

ARGOSI Evaluation Report

Nicola Whitton, Manchester Metropolitan University
March 2009

1.1 Introduction

This report aims to synthesise all of the evaluation activity that took place during the JISC-funded Alternate Reality Games for Orientation, Socialisation and Induction (ARGOSI) project, which ran from April 2008 to March 2009.

An Alternate Reality Game (ARG) is comprised of an ongoing, and evolving, storyline that unfolds over time as players solve various collaborative puzzles and challenges. The rationale for using an ARG for student induction was that it provided a purposeful collaborative learning space in which students could learn new skills in a non-threatening environment. The project developed a game, ViolaQuest, which ran from September to December 2008.

The project aimed to develop an ARG, to be used during student induction, that:

1. enabled students to meet and work with other people;
2. supported students in finding their way around a new city;
3. provided a purposeful set of activities for learning about basic information literacy skills;
4. was an enjoyable and motivating form of induction.

The set of evaluation activities described in this document can be used to determine the degree to which these objectives were met, as well as lessons learned from the project. There were three phases to the evaluation: first, game testing took place during the game design and development itself, secondly evaluation and market research took place during the time that the game was running, and finally summative evaluation took place at the end of the game.

The purpose of the early evaluations was to make game as fit for purpose as possible and ensure that any issues or bugs, in the game design itself or software, were ironed out early on. It was also important to gain feedback from users that could inform the development and running of the game (one of the advantages of running ARGs over a long time frame is that they can be adapted while they are running). The summative evaluations were used to examine what had worked in the game, and what could work differently, to create recommendations for further iterations of the game, and its use in other contexts. ARGs are a relatively new game genre and have been little used in education, so it was considered important to capture the areas where improvement was needed as well as the successful areas of the project.

The underlying philosophy of the evaluation described here is that of a user-centred approach. It was important that representative users were involved at all stages and that as many aspects of the game design were evaluated as possible, including testing the challenge design, the software usability and robustness, the range of functionality offered, and the playability and immersion of the game. The project worked within an ethos of continuous improvement, and it was recognised that not every aspect of the game design would be right first time, but that with a culture of openness and ongoing evaluation it would be possible to refine the game as it evolved and to document lessons learned for future alternate reality game development.

1.2 Research methods

The evaluations aimed to use as many methods as possible in order to triangulate the findings and ensure that data and conclusions were as reliable and valid as possible. A combination of qualitative and quantitative research methods were used, at three phases: diagnostic during the game development, formative as the game was running, and summative after the game was finished. Table 1 below details each of the activities that took place in each phase and the purpose of the phase.

Method	Phase	Purpose
Pilots	Diagnostic	To inform the design of the story, challenges and game.
Expert evaluations		
Usage statistics	Formative	To enable modifications and improvements to the game as it ran.
Market research		
Interviews/questionnaires	Summative	To highlight lessons learned from the game.
Team reflections		

Table 1: Research methods used to evaluate the ARGOSI project

Some aspects of the evaluation deviate from original methods outlined in the evaluation plan. Initially, the plan was to use face-to-face video reflections from players and undertake social network analysis of the players; interactions and relationships. However, this was not undertaken because the take up from the game was too low to make these activities worthwhile. Summative interviews and questionnaires were planned to gain feedback from the game players, but unfortunately the response rates from these activities were so low as to make them valueless. Instead, two additional evaluation activities were added: formative market research and a reflective group exercise. The market research aimed to see how effective the marketing methods had been and the team reflections added as a way of gaining feedback after the project, as an alternative to student feedback. While it is recognised that this is not as valuable as evaluations from the students themselves it was felt, pragmatically, to be of some value and preferable to carrying out no evaluation whatsoever.

1.2.1 Game pilots

The game was designed to consist of a set of core challenges, which had to be solved to find out the next segment of the narrative and let the mystery unfold, and a separate set of library challenges that focused on information literacy learning outcomes and were supplementary to the main plot. Each element was tested individually as part of the pilot.

The piloting phase of the evaluation consisted of three distinct pilots. The first pilot aimed to test the core game challenges and narrative structure in a face-to-face environment. The second involved running the core game challenges. The second pilot tested the complete game with challenges and narrative running online in software environment. The third pilot focussed on testing the information literacy challenges face-to-face. It was hoped that this fairly extensive testing would ensure the development of a game that was appropriate and fit for purpose.

Pilot 1

The first pilot aimed to test the six core challenges in a face-to-face environment, with challenges being made available through the software environment but also in paper form. Six student participants took part in the evaluation, who were recruited through the university's Student Ambassador Scheme and paid at the standard hourly rate, and the

evaluation ran from 10am to 3pm with lunch and refreshments provided. A laptop with internet connection was provided for each participant. The participants were first asked to work through the challenges and observed throughout. Some of the challenges could be completed in the space provided but others required the participants to navigate through areas of the city, and all were observed by one or more of the three evaluators present. After the challenges were complete, the participants took part in a short focus group. When designing the evaluation, it was not clear if a bigger group would be manageable for focus groups and six was deemed sufficient.

The group were extremely quiet at the beginning but as the day went on they bonded more as a group and seemed more confident to offer answers and suggestions.

The pilot highlighted a range of issues associated with specific challenges, which are not discussed in detail here. However, there were a range of key themes that emerged from the pilot, in particular the following issues:

- The wording of the challenges was not always as clear or unambiguous as had been intended.
- Initially students were provided with a map piece containing a post-it note containing a URL. They had to be prompted to access the web site and did not go there immediately on acquiring the map segment.
- The login page of the game site was difficult to use and needed a big revamp.
- Typing of answers was case-sensitive, which was misleading, particularly when many of the answers took the form of names.
- There were accessibility issues with the first challenge, which was originally based around misspelling, making it very difficult for people with dyslexia.
- For the challenge that involved a treasure hunt around Manchester, because they were familiar with the city, the participants tried to second-guess the challenge by going to places they knew already. This meant that they didn't really find an initial correct point of reference so ended up taking lots of false routes. Others who were not so clued up on the places and local geography would have started at the beginning and possibly have found this challenge easier.
- In some instances the participants said that the activities they needed to carry out (e.g. looking at the source of a web page) they considered to be 'cheating'. The boundary between game and real world was sometime unclear to the players and they lacked the necessary ground rules of the genre.

From the six participants, five were moderately positive about the game, but one was extremely positive and highly enthusiastic about its potential for learning. They described the games as 'challenging' and 'fun', although also 'frustrating', particularly those challenges that didn't seem fair or were perceived as being too difficult. They felt that they had learned how to think laterally, in particular 'thinking outside the box for answers' and particularly liked the elements of the game that were not computer-based and where they had to work in collaboration.

Pilot 2

The second pilot again focussed on testing the core challenges and narrative structure (once appropriate modifications had been made after the first pilot), but this time the game was played over a period of a week and fully online. Twelve people were recruited through word-of-mouth to take part in the game and ten of these were interviewed by telephone at the end of the pilot. It is recognised that because there were a self-selecting group of 'game-players' taking part in this pilot, their responses may not be representative of the student population. However, the main purpose of this pilot was to test the overall feasibility of the game when run remotely, so this was not really seen as an issue.

The majority of players said that they played for around three hours in total during the week, although two played for less than an hour and one reported playing for over eight hours. The amount of free time available appeared to be an important factor in determining how much time was spent on the game.

All players said that it was easy to get started, they knew what they had to do and that the user interface was very easy to navigate and interact with. Reactions to the difficulty of the first challenge were, however, polarised with some players finding it extremely easy and others finding it impossible to solve without considerable help. The importance of making the first challenge accessible was highlighted, as two players dropped out at this stage when they could not solve this challenge.

Players were asked for up to three words that described their experience of the game (these are shown as a word cloud in Figure 1 below). It is interesting that while 'challenging' and 'fun' are prominent, 'frustrating' is also, highlighting the importance of getting the balance of challenges right to keep people engaged.



Figure 1: Word cloud showing players experiences of Viola Quest

In general, there was a positive reaction to the challenges that involved walking around Manchester rather than simply working at a computer screen. Attitudes towards competition and the value of the game community were again mixed with some players seeing these elements as very important and others viewing them as peripheral. All players said that the narrative was good to hold the challenges together but not a crucial element of the game. Overall, it appeared that there were six elements that motivated the different players to different extents:

- **Completion.** Some players simply wanted to complete the game and achieve all the tasks or challenges.
- **Competition.** Some players were motivated by competing against others.
- **Narrative.** The ongoing story was seen as more integral by some players.
- **Puzzle-solving.** The ongoing puzzles, riddles and challenges were seen as motivational for their own sake by many players.
- **Community.** The community elements and discussion boards were important for some players.
- **Creativity.** The opportunity for players to be creative, either through creative problem-solving or the creation of artefacts was important for some.

What these different motivators have highlighted is that people will engage in the game for a variety of reasons and it is important, when designing a game of this type, to try and include elements that will appeal to all different motivations.

Pilot 3

The third, and final pilot, aimed to evaluate the set of nine additional library challenges that were created to meet specific learning outcomes in the information literacy curriculum for library induction. This pilot was also useful in establishing the relative difficulty of the challenges, which were hard and which were easier, in order to present a gradually increasing difficulty of challenge to the players.

This pilot was again carried out face-to-face, over an afternoon period, during which the participants were asked to work through the challenges and observed and the session finished with a focus group. Five participants took part, who were again recruited through the Student Ambassador Scheme and paid at the standard rate. The group immediately formed themselves into two pairs and a single individual; the pairs engaged in a fair amount of talking and collaborating and tackled the challenges sequentially from start to finish, while the single individual took a more random approach to the challenges.

Although there were a large number of specific comments on individual challenges, there were a number of general issues that emerged:

- More help and support was needed for the challenges, particularly additional resources such as information on how to evaluate web sites, and how to correctly reference different publications.
- Many of the challenges were too open-ended and more guidance was needed in which web resources and databases to use, in particular for information seeking.
- Some of the challenges needed to be simplified as they were expecting players to undertake laborious work that was not directly related to the learning outcomes (for example, in the challenge that required students comparing reading lists the total number of books on each list could be reduced to 10-15 rather than 30-40).

In general, the participants had no major problems in understanding what was required from each challenge or completing them successfully. Again, the feedback on how useful and engaging the activities were was positive overall.

1.2.2 Expert evaluations

The project benefited from an extremely experienced team from a wide variety of backgrounds, with experience in interface design, usability, game design and the development of digital narrative. The wide-ranging skills of the team was put to use in evaluating the challenges, narrative and software produced.

The game challenges were developed by members of the core team, based in Manchester. In addition to the three phases of testing that was undertaken with students, each of the challenges was made available to the wider ARGOSI team through an online social network and was tested remotely by at least one other team member. Feedback on instructions, design and difficulty was then incorporated into the design.

While the overarching storyline was developed by the whole team, the actual detailed plotline was created by the core team and was reviewed by members of the wider team, in particular the expert in digital narrative. Likewise the software developed and graphical artefacts designed were evaluated and tested at a number of stages of its development by other members of the team.

An open team ethos and attitude of collaboration and continuous improvement, where team members were happy to provide feedback to one another throughout the life cycle of the project. The online social network that was used was very helpful in facilitating this.

1.2.3 Usage statistics

The statistics of how many users has signed up to play the game and how many were engaging were monitored throughout the game, and amendments made to the game itself based on what was happening. In total, the game had 173 genuine players and 8 players who were either played by members of the project team, or by enthusiastic individuals who were known to the project team, for example, had taken part in one of the piloting phases.

Table 2 shows the distribution of players by the number of challenges completed.

Challenges completed	Number of players
21	1
20	1
17	2
10	1
5	1
4	4
2	3
1	10
0	150

Table 2: Distribution of players by challenges completed

It can be seen that there were a small number – 5 (3%) – who were highly engaged (10 or more challenges), slightly more – 18 (10%) – who were moderately engaged (1–9 challenges) and the majority – 150 (87%) who signed up but did not engage in the game. Players did not need to sign up to see the challenges or discussion, so it was perhaps strange that the number signing up but then not engaging was so high. Possible reasons for this could be that some players were told to sign up by a member of lecturing staff but did not want to engage, or that the initial challenge was too difficult for many and this was off-putting.

A similar picture emerged for the distribution of players by discussion postings, with the number of posts overall ($n=47$) being low and the majority of these posts ($n=32$) being made by two players; in total only another six players contributed to the discussion. Not all of the players who completed the most challenges made the most use of the discussion boards, which provides more evidence that different people take part in this sort of game for different reasons.

An interesting finding was the pattern of engagement throughout the game (shown in Table 3). Take up initially was very slow (ignoring the figure of 76 for the week starting 29 September as this is almost certainly due to the game being promoted by a specific lecturer in a classroom setting) but the number of players signing up grew steadily and continued even after the game had finished.

Week beginning	Number of sign-ups
15/09/08	0
22/09/08	1
29/09/08	76
06/10/08	6
13/10/08	1
20/10/08	8
27/10/08	5
03/11/08	1
10/11/08	1
17/11/08	5
24/11/08	13
01/12/08	16
08/12/08	8
15/12/08	7
22/12/08	3
29/12/08	3
05/01/09	6
12/01/09	4
19/01/09	2
26/01/09	3
02/02/09	4

Table 3: Number of sign-ups to Viola Quest over time

The game started during Freshers' week, a time when students are typically overwhelmed by the amount of competing demands on their time, so it could well be the case that students didn't feel ready to engage with the game until later on in the term. This certainly has implications for the most appropriate time to start a game of this type.

1.2.4 Market research

After two weeks into the game there had only been one player who had signed up so it was decided to undertake an additional evaluation activity of market research to examine the reasons that students were not engaging in the game. Four researchers stood outside the university library for two hours during lunchtime and asked passing students if they would be willing to answer a short questionnaire. The questions focussed on which of the marketing materials had been seen by the students and their attitudes towards playing the game.

In total, 96 students were interviewed and the response towards the game was generally positive. However, although some had previously seen examples of the Viola Quest marketing (see Table 4), very few had realised that it was part of a game or that they were expected to take action.

Type of marketing material	Number seen
Postcards	14 (15%)
Posters	11 (12%)
Stickers	10 (10%)
Leaflets	9 (9%)
Other (email / t-shirts)	5 (5%)

Table 4: Effectiveness of different marketing methods

Other general findings from the interviews were:

- Students would have been more keen to take part if the marketing had made it more explicit that this was a game, and it was clear what steps they were expected to take.
- They required motivation to take part, several suggesting that 'a prize' would be motivational.
- Many said that they did not have time to play 'a game' but would be more interested in taking part if they had realised that it would benefit their studies.

These findings are interesting in relation to the ARG aesthetic of 'this is not a game' that the Viola Quest game aimed to adhere to, and brings into question the appropriateness of this model in an educational context.

1.3.5 Interviews and questionnaires

It was originally planned that qualitative feedback would be gained from the students who had taken part in the game by using interviews and questionnaires. An email was sent to all players who had not engaged (i.e. had not completed any challenges) asking them to complete a short questionnaire; there were two responses. In addition, those players who had engaged were asked if they would be prepared to take part in a short telephone interview; in this instance the response was zero. In each case the initial email was followed up with a subsequent one but the response rates remained low.

A possible reason for the low take up was the timing of the emails (just before Christmas and in the New Year) when they may have taken a low priority. Another possibility is simply that is the majority of the players did not engage in the game so it was always unlikely that they were going to engage in the evaluation. Owing to the difficulty of getting summative feedback from the game players, an additional evaluation activity was included in the form of a team reflection. While, in an ideal world, student evaluative feedback would have been good, pragmatically this is all that was available.

1.2.6 Team reflections

In order to provide a summative view on the elements of the project that were successful, and the lessons learned from less successful aspects, the core team met for two hours once Viola Quest had finished in order to consider some of the questions surrounding the game. The group reflected, in particular, on what had gone well during the project and what the successes were, what worked less well, and what was learned from the project.

What the project did well

The ARGOSI team felt that, overall, the project did many things well:

- The team successfully completed the design, development, testing and running of a pilot alternate reality game to support student induction.
- Development of robust and reusable software for supporting alternate reality games (e.g. challenge management, user management, communication, leader boards).
- There was a good choice of initial team design at the project design stage, and selection of experts with appropriate skills and experience who all worked well together as a team. The initial face-to-face 'kick-off' meeting was particularly valuable for establishing an open and collegial working atmosphere.
- Communication was also good, both between team members, with the Programme Manager and with other project. This was achieved through a variety of communication channels but, in particular, the social network that was set up to support internal project communication was very useful.

- The initial user testing was very successful, with participants engaging and providing a great deal of valuable feedback on the game design and software implementation.
- The narrative, graphical artefacts and challenges created are appropriate, fit-for purpose and relatively easy to re-use.
- The dissemination throughout the project has been thorough and effective and members of the project team have been asked to speak at a wide variety of events.

What was learned from the project

A range of elements were highlighted that could be improved in future projects:

- It was apparent that timing is a critical factor and that launching the game during freshers' week (a time when students may already be overwhelmed and disorientated) was going to be problematic. Although it is clear what does not work well, the ARG format offers so many alternatives, in terms of timing and length that it is not clear what the best time for such an induction activity might be but there are lots of possibilities (e.g. pre-registration or in second semester).
- Marketing is very important. Although the project had quite a comprehensive marketing strategy it proved to be only reaching a small percentage of its target audience, and the majority of those who saw the marketing did not realise that it was either a game, or an activity that would help their studies.
- Engagement with the game (both in terms of initial sign-ons and those who took part in the challenges) was far lower than expected (although anecdotally the 13% of those who signed up becoming active players is fairly typical). Strategies for ensuring a higher rate of engagement should be considered (e.g. making a compulsory part of a course, linking to assessment, offering prizes, reaching a far larger potential target audience initially).
- It is difficult to get the right challenges to suit every level or ability, but it is better to err on the side of 'too easy' rather than 'too hard'. The initial challenges were too hard for many people, who quickly became disengaged and dropped out of the game. Additional easier challenges were added early on as well as making all of the library challenges available at once, which helped to keep players engaged.

1.3 Discussion

This final section of the evaluation report shall explore some of the big questions that have emerged through the different stages of the evaluation. It does not, at this stage, aim to offer answers, but merely to highlight a range of issues and considerations for the use of alternate reality games in education as their use becomes more widespread and practice becomes established.

The big questions regarding alternate reality games in education, which emerged from the evaluation of the ARGOSI project:

- How can students be encouraged to engage without making it compulsory? If the game becomes compulsory does it then cease to be an ARG but become something else? Is it possible to make a compromise between the notion of 'this is not a game' and something
- How large does the game have to be to be workable as a voluntary activity? If it can be assumed that only around 10% of players will become active, how many need to be reached in the first place? Will a single institution ARG be viable?

- At what point does the game become value for money? How high does participation need to be for it to be a worthwhile activity? How can the success of such an activity be measured in terms of costs/benefits?
- What is an effective marketing strategy? Marketing as a mystery or game will not necessarily appeal to all (or even many) students, but if it is marketed as an educational activity but the game then start to move away from being an ARG? Does this matter?
- ARGs support the autonomous student, but can they also be used to develop autonomy? How do they fit into a culture of learning that says “don’t do anything unless you’re told to do it”?
- How can the tension between the niche nature of the ARG be resolved with the inclusivity that is desirable in Higher Education? How can challenges be made accessible for all without spoiling the game for some?

As well as the big questions for the use of ARGs in education, the evaluation also identified a number of areas in which different implementations could be explored:

- Changes in timing and length of the game, exploring the potential of use pre-entry or in the second semester.
- Engagement with tutors and embedding into specific curricula.
- More explicit marketing around the game, in particular highlighting the link to learning outcomes.
- Considering providing extrinsic motivation (e.g. prizes or prestige).
- Make a larger number of challenges available from the beginning with more easy challenges available earlier.
- Focus on individual, online and creative challenges earlier in the game and use collaborative and physical challenges later.
- Increased involvement with and support from the students’ union (however timing of elections and changeovers makes this difficult).
- Exploring the potential of commercial sponsors in order to spread the risk, however this may have ethical implications.

To conclude, while there is considerable potential for the use of alternate reality games in education, it is certainly not a quick win. The game genre is relatively new and very new in the field of Higher Education and there is, as yet, very little established practice or evidence as to how they can work effectively. A major issue of any form of game-based learning is how to use a game in a context where outcomes matter (e.g. are linked to learning outcomes, assessment or formal education) and keeping the ‘fun’ of the game itself. This will be an issue for ARGs, particularly if the model of educational ARGs moves away from the typical aesthetic to a more structured and learning-focussed model.

While the complete ARG model may not be appropriate to be used wholesale for learning, there is certainly potential for using elements. For example, a focus on activity-based learning through challenges, the use of narrative structures to support learning, or the development of a collaborative learning community are all elements that ARGs support that could equally-well be implemented outside of the ARG context.

In all, perhaps the education community expects too much from learning games. In the commercial world, 90% of entertainment games fail and this is seen as an acceptable, and predictable, failure rate. Are Universities willing and able to take that level of risk?