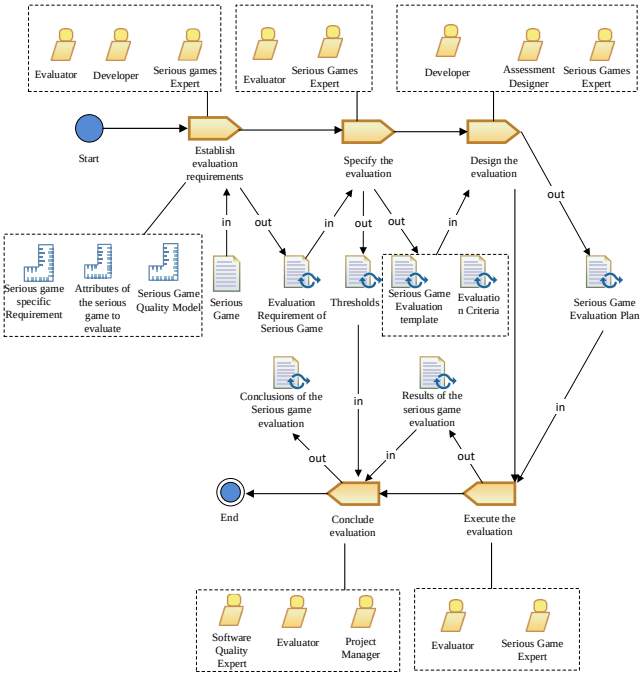

Fig. 1. Quality assessment method SG -QUAM

A. *Establish the assessment requirements*

This task receives as input an artifact: (1) Serious Game to evaluate and three guides: (1) Serious game specific Requirement, (2) Attributes of the serious game to evaluate and (3) Serious Game Quality Model.

The output of this task are the Evaluation Requirement of Serious Game, which were determined by the *Evaluator* in conjunction with the *Developer* and *Serious Game Expert* (see Fig. 2).

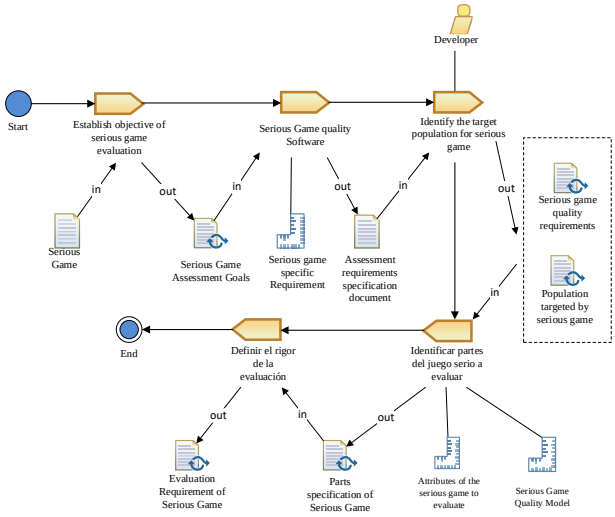

Fig. 2. Assessment requirements establishment phase.

B. *Specify the evaluation*

This process consists of selecting which modules of the serious game will be evaluated, based on the following artifacts: (1) Serious game quality requirements and (2) Evaluation Requirement of Serious Game. The thresholds for each metric are also defined, based on ISO / IEC 25022, which consist of defining quality measures in use for the characteristics defined in ISO / IEC 25010; and in IS0 / IEC 25023, which consists of defining quality measures to quantitatively evaluate the quality of the system and the software product in terms of characteristics and sub-characteristics defined in ISO / IEC 25010.

The output of this task are the Evaluation Requirement of Serious Game, which were determined by the *Evaluator* in conjunction with the *Developer* and *Serious Game Expert* (see Fig. 3).

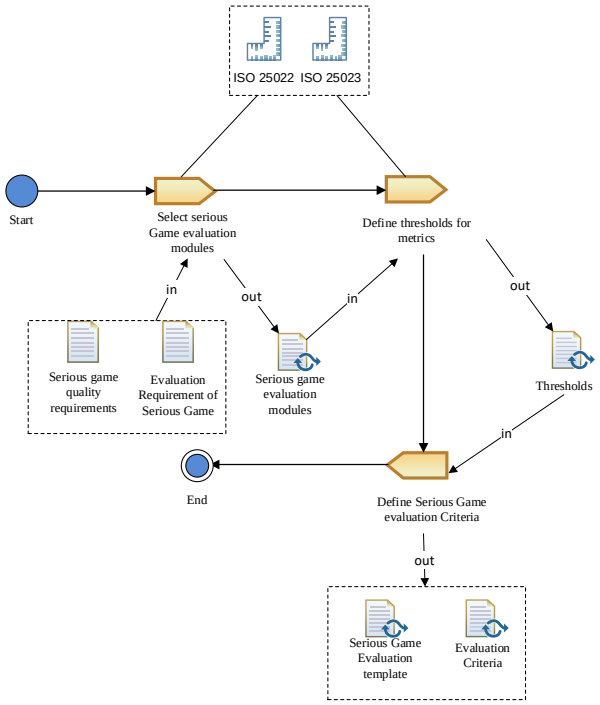

Fig. 3. Evaluation specification phase.

C. *Desing evaluation*

In Figure 4 this task is graphically expressed, it receives as input the Evaluation Criteria and Serious Game Evaluation template of the evaluation phase. In addition, the Serious Game Quality Model guide is used. The result obtained is a Serious Game Evaluation Plan.

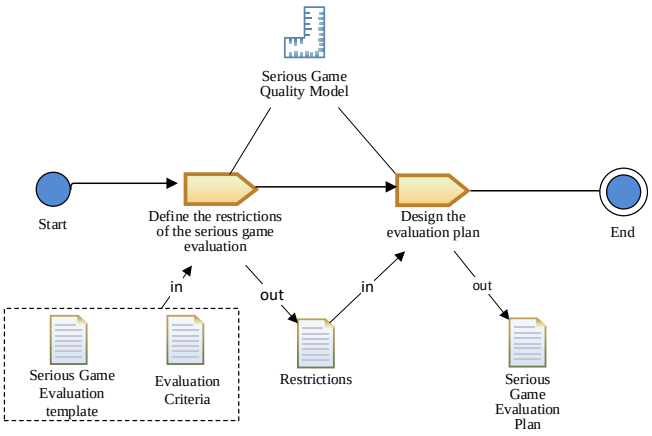

Fig. 4. Evaluation design phase.

D. *Execute the evaluation*

Once the artifacts of the previous process have been obtained, the Evaluator performs the evaluation of each selected component of the serious game and the results are analyzed together with the Serious Game expert. As a result of all this process, we have Results of the serious game evaluation (see Fig. 5).

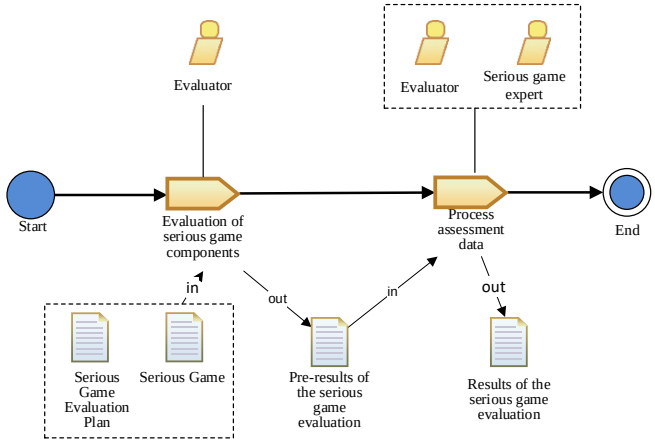

Fig. 5. Evaluation execution phase.

E. Conclude evaluation

The last process of the proposed model, consists of the presentation of the results obtained in the previous task to the Project Manager and the last analysis of the results, in which the thresholds obtained in the Evaluation Specification process are taken into account, therefore A document is obtained about the conclusions of the serious game evaluation process (Fig. 6).

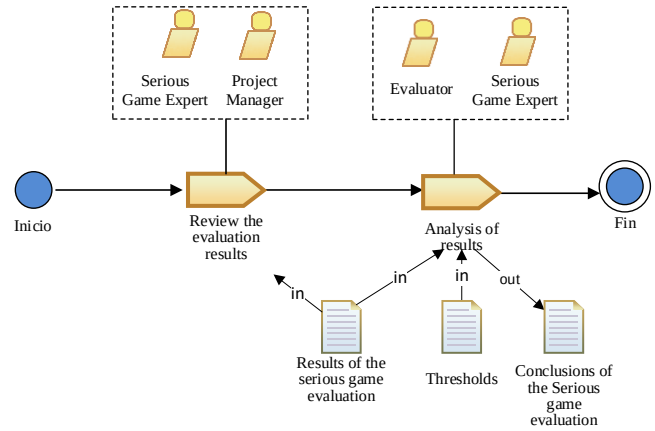

Fig. 6. Evaluation conclusion phase

IV. QUALITY MODEL

A. Product quality model.

This section describes the main sub-characteristics and attributes of certain metrics of the Serious Gaming Product Quality Model. This model has been adapted and extended from QSGAME-MODEL presented in [6]. This model uses the characteristics proposed in the ISO / IEC 25000 standard. These sub-characteristics have been adapted and some new ones have been defined within the usability and portability characteristics. The complete model is available at the following address: https://www.dropbox.com/s/o1vzphelufvvmue/Model%20of%20quality%20product.pdf?dl=0

The Learnability characteristic is analyzed below. Table II contains some of the sub-characteristics of this characteristic.

TABLE II. SUB-CHARACTERISTICS OF LEARNABILITY

Sub-characteristic	Attribute	Meaning
Simplicity	Navigability between functions available	Can you easily navigate between different functions of the serious game?
	Feedback messages	Are feedback messages useful for improving the learning Board Index?
Understandability	clear rules	To what extent the rules of the game state that serious actions are permitted?
	Independence	To what extent you can play independently without the need for assistance or support?

The second sub-characteristic to describe of the product quality model is that of Accessibility, in Table III, again there are some of the sub-characteristics and attributes for serious games.

TABLE III. SUB-CHARACTERISTICS OF LEARNABILITY

Attribute	Meaning
Accessibility for users with cognitive disabilities	To what extent can users with limited cognitive ability to successfully use the system?
alternative text	Are the roles serious games offer options so that any non-text content can be converted into other forms (descriptive Text, etc.).?
Availability of different languages	What languages are supported?

The last sub-feature to be described in this study is that of user interfaces aesthetics. Similarly, in Table IV, some of the sub-characteristics and attributes belonging to it are described.

TABLE IV. SUB-CHARACTERISTICS OF LEARNABILITY

Attribute	Meaning
Interface design	What proportion of interfaces use space, color and text correctly?
realistic appearance	To what extent is the interface design realistic and familiar to users?
aesthetic appearance	To what extent are acceptable user interfaces?

B. Use quality model

In this section some sub-characteristics of usability of the quality of use model are described, these have been compiled and adapted for serious games from the following studies [16] [17] [18] and ISO / IEC 25023. The model is available In the following link: https://www.dropbox.com/s/jv8nuc4mpdxrayu/model%20of%20quality%20in%20use.pdf?dl=0.

TABLE V. SUB-CHARACTERISTICS OF EFFECTIVENESS

Sub-characteristic	Attribute	Meaning
1.1.2 Effectiveness of the objectives	Clarity of the objective	Does the game provide clear targets and players can reach?
	completed goals	What is the average accuracy of completed goals?
1.2.1 Time task	Efficiency homework time	How long does the player complete a task, compared to planned?
1.2.2 Relating to the user	A user efficiency compared with an expert	What is so efficient for a player compared to an expert?

TABLE VI. SUB-CHARACTERISTICS OF EFFECTIVENESS

Sub-characteristic	Attribute	Meaning
1.2.1 Time task	Efficiency homework time	How long does the player complete a task, compared to planned?
1.2.2 Relating to the user	A user efficiency compared with an expert	What is so efficient for a player compared to an expert?

TABLE VII. SATISFACTION SUB-CHARACTERISTICS

Sub-characteristic	Attribute	Meaning
1.3.1 Playability	Player Experience	Does the experience is fun for the player?
	Immersion	Does the game encourage user involvement?
1.3.3 Design	Design user interfaces	Does the design of user interfaces are pleasing in appearance?

Some sub-characteristics of Effectiveness (Table V), Efficiency (Table VI) and Satisfaction (Table VII), are described in the tables mentioned.

V. RESULTS OF EXPERIMENTATION

To test the perceived efficacy of the SG-QUAM method, a quasi-experiment has been carried out, which follows the experimental process of Wohlin [19], the following steps are considered: 1) definition of the scope, 2) planning of the experiment, 3) operation, 4) analysis and interpretation, 5) presentation and packaging, 6) threats to validity.

1. *Product quality model*

The goal of the quasi-experiment has been defined according to the Goal Question Metric (GQM) paradigm proposed by Basili [20] (see Table VIII).

TABLE VIII. DEFINITION OF SCOPE

Analyze	Obtaining quality requirements for a serious game.
With the purpose of	Evaluate the method of evaluating a serious game in the requirements elicitation phase with respect to efficiency, perceived ease of use, perceived usefulness, and intended use in the future.
From the point of view of	The Software Quality Engineer
In the context	A group of software engineers

To test the perceived efficacy of the SG-QUAM method, a quasi-experiment has been carried out, which follows the experimental process of Wohlin [19], the following steps are considered: 1) definition of the scope, 2) planning of the experiment, 3) operation, 4) analysis and interpretation, 5) presentation and packaging.

According to the above, the research questions are:

- **RQ1:** Is the serious games quality assessment method perceived as easy to use and useful? If so, are users' perceptions the result of their performance when using the evaluation method?
- **RQ2:** Is there an intention to use the SG-QUAM method in the future? If so, are the usage intentions the result of the participants' perceptions?

2. *Planning the experiment*

To test the perceived efficacy of the SG-QUAM method, a quasi-experiment has been carried out, which follows the experimental process of Wohlin [19], the following steps are considered: 1) definition of the scope, 2) planning of the experiment, 3) operation, 4) analysis and interpretation, 5) presentation and packaging.

These research questions can be evaluated by testing the following hypotheses.

- $H1_0$: SG-QUAM in the requirements obtaining phase is perceived as difficult to use, $H1_0 = \neg H1_1$
- $H2_0$: SG-QUAM in the requirements obtaining phase is not perceived as useful, $H2_0 = \neg H2_1$
- $H3_0$: There is no intention to use SG-QUAM in the future, $H3_0 = \neg H3_1$
- $H4_0$: Perceived ease of use cannot be determined by current efficiency, $H4_0 = \neg H4_1$
- $H5_0$:The perception of utility is not determined by current effectiveness. $H5_0 = \neg H5_1$
- $H6_0$:Perceived utility is not determined by perceived ease of use $H6_0 = \neg H6_1$.
- $H7_0$: Intent to use is not determined by perceived ease of use $H7_0 = \neg H7_1$
- $H8_0$: Intent to use is not determined by perceived utility$H8_0 = \neg H8_1$.

The first research question will be determined by hypotheses 1, 2, 4 and 5, and the second research question by hypotheses 3, 4, 7 and 8.

Table IX shows the questions that have been defined to measure the variables based on perception. These should be answered on a Likert scale from 1 to 5 with the format of opposite questions, where 1 represents the value of least significance and 5 the value of greatest significance.

TABLE IX. QUESTIONS FOR THE MEASUREMENT OF PERCEPTION

Question	Positive declaration (5 -point scale)
PEOU1	The methodology in the phase of obtaining the requirements of the evaluation has seemed simple and easy to follow.
PEOU2	In general, the methodology in the requirements obtaining phase of the evaluation is easy to understand.
PEOU3	What languages are supported?
PEOU4	The requirements elicitation phase of the assessment method is easy to learn.
PEOU5	I think it would be easy to use the guide to obtain requirements to evaluate a serious game in the proposed methodology.
PU1	I believe that the phase of obtaining evaluation requirements would reduce the time and effort required for the evaluation of serious games
PU2	In general, I consider that the phase of obtaining the requirements of the evaluation method is useful.
PU3	I believe that the method is expressive enough to define how the measurement of the guides provided in the phase of obtaining requirements to evaluate a serious game will be carried out.
PU4	Using this method would improve my performance in obtaining requirements to evaluate a serious game.
PU5	In general, I think that with this method I can get the requirements to evaluate a serious game properly.
PU6	I believe that the guide for obtaining requirements to evaluate a serious game is useful in the phase of obtaining the evaluation requirements of the proposed methodology.
ITU1	If I were to use a method for obtaining serious game evaluation requirements in the future, I think I would consider this method.
ITU2	Should you need to perform requirements to evaluate a serious game, you would intend to use this method in the future.
ITU3	I would recommend using this requirement elicitation method to evaluate serious play.

3. *Operations*

In the quasi-experiment 16 software engineers participated, the material designed for this was placed on a web page for easy access. This can be found at the link below: https://sites.google.com/ucuenca.edu.ec/sg-quam/entrenamiento.

4. *Analysis and interpretation*

In order to accept or reject the hypotheses raised, the following was carried out: i) analysis of user perceptions, ii) analysis of user performance and iii) analysis of causal relationships.

i. Analysis of user perceptions

Figure 7 shows the box plots for the variables: perceived ease of use, perceived usefulness, and intention to use. In the three graphs it can be seen that the mean exceeds 3, which is the neutral value of the Likert scale.

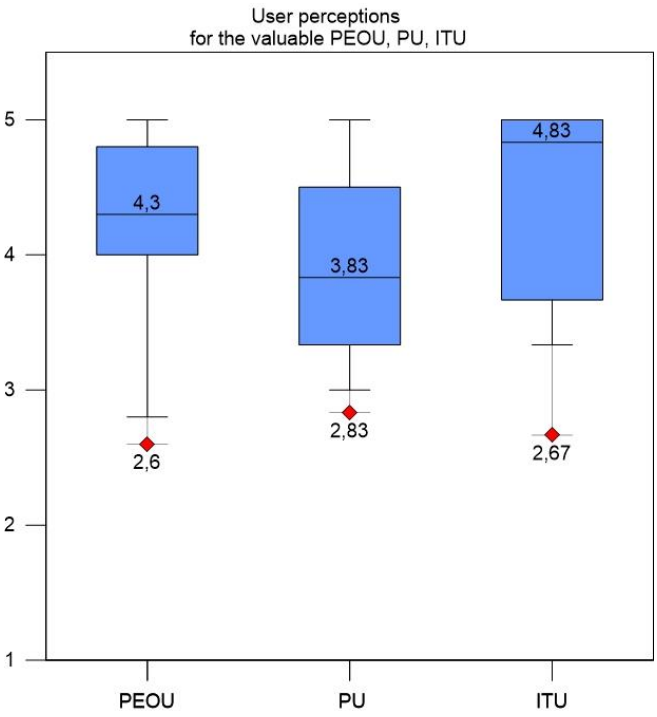

Fig. 7. Box plot for the PEOU, PU, ITU variables.

Table X shows the results of the Shapiro-Wilk test that has been applied to test the hypotheses H1, H2 and H3. If $p < 0.05$ the hypothesis is rejected with a significance level of 5%, this is only fulfilled by the ITU variable with $p = 0.02$. The three hypotheses are rejected because they have a mean value between 3 and 5, specifically the PEOU variable has a mean of 4.20, which indicates that the method was perceived as easy to use; the variable PU has an average value of 3.87, this means that the method is considered useful; and the ITU variable has a mean of 4.31, that is, it is intended to use the method in the future.

TABLE X. SHAPIRO-WILK TEST FOR SUBJECTIVE VARIABLES.

Var	Min	Max	Mean	Std. Dev.	Std. E,	Shapiro-Wilk test p-value
PEOU	2.60	5.00	4.20	0.71	0.18	0.045
PU	2.83	5.00	3.87	0.74	0.18	0.169
ITU	2.67	5.00	4.31	0.81	0.20	0.002

ii. User performance analysis

Table XI presents the descriptive statistic values for the performance-based variables. The mean effectiveness is 0.56, which represents that more than half of the participants were able to correctly execute the tasks. Efficiency has an average of 10.75 minutes, this variable was calculated taking into account the start and end time of each task. The minimum time is 4 minutes and the maximum 30, this great difference can be due to many factors such as experience or the description of the tasks of the experiment.

TABLE XI. DESCRIPTIVE STATISTICS FOR VARIABLES BASED ON USER PERCEPTION.

Variable	Min	Max	Mean	Std. Dev.
Effectiveness	0.23	0.83	0.56	0.17

Efficiency	4.00	30.00	10.75	5.91

iii. Analysis of causal relationships

Regression analysis has been used to evaluate the operationalization of the MEM, since the hypotheses to be tested are causal relationships between continuous variables. In this analysis, the levels of significance specified in table XII have been used.

TABLE XII. LEVELS OF SIGNIFICANCE.

Significance value	Rank
Not significant	p > 0.1
Low significance	p < 0.1
Medium significance	p < 0.05
High significance	p < 0.01
Very high significance	p < 0.001

- *Efficiency vs. perceived ease of use*

To check whether the perceptions of perceived ease of use are determined by efficiency, a simple regression model has been constructed where the independent variable is efficiency and the dependent variable is PEOU. From the analysis the following regression equation was obtained:

PEOU=5.035+(-0.078)*Efficiency

In table XIII it can be seen that the regression model with p <0.01 has a high significance. The R2 indicates that the efficiency variable explains 41.3% of the variance in PEOU, this means that the efficiency of the participants does influence their perceptions of ease of use. These results make it possible to reject the null hypothesis H40 and accept the alternative hypothesis, that is, it has been proven that the perceived ease of use is determined by efficiency.

TABLE XIII. SIMPLE REGRESSION BETWEEN EFFICIENCY AND PERCEIVED EASE OF USE.

Reg. Element	Coef (b)	Std. E.	Std. Coef	t	Sig (p)	R	R2
Constant	5.035	0.302		16.69	0.000		
Efficiency	-.078	0.025	-.643	-3.14	0.007	0.643	0.413

- *Effectiveness vs. perceived utility*

To check whether the perceptions of perceived utility are determined by effectiveness, a simple regression model has been constructed where the independent variable is effectiveness and the dependent variable is PU. From the analysis the following regression equation was obtained:

PU=3.716+0.281*Effectiveness

In table XIV it can be seen that the regression model with p> 0.1 is not significant. The R2 indicates that the effectiveness variable explains 0.4% of the variance in PU, this means that the effectiveness of the participants does not influence their perceptions of perceived utility. These results do not allow rejecting the null hypothesis H50 and accepting the alternative hypothesis, that is, it has been verified that the perceived utility is not determined by effectiveness.

TABLE XIV. SIMPLE REGRESSION BETWEEN EFFECTIVENESS AND PERCEIVED UTILITY.

Reg. Element	Coef (b)	Std. E.	Std. Coef	t	Sig (p)	R	R2
Constant	3.716	0.664		5.600	0.000		
Effectiveness	0.281	1.138	0.066	0.247	0.809	0.066	0.004

- *Perceived ease of use vs. perceived utility*

To test whether the perceptions of perceived utility are determined by perceived ease of use, a simple regression model has been constructed where the independent variable is PEOU and the dependent variable is PU. From the analysis the following regression equation was obtained:

PU=1.374+0.595*PEOU

In table XV it can be seen that the regression model with p <0.05 has a mean significance. The R2 indicates that the variable perceived ease of use explains 33.4% of the variance in PU, this means that the perceived ease of use of the participants influences their perceptions of perceived utility. These results make it possible to reject the null hypothesis H60 and accept the alternative hypothesis, that is, it has been found that PU is determined by PEOU.

TABLE XV. SIMPLE REGRESSION BETWEEN EFFECTIVENESS AND PERCEIVED UTILITY.

Reg. Element	Coef (b)	Std. E.	Std. Coef	t	Sig (p)	R	R2
Constant	1.374	0.957		1.436	0.173		
PEOU	0.595	0.225	0.578	2.648	0.019	0.578	0.334

- *Intent to use vs. perceived utility*

To check whether the perceptions of the intention to use are determined by the perceived utility, a simple regression model has been constructed where the independent variable is PU and the dependent variable is UTI. From the analysis the following regression equation was obtained:

ITU=0.928+0.874*PU

In table XVI it can be seen that the regression model with p <0.001 has very high significance. The R2 indicates that the variable perceived utility explains 63.1% of the variance in perceived utility, this means that there may be other factors that influence the intention of the participants to use the method. These results make it possible to reject the null hypothesis H70 and accept the alternative hypothesis, that is, it has been found that ITU is determined by PU.

TABLE XVI. SIMPLE REGRESSION BETWEEN PERCEIVED UTILITY AND INTENTION TO USE.

Reg. Element	Coef (b)	Std. E.	Std. Coef	t	Sig (p)	R	R2
Constant	0.928	0.704		1.318	0.209		
PU	0.874	0.179	0.794	4.889	0.000	0.794	0.631

- *Intent to use vs. perceived ease of use*

To check if the perceptions of the intention to use are determined by the perceived ease of use, a simple regression model has been constructed where the independent variable is

PEOU and the dependent variable is ITU. From the analysis the following regression equation was obtained:

ITU=1.000+0.789*PEOU

In table XVII it can be seen that the regression model with p <0.01 has high significance. The R2 indicates that the variable perceived ease of use explains 48.4% of the variance in the perceived utility, this means that the intentions of the participants to use the method in the future are determined by their perception of the ease of use of the method . These results make it possible to reject the null hypothesis H80 and accept the alternative hypothesis, that is, it has been verified that ITU is determined by PEOU.

TABLE XVII. REGRESIÓN SIMPLE ENTRE FACILIDAD DE USO PERCIBIDA E INTENCIÓN DE USO

Reg. Element	Coef (b)	Std. E.	Std. Coef	t	Sig (p)	R	R2
Constant	1.000	0.927		1.079	0.299		
PEOU	0.789	0.218	0.696	3.624	0.003	0.696	0.484

5. *Analysis and interpretation*

Table XVIII shows the results obtained for each hypothesis.

TABLE XVIII. EVALUATION SUMMARY

Hypothesis	Sig (p)	Significance	Action
$H1_0$	-	-	Rejected
$H2_0$	-	-	Rejected
$H3_0$	-	-	Rejected
$H4_0$	0.007	High significance	Rejected
$H5_0$	0.809	Not significant	Accepted
$H6_0$	0.019	Mean significance	Rejected
$H7_0$	0.000	Very high significance	Rejected
$H8_0$	0.003	High significance	Rejected

The results obtained have allowed answering the questions defined in the scope. For RQ1 it was determined that the SG-QUAM method is perceived as useful and easy to use, and for RQ2 it was determined that participants intend to use the method in the future.

The results of the regression analysis can be seen in Figure 8.

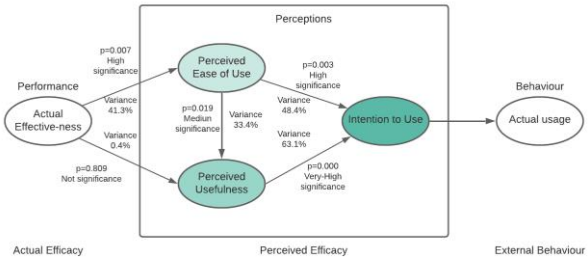

Fig. 8. Conclusions of the MEM application to SG-QUAM.

6. *Threats to validity*

The four types of threat described by Cook and Campbell in [21] are considered in order to explain the main problems that jeopardize the validity of the quasi-experiment.

i. Internal validity

One of the main threats to internal validity was the experience of the participants. To reduce this, before the experiment, a training consisting of similar activities was presented for the participants to have a better understanding of the method. In addition, the experiment was carried out with software engineers who have knowledge of software quality facilitating the understanding of the experiment.

ii. External validity

A threat to external validity is access to the serious game with which the evaluation was conducted. To avoid this problem, a game available on a web platform was selected, which was easy to access from any device, avoiding download problems. Another threat to external validity is the size and complexity of the tasks, so a level of complexity was proposed according to the time available, a pilot experiment was carried out to measure the approximate time it would take to solve the experiment.

iii. Construct validity

The main threat is the reliability of the questionnaire, so the Cronbach alpha reliability test was performed for each set of questions related to each subjective variable (PEOU, PU, ITU). It was obtained that the perceived ease of use has a value of $\alpha = 0.686$, the perceived utility has a value of $\alpha = 0.678$, and the intention of use has a value of $\alpha = 0.687$. These values are very close to the minimum accepted threshold that is $\alpha = 0.7$, here it is established that the questionnaire is reliable.

iv. Conclusion validity

Conclusion validity refers to threats that affect statistical conclusions. One of the main problems is the size of the sample, in this case, a homogeneous group of 16 participants was selected who managed to carry out the proposed tasks. For data collection, the same procedure was applied for all participants and it was ensured that the dependent variables are calculated using the same measurement function.

VI. CONCLUSIONS

This work has presented a model to evaluate the quality of serious games, specifically usability and portability; and a quality model in use. A method has also been developed to evaluate the quality of serious games called SG-QUAM, where the steps to be followed are detailed.

Also, a quasi-experiment has been presented with which the perceived efficiency of a group of students using SG-QUAM in the requirements obtaining phase has been evaluated. Analysis of the data reveals that the method was perceived by the participants as easy to use, useful and that they intend to use the method in the future. Regarding the open questions, most of the participants indicated that the method seemed adequate for the evaluation of cognitive reinforcement applications.

Finally, as future work, it is sought to apply the quality model to evaluate the quality of serious games and it is intended to validate the method with more participation of software engineers to later present a broad discussion among the authors.

ACKNOWLEDGMENT

This work is part of the software quality course at the University of Cuenca, for which we thank Dr. Priscila Cedillo, Ing. Cristina Sánchez, and Psic. Daniela Prado for the support and guidance provided.

REFERENCES

[1] L. A. Hirschfeld, "Is the acquisition of social categories based on domain-specific competence or on knowledge transfer?," in *Mapping the Mind*, Cambridge University Press, 1994, pp. 201–233.

[2] G. león, "Factores de riesgo para deterioro cognitivo y funcional en el adulto mayor," *Rev Med Inst Mex Seguro Soc*, vol. 47, no. 3, pp. 277–284, 2007, [Online]. Available: http://dspace.ceu.es/bitstream/10637/7040/1/Factores de riesgo asociados al deterioro cognitivo_prevención desde la oficina de farmacia_Tesis_María Teresa Climent Catalá.pdf.

[3] E. Stanmore, B. Stubbs, D. Vancampfort, E. Bruin, and J. Firth, "The effect of active video games on cognitive functioning in clinical and non-clinical populations: A meta-analysis of randomized controlled trials," *Neuroscience and Biobehavioral Reviews*, vol. 78. Elsevier Ltd, pp. 34–43, Jul. 01, 2017, doi: 10.1016/j.neubiorev.2017.04.011.

[4] L. K. P. Suryapranata, F. L. Gaol, B. Soewito, H. L. H. S. Warnars, and G. P. Kusuma, "Quality measurement for serious games," *Proc. - 2017 Int. Conf. Appl. Comput. Commun. Technol. ComCom 2017*, vol. 2017-Janua, pp. 1–4, 2017, doi: 10.1109/COMCOM.2017.8167098.

[5] A. Cevallos and M. Santorum, "A quality model for the serious games development process," in *Proceedings - 2019 International Conference on Information Systems and Software Technologies, ICI2ST 2019*, Nov. 2019, pp. 1–8, doi: 10.1109/ICI2ST.2019.00008.

[6] L. Garcia-Mundo, M. Genero, and M. Piattini, "Towards a Construction and Validation of a Serious Game Product Quality Model," in *2015 7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games)*, Sep. 2015, pp. 1–8, doi: 10.1109/VS-GAMES.2015.7295790.

[7] J. Anguera *et al.*, "Video game training enhances cognitive control in older adults," *Nature*, vol. 501, no. 7465, pp. 97–101, 2013, doi: 10.1038/nature12486.

[8] M. van der Kuil, J. Visser-Meily, A. Evers, and I. van der Ham, "A usability study of a serious game in cognitive rehabilitation: A compensatory navigation training in acquired brain injury patients," *Front. Psychol.*, vol. 9, no. JUN, Jun. 2018, doi: 10.3389/fpsyg.2018.00846.

[9] G. Hookham and K. Nesbitt, "A Systematic Review of the Definition and Measurement of Engagement in Serious Games," in *ACM International Conference Proceeding Series*, Jan. 2019, pp. 1–10, doi: 10.1145/3290688.3290747.

[10] "ISO 25010." https://iso25000.com/index.php/normas-iso-25000/iso-25010?limit=3&limitstart=0 (accessed Nov. 27, 2020).

[11] "ISO 25040." https://iso25000.com/index.php/normas-iso-25000/iso-25040 (accessed Dec. 04, 2020).

[12] "ISO/IEC 25010:2011(en), Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models." https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en (accessed Nov. 27, 2020).

[13] C. Wimmer, R. Schlögl, K. Kappel, and T. Grechenig, "Measuring mobile text entry performance and behaviour in the wild with a serious game," Nov. 2019, doi: 10.1145/3365610.3365633.

[14] A. Abdellatif, B. McCollum, and P. McMullan, "Serious games: Quality characteristics evaluation framework and case study," in *ISEC 2018 - Proceedings of the 8th IEEE Integrated STEM Education Conference*, Apr. 2018, vol. 2018-January, pp. 112–119, doi: 10.1109/ISECon.2018.8340460.

[15] A. Rienzo and C. Cubillos, "Playability and player experience in digital games for elderly: A systematic literature review," *Sensors (Switzerland)*, vol. 20, no. 14. MDPI AG, pp. 1–23, Jul. 02, 2020, doi: 10.3390/s20143958.

[16] J. L. González Sánchez, F. Montero Simarro, N. Padilla Zea, and F. L. Gutiérrez Vela, "Playability as extension of quality in use in video games," *CEUR Workshop Proc.*, vol. 490, 2009.

[17] C. Jennett et al., "Measuring and defining the experience of immersion in games," Int. J. Hum. Comput. Stud., vol. 66, no. 9, pp. 641–661, Sep. 2008, doi: 10.1016/j.ijhcs.2008.04.004.

[18] R. Ramadan and B. Hendradjaya, "Development of game testing method for measuring game quality," Mar. 2014, doi: 10.1109/ICODSE.2014.7062694.

[19] T. Gorschek, P. Garre, S. Larsson, and C. Wohlin, "A model for technology transfer in practice," IEEE Softw., vol. 23, no. 6, pp. 88–95, Nov. 2006, doi: 10.1109/MS.2006.147.

[20] F. Shull et al., "Knowledge-Sharing Issues in Experimental Software Engineering," in Empirical Software Engineering, 2004, vol. 9, no. 1–2, pp. 111–137, doi: 10.1023/b:emse.0000013516.80487.33.

[21] T. D. Cook and D. T. Campbell, "Quasi-Experimentation: Design and Analysis Issues for Field Settings." Houghton Mifflin, 1979, Accessed: Feb. 22, 2021. [Online]. Available: https://www.scholars.northwestern.edu/en/publications/quasi-experimentation-design-and-analysis-issues-for-field-settin.