USING EVALUATION AS A QUALITY ASSURANCE TOOL IN THE DEVELOPMENT OF SERIOUS GAMES

A Case Study based on the PRIME Game

Jannicke Baalsrud Hauge, Heiko Duin

BIBA, Hochschulring 20, D-28359 Bremen, Germany

Manuel Oliveira

KIT@Work, 3 Park Drive, NW11 7SG, U.K.

Keywords: Serious games, educational games, evaluation methodology.

Abstract: This paper describes how formal evaluation of learning outcomes is an important Quality Assurance (QA) tool in the development of serious games. The case study is based on the PRIME project, which aimed to develop a serious game for learning global strategic manufacturing. The paper discusses the benefits of implementing a formal evaluation for learning outcomes to validate the game, using the alpha prototype. The process allowed the identification of shortcomings and provided valuable feedback to the development activities of the serious game. In the particular case of PRIME, the QA process involved the evaluation within a classroom setting during a total of 6 weeks. Whilst the outcome of the evaluation methodology revealed interesting results, the most important was the rich data collated that contributed significantly towards the final version of the software.

1 INTRODUCTION

Manufacturing is a key generator of wealth and is still at the heart of the economic growth in industrialised economies. In recent years manufacturing in developed countries is undergoing profound changes bringing it from a resource based and centralised paradigm to a knowledge-intensive, innovation-based, adaptive, digital and networked one (e.g. Montorio and Taisch, 2006).

Within this complex and rapidly changing environment planning a suitable strategy concerning the manufacturing will be crucial for enterprises in order to face the ongoing transition while maintaining their competitive position. Unfortunately, strategic planning and strategic decision making is faced with some problems like incomplete knowledge about the exact situation in the market or simply time pressure for reacting on competitive challenges. Such a situation requires an experienced manager taking the decisions. The traditional way for getting strategy experience is learning by doing which takes quite a while for the involved managers. Within the EU funded project PRIME (Providing Real Integration in Multi-disciplinary Environments) a virtual business environment has been developed which acts as the basic setup for a serious game allowing the experimentation with strategic issues (Oliveira and Duin, 2007).

This paper presents the case study of performing Quality Assurance (QA) of the alpha prototype in the form of a formative evaluation of the targeted learning outcomes. This form of QA differs from the usual game development since the focus is achieving the desired learning outcomes rather than some other software quality parameter. The paper discusses the approach and presents some lessons learnt from the case study, which may be applicable for the development of other serious games aimed at training and competence development.

2 THE PRIME GAME

The PRIME project takes a different approach to competence development, using a serious game
(Wortley, 2007) to allow individuals to develop their competences in a soft failure environment with elastic timeline. The associated learning context is strategic global manufacturing, where it is necessary to address both the strategic and operational level.

The possibility of uploading seed scenarios, supported by the functionality of capturing the current state of the PRIME serious game, empowers individuals with the ability of rich experimentation where it is possible to return to critical events in the timeline and investigate alternate outcomes. This leads to the creation of a wealth of experience that is unfeasible to do in real life due to the prohibitive cost of failures that ultimately may lead to the demise of a business. One of the outcomes of using the PRIME serious game is that managers are given the opportunity to practice safely within a “sandbox” rather than perform in the real world. This promotes individuals to experiment and develop a positive attitude towards measured risk taking.

2.1 Description of the PRIME Game

An architectural overview of the PRIME serious game is represented in the diagram of Figure 1.

The Virtual Business Environment (VBE) is an alternate reality that is driven by a hierarchical simulation infrastructure where simulation nodes can be plugged-in. A simulation node is dynamic, which is governed by the centralised simulation clock, and defined by a set of one or more models. It is possible for a simulation model to be shared amongst one or more simulation nodes, as in the case of imposing rules and constraints.

The overall simulation model is complex and multi-faceted (Duin et al., 2007), with the following being part of the default configuration of the PRIME serious game:

- **Geographic Model.** The alternate reality is partitioned into geographic areas with different granularity, ranging from a continent to a region of a country. The scope is determined by the richness of the associated content. This model underpins many other models, such as the one encompassing the logistics networks and the one corresponding to supply of raw materials;

- **Innovation Model.** This model is responsible for the research output of all innovation activities within the VBE. The model takes a different approach to the usual branching models and provides some fuzziness to how innovation comes about, thereby introducing uncertainty;

- **Population Model.** This model encompasses both the consumer and worker dimension of the population. An agent based approach is taken to allow for interesting emergent behaviours.

- **Government Model.** This model is a set of rules and constraints that are organized as Tax (government fiscal plan), Economy (macro-economic factors such as unemployment) and Legislation (laws such as minimal wage);

The simulation model also addresses the internal operation of each business unit (organization), by modelling the company’s operational processes. The approach taken concerning organizational structure is to divide a business unit into the following departments (or Functional Units): Sales, Production, Human Resources, Finance, Product Development and Logistics.

It is possible for a Business Unit to have more than one site, each with different configurations, but by default there is always a headquarters.

Each player assumes the role of CEO of a Business Unit, being responsible for all the existing sites. In PRIME special consideration was taken to avoid the trap of micro-management, which destroys the “fun” factor of a game. Therefore, the player may engage with the Business Unit from a strategic perspective on a corporate level. However, PRIME allows a player to scope to site level for operation management. There are no constraints on how and which level to interact with.

A base system requirement was cross-platform to permit accessibility anywhere and anytime. As a result, the player can interact with the VBE via the PRIME client irrespective of the hardware device and its connectivity. So it could be a PDA through wireless network or a desktop computer via wired connection. As the simulation is not real-time, the user’s sensorial feedback cycle is quite lenient to problems that may emerge from poor connectivity or network distance to the cluster of servers hosting a particular VBE.

Although there is a system constraint of one user account per Business Unit, it is possible to have logical accounts. This supports the cases where a team of players manage a single Business Unit and each assumes a particular role (ie: director of a Functional Unit). In this case, some of the learning experience is attained by offline interactivity between members of a team.

2.2 Learning Objectives in PRIME

Similar to other simulation game engines, such as the renowned SimCity, the learning objectives are determined externally to the game. In the case of an individual player, they may establish their objectives...
To illustrate possible learning objectives that can be supported by the PRIME serious game, it is best to briefly describe some of the main objectives of the participating end-user organizations (total of six) within the consortium:

- **CRF.** This is a research centre associated to the FIAT group. Their main learning objective was for managers to understand the complexities of global logistics and their impact on the whole network. This essentially requires having scenarios in the PRIME serious game where multiple Business Units are part of different supply chains but share common logistic networks;

- **IAI.** This is an engineering company that provides products and services to the military. Their main learning objective is to train their managers in the effective implementation of the company’s proprietary methodology for product development. In this case, PRIME serious game is used specifically with the project management functionality;

- **Intracom.** This is a telecoms company. Their learning objective was to train product managers to interact with sales managers, collate information from research and make decisions on productisation of innovative technologies. In this case, the PRIME serious game was used by a team of players who made decisions within a Business Unit for launching a new product into the market;

- **KESZ.** This is a construction company. Their learning objective was for project managers to recognise the more profitable projects, thus being selective in their commitments. The aim was to support the entire process from securing a building
contract through a successful bid until the closing of the project with the final payment of the client. A key aspect was how to deal with unexpected events and the impact of choices (ie: selection of subcontractors) have in the cost/profit associated to the project.

- **LEGO.** This is a company that manufactures and sells toys. Their learning objectives involved their managers being able to develop the fine art of product portfolio management. This involved all functional units available within the PRIME serious game, namely product innovation, production, sales and strategic marketing, finance and human resources. The decisions made were more strategic, with outcomes taking a full 1-2 years before data is available for evaluation;

- **Siemens.** The particular company of the group is mainly involved in software development. Their learning objectives were mainly to train sales consultants and developers to develop possible scenarios associated to particular marketing strategies and the decisions on implementing particular functionality in the software. The PRIME serious game was mainly used in terms of project management coupled with Sales and Strategic Marketing.

In the particular case of the situation documented in this paper, it is important to denote that much of the PRIME-Time framework was adapted to fit the setting of an academic course. However, an entirely different evaluation methodology was designed and implemented to fit the particular requirements of the students and the teaching objectives of the course.

Evaluation is an essential step and should take place at all stages in the system life cycle. For evaluating the PRIME software during the development the following ISO standards were taken into account: ISO 9126, ISO 9242 and ISO 14598.

PRIME is using a two-step evaluation methodology to assure software quality. Within the first step a group of students is evaluating the software before in the second step the evaluation is performed by the real industrial end-users.

### 3 QUALITY ASSURANCE IN AGILE PROJECTS

#### 3.1 State-of-the-Art

Mnkandla and Dwolatzky (2006) observed that software quality is hard to define when dealing with Quality Assurance (QA) for agile software development processes. They defined a technique how to determine the factors of software quality to be improved by agile methods.

Abbas et al (2007) made the same observation and propose further research on defining an appropriate quality assurance model for agile development methods.

Both papers try to build a bridge between what is known from the quality area (e.g. ISO9000 ff) and agile development methods. ISO 9000 is concerned with quality assurance to provide confidence that a product will satisfy given requirements. Further, ISO 9126 categories attributes of software quality. These characteristics are mainly of technical nature. Not only the quality of the software itself is important, but also its integration, the quality in use needs to be evaluated according to ISO 14598. The main idea of evaluating the PRIME software with the help of the ISO standards is that this will ensure an objective and comparable evaluation.

#### 3.2 The PRIME Approach

Within PRIME, five agile processes were running in parallel to implement the serious game: vision management, evaluation, PRIME-Time, coding and design (Oliveira and Duin, 2007).

In PRIME, a software release is tested by two different groups: an internal group of gamers with individuals from developer partners and a group of surrogate users from any partner without strong gaming experiences.

The testing process in PRIME started very early, in order to receive feedback that can be used to guide the design at an early stage, so that when a prototype is available it is possible to measure whether user and organisational objectives have been achieved. In the early stages of the development and design process, changes are relatively inexpensive. The longer the process has progressed and the more fully the system is defined, the more expensive the introduction of changes is. However, as the first development milestone was achieved (release of the alpha version of the software), a more comprehensive testing framework was necessary, which would also encompass the evaluation of the effectiveness of PRIME as a serious game that supported competence development and knowledge acquisition.

For the alpha version, in addition to the testing mechanisms based on user surrogates, PRIME was deployed in a classroom setting of a university (Bremen University) for a more in-depth testing with a group of students (from economy and engineering courses). Evaluation elements were integrated into
the testing framework to capture how PRIME contributed to the learning experience of the individuals taking part of the course. The beta version would then be evaluated by the industrial end-users. The reason for this approach is that mid-level managers from industrial enterprises demand error free, resilient, stable and fully functional software even with a beta release. However, waiting until the beta version was available would have significant risks for the success of the project, even though frequently usability feedback was obtained from hands-on workshops involving user surrogates for each participating end-user organisation. Therefore, the adopted approach was to invite students from economy and engineering courses to partake in the evaluation of PRIME as it was used as a didactical tool in their practical work. Unlike the user surrogates, these students had no prior contact with PRIME, so all concepts and interaction paradigms had to be learnt. This had the following benefits:

- Students are more lenient concerning the stability of the software, albeit being as critical as mid-level managers. They can accept that an alpha release is not fully bug-free. They continue to focus on the functionality, even when the software crashes some times. Mid-level managers tend to not accept such behaviour. They claim that crashing software is detrimental to the cost effectiveness of their time. Therefore, a key contribution was the discovery of bugs and errors which have not been detected during testing.

- Most of the students have a gaming experience that is not wide-spread in middle management, thus many of the concepts and game mechanics were not entirely unknown, such as the case for assuming a role of CEO of a Business Unit within an alternate reality with a global sustainable economy.

- Students normally have a good intuitive understanding of usability and can wrestle more easily with the learning curve associated to software. This enabled the evaluation to provide feedback concerning the feasibility of the game-play and the interface design.

The purpose of the evaluation was twofold:

- To verify the concepts and design of the software.
- To check the functionality and the stability of the software within a real work environment (albeit it being a classroom setting).

The evaluation of the PRIME software, within the context of the classroom (20 master students of production and economics and engineering from the University of Bremen), is used to:

- Provide feedback for improving the design.
- Assess whether user and organisational objectives have been achieved.
- Monitor long term use of the product or system.
- Fulfilment of the given requirements.
- Ensuring that quality problems, esp. quality in use, will be recognised at an early stage.
- Recognising and removing errors/bugs in the code before the system is in use.
- A user may evaluate the suitability of a software product using metrics for quality in use.

In order to evaluate software it is necessary to select relevant quality characteristics. This should be done using a quality model which breaks software quality down into different characteristics. These can be further broken down into sub characteristics which have measurable attributes.

4 EVALUATION OF THE GAME

Even though the PRIME game is mainly developed to fulfil the needs of the manufacturing industry, the first main test and evaluation of the software were done by master students in production and economics engineering at the University of Bremen.

The University of Bremen has used different types of games for educating engineering students for several years. Even though the evaluation of the learning effects regarding soft skills needed in a distributed production are fairly positive, there is still potential for extend the classes with new games with additional features in order to show the students as many aspects of the working environment in distributed production as possible.

Taking into account that the PRIME software was the alpha version with some instability and limitations in terms of functionality, the evaluation process was executed with the support of developers involved in the software implementation.

4.1 Objectives of the Evaluation

As mentioned above, the PRIME game does not have any learning objectives in itself. Far more it is a game engine usable for training skills on strategic decision making. However, in order to evaluate whether the game is suitable to train such decisions, it was important to define some specific learning objectives and to embed the game in a specific curriculum. Therefore, the PRIME serious game was evaluated within a learning lab that supports the course. The main objectives were defined as:

- to learn new methods to better evaluate the consequences and impacts of a decision,
develop decision-making skills,

identify strategic decisions and develop an understanding for their importance,

understand the relevant factors in the decision-making process for various network configurations.

The learning lab used a blended learning concept (frontal teaching for the methods, which the then needed to apply in the game). There were 24 students starting, 18 passed. They had studied between 6 (1 Students) and 14 semesters (3 Students). Most of them had studied 8 semesters. The lab was organised as different workshops over a period of 6 weeks, and three hours each week (18h all together) and encompassed different phases for students and the workshops were organised as followed:

- Introduction to the Gaming Lab
- New Methods and Game Tutorial
- Additional Methods and Continuation
- of the Gaming Tutorial
- Additional Methods and Reflection of the Game
- Development of a Game Scenario
- Enactment of the Developed Scenario by playing the PRIME serious game.

Since the game was used by students, who did not all know very much about strategic decision making and methods supporting such decision making, theoretical background of such methods were introduced by the facilitator at the beginning of each workshop. Documentation on these methods was distributed to the students in advance.

Secondly, the PRIME serious game takes a holistic approach to the VBE, thus modelling multiple dimensions of operations within a business organization and their interaction with other organizations and the economic environment. This implies that the software is complex with a relatively steep learning curve, thus for beginners would require plenty of support in order to make their own models or just in order to be able to "establish" the given scenario in the simulator. Therefore, for the first three workshops, a comprehensive guideline was developed. In the first workshop, each user used a single player modus, for the other ones multi-player modus was planned and tested.

The outcome of the evaluation process within the laboratory environment would enable:

- To evaluate if it is possible for students to experience the impact of strategic decision making using the PRIME game.
- To support the implementation activities of the PRIME game prototype at an early stage, before the software were complete. This would give us the unique possibility to test the usability and the functionalities at an early stage and to give feedback to the developers if any changes were needed.

4.2 Evaluation Methodology

The approach used for this early evaluation of the PRIME game prototype was based on three different types of input:

- The observation and the exchange of information between the facilitator and the students;
- two questionnaires comprising questions on the functionalities, the utility and the usability of the software;
- the completion of quite extensive laboratory report comprising information on the involvement in the game, skills gained and past experiences. The students could also add comments and give feedback during a session in each class.

Table 1: Example of the Matrix for Functional Testing.

<table>
<thead>
<tr>
<th>Functionality</th>
<th>Workload</th>
<th>Execution</th>
<th>Design Problems</th>
<th>User Requirements</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Login</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creation of an Head Quarter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creation of the Finance Unit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creation of the Office Building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assignment of the finance operation processes to the building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creation of the Development Unit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assignment of the R&amp;D operation processes to the building</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis of the Development view</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management of Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creation of a new product</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creation of a Product Portfolio Management Unit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis of the Corporate Production view</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM to Product to the Product Portfolio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Additionally, the reports also contained information on how the applied the learned methods in the game as well as on the developed game scenarios, including their own goals and the fulfillment of these goals. The questions below shows some of the questions used for the evaluation of the utility of the game, which were supported by a Likert scale between 1 and 5 (corresponding to opposite extremes of the answer. The reports contained answers.

- Although the test cases only showed you a very small part of the planned functionality, based on this experience, please tell us how useful you think PRIME will be in your job or for doing training.
- Based on your experience with the test cases, how willing would you be to use PRIME?
Based on your experience with the test case, how well did the PRIME Client meet your expectations of a training tool?

Do you think you would prefer to use PRIME instead of other training / learning tools for strategic decision making in a business?

Do you think you would enjoy using PRIME?

Based on your experience with the test cases, how do you rate the usability of the PRIME Client?

Once you knew how to use the PRIME Client, do you think it would be useful in gaining experience in strategic decision making?

The questionnaire used for testing the functionalities were used the first and the second time, and only the functionalities which were already implemented were tested. Table 1 shows an example of the matrices for the functional testing.

5 RESULTS

The first statement which can be made is that the process was very challenging for the facilitator as well as for the students to use a partly developed prototype. Some features, which would have made it easier for the students to see whether they fulfilled their objectives and strategic aims, were affected by bugs or not were still not implemented. These type of limitations with the alpha version would be supplemented with offline measures to calculate and generate the necessary data. However, the involvement of the developers in the process helped smooth most of the problems and cases the challenges on the facilitator.

Secondly, the observation of the students trying to implement their scenarios made is clear, the learning curve associated to the PRIME software is steep and requires appropriate tutorials and documentation. As a result of the evaluation process, methodical and comprehensive video tutorials have been developed to support the users with the learning curve associated to the PRIME serious game.

The functional testing was mainly carried out in single player mode. The results showed that the functionalities implemented were working satisfactorily, but that a handbook or a FAQ would be required to support the player. This is a fundamental requirement that was identified, since there is no in-built help to support the player. The documentation support has improved with the release of the alpha version, and plans are being considered for the inclusion of personalised agents to act as advisors.

Although the PRIME serious game demonstrated reasonable stability during the single-user mode, more significant challenges emerged when the lab sessions started required the students to use the multi-player mode. This mode is essential for training of strategic decision making in a multi-stakeholder environment, thus approaching to what takes place in reality. Regrettably, multi-player mode was not working due to technical problems with the software. The culprit was the instability of the PRIME server that caused it to crash as soon as more than one person worked simultaneously on the same scenario. Quick analysis identified that one of the main reasons was the memory consumption on the server was very high. This was due to improper memory management on the server side caused by third party software for management of the database. This caused the same data to be loaded several times into memory, leading to numerous copies of the same data objects. Originally, it was intended that the server removes unused data from memory, but it was hard to recognize if the client still needed a specific data object or not, this was not done.

A major drawback was the lack of fault-tolerance capabilities within the PRIME server, so if it did experience a fault, it would be possible some form of recovery for the players to continue from a previous state. However, this was not the case so the players always needed to start from scratch again. Even though the students were very patient and tried several times, not only in the second but also in the third workshop, they never managed to successfully complete a scenario and then play, so that we had to stop using the game after the fourth workshop. We switched then to a paper based model of PRIME for the multiplayer interactions, which worked fairly well. This was a fundamental implementation flaw that has been addressed by increasing the stability of the PRIME server and by providing a toolbox that allows people to build and save scenarios. The saving functionality also allows to take snapshots of the current VBE.

Based upon this experience, most of the students concluded that they met the overall objectives of the course, and that they were partly able to experience the impact of their strategic decisions using the PRIME game (single player). In addition, most students admit to have been engaged with the PRIME serious game, despite the difficulties in achieving a flow that was not plagued by frustration. Therefore, the problems associated to software instability caused students to have mixed opinions about the use of PRIME serious game. This would depend on past experience using games.
6 CONCLUSIONS

It is important to observe that the aim of using an evaluation framework within a classroom environment was not to validate the use of PRIME serious game as an educational support tool, but to collate the data from the deployment of PRIME within a controlled environment. Therefore, it was a very important milestone. This early evaluation showed some important problems and some bugs in the software as well as some important input for what needs to be there in the future.

After the identification of the various problems, mainly those associated to the use of third-party software, major re-factoring took place and improved significantly the stability of the software. The problem with the memory management has been solved by restructuring the data model. Instead of having the server to load the data and transmit the data to the client, the clients load the data directly from the database themselves. Each client loads the data into its own memory on the client side, taking workload from the server. The client keeps track of the loaded data and releases it when it's not needed anymore. To prevent multiple clients from changing the same data at the same time, a locking mechanism was implemented. Further more, the server also saves lasts versions of the scenario, so that reloading is possible. Therefore the negative impact of faults has been minimised. Additionally several minor bugs have been corrected and the last short pre-test of the software shows that it works satisfactorily in multi – player mode.

It is significant to point out that the game design and conceptual model of the game was well accepted by the students, demonstrating the success of the cognitive modelling development approach of building the game-play (Oliveira and Andersen, 2007).

The major improvements to the stability of the PRIME serious game, supported by the documentation and tutorials, it was decided that PRIME would be used in the next lab course on strategic decision making too, which started at the beginning of November 2007 and has been deemed a great success.

For the development of serious games, the adoption of formative evaluation as a QA method allows concrete fine tuning with regards to learning outcomes. In the case of PRIME, the improvements targeted not only the software, but also the supporting methodology and the strengthening of the moderator role.

ACKNOWLEDGEMENTS

The authors thank the respective project partners, as well as the European Commission for all scientific, organisational and financial support. The PRIME project is partially funded under contract number FP6-016542 within the Sixth Framework Programme, Priority IST/NMP.

REFERENCES


