

# Evaluating the Effectiveness of Serious Games for Cultural Awareness: The Icura User Study

Michela Mortara<sup>1</sup>(✉), Chiara Eva Catalano<sup>1</sup>, Giusy Fiucci<sup>2</sup>,  
and Michael Derntl<sup>3</sup>

<sup>1</sup> CNR IMATI-Ge, Via de Marini 6, 16149 Genoa, Italy  
{michela.mortara, chiara.catalano}@ge.imati.cnr.it

<sup>2</sup> ORT-France, 16, villa d'Eylau, 75116 Paris, France  
giusy.fiucci@ort.asso.fr

<sup>3</sup> Advanced Community Information Systems (ACIS),  
RWTH Aachen University, Informatik 5, Ahornstr. 55, 52056 Aachen, Germany  
derntl@dbis.rwth-aachen.de

**Abstract.** There is an increasing awareness about the potential of serious games for education and training in many disciplines. However, research still witnesses a lack of methodologies, guidelines and best practices on how to develop effective serious games and how to integrate them in the actual learning and training processes. This process of integration heavily depends on providing and spreading evidence of the effectiveness of serious games. This paper reports a user study to evaluate the effectiveness of *Icura*, a serious game about Japanese culture and etiquette. The evaluation methodology extends the set of instruments used in previous studies by evaluating the effects of the game on raising awareness, by avoiding the selective attention bias and by assessing the medium-term retention. With this research we aim to provide a handy toolkit for evaluating the effectiveness a serious games for cultural awareness and heritage.

## 1 Introduction

There is an increasing awareness about the potential of serious games (SGs) for education and training in many disciplines such as healthcare, military/defense, politics, corporate applications and industry. However, much has still to be done to integrate serious games in the actual learning and training processes. Such integration heavily depends on revealing, providing and spreading evidence of their effectiveness, which is one of the main goals of the European network of excellence GALA [1]: the project explores different application domains in several special interest groups, including humanities and heritage [2].

Cultural content is very diverse: on the one side, there is the physical, or “tangible” cultural heritage such as historic sites and buildings, monuments, documents, works of art, machines, and even the natural environment which represents the setting where a society exists (or existed) and which influences its evolution and customs. On the other side, there are many further factors which deeply characterize a culture and have a non-physical nature, namely the “intangible” cultural heritage.<sup>1</sup> These factors include social

<sup>1</sup> <http://www.unesco.org/culture/ich/doc/src/01851-EN.pdf>

values and traditions, customs and practices, philosophical values and religious beliefs, artistic expression, language, folklore and rules of behavior in a society, without forgetting the influence of past events on that society. Intangible heritage is particularly difficult to preserve, and we believe serious games have the potential to maintain and communicate effectively especially this immaterial legacy. In fact, they are able to recreate accurately not only a physical setting but rather provide a comprehensive experience including spoken language, traditional music, and aesthetic elements.

An extensive and up-to-date survey on the serious games proposition in the cultural sector is given in [3], which highlights the educational objectives of games, analyzes the complex relations between genre, context of use, technological solutions and learning effectiveness, and finally identifies and discusses the most significant challenges in the design and adoption of educational games in cultural heritage. Above all, the main lack is a formal and methodological assessment of the effectiveness of serious games as learning tools with respect to other traditional means. In fact, the most common approach to analyze the effectiveness of SGs in the humanities and heritage field (but not limited to this domain) is a simple test after the game session (e.g. Playing History [4]) or the comparison of pre and post test results to measure learning effects. The latter approach has been used in a former evaluation of *Icura* [5], *Travel in Europe* [6], *ThIATRO* [7, 8] to give a few examples. In some cases, users are split into a group of people who actually play the game (experiential group), and a control group, who learns the same educational content by other means, typically an oral presentation. Users are randomly assigned to avoid allocation biases. This approach allows not only to estimate the game efficacy per se, but also to compare the game learning impact with other traditional teaching methods. Concerning the content of the tests, they typically include direct questions related to the learning content provided by the game.

An interesting issue related to the pre- and post-tests is the so called “selective attention” phenomenon: in the case of *Icura*, the authors in [5] argue that since the pre- and post-tests comprise the same twelve questions, it is possible that “the player may remember those questions and, wittingly or unwittingly, pay more attention to find the correct answers during gameplay”. They suggest to change the question ordering or to design a pre-test which comprises several questions that are not actually answered during the game.

The game evaluations often also cover the aspect of user satisfaction, with a set of questions in the post-test focused on the game usability and the user engagement and fun, whereas a deeper analysis of user experience emerge only in those few cases where particular I/O devices are involved (for instance, [9]).

The post-test is always done straight after the experience. According to [10] “games have the power to create this kind of long-term knowledge by connecting the learning content with meaningful actions”. However, none of the state-of-the-art serious games evaluated the impact on long-term knowledge.

The educational objectives clearly affect the set-up of a proper evaluation methodology. Specifically, games for humanities and heritage often aim at impact the affective sphere of players beyond communicating knowledge but comprehensive studies including the affective impact are still lacking; the specific impact with respect to different levels of cognitive and affective gain (e.g. following Bloom’s taxonomy [11]) should be devised starting from the specific educational goals of serious games.

In the perspective of a complete assessment, this paper contributes with a practical user study: we selected *Icura* as an interesting representative of serious games for cultural awareness, and based our study on the assumptions mentioned above, trying to overcome some limitations of the original evaluation. In the next section, we briefly describe the game features, while major details can be found in [5, 10]. In Sect. 3, the experimental setting is described, while Sect. 4 discusses the experimental results. Finally, Sect. 5 concludes the paper.

## 2 About *Icura*

*Icura* is an adventure game released in 2010 by the Electronic Commerce group at Vienna University of Technology. The player embodies an Austrian tourist arriving in Japan for the first time. He has been in contact with his Japanese host through the couch-surfing network and he is supposed to stay at his house. Unfortunately, he did not manage to get his precise address and after his arrival has to find out where exactly his host lives. He has got only a print-out of his e-mail with a few information about the Japanese culture, and then the player has to find his host accomplishing several sub-tasks.

The game takes place in a 3D virtual world representing a fictive isle in Japan. The environment is realistic, as buildings, temples, aesthetic styles and natural landscape are Japanese-style, and a soft traditional music plays in the background. Playing the game and exploring the environment, the player learns about Japanese culture and etiquette, which can either raise cultural interest or support a real pre-trip planning. To make a few examples, the player has to learn about *tsutsumi*, the art of gift-wrapping; to manage the right salutation for each kind of person; or to follow the adequate behavior in a syncretic temple.

From the technological point of view, the game was developed using the Torque Game Engine Advanced, an open source system offering support to the creation of 3D models, graphics, sound and light effects for games and simulations. It runs on Windows with no special system requirements. It is single player with no social mechanisms; the learning curve is quick and the player is typically operative in less than 5 minutes. The effective learning time for a first-time player is typically less than one hour. The player engagement is maintained by eye-catching graphics, clear sub-goals, the final score sheet, and the Information Agent (IA), a kind of virtual tutor who provides information, hints and feedback avoiding the player to get stuck in the game. The target audience of the game is the general public, there are no knowledge prerequisites to playing the game.

From the pedagogical point of view, the adventure genre and the 3D setting are suitable for implementing the constructivist and the learning-by-doing models: the 3D immersive environment represents a tangible learning context where the player can actively build his/her own knowledge through exploration, manipulation and interaction. The learning content is organized into small units related to sub-tasks, from simple to complex. In order to fulfill tasks the player has to interact with non-player characters (NPCs), gaining new knowledge and applying it to solve puzzles and advance in the

game. Some subtasks require to combine objects or use objects in the scene (e.g. use the steam from a teapot to take a sticker off a wallet).

The game is complemented by a pre- and a post-test: 12 multiple choice quiz questions about Japanese culture are asked before and after the game session. The entry test is a measure to determine the player's level of preliminary knowledge but does not serve to personalize automatically subsequent gameplay. After the session, the player receives a final score and a short summary displaying the wrong answers from the post-test and the time taken to complete the game. Both the final summary and the feedback provided by the Information Agent during the game follow the reflection principle: Sheng et al. [12] point out the importance of including an opportunity for the players to stop and think about what they are learning.

A first evaluation of the game effectiveness and of the user satisfaction with the game have been reported by the developers themselves in [5]. *Icura* has been evaluated at Vienna University of Technology involving 20 people aged between 21 and 43 years, mostly expert gamers. The evaluation was carried out in five steps: a pre-questionnaire collecting demographic data and user proficiency with computer games; the pre-test assessing the starting level of knowledge about Japanese culture; the game session; the post-test with the same questions as the pre-test to assess improvement of knowledge about Japanese culture; a post-questionnaire to determine the overall satisfaction with the game. They found that, on average, 5.05 correct answers were given in the pre-test against an average of 10 in the post-test, concluding that *Icura* successfully communicates information about Japanese culture and etiquette.

### 3 Experimental Setting

In the present study we aimed to go beyond the previously existing evaluation with respect to three issues. First, we wanted to address a higher number and higher diversity of study participants and avoid the selective attention bias. Second, we wanted to evaluate to what extent *Icura* is capable of transmitting higher level knowledge rather than just information and to actually raise interest in Japanese culture at affective level. Finally, we also mention preliminary results about the assessment of medium-term retention.

#### 3.1 Target Population

In order to increase the number of *Icura* players, we took advantage of the GALA network to attract a wider group of participants. However, since the game is designed for the public at large, we also paid attention to include players beyond the university context, but who can identify themselves with the main character (i.e., a young man travelling by couch-surfing). We were able to attract 61 volunteer players to participate in this evaluation study: 27 from CNR-IMATI Genova, 17 from RWTH Aachen University, and 17 from ORT FRANCE, Paris. One person did not complete the tasks assigned.

### 3.2 Design of the Evaluation Procedure

As the developers of *Icura* pointed out in [5], the pre- and post-tests included the same questions, and they tried to counteract the selective attention bias by presenting them in a different order in the post-test, which was useful to some extent but could not completely exclude a bias. As an alternative, they also suggested including many irrelevant questions in the pre-test (questions which will not get an answer in the game) so that the player is unable to focus his/her attention on searching for specific answers [10]. Instead, our strategy was to use different questions in the pre-test. Therefore, new pre-test questions were designed just to identify each participant's average knowledge of the Japanese culture and have the same complexity of those of the post-test. We are aware of losing the chance of directly comparing the correct answers from the pre- and post-tests; however, a relevant increase has been already proved by [5]. We argue that this approach is reasonable because our aim was not just assessing a quantitative measure of the increase of knowledge but also to evaluate whether higher-order knowledge was transferred and whether the game was successful in raising awareness about the Japanese culture.

To achieve this aim, we defined questions with respect to the expected educational goals with reference to the different levels of the Bloom's taxonomy [11], which is a classification of learning objectives organized into three "domains": Cognitive, Affective, and Psychomotor. Each domain includes different ordered knowledge and skills levels. The learning objectives of *Icura* can be recognized in the Cognitive and Affective domains. The levels in the Cognitive domain are: Remembering, Understanding, Applying, Analyzing, Synthesizing and Evaluating. Accordingly, the expected learning outcomes are that the learner simply remembers data, or understands information, or, at higher levels, is able to apply the new knowledge, decomposes concepts to understand their structure and relations, builds a structure or pattern from diverse elements and eventually makes judgments about the value of ideas or materials. Categories in the Affective domain are: Receiving phenomena, Responding to phenomena, Valuing, Organizing and Internalizing values. Students reflect such levels of affection when: they show interest in the topic being taught; actively participate and show motivation; attach values to the educational messages, organize these values solving conflicts among contrasting ones into a unique value system; and finally internalize these values, behaving accordingly.

We matched the educational objectives of the game against the Bloom's taxonomy and investigated proper ways to assess the learning impact with respect to the specific levels of cognitive/affective gain each objective refers to. In Table 1 an excerpt of our approach is shown. In collaboration with the game developers, we mapped each intended learning goal with a set of expected outcomes, each one related to a specific level in the Bloom's taxonomy. The assessment of each expected learning outcome is made accordingly: multiple choices questions for outcomes related to Cognitive/Remembering level; open answers about outcomes related to higher Cognitive levels; observation during the whole session, interviews and open answers for outcomes related to the Affective domain.

**Table 1.** Educational objectives of *Icura* matched against Bloom's taxonomy.

Learning goal	Type of learning	Expected learning outcome	How to assess
Learn about the Japanese culture and etiquette	Cognitive (Remembering)	The learner remembers the Japanese word for "gift-wrapping"	<i>Multiple-choice (MC) Question: Who or what is "Tsutsumi"?</i>
		The learner remembers the Japanese word for say "hello"	<i>MC Question: How do you say "Hello" in Japanese?</i>
		The learner recalls the most popular religions in Japan	<i>MC Question: Which religions are prevalent in Japan?</i>
	Cognitive (Understanding)	The learner understands that saying "no" is impolite	<i>MC Question: What do you have to keep in mind when talking to a Japanese man or woman?</i>
		The learner understands that keep off shoes is a sign of respect	<i>MC Question: When you are invited as a guest in Japan, what do you have to keep in mind?</i>
	Cognitive (Applying)	The learner is able to use the correct salutation	<i>Accomplishment of the game task and MC Question: You are going to talk to a senior person named Shotaro and you want to show high respect for him. Which salutation is best to choose in this situation?</i>
	Cognitive (Analyzing)	The learner is able to identify the main characteristics of the Japanese culture	<i>Open Answer (OA) Question: How would you describe in 3 adjectives the Japanese culture?</i>
Cognitive (Synthesizing)	The learner is able to generalize and deduce new facts about the Japanese culture	<i>OA Question: What can you deduce from the facts you learnt in the game about the Japanese culture?</i>	
Cognitive (Evaluating)	The learner is able to compare the Japanese culture against his own	<i>OA Question: How would you compare the Japanese culture to the Western?</i>	
Raise awareness about the Japanese culture	Affective (Responding to phenomena)	The learner shows new interest towards the principles of the Japanese culture and behavior, he is willing to respond and take the questionnaire, and even to find out more by himself	<i>Observation during the experience; Likert question: The game is able to motivate the user on the learning topic; OA Question: Did the game increase your interest towards Japan/Japanese culture? Would you read more about Japan? Would you visit it?</i>

*(Continued)*

**Table 1.** (Continued)

Learning goal	Type of learning	Expected learning outcome	How to assess
	Affective (Valuing)	The learner identifies positive (or negative) aspects in the Japanese culture and behavior (with respect to our own culture)	<i>OA Questions (if the respondent spontaneously assign positive/negative connotation): How would you describe in 3 adjectives the Japanese culture? What can you deduce from the facts you learnt in the game about the Japanese culture? How would you compare the Japanese culture to the Western?</i>
	Affective (Internalizing values)	He changes his mind about japan, and maybe now he wants to apply some behavior in his own life or he would like to visit Japan.	<i>OA Question in the long-term post-test: Did you take any concrete action to explore further the Japanese culture? If so, which?</i>

**3.3 Assessment of Medium-Term Retention**

The authors in [5] identified the lack of evaluation of long-term retention as the major open issue in their work. The evaluation took place right after the game session, while they expressed the wish to repeat the test after one year or later.

For the sake of this paper we repeated the post-test questionnaire after 6 months approximately since the game session, leading to a medium-term analysis of retention. We did not send out the post-test as it was but rather added new open questions to assess whether the game has really influenced the players’ perception of Japan and whether they actually put in practice what they previously declared to (e.g., whether they have recommended the game to a friend, or whether they have looked for more information about Japan). We were able to collect 16 answers so far.

**4 Experiment**

Different evaluation sessions were conducted in Genova, Paris and Aachen in the period from December 2012 to April 2013. When all players had the game installed on their computers, the instructor briefly introduced the game controls, i.e. how to move the player character, what actions the mouse clicks perform, and similar hints that do not reveal anything about the objective or solution to the game. Players played on their own or in couples, depending on the number of the available computers. In the following, we report about the demographics of the participants and we summarize the results coming from the different phases of the experiment: the multiple-choice questionnaires before and after the gaming session (pre- and post-tests), the questionnaire with Likert scaling about game analysis, and the open questions about the affective impact of the game.

## 4.1 Demographics

In total 61 volunteers participated in the evaluation sessions. The participant sample consisted in 24 university students (39 %), 11 teachers (18 %), 15 researchers (25 %) and 11 individuals with other profiles (18 %), ageing from 21 to 67 (mean = 32,8). 19 players were female (31 %) and 42 were male. About their gaming expertise, 15 % declared not to play any digital games; 54 % consider themselves casual gamers; 26 % play games regularly and only 5 % are hard-core gamers.

The mean level of familiarity with Japanese culture was rated at 1.88 on a Likert scale from 1 (poor) to 5 (excellent), indicating rather poor familiarity on average. The level of interest in Japanese culture was rated slightly higher (mean = 2.95) on the same scale. Two participants were Japanese.

## 4.2 Pre and Post Tests

Considering the 60 users who completed both pre- and post-tests, we can report an average of 7.36 out of 13 correct answers in the pre-test and 10.81 out of 13 on the post-test, with an average increase in the number of correct answers from the pre- to the post-test of 3.63 (see the histogram in Fig. 1). However, such data cannot reasonably be directly correlated to learning gain since the questions in the two tests were different.

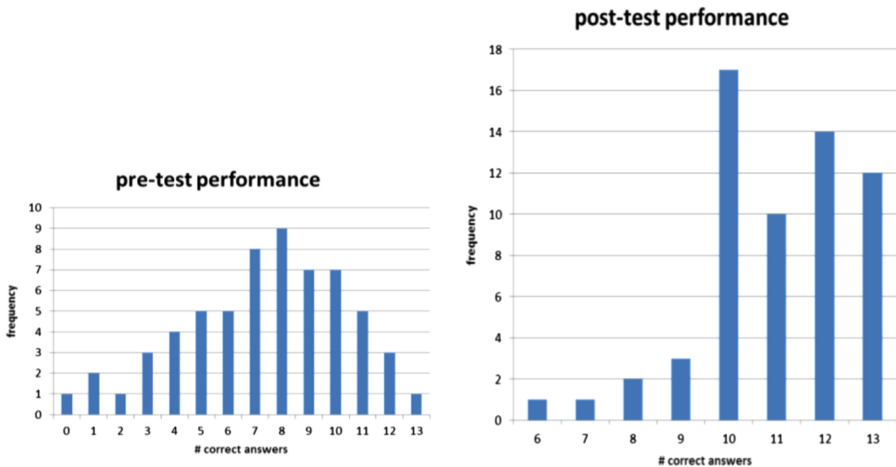


Fig. 1. Pre-test and Post-test performance

As expected, there was a positive correlation between the ratings of familiarity with Japanese culture and the number of correct items scored in pre-game test. However we note how the higher increase was achieved by students with the lowest previous knowledge or interest in Japan: 15 players answered correctly to less than 6 out of 13 questions in the pre-test (average of 3.6 correct answers). In the post-test, those users reached an average of 9.73 correct answers (average increase of 6.13). No participant scored worse in the post-test than in the pre-test. In the light of the rather moderate

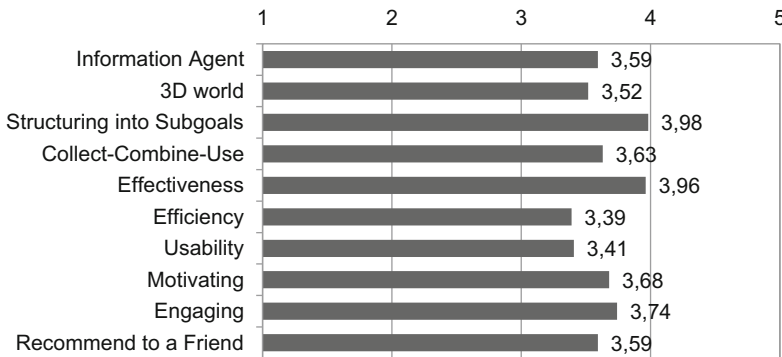


interest in Japanese culture as obtained in the pre-game survey, this is an encouraging result. Therefore, we can confidently support conclusions in [5, 10] about *Icura*'s educational potential.

Observing the participants during the game sessions, the instructors noticed high interest and motivation, the participants remained very focused even in those cases when they played in pairs and evidently had fun: some mentioned that it was fun because of the resemblance of the experience with “Monkey Island” (cited in the game dialogues). They were also motivated on the learning topic: for instance, students asked to know the correct answers to the pre-test.

### 4.3 Game Analysis Survey

The game analysis survey administered after the playing session focused on two aspects (see also Fig. 2): first, the factors that, according to the players, contributed most to determine the learning impact (“Rate how much you feel the following components helped you learning the educational content of the game”); second, the evaluation on a five point scale of the dimensions: effectiveness, efficacy, usability, ability of the game to motivate on the learning topic, and engagement. Finally, we asked if the respondent would recommend the game to a friend.



**Fig. 2.** Average ratings of game aspects in the game analysis survey on a Likert scale from 1 = poor to 5 = excellent.

About the game components, we found that the Information Agent was rated 3,59 on average but with large score variation (standard deviation 1,17): users report on the one side that it is helpful in avoiding deadlocks, on the other side it interrupts the flow and the text was very small. The comments provided with the rating indicate that many participants “did not pay much attention” or felt that the instructions by the agent were not clear.

The structuring of goals into sub goals was rated most helpful (mean = 3,98), and appreciated the collect-combine-use mechanism of interaction with objects (mean = 3,62) to solve quests – they were aware they effectively learnt from that mechanism. However, carrying out the combine operation by selecting one object and then right-clicking on the other one was considered cumbersome (this is reflected in a low score in usability).

Feedback about the 3D world and the usability were highly diverse (mean = 3,51 and 3,40, respectively, standard deviation = 0,96 for both), evidently depending on users' level of expertise with playing digital games. Mid-core and expert gamers rated the 3D world very positively (e.g. "it looks nice" and "immersive and authentic"). They felt the game was quite basic with simple rules, easy and with low cognitive load. They found the tasks very easy and suggested to have a longer version, adding tasks not strictly related to educational content and including reward videos which could give more insights on the Japanese culture. Many would have preferred to listen to dialogues (and not just to read them) in order to understand the right pronunciation of Japanese terms. Interestingly, some would appreciate the serious game as a massive multiplayer online experience, which would foster a wider diffusion, and some others wish increased realism through replicas of real geographic locations.

The overall feeling of the participants who seldom play digital games (including the majority of teachers participating in the Paris session) was that the 3D did not facilitate the learning since movements were difficult. They suggested a 2D setting instead, "maybe a point and click type 2D game would even be more suitable for the learning purpose". Some teachers suggested to add more content and to have more learning objectives.

All the profiles however found not intuitive the combine action. Overall, they noticed a good balance between learning and fun and they enjoyed the game.

The effectiveness of the game to reach the learning objectives was rated very positively with the highest average rating of 3,96 and lowest standard deviation (0,82). Participants responded that the game helped them learn about Japanese culture and some Japanese words through the dialogues and interaction with the game; they became aware of the main information a person should know when first going to Japan and acquired some cultural and language skills through the game.

The efficiency was rated lower at 3.38. In the post-game survey many participants noted that the game would be improved if there were more "useful information" during the gameplay, and having "more hints" available. Some explicitly mentioned the walking around as a major source of frustration since the "walking distances [...] unnecessarily increase the time spent" with the game. During the gaming session it was evident to the instructors that many participants had troubles with understanding the game instructions and controlling the interactions in the collect-combine actions, for example, "how to wrap the present". Not surprisingly then, the usability was also rated moderately with a mean of 3.40. The motivating aspect, that is, the game capability of raising interest in the learning topic was rated fairly with an average of 3.67 as well as and the level of engagement in the gameplay (3,74). Only 9 participants would not recommend the game to a friend. Finally, as evident from Fig. 3, players mainly regard

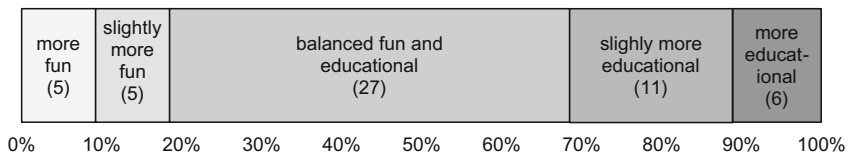


Fig. 3. Perception of balance between fun and education in the game

*Icura* as well balanced between education and fun, with 17 participants leaning more towards educational and 10 participants leaning more towards fun.

#### 4.4 Higher-Level Knowledge Assessment

From the perspective of Bloom's taxonomy, the responses to open-ended questions eventually show that the players were able not only to remember and apply new knowledge by solving the required tasks, but could also analyze and synthesize concepts about the Japanese culture: they describe it as "traditional, polite, formal" and from what they learnt in the game they deduce "it is a culture based upon respect", "ceremonial", "it is based on deeply established traditions and people are willing to preserve them". At Evaluation level, they were able to compare Japanese and western traditions that are perceived as definitely different and more flexible: "the Japanese culture seems to me more balanced and calm, while the western is more future-oriented and dynamic". At Affective level, the game reinforced the idea of a real visit in those already fascinated by Japan (one actually spent holidays in Japan after the session), while did not impact those with low interest; however, all were able to attach values to the Japanese behaviour they were exposed to ("the Japanese culture always pays attention to the preservation of the past and of traditions, while the western risks to lose the precious lessons from the previous generations"). We can conclude that *Icura* effectively stimulates higher-level knowledge and has impact at the affective level.

#### 4.5 Discussion

Generally, the level of satisfaction is related to the ability of the players in games: casual/mid-core gamers appreciated *Icura* more than non-gamers, who had to concentrate on the 3D environment and interaction first, and hard-core gamers, who found the game to be too simple.

From the conducted analysis we also get some evidence about what is the most effective mechanisms to trigger learning. In *Icura* learning is facilitated by one or more of these modalities: through information given by the Information Agent; through information given within dialogues with NPCs; through information given by textual resources in the inventory; and through educational content embedded as sub-tasks. Coupling each multiple-choice question in the post-test with the modalities the corresponding information is conveyed in the game and looking at the number of correct answers for each question, we can conclude that the Information Agent was actually the least effective means, while facts related to sub-tasks were highly successful. In fact, five questions about data given just by the Information Agent got 13, 24, 11, 31 and 4 wrong answers (over 60) respectively, while questions related to content embedded as sub-tasks rated 2 and 0 wrong answers. Other topics were treated by several means (e.g., information in the inventory and IA, IA and dialogue) had 38 wrong answers over 6 questions.

Compared to traditional means, respondents noted that a book can transmit more information in a direct, quicker way, but the game lets you experience things in a fun way, and this aspect may make the difference in the long-term retention ("it is so fun it

is impossible to forget”); interestingly, they noted that engagement makes the game the perfect tool when the learning/training has to be repeated over time. Teachers involved in the Paris session generally agreed that the game is better and more visual compared to other traditional educational tools like books and videos. They were also homogenous in thinking that the game has to be complemented with traditional education in a blended learning approach. An idea of complementing the game was also to invite some Japanese descendant into the classroom.

Concerning the medium-term retention, from the limited feedback we collected so far, we can observe that the average number of correct answers for those people only slightly decreased (from 11,5 to 10,18). Individually, three people did not change their scores and two people even increased their correct answers, while the other eleven respondents got a lower score than in the post-test. The question causing most errors was “how to say ‘yes’ and ‘no’ in Japanese” (11 errors over 16 respondents), a notion transmitted in the game by the Information Agent only. Questions about topics addressed both by the Information Agent and within dialogues witness a decrease of 3–4 correct answers. The two questions having no decrease in score are related to tasks in the game (the right behavior when entering a temple or paying a visit to a private house). Our impression is that a game can be considered comparable with traditional means when transmitting textual information, while it reaches its full potential when exploits the interaction capabilities proper of this medium.

Concerning their change of attitude, 10 respondents over 16 declared the game did increase their interest in Japan, and of those, 3 did concrete actions due to the game play (i.e., they looked for more information about Japan on their own). Almost everyone declared to have recommended the game to a friend.

## 5 Conclusion

In this paper, we presented a user study of a serious game belonging to the cultural awareness category where we gave a more extensive evaluation of the learning impact with respect to the state-of-the-art. The evaluation methodology was extended in multiple ways: we involved more players; we designed a slightly different experimental setting to overcome the selective attention bias and to assess both cognitive and affective gain; and we tackled the question of assessing medium-term effects and retention. Currently, the collection of medium-term data is ongoing; however, preliminary results are encouraging.

In the challenging perspective of conceiving formal methods to assess the learning impact of serious games, we have shown evidence of the limitations of using just pre- and post-game questionnaires; a complete analysis of game effectiveness for topics related to cultural awareness has to include the evaluation of higher level knowledge and impact at affective level, since those cognitive levels are the most peculiar ones of such applications. In fact, the main goal of *Icura* was to inform about the Japanese culture to stimulate curiosity and even a visit in the country.

Moreover, the outcome of our evaluation proved experimentally that the vantage point of serious games with respect to traditional learning means is the direct interaction of the learner/player with the learning material. It appears evident, and confirmed also by the explicit players’ opinion, that game mechanics play a crucial role: the

educational objectives should be embedded as game tasks and not simply transmitted by textual information. To understand deeply the relation between specific design choices and learning effectiveness ad hoc games have to be deployed, where different features can be turned on and off to be tested with a randomized control trial.

Finally, in the perspective of the learning evaluation, we believe that it is key to identify all the educational goals unambiguously from the early stages of the game design, both the specific target knowledge and the intangible values such as raising awareness and supporting motivation. This will drive, not only the suitable game mechanics to transmit the different levels of knowledge to the players, but also the definition of the data to be collected, of the proper analysis tools and consequently of the suitable learning metrics for the following phases of the game evaluation.

**Acknowledgments.** This research was supported by the European Commission in the Seventh Framework Programme Network of Excellence “GALA: Games and Learning Alliance” (<http://galanoe.eu>), contract no. 258169, FP7-ICT-2009.4.2. The authors thank Dieter Merkl for his support and Josef Froschauer for his collaboration in classifying *Icura*’s educational objectives and all the participants to this user study.

## References

1. EU NoE GALA – Games and Learning Alliance - contract n. 258169, FP7-ICT-2009.4.2. [www.galanoe.eu](http://www.galanoe.eu)
2. Mortara, M., Bellotti, F., Berta, R., Catalano, C.E., Fiucci, G., Houry-Panchetti, M., Petridis, P.: Serious games for cultural heritage: the GaLA activities. In: VAST 2011: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage. Prato (FI), Short and Project Papers, pp. 69–72, October 2011
3. Mortara, M., Catalano, C.E., Bellotti, F., Fiucci, G., Houry-Panchetti, M., Petridis, P.: Learning cultural heritage by serious games. Elsevier J. Cult. Heritage. doi:10.1016/j.culher.2013.04.004 (in press)
4. Playing History website. <http://www.playinghistory.eu/front>
5. Froschauer, J., Seidel, I., Gartner, M., Berger, H., Merkl, D.: Design and evaluation of a serious game for immersive cultural training. In: 16th International Conference on Virtual Systems and Multimedia (VSMM), pp. 253–260 (2010)
6. Bellotti, F., Berta, R., De Gloria, A., D’Ursi, A., Fiore, V.: A serious game model for cultural heritage. ACM J. Comput. Cult. Herit. **5**, 4 (2012)
7. Froschauer, J., Merkl, D., Arends, M., Goldfarb, D.: Art history concepts at play with ThIATRO. ACM J. Comput. Cult. Herit. **6**(2) (2013) (to appear)
8. Froschauer, J., Arends, M., Goldfarb, D., Merkl, D.: Towards an online multiplayer serious game providing a joyful experience in learning art history. In: Third International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES) (2011)
9. Seidl, M., Judmaier, P., Baker, F., Egger, U., Jax, N., Weis, C., Grubinger, M., Seid, G.: Multi-touch rocks: playing with tangible virtual heritage in the museum - first user tests. In: VAST 2011: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage. Prato (FI), Short and Project Papers, pp. 73–76, October 2011
10. Froschauer, J.: Serious Heritage Games: Playful Approaches to Address Cultural Heritage. Ph.D. Dissertation, Faculty of Informatics, Wien University of Technology, May 2012

11. Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R.: Taxonomy of educational objectives: the classification of educational goals. Handbook I. Cognitive Domain, Longmans, Green, New York (1956)
12. Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L.F., Hong, J., Nunge, E.: Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In: 3rd Symposium on Usable Privacy and Security (SOUPS '07), pp. 88–99. ACM, New York (2007)