International Journal of Serious Games

ISSN: 2384-8766
https://journal.seriousgamessociety.org/

Article

Designing a framework and validating a tool for evaluating the educational quality of serious games: a meta-synthesis

Rasoul Bakhtiari[1], Zahra Habibzadeh[1]

[1]*Bu-Ali Sina University, Department of Education, Hamedan, Iran, rbakhtiari1996@gmail.com*

Keywords:

Serious Games
Educational Quality Evaluation
Framework Design
Tool Validation
Confirmatory Factor Analysis
Meta-synthesis

Received: November 2022
Accepted: March 2023
Published: June 2023
DOI: 10.17083/ijsg.v10i2.576

Abstract

In recent decades, at the same time as the quantitative growth in the industry of serious games, its quality has also been the focus of investors, researchers, and developers. Therefore, the purpose of this research is to design a framework and validate a tool for evaluating the educational quality of serious games (i.e., a questionnaire). Evaluation frameworks and questionnaires are fundamental tools for designing and developing serious games. The method of this research was meta-synthesis with a mixed approach. Based on this, 5807 articles were identified during the years 1995 to 2021, and finally, 29 articles were selected for analysis. By analyzing these articles, basic dimensions, components, and indicators were extracted and turned into an evaluation tool using thematic analysis with the purpose of validation. First, the content validity of this tool was obtained by purposive sampling technique with the participation of 30 serious game experts. Then, for the validity of the constructs, a sample equal to the community was assumed, and the tool was sent to the members of the National Computer Games Foundation, and 537 people participated in completing the tool. Finally, the data were analyzed using confirmatory factor analysis. The findings of first and second-order factor analysis confirmed all dimensions, components, and indicators of the tool with a factor load (above 0.40) and a significant coefficient (above 1.96). Therefore, this meta-synthesis led to the design of the framework and the validation of the educational quality evaluation tool of serious games with 4 dimensions, 25 components, and 138 indicators. We argue that the proposed framework and tool are able to evaluate the educational quality of a serious game and cover its design project end to end. We thus propose them as a suitable resource for researchers and developers of serious games.

1. Introduction

In recent years, the number of serious games and their users is increasing day by day [1], which shows the high social influence of these games. For this reason, criticisms of the educational quality of serious games have increased [2], [3], [4], [5]; Because most serious games are evaluated based on graphical and technical quality, not in terms of content and design based on their intention and market [6], [7], [8]. They are not just a simple type of software, as their name suggests, they are considered a special and serious product that can have a deep impact on many users in different fields. Hence, the evaluation of the educational quality of serious games is different from other software and games [9]. The main difference between these games is their non-technical (serious) aspect, which separates them from other software and games. Therefore, to evaluate the educational quality of serious games, other dimensions, components, and indicators are needed.

Parallel to quantitative development and increasing demand, serious games have been accompanied by product and research in various fields such as military [10], health [11], education [12], and business [13], where addressing educational quality and components has become the main concern of investors, developers, and researchers. A concern that has been investigated in several studies [e.g., 14-15-16], but each of them alone has not been able to cover all factors and needs a comprehensive summary and classification.

Therefore, the current research is trying to design a framework and a tool for evaluating the educational quality of serious games. In this research, with the meta-synthesis of the literature on the evaluation of the educational quality of serious games, a framework and a tool for the evaluation of serious games are introduced and presented. The purpose of designing the framework and making the tool is to help the knowledge of serious games and the design and evaluation of principles and to promote the development and use of these games.

1.1 Serious games

Serious games are a sub-branch of the family of digital, computer, and video games currently of vital importance [17], [18], because they can be meant to achieve personal and organizational goals. Initially, they were provided for training people to perform tasks in specific jobs, such as training army personnel. Then, they evolved alongside the rapid expansion of other types of games and new devices such as smartphones, tablets, and various consoles. For serious games, different definitions have been presented by different authors [4]-[13]-[19], [20], [21], [22]. In summarizing the definitions, it can be argued that the term "serious game" was introduced by Abt [19], and it is considered a kind of digital game with different purposes, which primarily focuses on learning and improving the skills and performance of learners rather than pure entertainment, and ultimately provide many opportunities to connect training to learners' daily life experiences in a variety of markets.

Serious games "can be used as learning media because of several reasons, such as the ability to support constructive, experiential, situated, and procedural learning; engage and motivate players, and promote self-regulated learning" [23]. As learning media, serious games "need to have ability to motivate its player in order to play the game until the end so the player can finish the game while understanding learning materials given in the game" [23]. They are promoting and opening new horizons for active learning and providing a learning-by-doing experience [24]. Serious games also allow for multiple forms of evaluation, which can be part of a well-structured research protocol [25].

Since the development of serious games is difficult and time-consuming, for the design of these games, a comprehensive framework that includes all aspects of game development and a tool that enables evaluation is needed to make the development process more targeted. Of course, in many studies, serious game design and analytical guidelines are mentioned in the best form [e.g., 26-27-28-29-30-31-32], but none of these studies have provided a framework

or tool that can identify or evaluate dimensions, components, and indicators. Several studies used this method of dimensions, components, and indicators to introduce and identify concepts [e.g., 33-34-35]. The dimensions are the specifiable aspects of a concept [36] which are obtained from the existing theories (foundations) about that subject. Components are the elements or constituents of the dimensions without which we cannot analyze the dimensions. Indicators are more concrete aspects of dimensions [36]. They are more specific and are often what we observe in the world around us. Indicators are a tool for measuring and converting the qualitative concept to a quantitative variable.

1.2 Educational quality evaluation

The term quality evaluation can be used for two purposes; to judge the quality and to improve the quality. The first purpose is retrospective and formal and has a summative function. In contrast, the second purpose is forward-looking and relatively informal and has a formative function [37]. Cronbach [38] described Educational Evaluation as "Education evaluation is the process of information gathering and treatment necessary to make a decision for an education programme". For the serious game, educational quality can be considered as a criterion that, if used in certain conditions, must meet the specified and implied technical and non-technical needs. Hence, the evaluation of the educational quality of serious games is the process of gathering information that is done to judge the quality of serious games and improve (treat) their quality.

The evaluation of serious games is different from the evaluation of leisure games (digital, computer, and video) [39]. So far, great efforts have been made to evaluate the educational quality of serious games, and resources have been spent on research and evaluation of these games. As a result, the number and quality of serious game evaluations are growing, but there are still considerable weaknesses, two of which are the lack of comprehensive and multipurpose frameworks for comparative and longitudinal evaluation and the paucity of valid tools for serious games [2]-[5]-[40], [41], [42]. Therefore, the current research aims to design a framework and a tool that includes all dimensions, components, and indicators for evaluating the educational quality of serious games. The main question "When examining the educational quality of games, what dimensions, components, and indicators should be examined?" and the following sub-questions are formulated to achieve the main goals of this research.

1. What are the dimensions, components, and indicators of the educational quality of serious games?
2. What is the validity of the dimensions, components, and indicators of the educational quality of serious games?

2. Methods

The current research was carried out with a practical purpose, Meta-synthesis method, and sequential mixed approach (of qualitative and quantitative type). Meta-synthesis is a process that identifies and evaluates relevant qualitative studies. Then, it combines the findings of these studies to create a new interpretation [43]. In this research, the seven-step meta-synthesis process of Sandelowski and Barroso [44] was used to design the framework and validate the tool for evaluating the educational quality of serious games, because the process used by Sandelowski and Barroso is one of the most prominent methods for meta-synthesis and provides better results than other models (see Fig. 1).

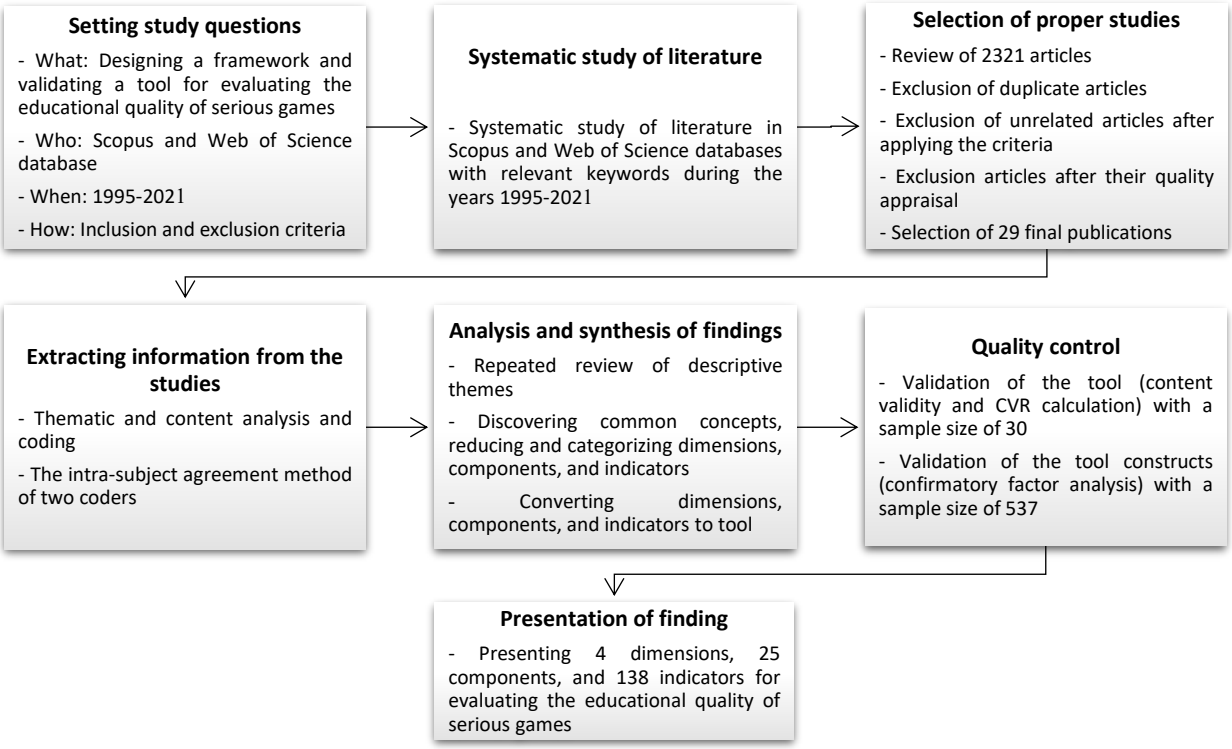

Figure 1. The seven-step meta-synthesis method [44] and the research implementation process

2.1 First step: setting study questions

The first step of setting the research question is to focus on "What". The main goal of this research is to extract the dimensions, components, and evaluation indicators of the educational quality of serious games, which is set by answering the following questions:

1. "Who", which defines the studied community. In this research, two databases, Scopus and Web of Science, were examined Because these two databases are reliable and well-known international reference databases with peer review. Their main focus is on mature research, not conference papers, etc. These two databases also cover a wide range of different fields and have published the majority of articles evaluating the educational quality of serious games.
2. "When", which represents the time period. In this research, the time period was limited to publications from 1995 to 2021, the period in which most of the research evaluating the educational quality of serious games has been produced.
3. "How" indicates the type of method or criteria based on which sources are selected or excluded from the meta-synthesis process. This research's inclusion and exclusion criteria were applied in two steps (Tables 1 and 2).

Therefore, the main question of the research was raised as follows: When examining the educational quality of games, what dimensions, components, and indicators should be examined?

2.2 Second step: Systematic study of literature

The statistical population of this research was all published articles on the evaluation of the educational quality of serious games. Keywords were searched in Scopus and Web of Science databases on 13/10/2021.

{(Framework* OR Model* OR Pattern* OR Approach*) AND (Evaluat* OR Assess* OR Measur* OR Analy* OR Design* OR Develop*) AND (“Education* qualit*” OR “Instruction* qualit*” OR “Pedagog* qualit*” OR “Learn* qualit*” OR Qualit* OR Criteria OR Standard*

OR Feature* OR Characteristic* OR Attribute* OR Mechanic* OR Dimension* OR component* OR indicator*) AND ("Serious gam*" OR "Education* gam*" OR "Instruction* gam*" OR "Pedagog* game*" OR "Learn* game*")}

The search result was a significant list of various documents, including 5807 documents related to relevant keywords.

2.3 Third step: Selection of proper studies

Considering that many documents obtained from Scopus and Web of Science databases only contained search keywords and did not match the research criteria, it was necessary to identify and select relevant documents from among the documents that fully cover the research criteria. Hence, the inclusion and exclusion criteria of the research were applied in two steps. The first step was at the very beginning of the search after the initial review of keywords when the criteria were applied in both databases (Table 1). The second step was after removing the duplicate articles, which were used both for the title and abstract, and for the full text (Table 2). In this section, EndNote X9 software was used to organize the review.

Table 1. The inclusion and exclusion criteria of the first step

Criteria	Include	Exclude
Language	English	Non-English
Date	From 1995 upwards	Before 1995
Type	Peer reviewed journal articles	Books, Book chapters, Proceedings, Editorials, PhD dissertations, MSc thesis, Commentaries, Conference, Others

Table 2. The inclusion and exclusion criteria of the second step

Criteria	Include	Exclude
Full-text	Yes	No
Relevance	Covers topic variables	Not relevant
Game type	(Digital, Video, Computer) Serious Games	Non-(Digital, Video, Computer) Serious games (Game, Game-based learning, Educational game, Instructional games, Learning games, Digital learning, Gamification, Tabletop games, Board game, Digital game, Video Game, Computer game)
Community and sample	Normal and healthy individuals	Abnormal and unhealthy individuals (Rehabilitation, Intellectual disability, Cognitive disability, learning disability, Autism spectrum disorder, Social anxiety disorder, Dementia, Children with speech and language delay, Neurorehabilitation, Visually impaired, Parkinson's Disease)
Study design	Conceptual / Theoretical / Non-empirical / Non-experimental	Empirical
Analysis	Qualitative/Mixed	Quantitative
Technology (platform)	Mobile / PC / Console / Virtual Reality (VR)/Online / Offline	-
Educational level	All levels	-

The second coder was used to determine the reliability of the articles. This coder was with the researcher in all steps of the review and selection of appropriate sources. Cohen's kappa coefficient was used to calculate the agreement coefficient between the two coders. The result of Cohen's kappa agreement coefficient between two coders in the first step (reviewing the title and abstract of the articles) was 0.94 and in the second step (reviewing the full text of the articles) it was 0.87.

For the validity review of all the primary articles (31 articles) prepared by the systematic literature review, the tool used in the study of Theelen et al. [45] was used. This questionnaire has separate criteria for both qualitative and quantitative studies. By means of this

questionnaire, 2 articles were excluded from the research, and finally, 29 articles remained in the final analysis (Table 3).

Table 3. Validity of the articles under review

	0 no elaboration	**1 some elaboration**	**2 good elaboration**	**3 extensive elaboration**
Criteria for qualitative studies				
Study methodologically is clear		6	15	10
Study theoretically substantiated		5	14	12
Ethical process transparent		8	19	4
Researcher(s) relation to participants are clear	7	10	11	3
Researchers(s) relation to the data are clear		3	20	8
Researcher(s) take a critical stance towards own research		8	21	2
Congruence between methodology and methods used for data collection, analysis, and interpretation		2	21	8
Participant involvement in data interpretation	7	17	7	
Limitations voiced	11	7	4	9
Criteria for quantitative studies				
Is the source population or source area well described?		1	5	5
Were interventions and comparisons well described and appropriate?		1	4	6
Were outcome measures reliable?		1	9	1
Were outcomes relevant?			9	2
Were the analytical methods appropriate?		2	6	3
Are the study results internally valid (i.e. unbiased)?		2	7	2
Are the findings generalizable to the source population (i.e. externally valid)?		2	7	2

Figure 2 shows the process of reviewing and selecting sources in full and in summary.

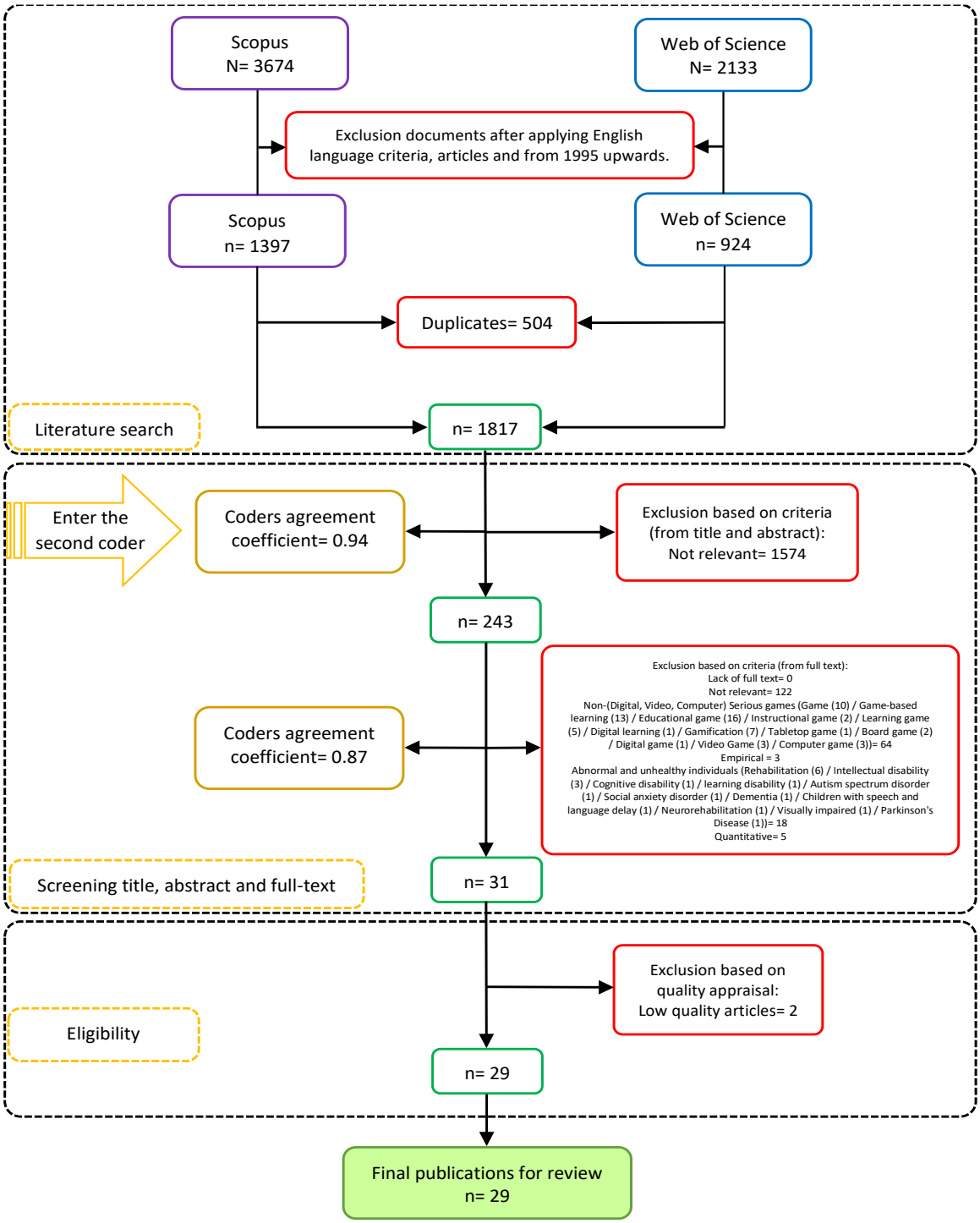

Figure 2. Algorithm of the third step of meta-synthesis

2.4 Fourth step: Extracting information from the studies

Before synthesis, first, the final articles (29 articles) were read and re-read to gain an initial understanding. According to the first step of thematic and content analysis, all results and findings were inductively coded line by line according to meaning and content. In the second step, these codes were grouped to create descriptive themes in a pre-designed table, including dimensions, components, and indicators. To help the transferability of the findings, the intra-subject agreement method of two coders was used [46], and the second coder was requested to code the final articles and carefully control and verify the process of creating themes. The intra-subject agreement percentage was 96.98 (Table 4). Considering that this reliability rate is more

than 0.70, it can be stated that the reliability percentage between the two coders in this study is confirmed.

Table 4. Validity of the articles under review

Row	Studies	The total number of extracted concepts	Number of concepts agreed upon	Number of concepts of disagreement	Coefficient of agreement between two coders
1	[14]	52	49	3	94.23
2	[47]	9	9	0	100
3	[15]	71	67	4	94.36
4	[48]	11	11	0	100
5	[27]	8	8	0	100
6	[49]	18	16	2	88.88
7	[50]	18	18	0	100
8	[51]	19	19	0	100
9	[16]	37	34	3	91.89
10	[52]	26	26	0	100
11	[28]	42	40	2	95.23
12	[53]	17	17	0	100
13	[54]	21	19	2	90.47
14	[55]	105	102	3	94.14
15	[56]	10	10	0	100
16	[57]	14	14	0	100
17	[1]	81	76	5	93.82
18	[58]	12	12	0	100
19	[29]	8	8	0	100
20	[59]	14	14	0	100
21	[60]	125	118	7	94.40
22	[30]	5	5	0	100
23	[61]	485	472	13	97.32
24	[31]	89	89	0	100
25	[3]	96	94	2	97.91
26	[62]	24	22	2	91.66
27	[63]	18	18	0	100
28	[32]	3	3	0	100
29	[64]	7	7	0	100
Total		1445	1397	48	96.98

2.5 Fifth step: Analysis and synthesis of finding

Qualitative meta-synthesis, like any qualitative analysis, requires the final synthesis of findings [44]. At this step, the researcher carefully and considers the concept of each of the dimensions, components, and indicators extracted in the fourth step has tried to discover the common concepts between them. Descriptive themes were repeatedly reviewed and a comprehensive framework and classification of the dimensions, components, and indicators of the educational quality of serious games were identified. In this step, analytical findings were produced, which provided key findings of studies and new perspectives of this field, and in this way, 4 main dimensions, 25 components, and 139 indicators were determined and turned into a questionnaire. NVivo 12 software was used for better and more regular coding.

2.6 Sixth step: Quality control

2.6.1 Validation of the tool (content validity)

After extracting the dimensions, components, and indicators for evaluating the educational quality of serious games and reaching the comprehensive framework, a tool (questionnaire) was prepared. In order to determine the content validity, the prepared tool was sent to 30 experts using a purposeful sampling method (judgment-type). 12 of these experts were female and 18 were male. Their education level was Ph.D. Their academic rank was a professor (2 people), associate professor (11 people), assistant professor (15 people), doctoral student (2 people),

and studied in the fields of software engineering (9 people), artificial intelligence (5 people), educational technology (5 people), computer architecture (5 people), information technology (4 people), computer engineering (1 person), curriculum studies (1 person).

The minimum acceptable value of the content validity ratio (CVR) for 30 experts is 0.33. The CVR calculation of the content validity of the prepared tool items showed that only the indicator (game structure) has a low validity (-0.33), which should be removed for the next step, which is the confirmatory factor analysis, and the rest of the indicators had a good validity.

2.6.2 Validation of tool constructs (confirmatory factor analysis)

The confirmatory factor analysis method was used to validate the constructs obtained in the previous section. In this research, 30 hypotheses (one hypothesis for dimensions, four hypotheses for components, and twenty-five hypotheses for indicators) were proposed to perform confirmatory factor analysis, and the relationship between dimensions (as independent variables) with their sub-set components (as dependent variables) and with the indicators of the sub-set of components (as observed variables) was examined. The necessary information for this validation was obtained by means of a questionnaire from 537 experts in the field of serious games (experts, researchers, university professors, and developers of serious games) who work in the field of research and development of serious games. The questionnaires were distributed in such a way that first, based on the database that Iran's National Computer Games Foundation had (over 3000 people), an SMS with a link to the questionnaire was sent to experts across the country. Then, in addition to the text message, the National Computer Games Foundation shared the questionnaire on its Telegram channel (titled Game Research (link t.me/gameresearch) and with more than 1400 followers). In addition, the researcher also shared the questionnaire with relevant professors and experts, and channels related to the subject. It is worth mentioning that in the text and introductions of the questionnaire, those who were actively focused on the field of serious games for at least two years were requested to complete the questionnaire.

It has been stated that in confirmatory factor analysis, the minimum sample size is determined based on factors, not variables. Therefore, if confirmatory factor analysis is used, about 20 samples are needed for each factor (latent variable) [65]. Another researcher recommends the recommended sample size of confirmatory factor analysis for 10 factors (hidden variable) equal to 200 people [66]. This research has taken the opinion of experts in relation to the three sections of dimensions, components, and indicators, and therefore the number of factors in each section was different. The dimensions section included 1 factor, the components section included 4 factors, and the indicators section included 25 factors. Therefore, a minimum of 250 and a maximum of 500 people were needed for the research indicators section (25 factors), and this research was able to provide 537 people, which is the maximum sample required, by placing the whole community as a basis. The overall reliability of the tool was calculated using Cronbach's alpha method equal to 98.0.

2.6.3 Demographic information of the examined sample

By examining the information obtained from the questionnaire, the frequency of the subjects' characteristics was obtained according to Table 5.

Table 5. Frequency distribution of subjects' characteristics

Variable	Spectrum	Frequency	Frequency percentage
Gender	Female	248	46.18
	Man	289	53.82
Level of education	Ph.D	196	36.50
	Master's degree	212	39.48
	Bachelor's degree	129	24.02
	Computer Engineering	147	27.37

Field of Study	Information Technology	51	9.50
	Cognitive Sciences	38	7.07
	Educational technology	36	6.70
	Multimedia systems	29	5.40
	Media management	25	4.65
	Counseling	21	3.91
	Educational Psychology	20	3.72
	IT management	19	3.54
	Education Management	16	2.98
	Communication Sciences	16	2.98
	Media and information management	14	2.61
	Cognitive rehabilitation	11	2.05
	Educational evaluation	11	2.05
	Production of computer games	10	1.86
	Architecture	9	1.67
	Entrepreneurial management	9	1.67
	Technology	7	1.30
	Psychology	7	1.30
	Information technology engineering	6	1.12
	Cognitive psychology	6	1.12
	Higher Education	6	1.12
	Educational technology in medical sciences	5	0.93
	Entrepreneurship	5	0.93
	Health information technology	4	0.74
	Curriculum	3	0.56
	English language	2	0.37
	Art	2	0.37
Academic rank	Full professor	5	0.93
	Associate Professor	37	6.89
	Assistant Professor	78	14.52
	Doctoral, Master's and Bachelor's degree student or graduate	102	18.99
Activity history	Developer, expert and researcher of serious games	315	58.66
	10 to 15 years	81	15.08
	6 to 9 years	192	35.75
	2 to 5 years	264	49.16

For the reliability of the measurement tool, the questionnaire was first given to a group of 30 people in a simple random manner, then it was collected and the overall Cronbach's alpha coefficient was 0.98. Then, the questionnaire was made available to the participants for confirmatory factor analysis by Iran's National Computer Games Foundation and the researcher. SPSS 26 and LISREL 8.80 software were used for analysis.

Table 6. Factor loading values, significance coefficient, average variance extracted, Cronbach's alpha and compositional reliability of dimensions, components, and indicators for evaluating the educational quality of serious games

factors	Hypothesis and items		Standard coefficient (factor load)	Significance coefficient	Average variance extracted (AVE)	Cronbach's alpha	Compositional reliability
Dimensions	Hypothesis one: To evaluate the educational quality of serious games, four main variables (dimensions) {preparation (A), pre-production (B), production (C), and post-production (D)} are used.						
	A	Preparation	0.59	0.00	0.39	0.71	0.72
	B	Pre-production	0.57	9.14			
	C	Production	0.69	9.95			
	D	Post-production	0.64	9.75			
components	Hypothesis two: To measure the preparation (A) dimension, five variables (components) Aa1, Aa2, Aa3, Aa4, and Aa5 are used for the technical (game) sub-dimension and five variables (components) Ab1, Ab2, Ab3, Ab4, and Ab5 are used for its non-technical (serious) sub-dimension.						
	Aa1	Game idea / Context	0.47	0.00	0.30	0.81	0.81
	Aa2	Game aims	0.52	8.40			
	Aa3	Ludic game script	0.47	7.90			

	Aa4	Planning and risks	0.53	8.55			
	Aa5	Resources, costs and budget estimate	0.43	7.44			
	Ab1	Problem	0.59	8.98			
	Ab2	Serious aims	0.61	9.14			
	Ab3	Target learner profile	0.66	9.49			
	Ab4	Fundamentals, domains and models or theories	0.58	8.95			
	Ab5	Content and serious resources	0.59	9.03			
	Hypothesis three: Five variables (components) B1, B2, B3, B4, and B5 are used to measure the pre-production (B) dimension.						
	B1	Choose name / engine / genre / technology or platform	0.59	0.00	0.43	0.79	0.79
	B2	Goals	0.63	11.18			
	B3	Concept development	0.69	11.92			
	B4	Content and activities	0.65	11.46			
	B5	Designing	0.70	12.00			
	Hypothesis four: ten variables (components) C1, C2, C3, C4, C5, C6, C7, C8, and C9 are used to measure the production (C) dimension.						
	C1	Objectives	0.52	0.00	0.34	0.82	0.82
	C2	Aesthetics and graphics	0.56	9.52			
	C3	Implementation of mechanics	0.61	10.00			
	C4	Experience and feeling	0.64	10.31			
	C5	Adaptivity	0.62	10.08			
	C6	Actions	0.61	10.05			
	C7	Tools	0.52	9.09			
	C8	Usability and function	0.58	9.76			
	C9	Prototype	0.61	10.01			
	Hypothesis five: A variable (component) D1 is used to measure the post-production (D) dimension.						
	D1	Implementation and evaluation	1.00	31.47	1.00	1.00	1.00
indicators	Hypothesis six: four variables (indicators) Aa1a, Aa1b, Aa1c, and Aa1d are used to measure the game idea/context (Aa1) component.						
	Aa1a	New idea	0.57	0.00	0.42	0.74	0.74
	Aa1b	Detailed explanation of the idea	0.72	11.07			
	Aa1c	Game world	0.58	9.76			
	Aa1d	Game outline	0.71	10.95			
	Hypothesis seven: Two variables (indicators) Aa2a and Aa2b are used to measure the game aims (Aa2) component.						
	Aa2a	Game goals	0.85	0.00	0.67	0.81	0.80
	Aa2b	Game objectives	0.79	11.60			
	Hypothesis eight: four variables (indicators) Aa3a, Aa3b, Aa3c, and Aa3d are used to measure the ludic game script (Aa3) component.						
	Aa3a	Narration / story (story line)	0.65	0.00	0.47	0.78	0.78
	Aa3b	The main roles and characters in the game and their characteristics	0.72	13.03			
	Aa3c	Gamification techniques	0.74	13.20			
	Aa3d	Explain how to create flow in the game	0.64	12.01			
	Hypothesis nine: Two variables (indicators) Aa4a and Aa4b are used to measure the planning and risks (Aa4) component.						
	Aa4a	Specify schedules and coordination	0.87	0.00	0.66	0.79	0.79
	Aa4b	Identify risks	0.75	13.95			
	Hypothesis ten: Five variables (indicators) Aa5a, Aa5b, Aa5c, Aa5d, and Aa5e are used to measure the resources, costs, and budget estimate (Aa5) component.						
	Aa5a	Identify resources (financial) according to the goals of the game and how to obtain resources	0.65	0.00	0.40	0.82	0.83
	Aa5b	Explain how the program handles and adapts to resource loss and backlog	0.68	13.27			
	Aa5c	Game development technical teammates	0.72	13.78			
	Aa5d	Identify sponsors and estimate the required costs and budget	0.74	14.06			
	Aa5e	Market analysis and how the product is delivered to the learner	0.70	13.47			
	Hypothesis eleven: four variables (indicators) Ab1a, Ab1b, Ab1c, and Ab1d are used to measure the problem (Ab1) component.						
	Ab1a	Problem identification and definition	0.59	0.00	0.47	0.78	0.78
	Ab1b	Problem disciplines	0.72	12.01			

Ab1c	Needs analysis based on target learners	0.71	11.96			
Ab1d	Serious outline	0.72	12.02			
Hypothesis twelve: four variables (indicators) Ab2a, Ab2b, Ab2c, and Ab2d are used to measure the serious aims (Ab2) component.						
Ab2a	Serious goals	0.73	0.00	0.58	0.84	0.85
Ab2b	Serious objectives	0.75	16.28			
Ab2c	Cognitive, skill-oriented and emotional goal setting	0.77	16.67			
Ab2d	The answer to why the game is needed	0.80	17.31			
Hypothesis thirteen: Three variables (indicators) Ab3a, Ab3b, and Ab3c are used to measure the target learner profile (Ab3) component.						
Ab3a	Target learners and their gender, age, nationality, culture, and geographical location	0.70	0.00	0.51	0.76	0.76
Ab3b	Previous experiences/skills	0.73	14.21			
Ab3c	The answer to whether the game is played alone or as a team	0.71	13.95			
Hypothesis fourteen: Four variables (indicators) Ab4a, Ab4b, Ab4c, and Ab4d are used to measure the fundamentals, domains, and models or theories (Ab4) component.						
Ab4a	Philosophical, cognitive, psychological, social, cultural, and religious foundations	0.68	0.00	0.51	0.81	0.81
Ab4b	Cognitive, emotional, and psychomotor domains	0.69	13.59			
Ab4c	Learning strategies	0.75	14.57			
Ab4d	Learning theories (including activity theory, empirical learning, flexible learning, or constructivist theories)	0.75	14.52			
Hypothesis fifteen: Six variables (indicators) Ab5a, Ab5b, Ab5c, Ab5d, Ab5e, and Ab5f are used to measure the content and serious resources (Ab5) component.						
Ab5a	Availability of pedagogical and learning resources	0.74	0.00	0.49	0.85	0.85
Ab5b	Syllabus and clarity of content selection criteria	0.75	16.50			
Ab5c	Concept knowledge	0.59	18.00			
Ab5d	Predict learning activities	0.78	17.12			
Ab5e	Domain Expert teammates	0.61	13.32			
Ab5f	Participants required to test prototypes	0.71	15.62			
Hypothesis sixteen: Four variables (indicators) B1a, B1b, B1c, and B1d are used to measure the choose name/engine/genre/technology or platform (B1) component.						
B1a	Choosing a suitable name for the game	0.68	0.00	0.54	0.82	0.82
B1b	Choosing an engine that fits your goals	0.77	15.13			
B1c	Choosing a genre that fits your goals	0.74	14.74			
B1d	Choosing the right technology or platform for your goals	0.74	14.74			
Hypothesis seventeen: Five variables (indicators) B2a, B2b, B2c, B2d, and B2e are used to measure the goals (B2) component.						
B2a	Goals consistent with the right approach	0.73	0.00	0.54	0.85	0.85
B2b	Matching and linking game goals with serious goals	0.75	16.29			
B2c	Ethical, feasible (realistic), and adjustable goals	0.72	15.67			
B2d	Short, medium, and long-term goals	0.74	16.05			
B2e	Determining and predicting scores, tasks, and competitions	0.74	16.04			
Hypothesis eighteen: Five variables (indicators) B3a, B3b, B3c, B3d, and B3e are used to measure the concept development (B3) component.						
B3a	Core (game idea)	0.74	0.00	0.49	0.83	0.82
B3b	Features (core amplifiers)	0.63	13.61			
B3c	Description of the environment and the world of the game	0.65	14.04			
B3d	Pedagogical integration of context with the needs of target learners	0.77	16.42			
B3e	Learning and knowledge foundation	0.69	14.77			
Hypothesis nineteen: Five variables (indicators) B4a, B4b, B4c, B4d, and B4e are used to measure the content and activities (B4) component.						
B4a	Linking of game content to serious content	0.66	0.00	0.50	0.83	0.83

B4b	Accuracy, clarity, and coherence of the content	0.73	14.28			
B4c	The fun content of the domain	0.69	13.69			
B4d	Specific features of activities	0.73	14.28			
B4e	Identify skills/behaviors that should be represented/practiced through the game	0.72	14.20			
Hypothesis twenty: Ten variables (indicators) B5a, B5b, B5c, B5d, B5e, B5f, B5g, B5h, B5i, and B5j are used to measure the designing (B5) component.						
B5a	Turning ideas into product production roadmaps along with pedagogical scenarios	0.71	0.00	0.49	0.91	0.91
B5b	Ludic game script (gamification)	0.78	17.20			
B5c	Rules	0.68	15.09			
B5d	Link learning mechanics and game mechanics	0.67	14.83			
B5e	Incentives and rewards	0.64	14.23			
B5f	Learners' interactions, interfaces, and experiences	0.72	15.84			
B5g	Narrative/story events	0.72	18.96			
B5h	Protagonist and secondary characters	0.67	14.94			
B5i	Aesthetics	0.72	16.07			
B5j	The sequence of steps, game level difficulty rating, and game guide	0.71	15.79			
Hypothesis twenty-one: Two variables (indicators) C1a and C1b are used to measure the objectives (C1) component.						
C1a	Alignment of game objectives with serious objectives	0.82	0.00	0.58	0.73	0.73
C1b	The predetermination of short, medium, and long-term objectives of the game	0.70	12.47			
Hypothesis twenty-two: Three variables (indicators) C2a, C2b, and C2c are used to measure the aesthetics and graphics (C2) component.						
C2a	Representations (including scenes, visual appeal of the environment, characterization, descriptions, roles, and similarity to reality)	0.69	0.00	0.48	0.73	0.73
C2b	Action (including sound, animation, motion, and graphic style)	0.66	12.23			
C2c	GUI indicators (including art, music, and video)	0.72	12.86			
Hypothesis twenty-three: A variable (indicator) C3a is used to measure the Implementation of mechanics (C3) component.						
C3a	Alignment of learning mechanics (including instructional, guidance, participation, generalization / discrimination, observation, explore, identify, plan, objectify, hypothesis, motivation, ownership, responsibility, accountability, incentive, discover, competition, repetition, demonstration, tutorial, action / task, feedback, question and answer, experimentation, reflect / discuss, analyze, imitation, shadowing, modelling, simulation, assessment) with game mechanics (including fun, challenge, behavioral momentum, rewards / penalties, pavlovian interactions, urgent optimism, communal discovery, strategy / planning, story, cooperation, pareto optimal, feedback, protege effects, mini-games, design / editing, realism, ownership, role play, virality, cascading information, collaboration, competition, cut-scenes, action points, levels, tokens, questions and answers, game turns, selecting / collecting, resource management, capture / eliminate – quick feedback, goods / information, time pressure, tutorial, tiles / grids, infinite gameplay, appointment, movement, assessment, status, simulate / response)	1.08	0.00	1.16	1.16	1.16

Hypothesis twenty-four: Ten variables (indicators) C4a, C4b, C4c, C4d, C4e, C4f, C4g, C4h, C4i, and C4j are used to measure the experience and feeling (C4) component.						
C4a	Engagement	0.66	0.00	0.50	0.91	0.91
C4b	Motivation	0.65	13.84			
C4c	Immersion	0.74	14.85			
C4d	Flow and Challenge	0.70	14.26			
C4e	Ease of use	0.69	14.12			
C4f	Affect and tension	0.73	14.76			
C4g	Feeling (Including a sense of self-control, game control, camera look, animation and particle, information clarity, standards, and fluidity)	0.71	14.43			
C4h	Problem-solving/learning questioning and decision-making (meaningful, meaningless, overt, and blindly)	0.75	15.05			
C4i	Support for social interactions (including personal skills, interpersonal skills, practical ethics, and social awareness)	0.72	14.66			
C4j	Cognitive load (including effort rating, difficulty rating, and response time)	0.71	14.41			
Hypothesis twenty-five: Five variables (indicators) C5a, C5b, C5c, C5d, and C5e are used to measure the adaptivity (C5) component.						
C5a	Cognitive adaptation	0.72	0.00	0.49	0.83	0.83
C5b	Motivational adaptation	0.71	15.09			
C5c	Difficulty level	0.66	15.06			
C5d	Intervention	0.72	15.27			
C5e	Game pacing	0.69	14.77			
Hypothesis twenty-six: Ten variables (indicators) C6a, C6b, C6c, C6d, C6e, C6f, C6g, C6h, C6i, and C6j are used to measure the actions (C6) component.						
C6a	Tutoring	0.64	0.00	0.50	0.91	0.91
C6b	Inform learner of objective	0.68	13.56			
C6c	Stimulate recall of prior learning	0.73	14.52			
C6d	Entity, movement, and demonstrate	0.71	14.18			
C6e	Present material, repetition, review and time-related	0.73	14.39			
C6f	Diversity of activities and information and info chunking	0.72	14.25			
C6g	Qualitative and quantitative assess performance	0.71	14.05			
C6h	Reward good performance and sanction bad performance	0.72	14.25			
C6i	Scaffold	0.73	14.50			
C6j	Suggest improvements and support recovery from errors	0.72	14.34			
Hypothesis twenty-seven: Ten variables (indicators) C7a, C7b, C7c, C7d, C7e, C7f, C7g, C7h, C7i, and C7j are used to measure the tools (C7) component.						
C7a	User interface (including color proportion, size proportion)	0.52	0.00	0.47	0.90	0.90
C7b	Rules (including interaction, scoring, rewards, and penalties)	0.73	11.68			
C7c	Objects (including 2D/3D space, avatars, cards, gifts, goods, grids, information, modifiers, non-playing characters, tiles, tokens, and virtual money)	0.71	11.50			
C7d	Feedback (including diagnostic, formative and summative feedback, providing appropriate, immediate, no-delay and careful feedback on various aspects)	0.71	11.52			
C7e	Chance/randomness, deadlines, and choices	0.70	11.43			
C7f	Multimedia	0.67	11.12			
C7g	Textual and graphical information	0.73	11.69			
C7h	Questions & answers, practice, discussion, and checklists	0.65	10.99			
C6i	Help text, help, warning messages, and tips/assistance	0.69	11.38			

C7j	Services (including chatting, ranking, and leaderboard)	0.69	11.33			
Hypothesis twenty-eight: Five variables (indicators) C8a, C8b, C8c, C8d, and C8e are used to measure the usability and function (C8) component.						
C8a	Comprehensibility and Learnability	0.64	0.00	0.52	0.84	0.84
C8b	Ability to operate and service	0.77	14.34			
C8c	Accuracy	0.73	13.77			
C8d	Security	0.70	13.35			
C8e	Commands, motion sensor, keyboard, and appearance	0.75	14.08			
Hypothesis twenty-nine: Five variables (indicators) C9a, C9b, C9c, C9d, and C9e are used to measure the prototype (C9) component.						
C9a	Assess the alignment of game design and content elements with serious design and content elements	0.71	0.00	0.50	0.83	0.83
C9b	Design prototype	0.68	14.40			
C9c	Application prototype	0.71	14.90			
C9d	Alpha test	0.75	15.69			
C9e	Beta test	0.68	14.33			
Hypothesis thirty: Twenty variables (indicators) D1a, D1b, D1c, D1d, D1e, D1f, D1g, D1h, D1i, D1j, D1k, D1l, D1m, D1n, D1o, D1p, D1q, D1r, D1s, and D1t are used to measure the implementation and evaluation (D1) component.						
D1a	Start and introduction	0.60	0.00	0.27	0.88	0.88
D1b	Narrative/story (including reasonable game narrative/story, attractive story/narrative content, complete quest, complete side quests, get acquainted with narration/story, reach narrative/story end)	0.52	10.21			
D1c	Concentration (including attract learner, concentrated attention, deep engagement, animations, color schemes, informative screens, resource management, sound effects and music, tactile response, and virtual or augmented reality)	0.53	10.35			
D1d	Challenges (including adaptive difficulty, difficulty curve, difficulty selection, number Of attempts, artificial intelligence (automatic challenges))	0.56	10.85			
D1e	Control (including error recovery features, fluid animations, pause system, player character acknowledgment, responsive sound effects, save and load system, tactile feedback)	0.51	10.00			
D1f	Clarity of objective criteria (including character dialog, easily accessible help screen, in-game hints and tips, objective log, maximize performance, maximize the score, success level)	0.48	9.57			
D1g	Social interaction (including communication system, competitive and co-op mode suggested elements In-game access to official forum, player ranking and scoring, resource exchange and sharing, discussions, surveys, tests)	0.50	9.90			
D1h	Content and learning activities (including content quality, balanced presentation of ideas, content coherence and learning activities, propaganda and information, ethical, cultural, and social issues (Including personal skills, interpersonal skills, practical ethics, and social awareness), duration of in-game activities, original assignments, diversity of activities)	0.57	10.93			
D1i	Rules (including flexible rules, changing the rules according to the circumstances of the audience, complete information, incomplete information, competition, game modes, game	0.55	10.67			

	master/referee, multiplayer, zero-sum / non-zero-sum)		
D1j	Scaffolding (including info chunking, alternating turns, game period, meta-game, puzzles, quest/problem, time of access, complexity, the structure of game levels, difficulty adaptation, balancing and challenging the learner, limitations, self-assessment, guidance)	0.50	9.86
D1k	User experience and feeling (including motivation (motivating and stimulating learning, encouraging play by embedding challenges within the game), engagement (including engaging the learner in analyzing, interpreting and simplifying concepts, engaging the learner in a variety of tasks), flow, immersion (including animations and visual effects, sound effects and music, storytelling elements), affect, fun, tension, confusion, frustration, anger, pleasure (including establish an emotional connection, competitiveness, excitement, sense of control, providing control of the game and freedom of action for the player), tools usage, ease of use, competition (including first person, last person), explore/navigation (including emotion perception situations, self-navigating situations, positions for finding new opportunities), own, plan/strategy, tactical maneuver, tasks (including collect resources, management resources, collect information, solve the puzzle, capture, design, destroy, edit, eliminate, exchange, trade virtual items, create, match, remove, select), achievements)	0.48	9.50
D1l	Adaptation (including cognitive and motivational adaptation, difficulty level adaptation, intervention, game pacing (role play speed))	0.44	8.90
D1m	Usability (including learnability, operational capability, serviceability, ease of use, reusability, and security)	0.48	9.48
D1n	Interaction (including user interface, click location, click speed, customize, active participation, attention to the rules of collaborative learning and the approximate extent of learners' growth, attention to the in-game discussion, movement (including avoid, collide, move, evade, rotate, shoot, target, teleport, visit), time (including manipulate time, start/stop time, advance game period), problem-solving (including problems, puzzles, providing hypothetical opportunities, creating ambiguous and questionable situations, providing the necessary information to solve the problem, situations for implementing conjectural solutions))	0.49	8.70
D1o	Actions (including remembering (define, describe, draw, find, identify, imitate, label, list, locate, match, memorize, name, observe, read, recall, recite, recognize, relate, reproduce, select, state, write, tell), understanding (compare, convert, demonstrate, describe, discuss, distinguish, explain, explore, find more information about, generalize, interpret, objectify, outline, paraphrase, predict, put into own words,	0.57	10.88

	relate, restate, summarize, translate, visualize), applying (apply, calculate, change, choose, classify, complete goal, complete, construct, examine, experiment, illustrate, interpret, make, manipulate, modify, perform action / task, produce, put into practice, put together, show, solve, translate, use), analyzing (advertise, analyze, categorize, compare, contrast, deduce, differentiate, discover, distinguish, examine, explain, identify, investigate, separate, subdivide, take apart), evaluating (argue, assess, choose, critique, debate, decide, defend, determine, discuss, estimate, evaluate, judge, justify, prioritize, rate, recommend, review, select, value, verify, weigh), creating (add to, build model, combine, compose, construct, create, design, devise, forecast, form goal, formulate, hypothesize, imagine, invent, originate, plan, predict, propose))		
D1p	Presentation design (including feedback (including diagnostic, formative, summative, action, explanatory, practical, no-delay and careful, learner ranking and scoring, progress log, progress-related cutscenes, responsive interactions, victory and defeat screens, achievements, leaderboards, performance meters, performance record, points, progress bars, rewards, status levels, rewards (including providing short, medium and long term rewards), penalties), aesthetics (including art, cartoons, environments and characters, diagrams, displays, graphed information, graphics, graphs, illustrations, maps))	0.55	10.61
D1q	Fidelity (including physical fidelity (including conditions, states, features, attributes, phenomena, specifically, size, color, shape, quality, temperature, and density), functional fidelity (including information, functional options, learning content, stimulus, responses, feedback), sensory fidelity (including visual, auditory, olfactory, tactile effects, spatial and kinesthetic perception), conceptual fidelity (including abstract theories, knowledge, content, functions, rules, relationships), emotional fidelity (from including emotions of pleasant, excited, stressful, anxious, nervous))	0.51	9.95
D1r	Internal support (including advice and assistance, guide character, checklists/task lists, tips, tutorials, warning messages, facilitating, and various resources)	0.57	11.01
D1s	External support (including contact number, email, and website)	0.51	9.97
D1t	Maintainability (including testability, analyzability, changeability, stability, and extensibility)	0.52	10.22

*Note: All factor loadings are significant at the 95% confidence level: ($p < 0.05$) and ($t > 1.96$)

As the data in Table 6 shows, the factor loadings (standard coefficient) obtained for all dimensions, components, and indicators are more than 0.40 and are at a significant level (significance coefficient) of less than 0.05 ($p < 0.05$, all t values are greater than 1.96), we conclude that the construct validity of all dimensions, components, and indicators are confirmed. The value of compositional reliability and Cronbach's alpha of all dimensions,

components, and indicators was more than 0.70, which means that they have good reliability. The average variance extracted, which measures the convergent validity of each scale, was obtained from a minimum of 0.27 to a maximum of 0.67, which in total are suitable values and confirmed for convergent validity. Overall, the results show the reliability and validity of all dimensions, components, and indicators.

Table 7. Fit indexes of confirmatory factor analysis of dimensions, components, and indicators for evaluating the educational quality of serious games

Dimensions, components and indicators	Fit indexes, processing range and calculated fit						
	X^2/DF	RMSEA	GFI	CFI	NNFI	NFI	IFI
	<5	<0.08	<0.05	>0.9	>0.9	>0.9	>0.9
The dimensions of evaluating the educational quality of serious games	4.06	0.07	0.99	0.99	0.96	0.98	0.99
Components of preparation (A), pre-production (B), production (C), and post-production (D) dimensions	2.81	0.06	0.90	0.97	0.96	0.94	0.97
Indicators of the preparation (A) dimension	1.78	0.38	0.90	0,99	0.98	0.97	0.99
Indicators of the pre-production (B) dimension	1.56	0.32	0.93	0.99	0.99	0.98	0.99
Indicators of the production (C) dimension	1.54	0.32	0.88	0.99	0.99	0.97	0.99
Indicators of the post-production (D) dimension	1.17	0.02	0.96	1	1	0.97	1

*Abbreviations: X^2/DF, Normal theory weighted least squares chi-square/Degrees of freedom; RMSEA, Root mean square error of approximation; GFI, Goodness of fit index; CFI, Comparative fit index; NNFI, Non-normed fit index; NFI, Normed fit index; IFI, Incremental fit index.

Thus, as the fit indexes in Table 7 show, the data of this research has an ideal fit with the factor constructs and the theoretical foundation of the research, and this indicates the high reliability of the constructs.

2.7 Sixth step: Quality control

In the last step, the findings from the previous steps are presented, where the research literature reflected 4 general dimensions that can be evaluated through 25 components and 138 indicators (see Table 6).

3. Findings

In Table 6, the dimensions, components, and indicators extracted from the literature were discussed. These components and indicators were classified and confirmed in 4 main dimensions (1. preparation 2. pre-production 3. production 4. post-production).

The first dimension (step) of the proposed framework is preparation. The preparation dimension contains two technical (game) and non-technical (serious) aspects, which are the arms of serious game development. This step can be more or less overwhelming depending on the size of the project. It is very creative and can be very time-consuming. The second dimension identified in the framework for evaluating the educational quality of serious games is pre-production. The main purpose of this dimension when developing a serious game is to "technical (game) and non-technical (serious) link". The third dimension of the proposed framework is production. The goal of the production dimension in the development process of a serious game is "technical (game) and non-technical (serious) alignment" [40]. In this dimension, the developers try to align the technical and non-technical features so that the learners are interested in the educational content of the game which leads to the improvement of their learning results. The last dimension or step is after production. This step includes implementation and evaluation. After the release of the game, it is necessary to evaluate the final product with an appropriate methodology (appropriate research method, appropriate community and sample, and implementation or integration of the game in the field to be used) and data collection and analysis (appropriate collection methods and appropriate analysis

methods). In Figure 3, the dimensions and components of the educational quality evaluation framework of serious games (apart from the indicators) can be seen.

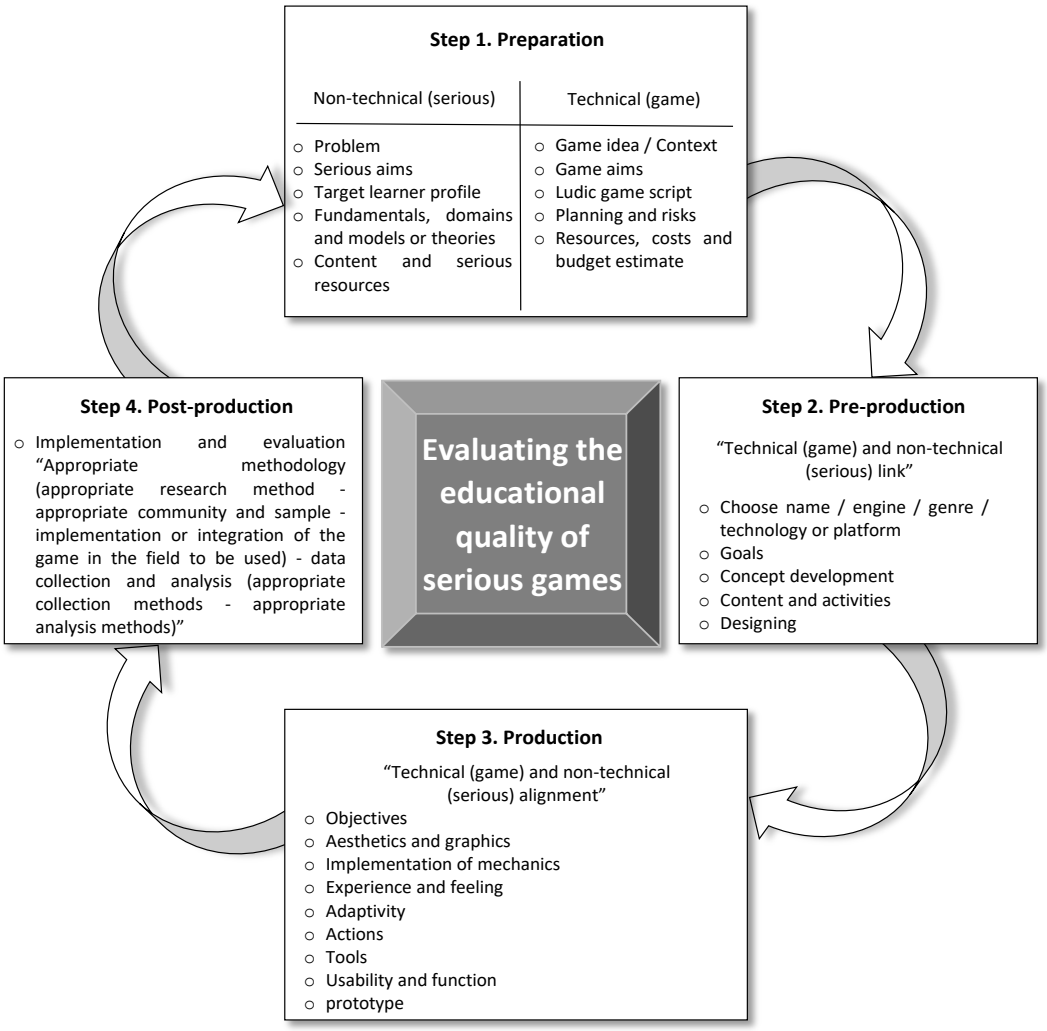

Figure 3. Dimensions and components of the educational quality evaluation framework of serious games.

4. Discussion and conclusion

In order to design a framework for evaluating the educational quality of serious games, this research first studies articles and examines the theoretical foundations and previous research, and then uses the meta-synthesis method to identify dimensions, components, and indicators in the field of serious games. The results of the fourth and fifth steps of meta-synthesis (extracting information from the studies and analysis and synthesis of findings) during several steps of coding led to the identification of 4 main dimensions of preparation, pre-production, production, and post-production, 25 components and 139 indicators. In the continuation of the research and the sixth step of meta-synthesis (quality control), the primary tool was designed based on the data of the previous step and sent to experts, professors, and developers to measure content validity. Content validity led to the removal of one indicator. Then, confirmatory factor analysis was used to check the validity of the introduced constructs. In order to perform confirmatory factor analysis, 30 hypotheses (one hypothesis for dimensions, four hypotheses for components, and twenty-five hypotheses for indicators) were proposed in this research. The

first and second-order factor analysis confirmed all factor loadings and significant coefficients. Therefore, the proposed framework was approved with 4 dimensions, 25 components, and 138 indicators.

In relation to supporting the findings of this research, as all the dimensions, components, and indicators have been obtained from the research literature, there is much research evidence that confirms and supports the findings of the current research. The evidence in the dimensions section, in accordance with the findings of the current research, shows that the dimensions of the educational quality evaluation of serious games include preparation, pre-production, production, and post-production [e.g., 1-3-14-55-59-63]. Also, research evidence [e.g., 1-3-15-16-52-61] confirms and supports the findings related to the components and indicators section of the current research.

The new achievement of our research is to resolve the gap created in the literature on the educational quality of serious games. Our research, unlike the previous researches, each of which pointed to some stages of educational quality evaluation in the development of serious games, by presenting a new framework and tool, covers from the beginning to the end of a serious game design project and its educational quality evaluation, in a way that may to some extent guarantee the educational quality of serious games. In this way, all the development steps of a serious game and its dimensions, components, and indicators have been identified and presented. The evaluation of the preparation dimension affects pre-production, pre-production on production, production on post-production, and post-production on preparation and other steps. Therefore, we hope that the proposed framework and tool will encourage investors, developers, and researchers to develop serious games with higher educational quality in the future.

Acknowledgments

In the end, we are obliged to thank Iran's National Computer Games Foundation for distributing the tool of this research among the experts in the field of serious and computer games.

Conflicts of interest

There are no conflicts of interests to disclose.

References

[1] D. Ávila-Pesántez, L. A. Rivera, and M. S. Alban, "Approaches for serious game design: A systematic literature review," ASEE Comput. Educ. J., vol. 8, no. 3, 2017.

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

[3] I. Mayer et al., "The research and evaluation of serious games: Toward a comprehensive methodology," Br. J. Educ. Technol., vol. 45, no. 3, pp. 502–527, 2014. https://doi.org/10.1111/bjet.12067

[4] C. S. Loh, Y. Sheng, and D. Ifenthaler, "Serious games analytics: Theoretical framework," in Serious games analytics, Springer, 2015, pp. 3–29. https://doi.org/10.1007/978-3-319-05834-4_1

[5] D. Charsky, "From edutainment to serious games: A change in the use of game characteristics," Games Cult., vol. 5, no. 2, pp. 177–198, 2010. https://doi.org/10.1177/1555412009354727

[6] W. Westera, R. J. Nadolski, H. G. K. Hummel, and I. G. J. H. Wopereis, "Serious games for higher education: a framework for reducing design complexity," J. Comput. Assist. Learn., vol. 24, no. 5, pp. 420–432, 2008. https://doi.org/10.1111/j.1365-2729.2008.00279.x

[7] M. B. Carvalho et al., "A case study on service-oriented architecture for serious games," Entertain. Comput., vol. 6, pp. 1–10, 2015. https://doi.org/10.1016/j.entcom.2014.11.001

[8] C. Kleinschmidt and M. Haag, "Evaluation of Game Engines for Cross-Platform Development of Mobile Serious Games for Health.," in eHealth, 2016, pp. 207–214. https://doi.org/10.3233/978-1-61499-645-3-207
[9] S.-J. Yoon and H.-S. Park, "Development of a Quality Evaluation Standard for Educational Serious Games," J. Inf. Commun. Converg. Eng., vol. 11, no. 2, pp. 103–111, 2013. https://doi.org/10.6109/jicce.2013.11.2.103
[10] J. A. DeFalco et al., "Detecting and addressing frustration in a serious game for military training," Int. J. Artif. Intell. Educ., vol. 28, no. 2, pp. 152–193, 2018. https://doi.org/10.1007/s40593-017-0152-1
[11] M. Suppan et al., "Nationwide Deployment of a Serious Game Designed to Improve COVID-19 Infection Prevention Practices in Switzerland: Prospective Web-Based Study," JMIR serious games, vol. 9, no. 4, p. e33003, 2021. https://doi.org/10.2196/33003
[12] P. C. Ferreira et al., "Exploring empathy in cyberbullying with serious games," Comput. Educ., vol. 166, p. 104155, 2021. https://doi.org/10.1016/j.compedu.2021.104155
[13] C. Tanner, D. Schmocker, J. Katsarov, and M. Christen, "Educating moral sensitivity in business: An experimental study to evaluate the effectiveness of a serious moral game," Comput. Educ., vol. 178, p. 104381, 2022. https://doi.org/10.1016/j.compedu.2021.104381
[14] D. Jaccard, L. Suppan, E. Sanchez, A. Huguenin, and M. Laurent, "The co. LAB Generic Framework for Collaborative Design of Serious Games: Development Study," JMIR serious games, vol. 9, no. 3, p. e28674, 2021. https://doi.org/10.2196/28674
[15] J. L. F. de Almeida and L. dos Santos Machado, "Design requirements for educational serious games with focus on player enjoyment," Entertain. Comput., vol. 38, p. 100413, 2021. https://doi.org/10.1016/j.entcom.2021.100413
[16] P. Caserman et al., "Quality criteria for serious games: serious part, game part, and balance," JMIR serious games, vol. 8, no. 3, p. e19037, 2020. https://doi.org/10.2196/19037
[17] Y. Zhonggen, "A meta-analysis of use of serious games in education over a decade," Int. J. Comput. Games Technol., vol. 2019, 2019. https://doi.org/10.1155/2019/4797032
[18] A. Min, H. Min, and S. Kim, "Effectiveness of serious games in nurse education: A systematic review," Nurse Educ. Today, vol. 108, p. 105178, 2022. https://doi.org/10.1016/j.nedt.2021.105178
[19] C. A. Clark, "Serious games," New York Viking, 1970.
[20] M. Zyda, "From visual simulation to virtual reality to games," Computer (Long. Beach. Calif)., vol. 38, no. 9, pp. 25–32, 2005. https://doi.org/10.1109/MC.2005.297
[21] T. Marsh, "Serious games continuum: Between games for purpose and experiential environments for purpose," Entertain. Comput., vol. 2, no. 2, pp. 61–68, 2011. https://doi.org/10.1016/j.entcom.2010.12.004
[22] G. P. Papanastasiou, A. S. Drigas, and C. Skianis, "Serious Games in Preschool and Primary Education: Benefits And Impacts on Curriculum Course Syllabus.," Int. J. Emerg. Technol. Learn., vol. 12, no. 1, 2017. https://doi.org/10.3991/ijet.v12i01.6065
[23] L. K. P. Suryapranata, B. Soewito, G. P. Kusuma, F. L. Gaol, and H. L. H. S. Warnars, "Quality measurement for serious games," in 2017 International Conference on Applied Computer and Communication Technologies (ComCom), 2017, pp. 1–4. https://doi.org/10.1109/COMCOM.2017.8167098
[24] R. Berta, F. Bellotti, E. van der Spek, and T. Winkler, "A tangible serious game approach to science, technology, engineering, and mathematics (STEM) education," Handb. Digit. Games Entertain. Technol., pp. 1–22, 2016. https://doi.org/10.1007/978-981-4560-52-8_32-1
[25] T. C. Clapper, "Serious games are not all serious." SAGE Publications Sage CA: Los Angeles, CA, 2018.
[26] A. Dimitriadou, N. Djafarova, O. Turetken, M. Verkuyl, and A. Ferworn, "Challenges in serious game design and development: Educators' experiences," Simul. Gaming, vol. 52, no. 2, pp. 132–152, 2021. https://doi.org/10.1177/1046878120944197
[27] M.-A. Maheu-Cadotte et al., "Involvement of End Users in the Development of Serious Games for Health Care Professions Education: Systematic Descriptive Review," JMIR serious games, vol. 9, no. 3, p. e28650, 2021. https://doi.org/10.2196/28650
[28] X. Ye, P. Backlund, J. Ding, and H. Ning, "Fidelity in simulation-based serious games," IEEE Trans. Learn. Technol., vol. 13, no. 2, pp. 340–353, 2019. https://doi.org/10.1109/TLT.2019.2913408

[29] F. Ulrich and N. H. Helms, “Creating evaluation profiles for games designed to be fun: An interpretive framework for serious game mechanics,” Simul. Gaming, vol. 48, no. 5, pp. 695–714, 2017. https://doi.org/10.1177/1046878117709841
[30] Y. Wang, P. Rajan, C. S. Sankar, and P. K. Raju, “Let them play: the impact of mechanics and dynamics of a serious game on student perceptions of learning engagement,” IEEE Trans. Learn. Technol., vol. 10, no. 4, pp. 514–525, 2016. https://doi.org/10.1109/TLT.2016.2639019
[31] S. Arnab et al., “Mapping learning and game mechanics for serious games analysis,” Br. J. Educ. Technol., vol. 46, no. 2, pp. 391–411, 2015. https://doi.org/10.1111/bjet.12113
[32] P. Rooney, “A theoretical framework for serious game design: exploring pedagogy, play and fidelity and their implications for the design process,” Int. J. Game-Based Learn., vol. 2, no. 4, pp. 41–60, 2012. https://doi.org/10.4018/ijgbl.2012100103
[33] J. Brinckmann, Competence of top management teams and success of new technology-based firms. Springer, 2008. https://doi.org/10.1007/978-3-8350-5503-2
[34] H. Danaee Fard and A. A. Anvary Rostamy, “Promoting public trust in public organizations: Explaining the role of public accountability,” Public Organ. Rev., vol. 7, pp. 331–344, 2007. https://doi.org/10.1007/s11115-007-0041-4
[35] J.-M. Trujillo-Torres, H. Hossein-Mohand, M. Gómez-García, H. Hossein-Mohand, and F.-J. Hinojo-Lucena, “Estimating the academic performance of secondary education mathematics students: A gain lift predictive model,” Mathematics, vol. 8, no. 12, p. 2101, 2020. https://doi.org/10.3390/math8122101
[36] A. Bryman, Social research methods. Oxford university press, 2016.
[37] R. Barnett, “Power, enlightenment and quality evaluation,” Eur. J. Educ., vol. 29, no. 2, pp. 165–179, 1994. https://doi.org/10.2307/1561639
[38] L. J. Cronbach, “Essential of Psychology Testing,” New York: Harper & Row, 1984. https://doi.org/10.1177/014662168400800314
[39] A. J. Abdellatif, B. McCollum, and P. McMullan, “Serious games: Quality characteristics evaluation framework and case study,” in 2018 IEEE Integrated STEM Education Conference (ISEC), 2018, pp. 112–119. https://doi.org/10.1109/ISECon.2018.8340460
[40] W. H. Tan and I. Boo, “A Provisional framework for multimodal evaluation: Establishing serious games quality label for use in training and talent development,” in Game-based assessment revisited, Springer, 2019, pp. 163–183. https://doi.org/10.1007/978-3-030-15569-8_9
[41] D. Gee et al., “Assessing serious games: The GRAND assessment framework,” 2014. http://dx.doi.org/10.11575/PRISM/35635
[42] K. Mitgutsch and N. Alvarado, “Purposeful by design? A serious game design assessment framework,” in Proceedings of the International Conference on the foundations of digital games, 2012, pp. 121–128. https://doi.org/10.1145/2282338.2282364
[43] J. Thomas and A. Harden, “Methods for the thematic synthesis of qualitative research in systematic reviews,” BMC Med. Res. Methodol., vol. 8, no. 1, pp. 1–10, 2008. https://doi.org/10.1186/1471-2288-8-45
[44] M. Sandelowski and J. Barroso, Handbook for synthesizing qualitative research. springer publishing company, 2006.
[45] H. Theelen, A. Van den Beemt, and P. den Brok, “Classroom simulations in teacher education to support preservice teachers’ interpersonal competence: A systematic literature review,” Comput. Educ., vol. 129, pp. 14–26, 2019. https://doi.org/10.1016/j.compedu.2018.10.015
[46] M. Banerjee, M. Capozzoli, L. McSweeney, and D. Sinha, “Beyond kappa: A review of interrater agreement measures,” Can. J. Stat., vol. 27, no. 1, pp. 3–23, 1999. https://doi.org/10.2307/3315487
[47] M. Høiseth, O. A. Alsos, S. Holme, S. Ek, and C. T. Gabrielsen, “Serious Game Design to Support Children Struggling with School Refusal,” Int. J. Serious Games, vol. 8, no. 2, pp. 109–128, 2021. https://doi.org/10.17083/ijsg.v8i2.416
[48] M. N. Katsantonis, I. Mavridis, and D. Gritzalis, “Design and Evaluation of COFELET-based Approaches for Cyber Security Learning and Training,” Comput. Secur., vol. 105, p. 102263, 2021. https://doi.org/10.1016/j.cose.2021.102263
[49] G. Kalmpourtzis and M. Romero, “Constructive alignment of learning mechanics and game mechanics in Serious Game design in Higher Education,” Int. J. Serious Games, vol. 7, no. 4, pp. 75–88, 2020. https://doi.org/10.17083/ijsg.v7i4.361

[50] M. Carrión-Toro, M. Santorum, P. Acosta-Vargas, J. Aguilar, and M. Pérez, "iPlus a User-Centered Methodology for Serious Games Design," Appl. Sci., vol. 10, no. 24, p. 9007, 2020. https://doi.org/10.3390/app10249007

[51] K. Omari, M. Moussetad, E. Labriji, and S. Harchi, "Proposal for a New Tool to Evaluate a Serious Game," Int. J. Emerg. Technol. Learn., vol. 15, no. 17, pp. 238–251, 2020. Retrieved February 28, 2023 from https://www.learntechlib.org/p/218030/.

[52] J. Moizer et al., "An approach to evaluating the user experience of serious games," Comput. Educ., vol. 136, pp. 141–151, 2019. https://doi.org/10.1016/j.compedu.2019.04.006

[53] C. Meftah, A. Retbi, S. Bennani, and M. K. Idrissi, "Mobile Serious Game Design Using User Experience: Modeling of Software Product Line Variability.," Int. J. Emerg. Technol. Learn., vol. 14, no. 23, 2019. https://doi.org/10.3991/ijet.v14i23.10899

[54] J. W. Tan and N. Zary, "Diagnostic markers of user experience, play, and learning for digital serious games: a conceptual framework study," JMIR serious games, vol. 7, no. 3, p. e14620, 2019. https://doi.org/10.2196/14620

[55] D. Avila-Pesantez, R. Delgadillo, and L. A. Rivera, "Proposal of a conceptual model for serious games design: a case study in children with learning disabilities," IEEE Access, vol. 7, pp. 161017–161033, 2019. https://doi.org/10.1109/ACCESS.2019.2951380

[56] R. F. Mulcahy, R. Russell-Bennett, N. Zainuddin, and K.-A. Kuhn, "Designing gamified transformative and social marketing services: an investigation of serious m-games," J. Serv. Theory Pract., 2018. https://doi.org/10.1108/JSTP-02-2017-0034

[57] A. Gauthier and J. Jenkinson, "Designing productively negative experiences with serious game mechanics: Qualitative analysis of game-play and game design in a randomized trial," Comput. Educ., vol. 127, pp. 66–89, 2018. https://doi.org/10.1016/j.compedu.2018.08.017

[58] N. Lytle, M. Floryan, and D. Amin, "Experience, experiment, evaluate: A framework for assessing experiential games," Int. J. Serious Games, vol. 4, no. 1, 2017. https://doi.org/10.17083/ijsg.v4i1.128

[59] S. Arnab and S. Clarke, "Towards a trans-disciplinary methodology for a game-based intervention development process," Br. J. Educ. Technol., vol. 48, no. 2, pp. 279–312, 2017. https://doi.org/10.1111/bjet.12377

[60] L. Shoukry and S. Göbel, "Reasons and Responses: A multimodal serious games evaluation framework," IEEE Trans. Emerg. Top. Comput., vol. 8, no. 1, pp. 245–255, 2017. 10.1109/TETC.2017.2737953

[61] M. B. Carvalho et al., "An activity theory-based model for serious games analysis and conceptual design," Comput. Educ., vol. 87, pp. 166–181, 2015. https://doi.org/10.1016/j.compedu.2015.03.023

[62] C.-H. Su, K. T.-K. Chen, and K.-K. Fan, "Rough set theory based fuzzy TOPSIS on serious game design evaluation framework," Math. Probl. Eng., vol. 2013, 2013. https://doi.org/10.1155/2013/407395

[63] J. Torrente et al., "Development of game-like simulations for procedural knowledge in healthcare education," IEEE Trans. Learn. Technol., vol. 7, no. 1, pp. 69–82, 2013. https://doi.org/10.1109/TLT.2013.35

[64] L. A. Annetta, "The 'I's' have it: A framework for serious educational game design," Rev. Gen. Psychol., vol. 14, no. 2, pp. 105–113, 2010. https://doi.org/10.1037/a0018985

[65] D. L. Jackson, "Revisiting sample size and number of parameter estimates: Some support for the N: q hypothesis," Struct. Equ. Model., vol. 10, no. 1, pp. 128–141, 2003. https://doi.org/10.1207/S15328007SEM1001_6

[66] R. B. Kline, "Principles and Practice for Structural Equation Modelling (3rd Eds)." New York: The Guilford Press, 2011.