A framework for In-process Assessment in Games and Simulations

Steve QIAN Xin and YEO Gee Kin
School of Computing
National University of Singapore
Computing 1
13 Computing Drive
Singapore 117417

Keywords: Gaming & Simulation; In-process Assessment; Evidence-centered design; Performance scoring; Game immersion.

Abstract

The increasing use of information technology in education opens up a new dimension of learning experience. Traditionally, end-of-game debriefings, out-process question-and-answer tests and stand-aside observations are conducted as means of assessments in gaming and simulation. Since committed game immersion is a positive element that engages players, assessment is preferable to be conducted in-process of any game play. Game designers are skilled in creating environments that are more engaging and educators are skilled in creating environments that foster learning. Today's demand requires that they work together to make games that are immersive learning environment. In this paper, we describe a framework to design and deliver in-process assessments of gamers. The framework follows an evidence-centered design to devise the performance scoring process which involves response processing and summary scoring. We illustrate the framework with a prototype of a simple game.

Introduction

The increasing use of information technology in education opens up a new dimension of learning experience. The use of gaming and simulation has become more widespread. Educational values in gaming are more clearly identifiable. Simulation and games make classroom teaching more fun and engaging. However, educators are still experimenting on how to integrate them into the learning process. Assessment is one of the important

components in learning. Pressing educational concerns have driven the need for improved assessment and the need for new kinds of assessment in gaming and simulation. Traditionally, end-of-game debriefings, out-process questions-and-answers tests and stand-aside observations are conducted as means of assessments. Since committed game immersion is a positive element that engages players, assessment is preferable to be carried out within the process of game play.

Advances in cognitive psychology deepen our understanding of how students gain and use knowledge (Anderson, 2004). Advances in technology make it possible to capture more complex performances in assessment settings by including, for example, simulation, interactivity, collaboration, and constructed response. Automated methods have become available for parsing complex tasks and identifying educationally-meaningful features of them (Robert J. Mislevy, Russell G. Almond, Janice F. Lukas, 2004). The challenge is in knowing just how to put all this new knowledge to work to best serve the purposes of in-process assessment.

Game designers are skilled in creating environments that are fun-filled and educators are skilled in creating environments that foster learning. Today, they are required to work together to make games that are immersive learning environment. The objectives of educational games emphasize heavily on imparting knowledge, skills and attitude. It is very important for game developers to adopt a framework that suits the purpose well. In this paper, we discuss the development of a framework to design and deliver in-process assessments of gamers that will offer flexibility to game administrators to customize. It is hoped that the insights would be helpful to others in incorporating in-process assessments in gaming and simulation in general.

In-process assessment framework - Scoring System

A scoring system maps the observations in a given task situation to certain interpretation. It contains rules for extracting bits of evidence from individual actions as observation, and a weighted score model for synthesizing the information from observation to interpretation. Interpretations are our understanding of the player for their actions in the game. They are

quality expressions that should reflect the learning objectives of the game rather than the gaming objectives. A feedback loop can also be incorporated to provide indication to the player about his/her performance or to affect the flow of game. A summary of all the interpretations can be presented at the end of the game to both the player and the game facilitator. The Scoring System agrees to the Evidence-Centered Design (Robert J. Mislevy, Russell G. Almond, Janice F. Lukas, 2003) methodology, which underscores the central role of evidentiary reasoning in assessment design.

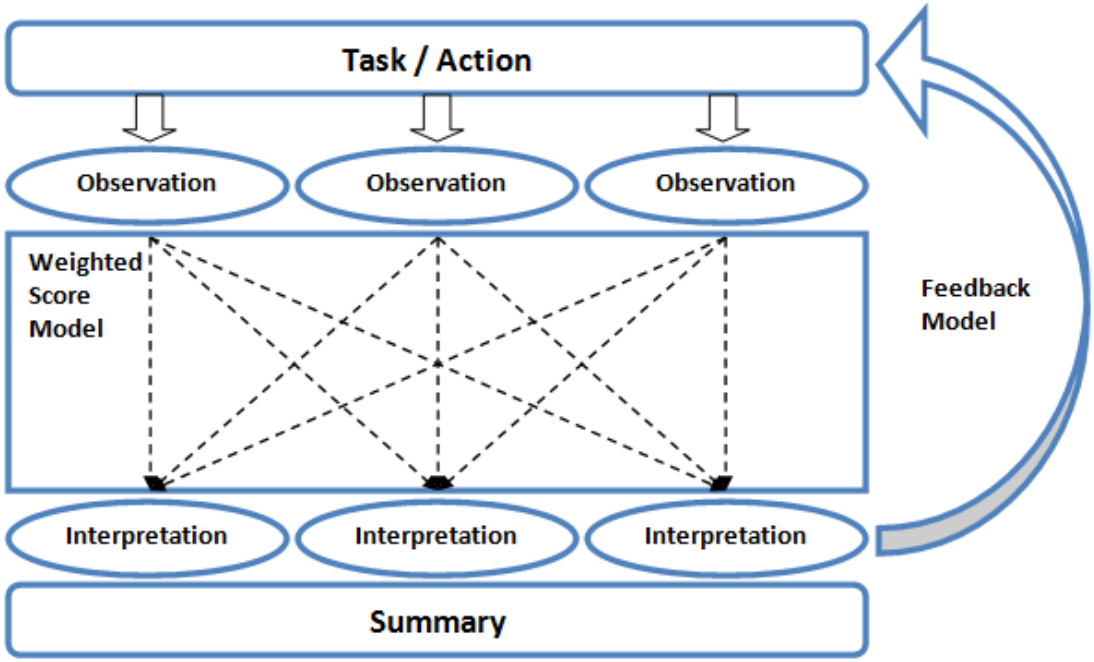

FIGURE 1 SCORING SYSTEM

Weighted Score Model

The weighted score model of the scoring system contains models specifying how the observation contribute to certain interpretations. Observations to different tasks are modeled as conditionally independent from those of other tasks. Observations within tasks may be modeled as conditionally dependent. Nonetheless, to model observations to different tasks as conditionally dependent is still possible, but it makes the model more complicated and even harder to justify. When observations are obtained for any task, they induce a likelihood function for one or more interoperations. The likelihood function is expressed as a weighted

score for simplicity and it is easy to understand by educators. Different observations from different tasks may carry different weighted score for different interpretations. The weighted score for observation that has no influence to particular interpretations is simply zero.

To describe the scoring system in detail, we will look at an example using the simplest weighted score model which could be presented in a table.

Task	Observation	Interpretation I	Interpretation II	Interpretation III	…
1	A	2	5	1	…
	B	3	2	4	…
	C	7	2	8	…
2	A	2	0	3	…
	B	9	5	7	…
	C	4	8	0	…
	D	0	3	6	…
	E	8	4	2	…
…	…	…	…	…	…
Total		235	482	562	

FIGURE 2 EXAMPLE OF TABLE FOR WEIGHTED SCORE MODEL

Generally, in any game or simulation, the player is required to perform different tasks. We label them from 1 to N in the task column in the table. But it does not mean that these tasks need to be performed in sequence. In general, these tasks can be performed in an order governed by game play at any time. As the player performs each task, we expect different observations from each action. These observations can be from different aspect of the action, for example *what* he/she does, *how* he/she does, *when* he/she does. One point to take note here is that, not performing or not doing anything can also be one of the observations. We label them A, B, C … in the observation column with respect to the task it is attached to. The number of observations may not be the same for each task.

Next we label the interpretation I, II, III … in each column in the table. A weighted score then can be assigned to each of the cell corresponding to the observation and interpretation base on cognition in the subject domain. Within one interpretation, the weighted score for one observation is in respect to the degree of importance compared to other observations. The

weighted scores between each interpretation column may not have influence to each other. Zero is assigned for case that is not important at all.

Summing up the weighted scores in each column gives the total possible score for that interpretation. On the other hand, the player may not perform every task, and some observations are not obtainable. Therefore, the player may not get exactly the total possible score. A percentage number can be used to represent the likelihood of the interpretation. It is constructed by dividing player's score obtained by the total possible score.

Using the simple evaluation rules presented in the table, the process of constructing the player's score can be done in-process. All of the evaluation is done by the program with no human intervention. The player may not need to perform additional task, and can enjoy the game as much as they can. Some might challenge the validity, accuracy of the weighted score assigned. However, we can just use the numbers as relative indications. Comparing the score or percentage score across all other players' records accumulated could give us an indication that how well the player stand relative to the others. With more advance statistical models, further study can be done to understand the particular gamer or a group of gamer with simpler profile. Furthermore, the weighted score model itself can be enhanced over time. With so much data being captured in the game, the educator could review the game play from time to time, and revise the weighted score model if needed. It is also possible that with models from the subject domain, the game can incorporate some mechanisms to automate the revision process. To the extreme, the game can learn and evolve itself to enhance the assessment or even the learning environment.

Feedback Model

Another useful attribute of in-process assessment in gaming and simulation is allowing a game to adapt to the player's behavior to give the player appropriate feedback. Players come to understand the connection between their in-game actions and the outcomes. The goal of decisional feedback mechanisms in games is to provide the player with the information they require to make informed choices. The placement of a feedback will also determine its function. When presented in immediate or near immediate response to input, it will be for the purposes of correction, reinforcement, memorization and persuasion. Else, at the end-point

where just before completing the game, it will be to evidentially support review, analysis, and/or accountability.

Real time feedback is one very important component enabled by the in-process assessment framework. Related interpretations can be grouped, and their percentage scores can be presented in a radial chart (aka spider chart), at real time to the player. An overlay of different colors can be used to represent different stage in the game, or compare two results between two different trials. These could also give an indication on how much the player had learnt and improved. Real time feedback gives confidence to player and encourages the player to constant improvement himself, thus making the game more engaging and promotes further learning.

Other than presenting the feedback to the player and educator, we can also use the information to change the behavior of the game. For examples, games can move on to reassert an interpretation or to provide extra task to help the player improve in area they are weak at. However, it requires extra effort at the task selection and design stage. It is also possible to use the feedback information to increase the difficulty of the game so that the game would appear to be more challenging for the better player, and improve the overall gaming experience.

A simple game prototype

So far, the in-process assessment framework was discussed at the conceptual level. In this section, we use an example to illustrate how it can be used to obtain the scoring system. The objective here is to show how a conceptual model, defined at a general level, can be used to describe the design and delivery processes of assessment in a game or simulation.

The game prototype is adopted from the paper, "Reducing the risk of becoming a victim of terrorism while on international business assignments" by Elango, Lee and Masoud. It presents an experiential exercise on how a businessperson, engaged in foreign travel, can decrease the risk of becoming an easy victim of terrorism or crime. (B. Elango, Lee A. Graf and Masoud Hemmasi, 2008).

The objective of the game is to make participants aware of the four stages of terrorist acts. When an individual understands these four stages, the chance of avoiding an "incident" is greatly increased. And to help participants understand the actions that could be taken to prevent, or at least, significantly diminish involvement in a terrorist act while traveling abroad. The target audiences of the game are business executives expecting to travel abroad, employees anticipating international assignment, individuals planning international vacations, students planning for international careers or study abroad and their family members.

The game can be accessed at http://xapp.appspot.com/terrorism/. It takes the form of an interactive exercise following a traveler's decision path in visiting a foreign country, to gain knowledge about terrorists' attack and ways in reducing the risk of falling victim to the terrorists. Along the way, player's decision option will be evaluated, and a final score will be shown at the end of the exercise.

Application of Scoring System

Task/Action

The Task in the game is to make decision for the game character. We first present the participants with a passage entitled "The four stages of terrorist acts" for them to have a basic understanding of the situation he/she should beware of. During the game, the participants will be presented with scenarios that reflect some of the situation mention in the passage. A decision has to be made at each scenario before proceeding to the next one.

Observations

The Observations of the game are made by taking note of the choice by the participants in each scenario. We also keep track of other parameters, such as the time taken to complete the tasks. The set of observations was kept in the database at the end of the game. These data served as raw data captured during the game play, which can be used to compare among participants, or revise the game design.

Interpretations

The learning objective of the game is to increase safety awareness and help participants understand the actions that could be taken to prevent involvement in a terrorist act. During

the game, we assess the level of safety awareness the participant has. There are three Interpretations: *property get stolen*, *encounter hostile event*, and *taken for hostage*. They reflect the severity of the impact to the game character's personal safety. The gamer is expected to use more caution while traveling next time if he or she does not do well in the game. These interpretations are also used to steer the flow and determine the ending of the game. Nevertheless, it is not necessary to align the game objective, such as game ending, to its interpretations.

Weighted Score Model

Scenario	Decision	Property got stolen	Encounter hostile event	Taken for hostage
1	17 to 19		1	
2	A			1
3	A			1
4	B	1		
5	A	1		
6	A or B or D		1	
8	A			1
10	A			1
12	A	1		
	B			1
14	B or C			1
15	A			1
16	A			1
17	A		1	
Total		3	3	8

FIGURE 3 TABLE OF WEIGHTED SCORE MODEL FOR THE GAME PROTOTYPE

The Weighted Score Model employed in the game is presented in Figure 3. As shown, not all the scenarios are accounted for, as some of those scenarios are proposed to create awareness and allow transition in the storyline. In the decision column are those choices that had significant implication. For simplicity, the weighted score contributed to each Interpretation is one. For choices that are not significant, zero assignment is omitted in the table. There are cases where a few decisions in one scenario may lead to the same interpretation, for example scenario 6 and 14. Also, different decisions in one scenario may lead to different interpretations. For example scenario 12, choice A contributes to Property got stolen, and

choice B contributes to Taken for hostage. This is an example to showcase the flexibility of the weighted score model. One thing to take note here is that, each interpretation can mean anything, and not necessary to be positively phrased. However, in case there is a change of the phrasing, we need to change the weighted scores to reflect the change in interpretation as well.

Feedback Model

The Feedback in the game is in the form of summary reading at the end of each section of the journey. There are four sections in the game: *Pre-departure Preparation*, *Airport Safety, Hotel Safety*, and *Destination Safety*. Each section consists of 3 to 6 scenarios. Summary at the end of each section covers advices by experts from many areas related to the theme of that section. Specific warning tips were also given to the gamer on certain decision he/she had made in the game. It is in the summary page that the gamer may gain additional knowledge on terrorist acts.

At the end of the game, the gamer is presented with an overall summary about his/her performance in the game. In terms of how many scenarios he/she had made a better decision, and also the likelihood that he/she might end up in a bad situation. A bar chart was presented to illustrate his/her learning progress in the game, comparing the percentage of better decision made between four sections in the game. We also allow the gamer to go back to the section that he/she would like to try again and learn more.

Future Work

The framework we proposed has only been tested with a simple game prototype. More work is needed to verify and enhance the framework. We would like to find out the usefulness of such framework to gaming and simulation development community. Collaboration with other ongoing development projects would be useful in finding other challenges in developing in-process assessment.

We would like to see how the framework can be applied to a whole different genre of games. The anticipated challenge faced by such investigation is to identify the learning objective; and based on the interpretations, determine the possible observations. The weighted score model

is hard to draft since it requires a lot of domain expertise. It is likely that new components are needed to add adaptability to the framework.

Further enhancement is needed for the game prototype we had developed. We need gather feedback on both the gaming and learning experience, such as level of interaction and degree of immersive experience.

Developments in gaming and simulation are rapidly growing. Both game designers and educational professionals must collaborate fully for a game to provide the most engaging and effective learning experience. In-process assessment is one very important part that is not commonly found in current educational games. It is hoped that our work could bring value to the community, bridging the gap between the game designer and educational professionals, foster next generation games that are fun, engaging and rich in educational value.

References

Anderson, J. R. (2004). *Cognitive psychology and its implications.* Worth Pub.

B. Elango, Lee A. Graf and Masoud Hemmasi. (2008). Reducing the risk of becoming a victim of terrorism while on international business assignments. *Simulation & Gaming* , Vol. 39, pp. 540-557.

Robert J. Mislevy, Russell G. Almond, Janice F. Lukas. (2003). A Brief Introduction to Evidence-Centered Design, CSE Report 632. Center for the Study of Evaluation(CSE).