# 9 Comparative study

This chapter describes one of the key pieces of work of this project: the set of comparative experiments that conclude this research. The purpose of these experiments was to compare two game-based learning activities, the Time Capsule and the Pharaoh's Tomb, the design and development of which are described in Chapters 6 and 7. These studies aimed to determine if there were any differences in educational effectiveness – using self-reported learning and engagement as indicators – between students using each of these activities as part of their course of study.

First, the design of the overall learning session, of which the game-based activity formed part, is described and differences between the two activities are considered in relation to the characteristics of games determined in Chapter 2. This is followed by a description of the experimental design and use of pilot studies. The final two sections of the chapter describe the results from the self-reported learning questionnaire and the engagement questionnaire.

## 9.1 Comparing the Time Capsule and the Pharaoh's Tomb

Two collaborative game-based activities were developed: the Time Capsule, a direct online translation of a face-to-face collaborative activity; and the Pharaoh's Tomb, a graphical multi-player adventure game. While it was considered that both of these applications were examples of game-based learning, using the inclusive definition from Chapter 2, the Pharaoh's Tomb was designed to exhibit all of the characteristics of games determined previously, whereas the Time Capsule was developed as a direct online translation of a more traditional activity and only exhibited some of them. The game characteristics (first shown in Table 2-1 in Chapter 2) are applied to the Time Capsule and the Pharaoh's Tomb in Table 9-1 below.

The Pharaoh's Tomb is a three-dimensional collaborative graphical adventure game that offers a range of challenges based around group problem-solving and has explicit group goals of returning an object to a certain place within the tomb and enabling the whole team to escape, and implicit rules as to what the team members can do (e.g. each player can carry only one object at a time).

| | Pharaoh's Tomb | Time Capsule |
|---|---|---|
| Competition | ✓ | × |
| Difficulty | ✓ | ✓ |
| Exploration | ✓ | × |
| Fantasy | ✓ | ✓ |
| Goals | ✓ | ✓ |
| Interaction | ✓ | ✓ |
| Outcomes | ✓ | × |
| People | ✓ | ✓ |
| Rules | ✓ | ✓ |
| Safety | ✓ | ✓ |

Table 9-1: Analysis of game characteristics (from Table 2-1) applied to the Pharaoh's Tomb and the Time Capsule applications

The Pharaoh's Tomb game is scored, which means that there are measured outcomes (as opposed to simply achieving the goal or not) and teams can compare themselves with other teams, introducing an element of competition. The game is not designed to support inter-group competition because this would be at odds with the collaborative nature of the exercise. It provides a fantasy environment of an Egyptian Pharaoh's tomb that can be navigated and explored; characters can interact with objects and gain feedback from the environment as well as from other players.

The Time Capsule, in contrast, is an interactive group negotiation activity that exhibits fewer characteristics of games. While the activity is difficult (certainly non-trivial) and students are presented with an explicit goal to agree on items to be placed in a time capsule and clear rules regarding the number of items and their total cost, there are no measured outcomes or scoring. Therefore, beyond achieving the goal or not, it is not possible for teams to compare themselves with others, removing the element of competition. The Time Capsule provides a fantasy scenario for the participants but it does not present an immersive world that can be explored as part of that scenario, although it does provide interactivity and feedback to actions. Neither activity is assessed, so, as far as is possible within a non-voluntary learning environment, participation in the activity is not deemed to have consequences in the real world.

Both game-based activities were designed to take the same time to complete and with the same set of learning outcomes and supporting materials. Each of the sessions was designed to fit into a one-hour time slot, because this fitted with the timing of lessons at the universities where the trials were undertaken. The timings of the sessions were as follows:

- Students are given a verbal introduction to the aims of the session and the nature of the research, including time for questions (at this stage participants were given the option to opt out of the research element if they wished); students read two-page briefing on learning outcomes (approx. 10 minutes).
- Students are asked to complete the background questionnaire (see Appendix 4).
- The game-based learning activity is played (approx. 30–40 minutes).
- Students are asked to complete the engagement questionnaire (see Appendix 22).
- There is a debriefing session in which students are asked to consider and discuss their team behaviours during the game and relate to the learning outcomes (approx. 10 minutes).
- Students are asked to complete the self-reported learning questionnaire (see Appendix 18).

Paper-based support materials were provided, including instruction sheets for the Time Capsule (Appendix 12) and Pharaoh's Tomb (Appendix 13), briefing on learning outcomes (Appendix 6) and the debriefing exercise (Appendix 10). The following section explains how the comparative experiments were implemented.

## 9.2 Experimental design

The population of students selected for this trial were undergraduate computing students, because it was expected that this would be a group in which there was a relatively high level of computer literacy and, being predominantly male and under 30 years of age, likely to have more experience with playing computer games (Entertainment Software Association, 2007); therefore it was hypothesised that there would be a lower impact on learning engagement of factors such as learning to interact with the interface or use the games themselves so any effects could be attributed to the design of the game. On a more practical level, it was a group of students on a course and in a department

to which the researcher had access. The experiment was designed so that the students were split into two groups, each group undertaking only one of the two activities; the relative levels of self-reported learning and engagement could then be compared for students undertaking each of the activities.

Initially it was planned to recruit students for this study from the pool of those taking group project modules in Napier University School of Computing as an extra-curricular activity to support their studies; however, it proved impossible to recruit enough students in this way to have a meaningful sample. In order to increase recruitment rates, funding was secured to enable payment for participation, but again recruitment rates were low and only 15 students took part – this was then used as a pilot study. Subsequently an opportunity arose to trial the activities with a group of 19 final year marketing students at the City University of Hong Kong. This second small-scale study was used to test the practicalities of carrying out the experiment in a classroom setting and to collect additional data.

Finally, co-operation was secured to embed the learning activity into a first year professional skills module so that it was then part of the core curriculum. On this third occasion 78 students took part. The fact that the game-based learning activities were embedded into the curriculum meant that ethical issues regarding the issue of informed consent and right not to take part in an experiment had to be considered. To address this, students were briefed about the nature of the research, provided with the opportunity to ask questions, and given the option not to complete the questionnaires if they wished; however, all were happy to do so. In addition, the fact that individuals were not self-selecting adds validity to the experimental design in that it removes any selection bias from undue representation of people who are motivated to play games. Table 9-2 below provides a summary of the three trials.

| | Date | *n* |
|---|---|---|
| Napier pilot (NP) | March 2006 | 15 |
| Hong Kong pilot (HK) | September 2006 | 19 |
| Napier trial (NT) | November 2006 | 78 |

Table 9-2: Summary of the experimental trials to compare the Time Capsule and the Pharaoh's Tomb

In each trial students were allocated to one of the two game-based learning activities – Pharaoh's Tomb or the Time Capsule. In the case of the first two pilots, members of the group were randomly allocated to one condition or other, and in the case of the final Napier trial the games were used in six separate tutorial classes, with three classes randomly allocated to each condition; the original allocation of students to tutorial groups was random. When a student logged in to the multi-user software engine he or she was automatically allocated to the next available game, which effectively meant that players were allocated to teams at random. The breakdown of participants for each of the two pilot studies is shown in Table 9-3.

| | Napier pilot (NP) | | Hong Kong pilot (HK) | | Napier study (NS) | |
|---|---|---|---|---|---|---|
| | **PT** | **TC** | **PT** | **TC** | **PT** | **TC** |
| Number | 8 | 7 | 12 | 7 | 42 | 36 |
| Average age (mode) | 20–29 | 30–39 | 20–29 | 20–29 | Under 20 | Under 20 |
| Gender split (m/f) | 7/1 | 5/2 | 7/5 | 2/5 | 36/6 | 30/6 |
| Computer game players | 100% | 71% | 92% | 71% | 93% | 89% |
| Motivated to learn with games | 88% | 71% | 100% | 71% | 50% | 61% |

Table 9-3: Summary of number of students and characteristics of students allocated to each condition in the comparative studies

It can be seen from the table above that the majority of each group had experience of playing computer games, which supports the hypothesis that they are very likely to be familiar with gaming interfaces but, interestingly, only 50% and 61% of the students taking part in the Napier trial said that they would be positively motivated to learn using games, which might indicate that although

the students in this group are able to interact with computer games, and enjoy using them in their leisure time, they may not wish to do so as part of their learning experience. In the next section, the results of the student self-reported learning questionnaire are discussed.

## 9.3 Evaluating learning

At the end of each of the learning sessions with one or other of the game-based learning applications, the students were asked to complete a questionnaire that examined what they felt they had learned during the session (see Appendix 18). In order to consider the reliability of this questionnaire (i.e. that students were not just answering what they thought they were expected to) two dummy statements were added to the questionnaire that did not match the intended learning outcomes from the session. The two dummy statements were:

- How to be an effective leader.
- Ways of generating new ideas.

Table 9-4 below shows the responses of the students who felt that that they had improved on each of the questions in the questionnaire. From the total of 112 students who took part in the trial, only 10% (11 students) responded that they had not improved to all of the questions in the self-perception questionnaire, which could be indicative of the fairly basic nature of the learning outcomes. This provides some basic validation that the students themselves perceived they had leaned something from the activity.

It can be seen from this table that responses to one of the dummy questions (How to be an effective leader) is rated considerably lower than for the other factors (< 30% combined), which provides some evidence that the students are not just responding in a way that they feel they are expected to. However, the other question (ways of generating new ideas) elicited a higher percentage of positive responses (44% combined), which is comparable with the other questions. On reflection, it is quite possible that, while this is not one of the intended learning outcomes of the session, the nature of the tasks does indeed require the generation of new ideas, so this could, in fact, be an unintentional learning outcome and not reflect on the validity of the scale. However, it should he highlighted that these questions are, at best, indicative of learning and their

use here is primarily as a triangulation measure to support any findings of the engagement questionnaire.

| | Percentage of students reporting an improvement (%) | | |
|---|---|---|---|
| | **Time Capsule** | **Pharaoh's Tomb** | **Combined** |
| How to make good decisions as part of a group. | 28 | 48 | 38 |
| What makes communication effective. | 44 | 60 | 52 |
| Constructive controversy as a way of making decisions. | 42 | 35 | 38 |
| What makes a group effective. | 53 | 53 | 54 |
| How to be an effective leader. | 30 | 27 | 28 |
| How to communicate with others in the future. | 42 | 63 | 53 |
| Ways of generating new ideas. | 34 | 52 | 44 |
| The importance of group reflection for effective groups. | 40 | 44 | 41 |
| How to contribute to group decision making in the future. | 52 | 45 | 47 |
| The roles that people take in teams. | 48 | 47 | 47 |
| The benefits of collaborating with others. | 44 | 50 | 46 |
| Contributing to make group work more effective. | 46 | 53 | 50 |

Table 9-4: Responses to self-reported learning questionnaire from groups using the Time Capsule and the Pharaoh's Tomb

In order to determine statistically whether there was any significant difference between the responses from the students who had used the Time Capsule and the Pharaoh's Tomb, a Chi-squared ($X^2$) analysis was undertaken using the SPSS statistical software. In order that the data could be analysed using this technique, they were first re-coded into two categories: 'improvement' and 'no improvement' which created a 2×2 contingency table using the two experimental conditions; as each of these tests is undertaken with a 2×2 table, Yates' continuity correction is applied to prevent overestimation of statistical significance (Field, 2005), and a two-tailed test is used because it is not hypothesised that one particular condition will be more favorable than the other, simply that there will be a difference. The results are shown in Table 9-5 below.

These data show that there is no significant difference in self-reported learning on the majority of questions; however, two questions show a difference at the

0.05 level of significance: how to make good decisions as part of a group, and how to communicate with others in the future. Both of these questions showed a significant difference in the number of respondents who reported having learned from the Pharaoh's Tomb as opposed to the Time Capsule. It is worth remembering, however, that at a 0.05 level of significance, 1 in 20 tests would be expected to show a false significance by chance, so that while these results might indicate self-reported learning is greater with the Pharaoh's Tomb, it is important not to read too much into them.

| | $X^2$ | *p* |
|---|---|---|
| How to make good decisions as part of a group. | 4.006 | 0.045 |
| What makes communication effective. | 2.136 | 0.114 |
| Constructive controversy as a way of making decisions. | 0.260 | 0.610 |
| What makes a group effective. | 0.010 | 0.919 |
| How to be an effective leader. | 0.008 | 0.928 |
| How to communicate with others in the future. | 0.058 | 0.044 |
| Ways of generating new ideas. | 2.810 | 0.095 |
| The importance of group reflection for effective groups. | 0.034 | 0.853 |
| How to contribute to group decision making in the future. | 0.281 | 0.596 |
| The roles that people take in teams. | 0.000 | 1.000 |
| The benefits of collaborating with others. | 0.195 | 0.659 |
| Contributing to make group work more effective. | 0.325 | 0.569 |

Table 9-5: The $X^2$ and *p* values for each of the questions in the self-reported learning questionnaire

The respondents were also asked if they would like to work with the same team again, and 85% of students who had used the Pharaoh's Tomb and 82% who used the Time Capsule said that they would like to work with the same team again. While this does not show any difference between the activities, it does provide some evidence that the students valued working in the team they were allocated and would be prepared to work in the same team in the future.

The next section follows on from this examination of self-perceived learning as a measure of educational effectiveness, to consider the differences in the levels of engagement as a comparative measure.

## 9.4 Evaluating engagement

This section presents and analyses the results from the questionnaire that was used to measure engagement (see Appendix 22). Before examining the comparative levels of engagement between the two experimental conditions (students using the Time Capsule and those using the Pharaoh's Tomb), the Cronbach's Alpha statistic for each experimental group was calculated using SPSS. First the responses were coded on a scale of 1 to 5 (the choice of scale was arbitrary as the questionnaire was not used to produce an absolute score but to compare ranked data), and nine questions were re-coded so that the direction of the scale was the same for all. A summary of the results is provided in Appendix 26. Cronbach's Alpha is a measure commonly used to assess the internal consistency reliability of several items in a questionnaire that are intended to be summed to make an overall score; Alpha should be positive and greater in value than 0.70 to provide good support for internal consistency reliability (Morgan, 2004). The results are shown in Table 9-6 below.

| | **Cronbach's Alpha** |
|---|---|
| **Napier pilot** | 0.897 |
| **Hong Kong pilot** | 0.759 |
| **Napier trial** | 0.860 |

Table 9-6: Cronbach's Alpha score for each use of the engagement questionnaire

In all three cases Alpha is positive and the reliability level is well above 0.7, which would indicated that the questions are internally consistent (i.e. they are all measuring the same concept).

The responses to each question in the questionnaire were summed to produce an overall engagement score for each individual. To see whether there is a significant difference between the levels of engagement of those students using the Pharaoh's Tomb game and those using the Time Capsule learning activity, the Mann–Whitney statistical test was used. This is an appropriate non-parametric statistical test to use when a single ranked variable is being analysed with different participants in two experimental conditions (Greene & D'Oliveria, 1993).

The hypotheses of this experiment are as follows:

$H_0$: There is no difference in engagement between the two conditions.

$H_1$: There is a difference in engagement between the two conditions.

As $H_1$ does not suppose a difference in engagement in a specific direction (i.e. that a particular condition is likely to show more engagement), a two-tailed test is used. The Mann–Whitney *U* statistic was calculated for each of the studies, which translates into a probability (*p*). These are shown in Table 9-7, where $n_1$ is the number of students who were in the Pharaoh's Tomb condition and $n_2$ is the number in the Time Capsule condition.

| | $n_1$ (PT) | $n_2$ (TC) | *U* | *p* |
|---|---|---|---|---|
| **Napier pilot** | 8 | 7 | 19.0 | 0.296 |
| **Hong Kong pilot** | 12 | 7 | 40.0 | 0.865 |
| **Napier trial** | 42 | 36 | 693.0 | 0.531 |
| **Total** | 62 | 50 | 1491.5 | 0.732 |

Table 9-7: Mann–Whitney *U* and *p* for each experimental condition

In none of these cases does *p* approach significance (at the 1% level of significance this would require a *p* value of less than 0.01), and therefore the null hypothesis ($H_0$) should be accepted; there is no significant difference in engagement overall between the students who used the Time Capsule and those who used the Pharaoh's Tomb.

However, engagement was hypothesised to be made up of five factors: challenge, control, interest, immersion and purpose, and the questionnaire was developed so that specific questions mapped to each of these factors. It was therefore also possible to use the Mann–Whitney statistical analysis to examine whether there was a significant difference in any of the factors between the two experimental conditions.

Before the analysis was carried out, correlations between the questions that were hypothesised to measure each factor were carried out using the Kendall's Tau statistic for measuring correlations between ordinal data. In the case of each factor, every question correlated with every other question that was intended to measure that factor (at the 0.01 level of significance in every case

except for two of the questions that measured immersion, which correlated at the 0.05 significance level). A detailed breakdown of these correlations is shown in Appendix 27. This provides evidence that the questions in the questionnaire that aim to measure each factor are internally consistent.

In order to generate a data set large enough for there to be sufficient variation in each of the factors for application of the Mann–Whitney statistical analysis, data from all three studies were aggregated. The totals for each factor were summed and tested, again using the Mann–Whitney test for significant differences between the conditions. The *U* and *p* values for each factor are shown in Table 9-8 below.

| | Challenge | Control | Interest | Immersion | Purpose |
|---|---|---|---|---|---|
| ***U*** | 1493.5 | 1135.0 | 1354.0 | 1288.0 | 1524.0 |
| ***p*** | 0.739 | 0.014 | 0.242 | 0.119 | 0.878 |

Table 9-8: Results of the Mann–Whitney test to compare the factors contributing to engagement across the two experimental conditions

This shows that there is no significant difference between the level of challenge, interest, immersion and purpose between the two activities. However, there is a significant difference, at the 0.01 level of significance, in perceived control between the two activities, with the Time Capsule activity being rated more highly for control. This is a particularly interesting finding as it provides evidence that, while the Pharaoh's Tomb was designed to provide an environment with many options and objects that could be manipulated, students actually felt a significantly higher level of control using the Time Capsule application. This could be due to the greater complexity in the interface of the Pharaoh's Tomb, or the fact that the Pharaoh's Tomb required three-dimensional spatial navigation skills, which left a small number of students unable to move around in the environment, whereas the time Capsule did not require the players to master navigation or interact with objects in a virtual environment.

In all, this chapter has presented the results of a series of comparative studies to examine differences in educational effectiveness, measured through self-perceived learning and engagement, between the Time Capsule negotiation

activity and the Pharaoh's Tomb adventure game. The statistical analyses showed that there were no significant differences between these two game-based activities in terms of either self-reported learning or engagement. However, the difference in perceived control is important when considering the extent to which game-based learning activities need to exhibit all of the characteristics of games to be effective learning environments. In the final concluding chapter, the implications of this, and the other research presented in this thesis, are discussed.