



# Development of a Quality Evaluation Standard for Educational Serious Games

Seon-Jeong Yoon<sup>1</sup> and Hee-Sook Park<sup>2\*</sup> *Member, KIICE*

<sup>1</sup>Division of Digital Contents, Dongseo University, Pusan 617-716, Korea

<sup>2</sup>Department of Computer Engineering, Pukyong National University, Pusan 608-737, Korea

## Abstract

Given the lack of suitable quality evaluation standards for educational serious games (designed for both entertainment and instruction), we designed a development framework for evaluation standards of educational serious games and proceeded to develop standards in accordance with our proposed procedure. Our standards were designed to evaluate the quality of both technical and non-technical elements of educational serious game software products. We conducted a survey on the need for individual elements of the standard. Participants rated the need for each element on a five-point Likert scale. We then performed a reliability analysis of the survey results. Based on the survey results, we established a final standard for quality evaluation composed of 9 main elements and 31 sub-elements. The results of our research will contribute useful information to users as well as to the developers of educational serious games.

**Index Terms:** Educational serious game, Evaluation elements, Metadata, Metrics, Quality standard

## I. INTRODUCTION

Serious games are computer games designed not only for entertainment but also for achieving a special purpose [1]. In comparison with other computer games, they have a short history, yet their popularity is quite apparent. Various fields are currently in the process of developing serious games. Further, the market volume and demand for serious games increases each year. In 2012, the world market volume of serious games is expected to reach US \$51.5 billion [2].

Today, serious games have diverse applications in fields such as the military, government, education, business, and health care, and in the future, their application is expected to expand to even more fields. A wide variety of terms have been used to describe learning programs based on games, including “edugame,” “edutainment,” and “game learning.”

The current trend is to represent these concepts under the term “serious game.”

Educational serious games were first developed to increase the effectiveness of student learning using the properties of conventional computer games, namely, that they are interesting and fun and have flow. Among the many types of serious games, educational serious games currently occupy the highest market share. According to 2009 Statistics, educational serious games occupied over 30% of the United States serious game market and comprised more than 50% of South Korea’s market [2].

Nevertheless, quality evaluation standards for educational serious games have not yet been established, either internationally or within Korea. As a result, users are inconvenienced in the process of selecting educational serious games. Additionally, game developers frequently

Received 07 January 2013, Revised 13 February 2013, Accepted 28 February 2013

\*Corresponding Author Hee-Sook Park (E-mail: [bg0007@hanmail.net](mailto:bg0007@hanmail.net), Tel: +82-10-4574-3858)

Department of Computer Engineering, Pukyong National University, 45 Yongso-ro, Nam-gu, Pusan 608-737, Korea.

**Open Access** <http://dx.doi.org/10.6109/jicce.2013.11.2.103>

print ISSN: 2234-8255 online ISSN: 2234-8883

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering

design serious game products without reflecting on specific quality characteristics or quality evaluation metrics. This can be a major cause of weakening market competitiveness of the products.

Although some existing software quality evaluation standards have been developed, they are not suitable for use in the quality assessment of educational serious games because they fail to take into account specific characteristics of such games.

Thus, we considered it necessary to develop a quality assessment standard for educational serious games. To this end, we designed a development framework and used it to identify quality evaluation elements [1]. The objectives of this study are as follows: Firstly, it seeks to provide criteria of quality assessment in order to aid users in their selection of educational serious games. Secondly, it aims to supply developers with standardized quality criteria for the production of high-quality educational serious games in a highly competitive market. The remainder of this paper is organized as follows. In Section II, we present an overview of related works and existing standards of software quality assessment. Section III explains the design of the quality evaluation elements. Then, in Section IV, we describe the development of our quality assessment standard model. In Section V, we state metrics of the quality evaluation. Finally, in Section VI, we discuss the final conclusions of our study.

## II. RELATED WORKS

This chapter presents an overview of serious games and the previous literature that has addressed them. It also gives a summary of the existing quality assessment standards.

### A. Serious Games

Serious games are designed for a special purpose in training or education. By playing serious games, users can experience both fun and effective learning of specific content material. The software design includes elements of both simple entertainment and pedagogy [3]. Serious games can be classified according to various criteria depending on the application purpose, market, game modes, target users, etc. Zyda [4] notes that serious games can be applied in fields such as health care, public policy, strategy, communications, defense, training, and education. Michael and Chen [1] categorize them into military, government, educational, corporate, or healthcare games, or “politics, religion, and art games” as a single category, according to the market majority. Now, a variety of serious game types have been developed, and positive awareness of and demand for them continue to increase.

### B. Standards of Quality Assessment

Software quality management has been classified into two types. The first type considers that if a software organization’s development method and procedure are appropriate, it can produce high-quality products. Therefore, this perspective emphasizes quality assessment of the development process itself. Yet, appropriate development methods and procedures do not guarantee product quality. Therefore, the second viewpoint emphasizes quality assessment of the product. Based on these viewpoints, two kinds of software quality evaluation models have been developed and applied in IEEE [5]. The former viewpoint is represented by international standards such as the Capability Maturity Model (CMM), ISO/IEC 15504 (Software Process Improvement & Capability Determination [SPICE]), and ISO/IEC 19796-1. Examples of the latter are ISO/IEC 9126, ISO/IEC 25051, ISO/IEC 25040, and ISO/IEC 25000. The CMM consists of five levels: Initial level, Repeatable level, Defined level, Managed level, and Optimizing level [6]. ISO/IEC 15504 standards identify five process categories: customer/supplier, engineering, supporting, management, and organization [7].

The ISO/IEC 19796-1 standard is a method of quality assessment concerning learning, education, and training. This process model consists of a needs analysis, framework analysis, conception/design, development/production, implementation, learning process, and evaluation/optimization. Each process is divided into 38 sub-processes [8].

The ISO/IEC 9126 standard includes 6 main characteristics and 27 sub-characteristics [9-11]. ISO/IEC 25051 is used for evaluation of software package quality; it conforms to the quality model of the ISO/IEC 9126 standard. ISO/IEC 25051 classifies quality requirements for product description, user documentation, and programs and data [12]. ISO/IEC 25040 covers steps in establishing evaluation requirements as well as specifying, designing, and executing the evaluation. Use of this standard model is appropriate for the quality evaluation of software products by developers, acquirers, and evaluators [13]. The ISO/IEC 25000 is a software evaluation model for integrating and enhancing existing international standards that have been developed for use in place of the ISO/IEC 9126 and ISO/IEC 25040 standards [14].

The TTAS.KO-11.0078 is for mobile game software, and TTAS.KO-11.0059, for Web-based software. These standards are representative of Korean quality evaluation standards for software.

The TTAS.KO-11.0078 categorizes elements as technical or non-technical. For technical elements, it provides a quality model composed of 6 main characteristics and 25 sub-characteristics. The technical quality model considers both general software attributes and mobile software

attributes. For non-technical elements, the TTAS.KO-11.0078 provides a quality model composed of 4 main characteristics and 11 sub-characteristics. The non-technical quality model can be used to evaluate both multimedia properties and service properties [15].

The TTAS.KO-11.0059 differs from other standards in that it considers security properties of the software. The model lists 6 main characteristics such as usability, stability, interoperability, efficiency, security, and maintainability, and 23 sub-characteristics [16].

### III. DESIGN OF THE QUALITY EVALUATION ELEMENTS

The major purpose of our work is to develop quality evaluation standards for educational serious game products. Therefore, in this chapter, we describe the necessary elements for development of quality evaluation standards for educational serious game products.

Educational serious games include features that are associated with multiple factors, including factors related to software, content, gaming, and pedagogy. Thus, the development of quality elements necessary to educational serious game quality evaluation standards should distinguish between technical and non-technical evaluation areas.

**Table 1.** Frameworks of development

Preliminary survey	Design of the quality evaluation elements	Development of the quality standard model	Evaluation and Testing	Complementary works
• Analysis of existing s/w quality standard	• Design of frameworks	• Survey of questionnaire & analysis of results	• Development of the quality evaluation tools	• Identification of insufficient elements
• Analysis of learning theories of constructivism	• Mapping of existing standards	• Verification of questionnaire reliability	• Selection of games for evaluation	• Complement/improvement
• Analysis of serious games	• Definition of the quality evaluation elements	• Determination of the final quality evaluation standard	• Organization of evaluation team	• Decision of the game's quality rating
• Analysis of characteristics of educational games	• Creation of the first draft of a standard	• Creation of metrics of the quality evaluation	• Simulation	

Technical evaluation areas should cover the necessary quality evaluation elements for functional or technical elements of educational serious game software. In contrast, non-technical evaluation areas should include quality assessment elements for features of contents, game, and pedagogy. We adopt the distinction between technical and non-technical in our development of quality evaluation elements.

#### A. Framework of Development

Table 1 shows a procedure required for developing a standard of quality assessment of educational serious games.

The required procedure for development for our standard consists of all five steps; however, in this paper, we discuss only the first three stages of research. We leave the remaining two steps of the procedure to be described in our next study.

#### B. Mapping of Existing Standards

For development of the quality evaluation standard, we began by mapping the existing standards, namely, the international standard ISO/IEC 9126 and the Korean standard TTAS.KO-11.0078, to ensure the objectivity and reliability of the elements used. We extracted the necessary elements based on the results of mapping and accepted the extracted elements into our standard. These elements were used to evaluate technical areas of educational serious games. The main concern of our research is to develop evaluation standards for the quality of educational serious game software products. Thus, the reasons behind our selection of ISO/IEC 9126 and TTAS.KO-11.0078 among several existing standards of quality assessment are as follows. First, we selected ISO/IEC 9126 because three other standards comply with it, and it is currently the most widely used as an international standard. We chose TTAS.KO-11.0059 because it contains many enhanced elements of security.

For this reason, we regarded it as a more reasonable choice than TTAS.KO-11.0059. Moreover, many elements of the TTAS.KO-11.0078 standard are consistent with the elements of the ISO/IEC 9126 standard. Second, we considered that the developed standard model will be used mainly in Korea, so it is practical to choose a Korean standard. Further, we hope to expand our standard into an international standard or, at least, to use it as a guide in the development of an international standard model. Therefore, it is wise to comply with one of the international standards. Table 2 shows the results of mapping of the two standards. These results will be used as elements for evaluating the technical quality of educational serious games.

**Table 2.** A result of mapping for ISO/IEC 9126\* and TTAS.KO-11.0078 [11, 15]

ISO/IEC 9126		TTAS.KO-11.0078	
Main characteristics	Sub-characteristics	Main characteristics	Sub-characteristics
Functionality	Suitability, Accuracy, Interoperability, Security, Functionality Compliance	Functionality	Suitability, Accuracy, Interoperability, Security, Concurrency
Reliability	Maturity, Fault Tolerance Recoverability, Reliability compliance	Reliability	Maturity, Fault tolerance Recoverability, Reliability, Compliance, Availability
Usability	Understandability, Learnability, Operability, Attractiveness, Usability Compliance	Usability	Understandability, Learnability, Operability, Attractiveness, Usability, Compliance
-	-	Serviceability	Accessibility, Diversity, Utilizability
Efficiency	Time Behavior, Resource Utilization, Efficiency, Compliance	Efficiency	Time Utilization, Resource Utilization
Maintainability	Analyzability, Changeability, Stability, Testability, Maintainability, Compliance	Maintainability	Changeability, Stability, Testability, Maintainability, Compliance
Portability	Adaptability, Installability, Co-existence, Replaceability, Portability, Compliance	Portability	Adaptability, Installability, Co-existence

\* © ISO/IEC 2013. All right reserved.

As shown in Table 2, there are six main elements common to both standards: functionality, reliability, usability, efficiency, maintainability, and portability. Additionally, the standards contain 19 sub-elements that are the same or similar. We extracted these 6 main elements and 19 sub-elements and incorporated them into our standard. In addition, we combined the Time Behavior of ISO/IEC 9126 and the Time Utilization of TTAS.KO-11.0078. We combined these terms as Behavior Utilization, which we arranged as a sub-element of Efficiency in our standard.

We added a new extensibility element as a sub-element of Efficiency in our standard. This element considers whether the software allows users to change the language settings of the game. An extensibility element is required for the following reasons. Some types of education are not limited to a specific country; for example, there is an

international demand for environmental education, math, and science. Educational serious games that can be played in different languages may be easily extended to different countries. Consequently, they accomplish the effect of One Source Multi-Use in a wide sense. Accordingly, we expect an increase in the game's global marketability as well as savings in development costs due to content reusability.

Additionally, we incorporated all of sub-elements of Service in the TTAS.KO-11.0078 standard by re-labeling them as Serviceability and placing them in our model as sub-elements of Usability.

### C. Definition of the Quality Evaluation Elements

The proposed standard in this paper contains 6 main elements and the 19 sub-elements derived from the results of previous mapping of ISO/IEC 9126 and TTAS.KO-11.0078. These elements are intended to be used in the evaluation of technical quality of educational serious games. We used the original definitions of elements extracted from the existing standards. They are as follows.

- 1) **Serviceability:** the product's ability to provide additional services, including answers to user questions and resolution of inconveniences
- 2) **Behavior Utilization:** the product's ability to handle a specific function (without problems) within a limited time
- 3) **Extensibility:** the product's ability to extend to various language versions easily

Elements of the non-technical evaluation areas should reflect appropriate characteristics of educational serious games. Therefore, to establish these elements, we referred to the learning theory of constructivism, international e-learning quality standards, and the curriculum of the Korean Ministry of Education, Science and Technology.

As the result, we determined the 3 main elements and 12 sub-elements of our standard. The elements of non-technical quality evaluation are defined as follows.

- 4) **Content Property:** used to evaluate the quality of multimedia factors and content factors of the educational serious game

**Fun:** how many fun (i.e., entertainment) factors are included in the educational serious game; for the purpose of encouraging repeated play

**Creativity:** whether the scenario, characters, and objects in the game are unique or innovative

**Harmony:** whether the game contents are appropriate for the game goals

- 5) **Special Purpose Property:** used to evaluate the quality of common game elements (e.g., example estimation, control, and ethics)

**Goal Feasibility:** the user's satisfaction level, goal achievement level, and learning effectiveness persistence after playing the game

**Ethicality:** whether the game material and contents do not pose moral or ethical problems

**Interactivity:** whether the functions for bidirectional interaction in the game are provided among users or between the user and system

**Adjustability:** whether the game has functions to prevent user's flow controllable by the supervisor or system.

**Check Study:** whether the game has functions to check the effects of learning or training.

- 6) **Educational Property:** used to evaluate the quality of elements associated with pedagogy

**Clarity of Targets and Content of Education:** whether the game conforms exactly with the standards set by the Korean Ministry of Education, Science and Technology for purpose, contents of education, and learners' selection criteria

**Difficulty Conformance:** how the game's level of difficulty and amount of learning compare to the Korean Ministry of Education, Science and Technology standard

**Self-Directedness:** whether the game contains the self-directed elements that encourage experienced users to play again

**Problem Solving/Inquiry Learning:** whether the game offers to variety of problem-solving methods and inquiry functions for users to solve a specific problem.

Accordingly, we proposed the first draft of a standard for the quality evaluation of educational serious games. The draft standard was composed of 9 main elements and 31 sub-elements.

#### IV. DEVELOPMENT OF THE QUALITY EVALUATION STANDARD MODEL

Before determining the final quality evaluation standard for educational serious games, we first verified the need for each element in the proposed a draft standard. To this end, we carried out a survey on the need for each element. Participants were asked to rate each element on a five-point Likert scale as follows: 1, very unnecessary; 2, unnecessary; 3, normal; 4, necessary; and 5, very necessary.

The survey participants were divided into an expert group and general users group. The expert group included instructors, game developers, and teaching-learning designers who had at least three years of experience in their field.

**Table 3.** Statistics on the technical quality evaluation elements

Item	N	Mean	Standard error	Standard deviation
Accuracy	12	4.67	0.142	0.492
Interoperability	12	4.42	0.229	0.793
Security	12	4.00	0.302	1.044
Behavior Utilization	12	4.42	0.193	0.669
Resource Utilization	12	4.00	0.213	0.739
Analyzability	12	3.67	0.256	0.888
Testability	12	3.58	0.229	0.793
Changeability	12	4.33	0.142	0.492
Stability	12	4.50	0.195	0.674
Extensibility	12	3.50	0.359	1.243
Adaptability	12	4.08	0.260	0.900
Installability	12	4.25	0.305	1.055
Co-existence	12	4.17	0.207	0.718
Fault Tolerance	12	4.00	0.174	0.603
Recoverability	12	4.17	0.241	0.835
Learnability	12	4.42	0.149	0.515
Operability	12	4.67	0.188	0.651
Attractiveness	12	4.42	0.229	0.793
Serviceability	12	3.92	0.260	0.900

The expert group was surveyed on all elements of the draft standard, including both technical and non-technical elements. They were also provided space at the end of the survey form to suggest additional elements for the quality evaluation.

The general users group was surveyed only on the non-technical elements, as we judged that they might have difficulty distinguishing specialized technical factors.

We used the SPSS statistical software program to analyze the survey data and calculate descriptive statistics, in order to determine which quality evaluation elements should be included in the final standard. Table 3 shows the descriptive statistics of the expert group responses to technical quality evaluation elements. As can be seen in Table 3, the mean statistics representing the need for technical elements are all higher than 3.50. This means that all technical quality evaluation elements of the draft standard were accepted into the final standard. No elements were eliminated. Table 4 lists the descriptive statistics of the expert group and general users' group responses to the non-technical quality evaluation elements.

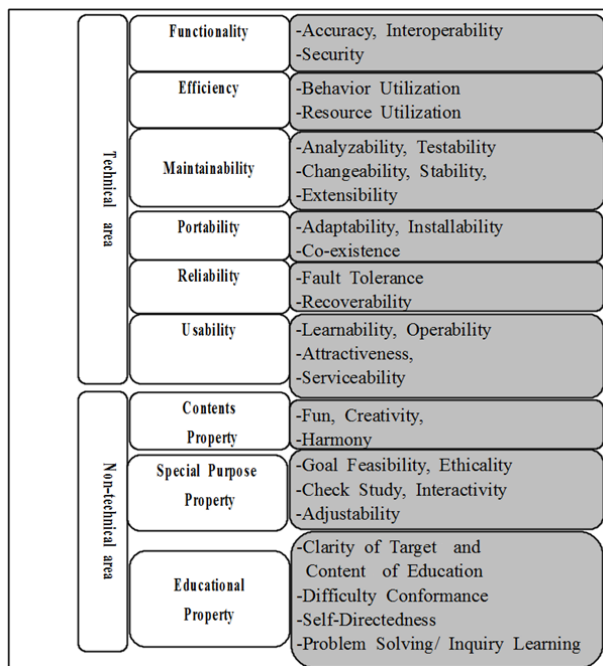
As Table 4 shows, the mean statistics representing the need for non-technical elements are all higher than 3.75 in the expert group and higher than 3.67 in the general users group. These results indicate all the proposed non-technical elements are required; therefore, they were all accepted into the final quality evaluation standard.

Consequently, the proposed final quality standard is composed of 9 main elements and 31 sub-elements. Some members of the expert group suggested additional quality

**Table 4.** Statistics on technical quality evaluation elements

Item	Expert group				General users' group			
	N	Mean	SE	SD	N	Mean	SE	SD
Fun	12	4.67	0.188	0.651	110	4.24	0.073	0.765
Creativity	12	4.50	0.151	0.522	110	4.02	0.082	0.857
Harmony	12	4.50	0.151	0.522	110	4.13	0.077	0.803
Goal Feasibility	12	4.33	0.284	0.985	110	3.99	0.098	1.027
Ethicality	12	4.58	0.193	0.669	110	3.85	0.095	0.997
Interactivity	12	4.25	0.179	0.622	110	3.95	0.076	0.799
Adjustability	12	3.75	0.351	1.215	110	3.68	0.098	1.031
Check Study	12	4.58	0.149	0.515	110	3.84	0.093	0.972
Clarity of Target and Content of Education	12	3.92	0.229	0.793	110	3.67	0.099	1.041
Difficulty Conformance	12	4.00	0.275	0.953	110	3.77	0.090	0.945
Self-Directedness	12	4.58	0.193	0.669	110	3.98	0.075	0.790
Problem Solving/Inquiry Learning	12	4.33	0.188	0.651	110	4.06	0.077	0.805

elements. First, they proposed generalization and metastatic elements to evaluate whether a user is able to apply the contents of learning to daily life after playing the game. Second, they proposed a satisfaction element for product price, and lastly, an automatic control function element to evaluate whether the game provides a function to automatically prevent excess use of inappropriate language in the user chat system. We will consider these suggested elements in greater detail in a later work. In this paper, we propose elements of the final quality evaluation standard, as shown in Fig. 1.

**Fig. 1.** Quality evaluation standards for educational serious games.

The most common method of verifying questionnaire reliability is to calculate internal consistency and the Cronbach's alpha coefficient. In general, a Cronbach's alpha value of greater than 0.6 indicates an acceptable level, and greater than 0.7 indicates the questionnaire is reliable [17-19].

We used Cronbach's alpha coefficient to confirm the reliability of our survey results. Here, we omitted the verification process for reliability of the technical evaluation elements. Because most of the technical elements were extracted from existing standards, they had already been verified in the development process of the original standards. Therefore, in this study, we only verified reliability of the non-technical elements.

The content property, special purpose property, and educational property were constructed as multi-item properties. Table 5 presents the results of analysis of reliability for each of the constructed items.

**Table 5.** Results of analysis of reliability

Construct	Item	N	Cronbach's $\alpha$ coefficient
Content Property	Fun	122	0.752
	Creativity		
	Harmony		
Special Purpose Property	Goal Feasibility	122	0.632
	Ethicality		
	Interactivity		
	Adjustability		
	Check Study		
Educational Property	Clarity of Target and Content of Education	122	0.760
	Difficulty Conformance		
	Self-Directedness		
	Problem Solving/Inquiry Learning		
	Problem Solving/Inquiry Learning		

**Table 6.** Metric of Fault Tolerance (© ISO/IEC 2013. All rights reserved) [11]

Main elements	Sub-elements	Metric of detailed evaluation	Formula	Interpretation of measured value
Reliability	Fault Tolerance	1. Count the number of breakdowns that occur with respect to the number of failures.	A Number of breakdowns	$0 \leq X \leq 1$ The closer to 1.0, the better
			B Number of failures	

According to the above analysis, the Cronbach's alpha coefficient of the special purpose property is 0.632, and the Cronbach's alpha coefficients of the remaining constructs are greater than 0.7. Therefore, all items were proven reliable.

## V. METRICS OF THE QUALITY EVALUATION

In the quality assessment of software products, the evaluator's subjective view often has a greater influence on the overall evaluation than objective considerations. To reduce this, we must design the quantitative evaluation metrics in such a way as to achieve objective and reliable

product evaluations. With this goal in mind, we establish detailed evaluation criteria of individual elements of the proposed standard. With the exception of the proposed elements in this paper, we adopt, intact, the metrics of the original standards as the evaluation criteria of all elements in our standard. Table 6 shows an example of an ISO/IEC 9126 metric for Fault Tolerance.

For the creation of metrics for both technical and non-technical elements proposed in this paper, we employ an alphabetical grading system. Each metric contains two to five items used to assess the quality of an element. Each element is assigned one of three grades in the overall evaluation: A (30 points), B (20 points), or C (10 points). Table 7 presents the metric of Extensibility as defined in this

**Table 7.** Metric of extensibility

Main elements	Sub-elements	Metric of detailed evaluation	Rating scale	Measurement method
Maintainability	Extensibility	1. The game has an extensible function to allow play in a different language. 2. The game can support three or more different languages.	<i>Grade A:</i> The game meets both criteria <i>Grade B:</i> The game meets one of the criteria <i>Grade C:</i> The game does not meet either criterion	-Measurement by Expert evaluators group - Results of simulated use of the program - Checklist

**Table 8.** Metrics of non-technical elements (some cases)

Main elements	Sub-elements	Metrics of the detailed evaluation	Rating scale	Measurement method
Special Purpose Property	Check Study	1. The game has the ability to assess learning achievement. 2. The game has the ability to check the effects of learning through feedback. 3. The game has the ability to give a reward or punishment for achievement of learning. 4. Rewards and punishments in the game are appropriate. 5. The game has the ability to give various rewards or punishments.	<i>Grade A:</i> The game meets all five criteria. <i>Grade B:</i> The game meets three of the criteria. <i>Grade C:</i> The game meets two or fewer criteria.	-Measurement by expert group -Measurement by general users group
Educational Property	Clarity of Target and Content of Education	1. The purposes of education in the game are clear. 2. Target users/learners in the game following the selection criteria presented by the Korean Ministry of Education, Science and Technology. 3. The game conforms to the aim of learning established by the Korean Ministry of Education, Science and Technology.	<i>Grade A:</i> Three items in the game meet all. <i>Grade B:</i> Two items in the game meet all. <i>Grade C:</i> Only one item in the game meets all.	-Measurement by educationalist group * Using checklist

paper. Table 8 shows a sample of a metric for educational elements. This metric is used to evaluate the quality of non-technical game components.

For a metric of evaluation for each sub-element of the main elements, we computed the average value of the results of evaluation.

We propose to express information on the quality of evaluation results for the main elements as follows. If an average value is 27 points or higher (i.e., at least 90% of the full 30 points), the element will receive a grade of A. A score of 24 to 26 points (i.e., 80%) indicates a B, and other point corresponds to a C.

## VI. CONCLUSIONS

Educational serious games currently occupy the largest portion of the serious game market. Therefore, we proposed quality evaluation standards to objectively and reliably evaluate the quality of educational serious games. In this paper, the proposed quality standards were designed to evaluate all aspects of the quality of both technical and non-technical elements of educational serious game software. We presented a development framework for quality standards and proceeded to develop the quality evaluation elements in accordance with our proposed procedure. We then added these developed elements into our standard. As a result, we have proposed, for the first time, draft standards for the quality evaluation of educational serious games. The standards are composed of 9 main elements and 31 sub-elements.

To decide which elements would be used for the final quality standard, we conducted a survey about the need for elements of our draft standard, targeting experts and general users. We determined the necessary elements based on descriptive statistics of the survey results.

Further, we performed a reliability analysis of the survey results to ensure the reliability of the questionnaire. The analysis showed no unnecessary or low reliability evaluation elements in the draft standard. Thus, we accepted intact all the elements outlined in the draft standards as elements of the final quality evaluation standard.

Quantitative and specific metrics are needed to objectively and reliably evaluate the quality of educational serious games products. Therefore, we developed the detailed metrics and presented examples in this paper. One of the main goals of our research is to supply information about the quantitative and quality evaluation of educational serious games to users, in order to aid their selection of educational serious games. Consequently, we expect that users will be able to utilize this information in order to easily and conveniently select high-quality educational serious games.

Our second goal is to provide game developers with quality standards for developing high-quality educational serious games. Accordingly, we anticipate that game developers will be able to produce competitive and user-favored, high-quality products using the information presented here.

In future research, we will carry out the remaining two steps of our framework and attempt to establish a complete quality standard. Additionally, we will devote more effort to expand our standard toward an international standard of educational serious games.

## ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. 2012-0003159).

## REFERENCES

- [1] D. Michael and S. Chen, *Serious Games: Games That Educate, Train and Inform*. Boston, MA: Thomson Course Technology, 2006.
- [2] Korea Creative Contents Agency, Serious game statues and demand survey [Internet], Available: [http://seriousgame.kocca.kr/info/report/1200861\\_3212.html](http://seriousgame.kocca.kr/info/report/1200861_3212.html).
- [3] X. Chen, "Influence from the serious game on mobile game developers' commercial strategies," in *Proceedings of the International Seminar on Business and Information Management*, Wuhan, China, pp. 207-209, 2008.
- [4] M. Zyda, "From visual simulation to virtual reality to games," *IEEE Computer*, vol. 38, no. 9, pp. 25-32, 2005.
- [5] IEEE, IEEE Standard for software quality assurance plans [Internet], Available: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1040117>.
- [6] D. E. Lowe and G. M. Cox, "Implementing the capability maturity model for software development," *Hewlett-Packard Journal*, vol. 47, no. 4, pp. 1-6, 1996.
- [7] Information Technology - Process Assessment. Part 1: Concept and vocabulary, ISO/IEC 15504-1:2004.
- [8] Information technology - Learning, Education and Training- Quality Management, Assurance and Metrics. Part 1: General approach, ISO/IEC 19796-1:2005.
- [9] Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE). System and software quality models, ISO/IEC 25010:2011.
- [10] Software engineering - Product quality. Part 2: External metrics, ISO/IEC TR 9126-2:2003.
- [11] Software engineering - Product quality. Part 3: Internal metrics, ISO/IEC TR 9126-3:2003.
- [12] Software engineering - Software product Quality Requirements



and Evaluation (SQuaRE). Requirements for quality of Commercial Off-The-Shelf (COTS) software product and instructions for testing, ISO/IEC 25051:2006.

- [13] Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE). Evaluation process, ISO/IEC 25040:2011.
- [14] Software Engineering - Software product Quality Requirements and Evaluation (SQuaRE). Guide to SQuaRE, ISO/IEC 25000:2005.
- [15] Telecommunications Technology Association, TTAS.KO-11.0078: the quality model for mobile game software [Internet], Available: [http://www.tta.or.kr/data/ttas\\_view.jsp?rn=1&pk\\_num=TTAS.KO-](http://www.tta.or.kr/data/ttas_view.jsp?rn=1&pk_num=TTAS.KO-11.0078)

11.0078.

- [16] Telecommunications Technology Association, TTAS.KO-11.0059: quality evaluation guideline for Web-based software [Internet], Available: [http://www.tta.or.kr/data/ttas\\_view.jsp?rn=1&pk\\_num=TTAS.KO-11.0059](http://www.tta.or.kr/data/ttas_view.jsp?rn=1&pk_num=TTAS.KO-11.0059).
- [17] G. A. Churchill and J. P. Peter, "Research design effects on the reliability of rating scales: a meta-analysis," *Journal of Marketing Research*, vol. 21, no. 4, pp. 360-375, 1984.
- [18] J. C. Nunnally, *Psychometric Theory*, 2nd ed. New York, NY: McGraw-Hill, 1978.
- [19] A. H. Van de Ven and D. L. Ferry, *Measuring and Assessing Organization*. New York, NY: John Wiley & Sons Inc., 1980.



**Seon-Jeong Yoon**

is a Professor at Division of Digital Contents of Dongseo University in Busan, Republic of Korea. She worked for Sammi Information System Co. She received B.S. degree 1983, M.S. degree in 1987 from Kyungsoo University, Busan, Republic of Korea, and Ph.D. degree from Dongseo University. She is interested in serious game design, mobile games, and game databases.



**Hee-Sook Park**

received the M.S. degree in the graduate school of Education of the University of Kyungnam in 1998, Republic of Korea and the Ph.D. degree in the Department of Computer Engineering of Pukyong National University in 2006, Republic of Korea. She is a part-time instructor at the Department of Computer Engineering of Pukyong National University. Her research interests include database performance, RFID technology, and quality assessment of serious games.