Astroversity

A Futurelab prototype research report

by Mary Ulicsak

December 2004

CONTENTS

"[F]un is not a property of software, but a relationship between the software and the user's goals at that moment"

(Steve Draper: Analysing Fun as a Candidate Software Requirement, Personal Technologies, 3, 3, pp117-122 1999)

Transcript conventions:
{} - simultaneously being spoken
[] - actions occurring in parallel to speech

1. EXECUTIVE SUMMARY

Astroversity is a game for students aged 12 to 14 played in groups of three. It is designed to give the students an opportunity to develop collaborative and scientific enquiry skills. It does not teach these skills, but the task is structured so that an individual cannot complete it in one attempt nor is it easy to solve by memory alone. The most effective solution requires three students to work together, systematically recording data and coming to a shared solution.

1.1 The prototype development process

The project was developed iteratively. The International Centre for Digital Content (ICDC) and Futurelab developed a puzzle that was trialled by students on paper. Once it was agreed that the task was of suitable difficulty and could be made engaging ICDC created a two-dimensional version in Director. In an iterative process this was tested and revised with observations and feedback from the students used to refine the text and format. These findings were used to create a three-dimensional version, which again was iteratively revised. Finally, this was tested within a classroom context over a period of three weeks.

The details of the various stages can be found in Table 1.

Prototype stage	**Dates**	**Sample**	**Participants**
Concept development	August 2003		ICDC, Futurelab
Paper prototype (PP1)	Two trials - September 2003	Bedminster Down Secondary School	6 students: 3 in Year 8 and 3 in Year 9
2D prototype (1)	Two trials - October 2003	Bedminster Down Secondary School	6 students: 3 in Year 8 and 3 in Year 9
2D prototype (2)	Three trials - October 2003	Bedminster Down Secondary School	9 students: 6 in Year 8 and 3 in Year 10
		Futurelab	3 staff
Review of toxins	October 2003		School of Chemistry in the University of Bristol, Futurelab
Story development	November-December 2003	Holly Lodge School	ICDC and students
2D prototype (3)	Two trials - December 2003	John Cabot City Technical College	6 students in Year 9
		Futurelab	6 staff
3D prototype (1)	One trial - January 2004	John Cabot City Technical College	Year 8: 2 groups of 3 girls, 4 groups of 3 boys, 1 boy working

			with 2 teachers
3D prototype (2)	Two trials - May 2004	Cotham Secondary School	2 Year 9 boys; 4 Year 9 boys and 1 girl
3D prototype - final trials	Three trials - June/July 2004	Monks Park	Year 9: 8 girls, 10 boys (in 6 groups)

Table 1: Summary of activities

In addition to these formal trials Futurelab sent comments on the Astroversity software to ICDC after each release.

1.2 Non-technical prototype description

The final prototype is a game for three 13-15 year-old students playing on a computer network. In the scenario they are students at the Astroversity, an orbiting space academy. They are currently undergoing search and rescue training, ie learning how to control a probe, infer information from sensors, and plot routes for a rescue vehicle. During this exercise an alien vessel crashes into the Astroversity releasing three toxic substances into the atmosphere. Bloppo causes brain swelling and eventual explosion. Moob increases the heart rate leading to a cardiac arrest. In contrast, inhaling gunk causes liquids to collect in the lungs leading to death by drowning. The robot headmaster instructs them to form groups of three and rescue their peers on various levels of the academy.

The most efficient way to rescue the casualties is to work together to identify a strategy for recording data, record this accurately, and in the later levels divide the sensors among their individual probes to cope with multiple toxins. Communication skills are enhanced as they must explain their findings, route suggestions and theories about toxicity levels and the relationships between the toxins to each other. They are aided by their chosen robotic tutors who offer suggestions and question them on strategy, use of data, and their performance as a group.

The groups can repeat the rescues in order to effectively save the casualty.

1.3 Key innovations of Astroversity project

The following aspects of the experience are novel in games for learning:

- the use of multiple methods of representation, requiring students to switch between a virtual online world and a paper-based representation which they create as a consequence of exploring this world
- the requirement that team members do not simply work together by fulfilling different roles in the same task, but have to contribute information to a single activity with a collective outcome
- the explicit encouragement of students to self-assess and reflect on the skills being developed within the task.

1.4 General findings from Astroversity research

The following are the key findings from the studies:

- students need a structured environment when practicing group skills or performing a task requiring scientific enquiry as these are not instinctive skills
- self-assessment is not performed reliably when not mediated by an external person

- well designed games for learning can be motivating and engaging, shown by more students turning up for the final session with the three-dimensional than the first despite the fact it was end of term.

1.5 Key learning findings and recommendations

1.5.1 For researchers

- a task can be structured to encourage reflection but direct written prompting is often ineffective
- non-textual interfaces are important
- the findings are highly dependent on the method of assessment; clear questions need to be set, are we looking at improved technique, engagement, motivation, or use of context?
- if Astroversity was to be a useful research tool then data logging is required to observe changes in strategy, performance and responses to reflection prompts.

1.5.2 For teachers

- Astroversity needs to be supported by a teacher or expert student, written online instructions about the concept and strategies appear to be ignored
- fitting in with broad curriculum aims does not limit the enjoyment of the activity within the classroom
- learning gains may be associated with listening and talking skills more than intended scientific enquiry
- the lack of recording of data means that progress cannot be reviewed and the students will have to be asked to record progress in terms of damage to casualties.

1.5.3 For ICDC

- as found in commercial games providing 'training' before the mission is beneficial
- Astroversity needs to be more portable; despite recommendations for machine specification (which is higher than some schools currently have), the software had difficulty in collecting machine addresses and setting up a network
- although the issue of gas diffusion has been glossed over the system still needs to improve the underlying physics - it is currently possible to make probe go through floors with sufficient velocity
- team selection process requires improvement - currently it is hard to identify what team you belong to once selected
- identifying current sensor selections, for example, what sensor has been chosen during and after the exploration and what it impacts
- the original proposal was to develop Astroversity into a Playstation game, further studies are needed to show that the concept of using physical and virtual maps works outside a classroom environment before this is commenced.

1.5.4 Policy makers

- the importance of developing a cross-curricula tool - Astroversity has a science focus but as observed by an English teacher, it can also be seen as a listening and speaking activity
- provision of a non-text based science activity benefits all students but in particular those with poor reading abilities
- games are not necessarily negative, although little improvement in strategy was noted the engagement and motivation of the students to attend lessons improved.

1.5.5 Industry

- studies show that schools are willing to have non-traditional tasks within a syllabus setting
- games can be overtly educational and still enjoyable within a formal setting
- the possibility of mixing various representations does not detract from enjoyment - but no information exists as to how this would transfer to an informal setting.

1.6 Futurelab recommendation for the next stage of Astroversity

As it stands Astroversity requires more puzzles before it becomes a viable commercial game. The puzzle implemented is engaging but limited. It is therefore suggested that minor changes be made to the current version, they are:

- addition of teacher guidelines
- amending narrative so that the robotic tutors are external and therefore cannot perform the task
- amending graph to show likelihood of survival rather than damage.

This is then made widely available so that anyone can download the instructions, paper map grids, and software and play it on their school networks. Information about Astroversity as a research project would also be released and if wished trainee teachers and other researchers could use it to investigate how to support collaboration, scientific enquiry etc in the classroom.

Finally, it would be interesting to do a comparison of the two-dimensional and three-dimensional versions of Astroversity. It appears that task focus occurs earlier in the two-dimensional groups though both are poor at systematically recording data. Groups of students could alternate so each do a different level in a different format. One research focus would be on the type of talk, apparent engagement and strategies in each environment leading to a discussion of the overall favourite format and why. Another research issue is the impact of the shared environment, ie seeing each other in landscape? It appears motivating but the studies indicate it does not lead to an increase of recording data or speed at task although the difference in the current format, you cannot see the others in the two-dimensional version, means further work is needed for a comparison to be made.

2. CONTEXT, CONCEPT AND OVERVIEW OF DEVELOPMENT

2.1 National Curriculum context for Astroversity

The English National Curriculum provides a context for the Astroversity project in setting out a requirement for young people to learn skills of scientific enquiry and speaking and listening at Key Stage 3. Scientific enquiry can be further defined as 'constructing explanations, interpreting evidence, defending and challenging claims, using and developing models, transforming observations into findings, and arguing for or against particular theories'. In respect of speaking and listening the curriculum requires young people to develop skills in group discussion and interaction (further information about the National Curriculum requirements and how it has been integrated into Astroversity can be found in the context paper in Appendix A).

2.2 Astroversity concept and research agenda

Astroversity was designed to address the specific question:

Is it possible to create a game that achieves the same level of engagement as mainstream games, but which enables children to learn about and practise collaboration and the processes of generating and testing hypotheses?

This question was generated from the hypotheses that games can act as intrinsically motivating environments, and that young people do not instinctively develop scientific enquiry and collaboration skills.

The question is addressed through the development of a game environment based on the principle of guided discovery. The software was designed to support the development of: (a) group skills, by providing a situation in which a solution can be found faster and more easily if the group members provide information in a way that can be understood by their peers; and (b) scientific enquiry skills, by requiring a knowledge of the environment to solve the task which can be most easily done through systematic exploring and recording then analysing the findings to determine the impact of combining the elements. It is not intended that the students should learn any specific 'factual information'. The system offers suggestions for methods and prompts the group to reflect on their strategies, but there is no 'right' answer.

In addition, all Futurelab prototypes are evaluated against three generic research questions, these are:

1. What does this prototype tell us about the best ways of designing digital resources for learning?
2. What does this prototype tell us about how informal learning processes can be transformed through use of these tools?
3. How does this prototype help us understand the potential of next generation technologies to create intrinsically motivating and engaging learning experiences?

2.3 Project team

Table 2 shows the people and their respective roles in the Astroversity project.

Name	**Title**	**Role**
Jim Turner	Content Manager, ICDC, John Moores University	Developing content and managing development of software
Mary Ulicsak	Learning Researcher, Futurelab	Developing content and managing software testing and analysis
Ged Lee	Headmaster, Hurst High School, Northumbria	Responsible for initial idea generation at ICDC and bringing Criterion onboard although not involved in Astroversity development
Professor Angela McFarlane	University of Bristol	Concept advisor
Dr Richard Joiner	University of Bath	Concept advisor
Adrian Hall	DfES	DfES provided direct funding for development of Astroversity through Futurelab

Table 2: People involved in Astroversity

Special mention also goes to Matt Southern, formerly at ICDC for his work on the initial proposal, and the teachers Ian Rodgers, Justin Slowey and Lorraine Stolarczyk for their participation in the various software tests.

The background story to Astroversity was developed by ICDC working with students from Holly Lodge School, a low achieving girls school from the local area. The Futurelab team focused on the task content and usability trials.

2.4 Background to scenario and game task development

Astroversity is an extension of the VMULE (Virtual Multi User Learning Environments) project. This was designed by ICDC and trialled by Futurelab in 2003. Building on this work, an initial proposal was generated for a scenario called October Island, a volcanic island in which a group of volcanologists have been trapped after the volcano erupted. The students were to work as a team to rescue the trapped scientists. In this process they were to be given distinct roles: navigator, sensor control and medic, through their choice of probe. During the rescue they could leave information for their team-mates in the island and rendezvous at headquarters to develop their strategies. This scenario was not completed due to the development costs involved in creating the island's environment with the non-standard elements, as shown in Figure 1 - although the principles of a collaborative game environment were maintained.

Figure 1: October Island environment

The next scenario considered was a chemical factory; this was then changed to a space station setting. The advantage of the latter, in addition to not giving chemists a bad name, was that the toxins could be made fictitious (the reasons for which we discuss in Section 3.3).

There were a number of constraints operating on the design of the task itself. First, the puzzle had to be perceived as an authentic task, in other words, that the data gathering and interpretation tasks should be seen as intrinsic to rather than an 'add-on' to the games play. The puzzle needed to require students to perform and interpret a data gathering exercise, use this to construct or revise a theory or explanation, and communicate or negotiate it with others in the group. The puzzle, moreover, needed not to be prescriptive. Rather than having an experiment presented and explicit instructions about what data to collect the students had to be free to construct their own method for solving the problem (although advice should be available).

2.5 Final scenario and task

The task in Astroversity appears simple, three students work together to rescue a casualty from what appears to be a large empty room. This room contains one or more colourless heavy fictitious gases that harm humans, with the only method of detection being a probe calibrated for that gas. The students must identify a path for a rescue vehicle to take so that the health of the casualty they are rescuing remains stable. In the puzzle each student selects a probe with one sensor so can identify the location of one type of gas. The students must determine the density of the gas for it to be toxic and how they will find a safe path, to do this they can: plot a safe route, mark danger areas, change the area that the sensor detects, make several explorations, try and solve the task in one iteration etc. In later levels the task becomes more complex as the further toxic gases introduced interact with each other. In these levels the students can choose to operate probes which detect the presence of different gases.

3. PROTOTYPE DEVELOPMENT

Astroversity was developed over the course of a year. This chapter details the major stages and findings from this process based in Bristol. They are included to show the iterative nature of the design and track the key detailed findings from the sessions with students.

Prototype stage	Dates	Sample	Participants
Paper Prototype (PP1)	Two trials - September 2003	Bedminster Down Secondary School	6 students: 3 in Year 8 and 3 in Year 9
2D prototype (1)	Two trials - October 2003	Bedminster Down Secondary School	6 students: 3 in Year 8 and 3 in Year 9
2D prototype (2)	Three trials - October 2003	Bedminster Down Secondary School	9 students: 6 in Year 8 and 3 in Year 10
		Futurelab	3 staff
2D prototype (3)	Two trials - December 2003	John Cabot City Technical College	6 students in Year 9
		Futurelab	6 staff
3D prototype (1)	One trial - January 2004	John Cabot City Technical College	Year 8: 2 groups of 3 girls, 4 groups of 3 boys, 1 boy working with 2 teachers
3D prototype (2)	Two trials - May 2004	Cotham Secondary School	2 Year 9 boys; 4 Year 9 boys and 1 girl
3D prototype - final trials	Three trials - June/July 2004	Monks Park	Year 9: 8 girls, 10 boys (in 6 groups)

3.1 Paper prototype

The puzzle was developed iteratively using an informant design approach. The initial puzzle described here was designed to investigate whether the students already possessed the scientific enquiry and collaborative skills that the intended prototype was trying to support; and if not, whether the proposed puzzle format supported their usage. In this iteration real toxins were used at the actual levels that they became toxic and real side effects were described.

Format:
This paper prototype comprised of three paper grids containing various numbers representing the density of ozone, carbon dioxide or heat in an area respectively. Each student had a sheet describing the levels needed for one of these substances to be dangerous for a human and the outcomes. The task was as a group on a separate map to plot a safe path from the start to the finish. Students were given a selection of coloured pencils and asked to work by themselves for five minutes on their own grid to get an overview of their data, see Figure 2. They were then

given the blank grid and 25 minutes to combine their findings on a single grid to plot a safe route.

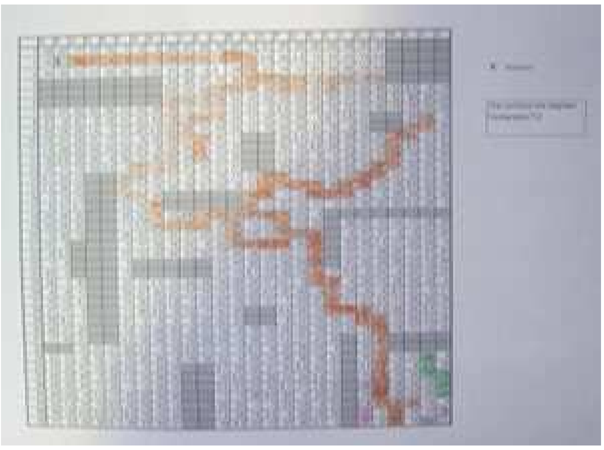

Figure 2: Student's attempt to identify safe route avoiding heat

Trials:
Six students, one group of three from Year 8 and one from Year 9 from Bedminster Down School were involved in these trials. They were observed during the task and given a short questionnaire to complete afterwards.

Findings:
The task set in this challenge is equivalent to having three toxins, thus the most effective method would be for the individuals to mark the danger areas for each toxin then determine a route that avoided these when the group came together. However, as shown in the example in Figure 2 the method employed by all students was to try and colour in a safe path on their map, starting from the marked 'Start' point and working towards the 'Exit'. Even when coming together and realising that the routes selected by each student were very different they did not change their strategy. Instead of individually identifying danger areas both groups continued to plot a route with one group member colouring in asking about each square, which they referred to after prompting by grid reference. These sessions suggested that students were not yet used to employing scientific enquiry strategies but could be systematic.

The paper study was also designed to observe current collaboration skills. The students did work together but at a cumulative level, ie, they took it in turns to contribute ideas and all participated but there were few explanations asked or offered, as shown in Transcript 1.

S: *Can you go across?*

L: *Hang on!... Can you go W36?*

S: *Yes*

L: *R...?*

R: *Yes*

L: *W... 35*

R: *Yes*

Transcript 1: Example of discussion in paper puzzle

When asked what sort of skills the puzzle developed in the short questionnaire the first group felt it was mental ability ("brains"), concentration and listening, while the older students felt it was teamwork. The students were also asked how well they listened and were listened to; three students believed they were listened to more than they listened, one felt that they listened more than listened to and the remaining two felt they listened and were listened to equally. This discrepancy did not add up in either group. This suggests the students are not used to reflecting on behaviour.

At this level when asked the appropriate age for the task the students disagreed. Those in Year 9 felt it was suitable for Year 9, while the younger students felt it was suitable for Year 6, Years 6 and 7, and Years 6 to 10.

Conclusions and recommendations:
In conclusion, the paper prototype testing showed that the students were motivated by the task - they wished to return to it after completing the short questionnaire. It showed that although they were systematic they did not apply scientific enquiry skills as they worked incrementally rather than thinking about the task. Moreover, although the students believed they needed collaborative skills (teamwork and listening) to complete the task it showed that the students' self-assessment needed supporting as the amount of listening was not in agreement.

3.2 Two-dimensional prototype

The next stage in the project's development was the creation of a two-dimensional prototype written in Macromedia Director. There were three iterations. Again the toxins were carbon dioxide, ozone and heat and the levels and side effects were described accurately. They are described in turn in this section.

Format Iteration 1:
This iteration consisted of one level all with three toxins present. It was possible to hypothesise about the toxins as each had a central high area with readings decreasing as the probe moved away from the central area in all directions. There were three stages. First the students could select any of the three probes for detecting the three gases, and read about the side effects. Then, also online, they could explore, this had no time limit and they could change the size of the scanner, allowing them to observe the average gas reading over one square, nine, or 25 as shown in Figure 3. They had blank grids to record their findings in this stage. After five minutes the team was given another blank paper grid to record their findings and had to come together to plot this final path on paper. They could refer back to the computer when they required additional data.

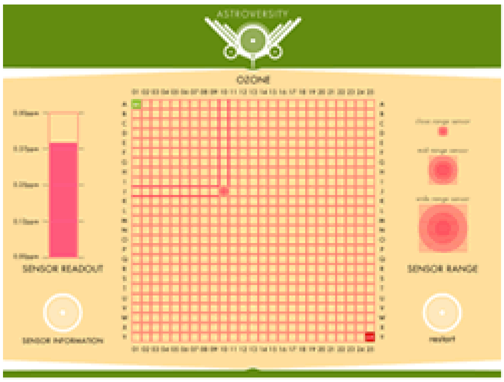

Figure 3: Iteration 1 - two-dimensional exploration of ozone levels

Trials:
Six students, three from Year 8 and three from Year 9 from Bedminster Down School tested the first two-dimensional version. Field observations were made, the paper maps were collected, and the sessions were tape-recorded. After this they were given a questionnaire and there was a debriefing session.

Findings:
As with the paper version the students worked individually until given the blank map - sometimes they did not talk at all, and one group had students choosing the same sensors. No student was observed using the larger sensors although two said they tried it and thought the

computer was broken. While plotting the final path again there was no defined lead role but whoever was marking the final map tended to ask the question about where was safe. In this case the values were not random, and one could hypothesise where the source of the gas or heat was from the surrounding values. However, there was no hypothesising going on by the students, they still tested square by square, but in the electronic version they automatically used grid references to identify where they were.

Year 9 thought the size of the grid was too big: "there were too many squares", yet Year 8 who managed to get much further through the maze did not comment this on. This suggests these responses were probably ability related. (Although it is interesting that the Year 9 students did not change the scanner size which would have reduced the area to be explored.)

Again the progress was incremental, with students trying to identify a safe path avoiding their hazard. There were fewer marks on the paper with the electronic version, as the students appeared to be trying to identify a safe route before recording findings.

This iteration confirmed that the students needed support in developing strategies for recording and constructing routes. Students agreed who would take which sensor, but not what would be recorded and they often did not realise they were looking at the same area. They did not automatically change scanner size; even when continuing after the interview when the purpose had been discussed. They still used the smallest size and continued testing square by square. It also highlighted that students were poor at reading the graph on the left of the screen, as they did not record what level was toxic. However, the onscreen representation made referencing easier, and they used grid references without prompting. There was no problem working on paper and on screen; they actually moved the paper round and used it to represent their progress.

Unsurprisingly the software was seen as something to be done in school, but in the paper-based debriefing only the boys felt they worked better as a group then they normally did. Only one girl felt that she listened and was not listened to - although she took the role of filling in the map and spoke most.

Recommendations:
The suggestions for the next iteration included:

- a reduction of written text. The students just wanted the answer at the end, this detracts from the enquiry side of the task but suggests that limited prompts about tackling the task would benefit the student
- being able to check information such as what sensor they have chosen.

Format Iteration 2:
Given these findings the software was revised in the second iteration. Again all three toxins were present but there were two levels; in the easy level there were multiple paths through that avoided the toxins, in the second only one path. The level and exploration time could be set before the game was started. The amount of time available in the first phase, exploration, was displayed on the screen. The exploration process, scanner sizes etc were identical to the first iteration, including individuals using paper maps to record their initial findings from their individual probes. However, now the rescue route could be plotted from one machine instead of on paper, as shown in Figure 4a. This meant feedback could be given, as when the rescue was executed the damage status as a result of the chosen route was displayed. A graphical screen further illustrated this; from Figure 4b below it is possible to see by the fifth waypoint the casualty has died.

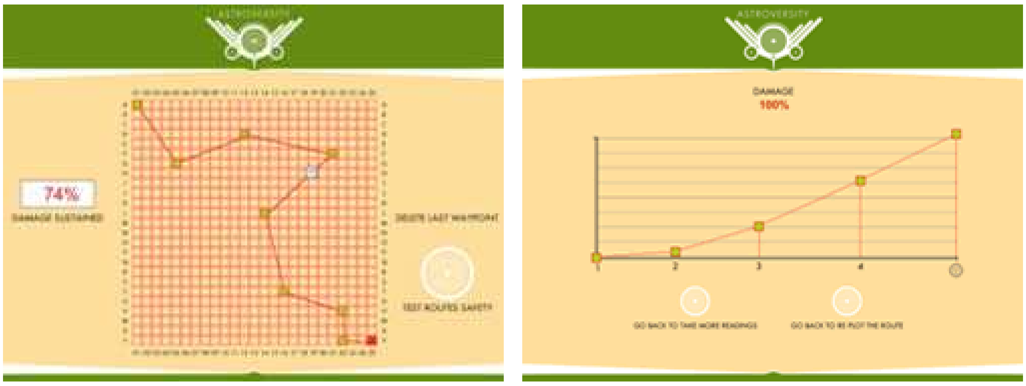

Figure 4: a) Way plotting screen; b) Graph of patient health during rescue

Trials:
Two Year 8 ICT groups (middle ability) and one Year 10 science group (low ability) from Bedminster Down School and three Futurelab staff tested this version of Astroversity. The paper maps were collected at the end and they were observed and tape-recorded. The students were asked to complete the same questionnaire as the previous iteration. Each trial lasted approximately half an hour.

Findings:
The majority could switch between paper and screen but were initially unsure about how to represent the data. The strategies that they used remained incremental, ie, marking safe cells in order. This suggests that scaffolding is required about alternative methods, such as marking only danger areas. When it was suggested that they mark unsafe rather than safe areas they proceeded to do so, without discussing or comparing other approaches. This suggests that there is a need for guidance.

The student groups all took at least three attempts to get a damage score less than 100% but this complexity was not detrimental. All were enthused by the task and all but one girl would like to play again - she found it too frustrating. The Year 10 students assumed that because the route they identified as individual was safe the survivor would be OK and took it in turns to suggest routes - regardless of the information about other substances recorded by their teammates.

One Year 8 student felt it was suitable for Year 9 to Year 13 students, one Year 8 and Year 9 and the last Year 7 to Year 12. One Year 10 student felt it was suitable for Year 7 to Year 12, the others Years 9, 10 and 11.

As before there was the same issue concerning listening. Six of the nine students believed they listened more than they were listened to, suggesting that asking students to reflect is beneficial in highlighting discrepancies.

Conclusions and recommendations*:*

1. The text was too complicated - it described the impact of freezing or burning for example in graphic detail. This implied that the initial amount of text should be reduced and a 'More' button introduced for those interested in the side effects.
2. Giving the students three minutes to use a probe initially so they could attempt the way plotting stage, reason what data they needed, then go back to the data collection phase with a longer time of five minutes was highly successful.
3. The timer was motivating as it gave a sense of urgency.
4. Guidance about strategies could be from a teacher or online expert, another role they could fulfil is to prompt for reflection on the data found. No student hypothesised or formulated a search strategy or that there might be a centre of the heat/gas source - whereas Futurelab staff did. This implied suggestions for how to approach the task would be useful, but not until after their first exploration. These prompts could include:

- being systematic when going through the environment - some paper maps only had several random blocks coloured in, implying reminding the students that recording all data would be beneficial
- stressing that multiple dangers exist in the feedback would help
- consider changing scanner size and use larger scale to identify areas for further investigation
- hypothesise about where the sources for heat, cold, ozone etc were.

It was suggested that examples of methods be given pictorially as well as in written form as the reading level of some students is poor, or that such techniques could be presented by the teacher prior to the exercise.

5. The grid references were useful and enabled the students to easily refer to the grid.
6. One group put the highlight was killing the casualty - they were the least systematic and just plotted points after the first probe mission rather than gathering more data. Another put killing the casualty as being the worst part just for balance. This highlights the benefit of giving students the opportunity to choose the purpose of the game - be it death and destruction or assistance, as it leads to engagement.
7. The students requested more animation of the face in Figure 4a when looking at their progress - although this was a great visual and they were counting aloud, even the Year 10 students, as the values decreased.
8. In this format the puzzle appeared to occupy an area between being a game and being homework.

Format iteration 3:
In the final two-dimensional iteration the software was further revised. Flexible timing was built into this version, that is, the time for exploration could be set prior to using the software. Three levels were introduced; Level 1 had one casualty and the a single toxin (ozone), Level 2 had one casualty and ozone and carbon monoxide toxins and the third level had two casualties and the toxins heat, ozone and carbon monoxide. Although six robot tutors were intended Dr Matius Stalker (Figure 5) was the only one implemented that offered advice on strategy, for example, on Level 1 he suggested that the students split the area. He introduced the task and the sensor types - although only the sensor that was valid for that level was available. The feedback was similar, but extended to show damage to the rescue vehicle as well as casualties, as shown in Figure 6a. After the rescue Dr Stalker gave advice based on performance, as in Figure 6b an early death was followed by the instruction for the group to share information for example.

Figure 5: Dr Matius Stalker

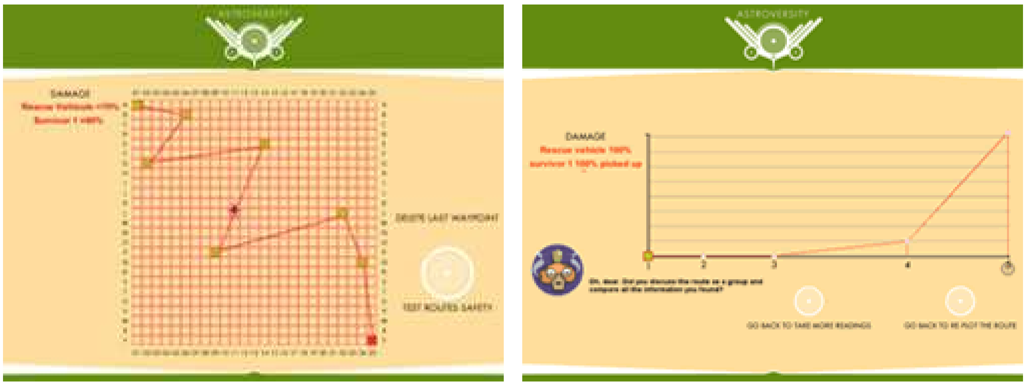

Figure 6: a) Revised way plotting screen; b) Graphical feedback

Trials:
Six people from Futurelab, and six Year 9 students selected by their teacher from John Cabot City Technical College tested this version. Those from Sir John Cabot were video recorded. The students were given the same questionnaire as those in the previous iteration.

Findings:
The students spent a minimal time responding to the questions given after they had finished playing although had as long as they liked. A typical response to the questionnaire by one of the Sir John Cabot groups is given in Table 3.

Question	**Student 1**	**Student 2**	**Student 3**
What was the amount of time you thought the others listened to you?	Half the time	One quarter of the time	Half the time
What was the amount of time you listened to the others?	Half the time	Three quarters of the time	Three quarters of the time

Table 3: Responses to self-assessment questions

The results showed that students believed they listened and were listened to all of the time or, as in Table 3, that they listened to more than they were listened to. In this example when asked to explain how this was possible Student 1 said they had "just put something"; Student 2 initially could not explain his answer, then justified his opinion by saying his teammates had not listened because they had not followed all of his instructions; while Student 3 eventually decided he had not listened to himself. These findings suggest that students do not think that these questions when written merit reflection. However, the act of explaining these discrepancies does cause reflection and it is hypothesised that going through the process of having to reflect then provide an explanation via external speech should lead to greater self-awareness and hence improvement in performance.

Although the greatest engagement appeared to be when following the route selected for the casualty where the students were concerned about the health score, there was little reflection about what damage meant. This was tested by looking for discussions about: "what's the difference to a person between 20% and 25%? Where is the advantage for taking a longer route that inflicts less harm than a shorter one?"

Two students believed it was suitable for Year 7 to Year 9 students, one Year 7 to Year 10 students, one Year 6 to Year 9, another Year 8 to Year 11, and Year 9 to Year 11. Apart from one student all believed that it was suitable for secondary school pupils.

Conclusions and recommendations
The key findings, any recommendations, and whether these were incorporated into the three-dimensional version are given below.

1. None of the students finished the task - but were keen to continue, despite the difficulty of the final level. This suggested that the task is motivating.
2. There is a lack of reflection on group skills and strategy until the final questionnaire given by the researcher. Astroversity could mechanise this questioning, self-assessment, and raising awareness of discrepancies and thus increase the times used. Based on individual responses the system could offer procedural prompting, that is, domain independent, generic or open questions such as "why are your opinions different?" as well as advice if the group agree that they did not have a good strategy. Wegerif believes that computer prompts are effective because they lead students to "engage in uninhibited debate amongst themselves between the prompts and responses of a computer in a way which would not be possible with a directive teacher" (1996, p53). Areas that would benefit from reflection include:
    - how good they were at recording the data?
    - how effectively did they use the data?
    - did they think their strategy for plotting a route was any good?
    - did they understand what they were meant to be doing within the team?
    - how much of the time did they listen to each other?
    - how much time do they believe they were listened to?
3. The software could then provide feedback consisting of comments or prompts asking for justification or highlighting differences in opinion, as in the session at John Cabot School about the amount of listening. Ideally the scores should be saved for later discussion. This was incorporated into the software but not made integral to the game, the results did not impact time given, score for rescue etc.
4. Further feedback is needed to support the students develop a systematic approach.

3.3 First three-dimensional prototype

At this stage it was agreed that alien fictitious toxins would replace heat, carbon monoxide and ozone. The reasons for this were:

- in order for the scenario to be realistic the gases would have to diffuse, this level of modelling meant building in complex formulae for diffusion rates according to the time and the movement of the probes which was outwith the scope of the funding available for this prototype
- the impact of the gases on a human body is often small and dependent on length of exposure as well as volume.

Alien gases had the advantage that they could be portrayed as static due to their heaviness, and that we could limit the impact to one area of the body making an analysis of what toxins were present easier and setting up relationships between the various gases.

Format Iteration 1:
In this iteration, each level of Astroversity has three stages: search, planning and rescue. In Level 1 there is only one toxin, bloppo. The students are expected to choose the same sensor for the probes and can decide how to record the data from the probes, whether to record safe or dangerous areas. The strategies of recording safe areas or recording dangerous areas will both work with only one toxin to consider. In the second level two toxins, bloppo and gunk, are present and students choose which sensors they want to use. Recording only a safe path is ineffective, as what is safe for bloppo may not be for gunk. In this case a more effective strategy is to record the dangerous areas. In the third level all three toxins are present and interact with each other. To rescue the casualty efficiently the students need to interpret the graph (which shows interactions of gases on the human casualty) as well as have a strategy for recording levels of gases. The safe path involves realising that the toxins interact, and that the casualty must be taken through the toxic areas in a specific order to be rescued safely.

When registering as a team the students were given a team name to select and one player was nominated as having the machine for way-plotting by the researcher before the game commenced. During the rescue the students are given a choice of six tutors who all gave advice: Prof Thingamy, Captain Bolt, Miss Appliance, Principle Diode, Prof Widget and Doctor Maestro. The type of advice can be logical, such as "divide the area", illogical, "I think you should be the leader", and contradictory - "choosing the shortest route regardless of exposure" or "creating a slow route by having more way points but avoiding all toxins". The students could choose separate tutors. At the end they were asked about the amount of time they spent listening to each other by the system rather than by the researcher. The recommendation that the tutors ask them to reflect on the graph, for example, what did 25% damage mean, was not incorporated. However, during the rescue stage initial comments from the two-dimensional version (see Section 3.2, Format Iteration 2) about the need for animating the head when watching progress led to the development of animated icons during the rescue phase.

During the rescue the student can see the position of their probe and sensor readings as well as the location of the Rescue Vehicle (RV) in the map on the bottom right. They can also see the casualty status - which is shown in the top right of the screen - see Figure 7a. The reason for the number of icons is that initially more than three toxins were to be introduced and this would be reflected in the readings. The feedback screen has similar information to the two-dimensional version but has the various parts of the body that could be impacted displayed, as shown in Figure 7b.

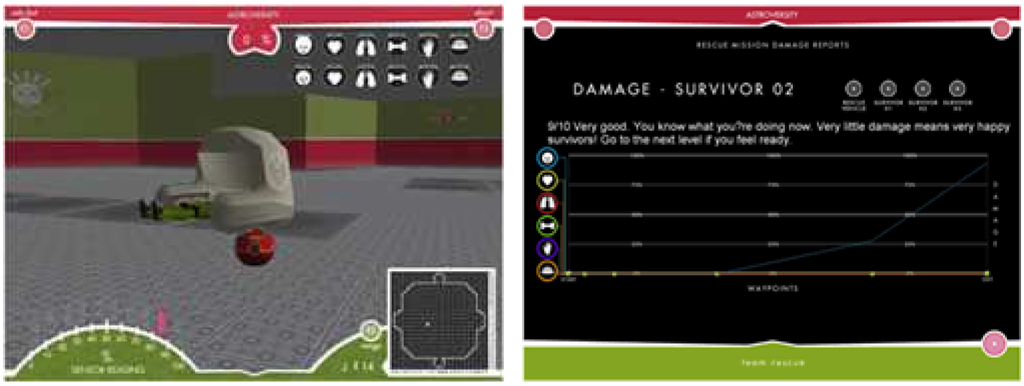

Figure 7: a) Rescue in progress; b) Graphical feedback (with spelling error)

Trials:
In January 2004 the software was tested with a middle ability Year 8 science class at John Cabot City Technical College. The players consisted of two groups of three girls, four of three boys and one boy with two teachers. The software was used for approximately 35 minutes. The students were given questionnaires (see Appendix D) and field notes were made. They were not asked about the age of the intended audience in this trial.

Technical difficulties:
The machine specification was sufficient but the trial highlighted the difficulty in having Astroversity run over an existing network. Technicians from Futurelab and ICDC both visited John Cabot CTC in order to set up Astroversity. Eventually the game was hard coded to force the network to be recognised. However, this still frequently caused the machines to crash once the waypoints had been plotted. The students were tolerant but frustrated by not being able to tell if their route was safe.

Findings:
The key findings from this trial highlighted that most students do not read on screen text. The researcher observed that four groups chose different sensors on level one despite only having one toxin present. In addition, when exploring there was confusion as to the purpose of the task, as shown in Transcript 2.

B1: *Everybody click more*
B3: [Reading instructions]

B2: *Go... go... go*
B1: *Oh next*
B2: *Wow, it's gone all amazing*
B3: *How do I move?*
B1: *There's a dead body - what do we do with it?*
B2: *Run it over*
B1: *I'm going to get you*
B2: *No way*
B1: *Bang! Have I got a gun on mine or something?*
B3: *What do I do?*
[Somebody external said read instructions]
B3: *I was reading them but he said to carry on...*
B3: *What shall we do with him?* [referring to body]
B2: *I think we should stand on him and all press enter*
B1: *How do you shoot I wonder?*

Transcript 2: Discussion within a group of boys starting Astroversity exploration phase

The session confirmed that students are poor at collaborating to achieve specific goals. There is plenty of talk, but the tape recording suggests that the groups are not very good at listening or helping each other. For example, one group called the Mad Medics believed they were doing pretty well, they liked that they could all be together on the same computer. However, the boy at the centre computer let his companion struggle to read the grid reference on the small displayed map instead of encouraging him to look at the large reference by the grid that gave this information. This could be because he was unaware of his companion leaning over to see what he was doing or just felt that he should realise. There were examples of team work, such as locating the body. At first they could not agree where the body was (U14 or V14), so they agreed to all meet at the body and compare readings - by all 'standing' on the body they decided it was U14. There then followed the exchange below:

B7: *Put an S for survivor*
B8: *No just colour in the box*
B9: *Put a 1 in the box... put sensor*
B8: *Now what do we do?*
B7: *There's one on the other side - turn around* [all laugh]

Transcript 3: Example of discussion

Despite this discussion when this group's grid sheets were examined they all had different methods for recording data. (Immediately after this extract this group got their probes to chase each other into the room at the end and try and jump, they also discovered camera views and the Rescue Vehicle in this play time.)

During this trial there were different techniques observed for playing Astroversity, usually involving simultaneous speaking - one may be reading the screen out loud, another saying that they should all press the button now, while the third is repeating "go, go, go". Similarly there are examples of groups that fail to listen, for example, one of the boys was giving instructions, such as "go through the door", they did, then the second wanted to know what they were meant to be doing, only one lad realised and said they needed to go round and take sensor readings. Another boy going "I've found it, I've found it" then interrupted them all.

The transcripts were analysed for group skills, such as justification, these were found but were often quite defensive, as in Transcript 4 describing their choice of route:

B10: *It's there you dickhead, there there there.*
B11: *Delete last waypoint*
B12: *No no no that's cos where all that gas is*
B11: *But we've got to go there to get all the people*
B10: *Exactly, so that's where all the gas is*
B11: *It's right a bit*
B12: *But if you go round the edge*
B10: *But that's where we started going all the way round, see S21 is here*
B12: *No that's in this corner as I started mapping it out... Can you start from that... No let's go round the edge*
B10: *But there's bloody gas up there*
Transcript 4: Working together to plot the waypoints

All the students (and teachers), and not just the Mad Medics, were poor at recording data. There did not appear to be any systematic means of searching through the level or dividing the area into sectors that they could all search. Nor agreeing a method for recording levels, safe and dangerous areas. When asked initially some said that it was because they had chosen the wrong sensor - but then did not seem to record in later versions. This did not seem to impact enjoyment.

Strategies for recording data were limited, one was marking danger areas: "but they didn't know it would be so much of it". They did exchange information, one group agreeing, for example, that the middle of the screen was full of gas. But this did not transfer to actually recording data. Although strategies were mentioned - "let's record all the places where the gas is less than 20" they weren't followed. Discussions overheard were sensible but often without conclusion, for example: "is a green reading safe?" They would announce cells to their teammates but this did not appear to translate into recording: "No just write the grid!". Another recorded strategy discussion was an argument about the logic of going round the edge, where one group argued that it was taking too long.

None of the groups reflected on the questions given at the end, they appeared separate to the task. The students answered and moved on - although no logs were available the time observed with the screen open was minimal. Some groups appeared to choose randomly, others answered as a group rather than individually - if this was the case they all gave the same answer. They did not answer the procedural prompts at all, preferring to continue with the game. However, there was a brief recap of what they should be doing before starting again: eg "We need to find the gaps". Then they reflected to produce some form of strategy: eg "Start in the corner and follow me", "Let's try that again and do it better".

Finally, the graphics during the rescue were well received, including: "Oh his brain's swelling up" "Just look at that brain - wow". However, the graph - an example is in Figure 7b - needed prompting to be viewed. Students tended to look at the score, eg 9 out of 10, rather than the fact they had caused 68% or so damage to the survivor's brains. Those groups that did look at the graphs got confused: "The rescue vehicle's good but I don't get this..." or "The green line's gone up, the brain damage?". This implied more support is needed to interpret the data. The number of possible areas for damage relate to the probes that were not included that impact other areas of the body.

Conclusions
In conclusion, Astroversity was engaging, despite technical difficulties. Students require support collaborating and developing scientific enquiry strategies but preferably not through written instructions, which are rarely read. Pictorial representations, such as the exploding

brain, were far more engaging. More support is required in establishing the task to prevent racing around aimlessly and making more apparent to the students the role of the sensor, grid locations etc. Students appeared to be able to switch from online to paper representations of the area, but did not maintain strategies for recording. Reflection is not instinctive and a prompt from a computer results in less discussion than from a person (in this case a researcher).

3.4 Usability testing of three-dimensional prototype

Format Iteration 2:
The software was revised on the basis of the previous trial of a three-dimensional version at Cabot. This version was more stable, ie did not crash immediately prior to the Rescue Vehicle moving, and an introductory movie and training session was introduced. This would give the students context and the opportunity to individually practice moving the probe, making readings, plot graphs etc. The structure is similar to that described in Section 4.2, but as in the previous iteration during the rescue the student can see the position of their probe and sensor readings as well as the location of the RV in the map on the bottom right. They can also see the casualty status - which is shown in the top right of the screen - see Figure 8a. The feedback screen was revised so that the students could see the damage scores and toggle the body parts to display all or single lines, as shown in Figure 8b.

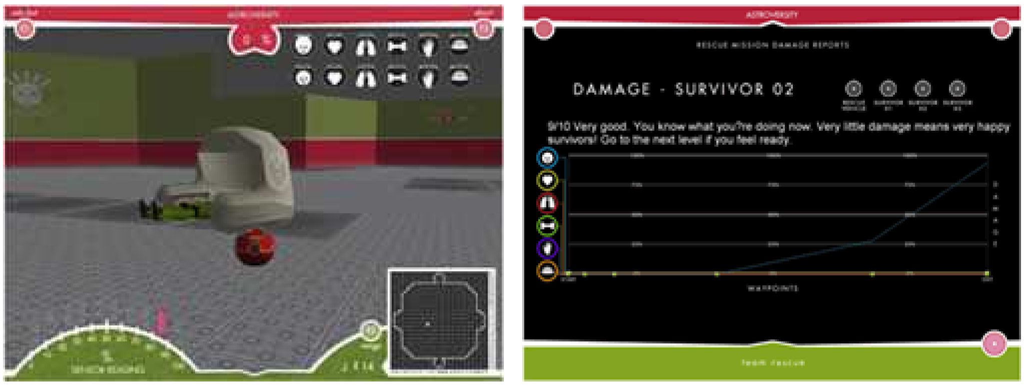

Figure 8: a) Rescue in progress; b) Graphical representation of performance

Trials:
This version of Astroversity was tested with six male and one female Year 9 students at Cotham School over two sessions; in the first session were two males, in the latter there was a boy-boy pair, a boy-girl pair and a boy working at the three computers. The sessions lasted approximately one hour.

The students all played computer games regularly and had a positive or neutral attitude towards science. Field notes were made and the sessions videoed. At the end they were asked to complete the standard questionnaire about game and science expertise and after each section - the introduction, training, and Level 1, asked to give three good and three negative aspects of Astroversity. In the hour available they completed Level 1 but were having difficulty with Level 2.

Findings:
Session 1
The training session confused one of the two students. He had no idea of what was happening and did not even pick up the dummy yet completed the task successfully and was given the disaster movie scenario.

Switching between the pen and paper and the screen seemed more problematic. In this first usability sessions neither participating student recorded more than ten marks on the paper, and had to be prompted to make any record for the training mission. They preferred to have races of their various probes. During their discussions to identify a rescue route they did not develop a shared method for recording data and used different pieces of paper when repeating the same level thus ignoring their previously gathered data. They did appear to develop individual strategies. GM appeared to use "S" and in Levels 1 and 2 plotted points consecutively to denote a path around the casualty. PL used ticks in Levels 1 and circles in Levels 2 that he put adjacently; in first attempt he recorded an area in the middle away from the body, in second attempt he explored a similar area to GM.

Neither boy said it was hard to record data, it appeared that they felt the exploration was more engrossing hence lack of recording. The lack of discussion about recording data, and lack of markings suggest that they have difficulty with the concept - but the lack of problems raised suggests that it may improve. However both would prefer to plot on screen.

Session 2
In the second session the two pairs found it far easier to record; without prompting by the researchers one student would explore and state whether a square was safe or dangerous which the other was meant to mark on the map. The recording strategies were more sophisticated when working in pairs. Two students recorded dangerous areas rather than just trying to identify a safe path, but only marked seven squares and for the first level one of these students had the scanner on wide. When asked what the impact of this was the student did not appear to understand the impact of averaging over a wider area.

The single player in the second session initially made few marks and then began to colour code; the victim was in green, the safe areas had ticks, the danger areas were crosses, and they were done systematically.

More reference was made to remembering rather than the marked paper grids: "I remember Q was quite high". None of them would record the path they plotted on the paper so there was no record of what they tried - which occurred within the two-dimensional versions.

General
When asked how old the participants should be all seven children thought the software was designed for Years 5 or 6 (primary school pupils aged 9-11) - yet all would play it again and they had trouble doing the puzzle and didn't complete the first level in the time allotted. This contradicted earlier findings from the two-dimensional iterations where the majority of students believed that Astroversity was more suitable for Years 7-10 (secondary school pupils aged 11-14). Nobody mentioned that it was childish in their responses, but they did spend more time looking at the tutor graphics - Captain Bolt was universally admired but his advice was never read aloud. One interpretation is that they are influenced by the graphics. The tutor characters are out of keeping with the rest of the game. This leads to questions about the design of the scenario and interface development work with students, as these discrepancies could have been identified earlier in the process. Despite this identification these characters remained in later versions.

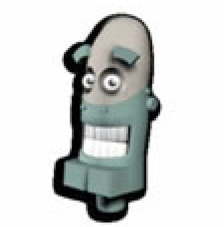

Figure 9: Captain Bolt

Despite the discrepancy between their age and that they believed Astroversity was appropriate for these sessions showed that the students were motivated by the task. In particular they were engaged with the idea of rescuing and the survival of the trapped student on Level 1 as shown in the transcript below.

GM: *We might have done it actually*
[Have to be prompted to click to do rescue]
GM: *Are we going to make it without causing the person severe brain damage?*
MS: *This is where we watch it*
PL: *It's looking good*
GM: *His heart doesn't look very good*
PL: *And his face doesn't look very good!*
MH: *And his brain isn't happy*

GM: *His brain is EXPLODING!*

MU: *Oh you killed her*
PL: *Good*
GM: *That's good... oh that's nice... I like that... whoa it blows the brain*
GM: *Lungs that's fine, it's just the brain, it starts going*

Transcript 5: Discussion of rescue

This rescue stage prompted most discussion, but not about strategies for future rescues. The students appreciate that the software is designed to support students' work together.

Recommendations:
The following recommendations were made.

- the cartoon appearance of the tutors is not in keeping with the rest of the game. There appears to be a danger that students will perceive the task as being for a younger audience because of this - up until this point the majority of students believe it is suitable for older audiences
- displaying the sensor readings for the cell that the probe is in during the rescue stage rather than that being experienced by the casualty is confusing
- as stated before the students should have more of an introduction when they are returning students to the disaster scenario. The jump from being a screen full of static to nothing wrong causes questions. "You know when the asteroid hits? Well it should go [explosion sound]". This was not changed
- further information is needed about the F3 function (this changes the area measured). Students understand it denotes they're reading from a wider area but not that it averages a greater number of squares and hence is less accurate. This could be in the form of a poster on the wall; a specific question from a tutor (but it would have to be a short question and there needs to be a response to ensure a misunderstanding is not appropriated); or mentioned in the training or F1 help
- the prompts needed rethinking; they had an impact if asked by a person but not as part of the game. The cause is probably the amount of text. As a research finding this is interesting as this type of procedural prompting has been shown to be effective when given by a computer but it clearly does not transfer from a system to support the group to a game scenario
- as in the two-dimensional and three-dimensional trials at Cabot there was confusion about the graph readings. What does 50% brain damage mean? If the team has a low damage score they are able to go onto the next level but they do not appear to reflect on the impact this would have on the casualties future life. This is not the main focus of Astroversity but should include conclusions such as: "She's survived but will never be able to talk again". This could require prompting by a teacher or some expert about the impact of such damage. Having one of the robot tutors highlight this aspect while the graph is displayed might be sufficient

- at this stage the tutors could also prompt for general reflection on the graph. The following suggestions were sent to ICDC but not incorporated. These questions don't have 'correct' answers but could stimulate discussion:
  - "What do the green points represent?"
  - "What is your hypothesis about bloppo impacting the brain?"
  - "What evidence can you provide me to support my theory that bloppo reacts with gunk?"
  - "What evidence can you provide me to support my theory that bloppo does not react with moob?"
  - "What can you tell about an area if it has a high reading?"
  - "When should you change sensor size?"
- written instructions about the registering and joining process were required as that was where most questions were asked.

It was discussed with ICDC that the students physically need to load sensor into probe, at the moment they just click on the first one and appear to forget which it is and what it senses. Perhaps movement would give a sense of ownership and reflection about what was to be measured. Then during the task the student should be able to check what sensor they have during the task and when they do the analysis.

4. 3D VERSION OF ASTROVERSITY TRIALLED IN SCHOOL

The software was revised over the summer. The major change was to the interface although the graphics for the tutors remained. The software no longer shows the probe sensor readings on the bottom left of the screen - which does not relate to the RV moving on the screen which is what the casualty is experiencing as previously the students. This was replaced in the next version by the icons, which were moved from the top right to bottom left of the screen and made larger (see Figure 12). As before, even if the final training mission was failed one was automatically allowed to continue to the main mission.

Other changes include incorporating comments from Futurelab staff that the characters should reflect ethnic diversity, thus Mary was replaced by Yuri etc. The ability to change sensor area using F3 was incorporated into the training and as a poster. The reflections prompts were replaced by shorter reworded versions. Finally, the instructions for joining and registering a team were revised and incorporated into this version by ICDC.

This section details the full background scenario and state of the final Astroversity software.

4.1 System requirements

Teams of three people are co-located but working on separate personal computers (PCs). The computers must have sufficient graphic capability and be networked, the minimum specification for a machine is:

- geforce 4 ti 4600
- ram 500mb
- pentium4 2ghz
- directX 9

In addition to the software the players have a paper map and pens to record their findings.

4.2 Story

The game takes place in the 'Astroversity', an orbiting space school of the future. After arrival denoted by an introductory movie the students are instructed in search and rescue. The school's robotic headmaster, Ed-1, introduces the various training modules the students will need. These introduce and give the students the opportunity to practice relevant skills needed for the later task, they are:

1. Mini rover control
2. Map reading
3. Sensor reading
4. Recording data
5. Way plotting
6. Rescue vehicle (RV) observance
7. Interpreting RV data

After the last module the students' screen appears disrupted. Ed-1 instructs all students to return to the control centre as the school has been struck by an alien craft. This has released three toxic substances, gunk (which causes the lungs to fill with liquid and eventual drowning), moob (which causes heart failure if sufficient is absorbed) and bloppo (which causes brain swelling and if in sufficient quantities explosion), into the atmosphere. The students must now form teams of three and rescue casualties from various levels within the 'Astroversity'.

In all the levels the same sequential procedure as in previous iterations is followed:

4.2.1 Search

Team members are given information on possible toxic substances detected and then must choose one sensor for their probe. Before exploring the students can select tutors to give advice on search and recording strategies but are free to choose their own method. Each group is given four minutes to explore. As they are in the same three-dimensional environment they can interact with their teammates. Each student has a different coloured probe that 'burns' different coloured gas so they can easily be distinguished. During the exploration the goal is to find the casualties and a safe route to the exit. Figure 10 shows the student exploring Level 1; they are in square I19, but can also see their location within the entire environment in the map in the bottom right, to the left the sensor shows the atmosphere in this square is 30% bloppo. They have 13 seconds left to explore.

To assist determining a route avoiding toxins the students are encouraged to record their findings on their personal paper-based maps - thus this stage requires developing search strategies and systematic recording and logging of data. Possible decisions include: should they divide the area into three and each explore one aspect? Should they do a quick reconnaissance with the sensor set to cover a large area to gain an overview and then focus on the most likely safe route? Should they look for danger areas or safe areas?

Figure 10: Exploring screen shot

4.2.2 Planning

After searching, the team are asked to gather round one machine to discuss their findings and identify a safe route for the rescue vehicle so that the casualty remains alive and healthy. In this stage the students must share the distinct knowledge each one has collected. Using the data gathered and recorded on their maps in the search stage together they hypothesise about best evacuation route. Figure 11 shows a group plotting in their suggested course. In later levels there are no routes free from toxins and the group must discuss risks, the number of turns and hence the speed through the Astroversity, and the impact of combining various toxins in order to create a route.

Figure 11: Wayplotting screen

4.2.3 Rescue

The plan agreed is then executed and the team can view the rescue vehicle and follow their route. During this stage the team receives feedback on the vehicle's success or failure in the form of a graphical health summary shown in the bottom right of the screen - in Figure 12a Yuri the survivor's head is beginning to explode by cell L14. Afterwards the casualties' health readings during the course of the rescue are represented in the form of a graph. The students can select which one or which group of sensor readings to display. From the graph the team must deduce when the toxins become poisonous and what happens to the casualty when various toxins interact - sometimes they nullify impacts while other levels may lead to a greater impact. For example, from Figure 12b students could determine how they killed Yuri. They could deduce that the area in front of the exit has a high level of gunk given the rapid increase in lung damage; therefore they can determine what level of gunk in the atmosphere is highly dangerous.

The graph also shows that gunk and bloppo interact. After exposure to bloppo, shown by the swelling brain, gunk will counteract the effect, however, it still damages the lungs. This occurred at the fourth waypoint. At this stage the team must individually decide what to do next. Do they have sufficient information and just need to replot the route, or should they go back and gather more data? They cannot continue until all three have chosen the same option. The group can only go on to a new level if they have successfully completed a rescue with minimal damage to the casualty, but they are not given information on what these levels are prior to the rescue.

Figure 12: a) Rescue in progress; b) Graphical representation of performance

4.3 Reflection

At various points the students are interrupted by a prompt to answer one question about their performance. As discussed in the recommendations of Section 3.2 these prompts were to:

- encourage the group to reflect on their collaborative and independent activities
- scaffold discussion and debate on collaborative issues as well as raise awareness of scientific enquiry strategies.

The questions are shown in Table 4.

How well did you record the data?
How effectively did the group use the data you all gathered?
Did you personally contribute your data and suggestions to the team?
How much of the time did you listen to each other?
How much time do you believe you were listened to?
Was your strategy for plotting a route any good?
How much of the time did you listen to each other while analysing the rescue performance?
How much time do you believe you were listened to while analysing the rescue performance?

Table 4: Reflection prompts

The students are asked to individually select one of five responses by clicking on it. Once all three have responded the system then generates a response based on the difference in self-assessments or the values of these assessments. For example, if each student believes that they listened all the time but their peers did not listen at all there would be a discrepancy. The software would respond with the following: "That's very strange - how come you as a group have different ideas about the amount of listening that went on? You will need to sort this out before your next rescue." All of the feedback is short to minimise the amount of text to be read.

4.4 Limitations of software

The Astroversity software had several limitations. The most important is the lack of logging. There is no record of the group's progress. Thus if the game were halted halfway through the group would have to restart the whole level the next time. However, as the casualty is stationary they can use the data gathered from the previous exploration. Although not designed for classroom use from a teacher viewpoint the lack of logging means that they have no record of the student's performance at the task, ie how many attempts were needed, how much damage was caused. From a researcher perspective the lack of logging means that the amount of time spent on each activity or the students self-assessments are not available for analysis.

A second error that has since been corrected is the progress needed before being allowed to continue. Initially even if you inflicted 80% of brain damage on the casualty you would be allowed to continue. This caused hilarity, as shown in the following statement:

1Bb: *Look, "you have succeeded with one of our best results"* [swivels monitor round to show 1Bb - both laugh] *...the dude died. Record more data?*

To stop students racing through they were instructed only to continue if they had less than 30% total damage. To assist them to keep track, and to overcome the problem that Astroversity has no recording mechanism, the students were asked to write down the damage after each level. It was also intended that these records act as a stimulus for students in a similar fashion to high score tables. Although they could continue it was suggested that they would be inspired to repeat a level in order to get a low damage score.

5. FINAL STUDY

5.1 Participants

The students were from the top Year 9 science group at Monks Park. Given the time of year student attendance was sporadic but 8 girls and 10 boys attended at least the first and last session. More students turned up to the final session than the first, with six students who had never used the software turning up for the final session. Another 11 attended at least one session. The 18 that attended more than one session formed six groups. The groups were self-selecting, and as far as possible remained the same. However, one group of girls worked in a four as in the first session the numbers did not divide accurately; there was a group of four boys in the third session. Groups 1, 3 and 5 were all boys, Groups 2 and 4 were all girls, and Group 6 had two girls and a boy. The participants are distinguished by labels, eg, G4c, is the third girl in Group 4, while B6c is the boy who is the third team member in Group 6, this is shown in Table 5.

Group 1	B1a	B1b	B1c
Group 2	G2a	G2b	G2c
Group 3	B3a	B3b	B3c
Group 4	G4a	G4b	G4c
Group 5	B5a	B5b	B5c
Group 6	G6a	G6b	B6c

Table 5: Team composition

The socio-economic background of the students is varied. In the school as a whole, of the 168 eligible for GCSE/NVQ qualifications 21.40% had special educational needs. There is an above average number of authorised absences and just above the national average for unauthorised absences. There are no A level students. 42% of students are awarded GCSEs at grades A*-C compared to a National Average of 52.9%. Although many have English as a second language in the school all this class were native English speakers. There was one student with learning difficulties, due to his dyslexia he has an assistant to read to him, his label was B3c.

5.2 Method

In their classroom the students were observed devising and performing an experiment over an hour lesson. They were given the equipment to make a pendulum and asked to investigate the impact of varying string length and weight on the time it took for one swing. They worked in groups of four students and were asked to record data they felt relevant and summarise their findings.

The students then used Astroversity for three sessions lasting approximately an hour each over two weeks. These sessions were held in Monks Park City Learning Centre (CLC). This had machines of sufficiently high specification and with a network that would support the teamwork aspect. In the first session they did the training exercise. In the second session they did Levels 1 and 2. In the third session the students started Level 3, but none completed this task.

The following week they were observed in their classroom devising and performing an experiment to determine the impact of changing trolley weight on speed at an angle of 20 degrees in an hour lesson. As far as possible they worked in the same groups as for the first session.

5.3 Data gathered

The students completed a questionnaire before and after using Astroversity. This asked for background information on game playing skill and frequency, what made someone good at science or working in a group, and a self assessment and reason of their own ability in these areas.

The teacher was informally interviewed at the end of each session and formally at the end of the study.

As Astroversity has no logging system the students were asked to record the damage they caused to the heart, brain, and lungs after each rescue attempt. This provided incomplete data as the students became involved in the activity and failed to record, or they did not want to record their poor performance.

The data recorded by the students on the map for each level was collected.

Five of the six groups were videoed when using Astroversity, one group of boys could not be filmed as one boy had not received permission from his parents. Approximately two hours in total of video was recorded. These were later transcribed. Field notes were made during each session.

Groups 1, 2, and 5 were informally interviewed during the task about their enjoyment of playing the game, and the best and worst aspects.

Videos and classroom observations were made of the students performing a scientific enquiry task before and after using Astroversity and the data they recorded within each experiment. As far as possible the groups were the same but not identical, thus all bar the group with the student who had not had permission granted were filmed for approximately five minutes.

5.4 Limitations and strengths of the research process

5.4.1 Study limitations

With respect to the study there were problems with the software installation. Although there were sufficient machines with the software the network would only permit 18 students to play in teams at any one time. This meant the students played in shifts.

The study was conducted in the last few weeks of term. The students were therefore not always focused on the subject and attendance in class was sporadic. This was especially noticeable in the science experiments. These were meant to corroborate the self-assessments. However, although they were asked to plan their experiment, perform it, record appropriate data, and draw a conclusion this was often not achieved. Possible explanations are the time of year (it was after the SAT exams and before the end of the summer term) and the short time allotted. According to the class teacher they work better over longer studies, the first lesson is spent investigating the kit and task, it is not until the second session that they start working. The data recorded therefore was not used.

Finally the results cannot be extrapolated from the small number of students.

5.4.2 Study strengths

The advantage of this testing approach was the integration into the classroom. As far as possible the class teacher ran the lessons with support from the researcher. It was done as a whole class exercise in a school using equipment currently available. This enabled the designers to see what kind of issues would be faced if this process were replicated.

5.5 Questionnaire results

5.5.1 Ability at computer games

The students had varying self-assessments of their ability and experience at computer games; these are summarised in Table 6.

Ability	Very good		Good		OK		Poor	Very poor		
Frequency	Every day	Every week	Every week	Every month	Every week	Every month	Very rarely	Very rarely	Very rarely	Never
Boys	1	1	7	1						
Girls			1		2	1	1	1	1	1

Table 6: Self assessment of computer game ability and frequency played

This table shows that the boys that participated throughout the study believe they are better at computer games than the girls and play them more frequently.

5.5.2 Understanding of group work

The students were also asked to complete the following statement:

"To be good at working in a group you must…"

The 18 students came out with 25 reasons, relating to:

Working in teams ("work together", "work with people, not on your own", "be able to participate", "cooperate")	(7)
Communication ("communicate", "be able to express your views freely")	(6)
Listening ("be prepared to listen to others", "listen to other members of the group")	(4)
Sharing ideas ("be able to accept other people's opinions", "able to speak out/tell your ideas")	(3)
State ("have a mature, open mind...")	(1)

The 8 girls came up with 15 reasons; the 10 boys came up with 10. The reasons given suggest that the students are aware of the words that describe the type of behaviour that leads to an effective team.

5.5.3 Understanding of science

The students were also asked to complete the following statement:

"To be good at science you must…"

The 18 students came out with 23 reasons, relating to:

Listening ("pay attention", "listen in class")	(8)
Understand ("be able to understand the things surrounding you", "try to understand")	(6)
Scientific approach ("have a scientific brain", "good logic skills")	(3)
Learn	(2)
Study	(1)
Work hard	(1)
Enjoy	(1)

The 8 girls came up with 13 reasons; the 10 boys came up with 10. Interestingly the most frequently referenced skills are more to do with behaviour acceptable in a classroom than those required by scientists such as being methodical, developing hypotheses, or making observations. This suggests that students are not encouraged to act 'as scientists' in the classroom, as recommended by the Beyond 2000 report (Millar and Osborne 1998).

6. RESULTS AND ANALYSIS

Given that Astroversity is designed to encourage scientific enquiry and collaboration skills the results and analysis are divided into these categories. The analysis of the development of scientific enquiry skills is based upon:

- student self-reports in interviews and questionnaires
- observation of the methods used to record data
- the discussions they have around the exploration and review stage
- the efficiency of their rescues.

Changes in group skills are examined through:

- the students' self-reports
- the talk between the students
- their responses to reflection prompts

This structure allows us to analyse the impact of tutors. In addition this section will look at the engagement of the students with the software and their beliefs about the purposes of the software.

The aim of gathering multiple data sets is to corroborate the findings. Ideally there would be a record of rescue routes and damage caused within the rescue and responses to the various reflective questions. The software did not allow this. Thus the students were asked to record their own progress. Unfortunately, there is a degree of unreliability in the recording of progress. If the sheets given to the students to fill in with the amount of damage were completed after each attempt it would imply that most students scored less than 30% damage on the first or second attempt. From the field observations this is not the case. In addition, there is boasting between groups that they have caused 100% brain damage even though this is not recorded.

6.1 Scientific enquiry skills

To assess the extent to which the software usage encouraged the development of scientific enquiry skills the data used are the students self-reports of method - asked after the final Astroversity session, the maps they record their results on over the sessions where collected, and the discussions they have about the software. It also includes the responses to the tutors. These were meant to provide advice and cause the students to question their approach; for example, they may offer conflicting advice, one may say taking the most direct route is more efficient, another that avoiding all gas, regardless of length is preferable. The students could choose separate tutors and receive different suggestions; the issue is whether these were incorporated into the teams approach.

6.1.1 Data gathered

Self reports for ability at science

This self-assessment phase indicates whether the students believe that using Astroversity had any impact on their ability at science. This task could have been made more useful if it was refined and students had been asked about their attitudes to conducting and drawing conclusions from an experiment. However, the question was broad so as not to be leading.

The students were asked to fill in the gaps of the following statement before and after using Astroversity:

"I am............ at science because............"

From Table 7 only one student believed that they were bad at science; this is unsurprising given that it is the top group. Of the 18 students that attended all sessions 12 (seven girls and five boys) (66%) do not change opinion of ability, one boy (6%) believes they got worse, and four boys and one girl (28%) think they improved.

	OK to Good	Alright to Good	Bad to OK	Not good to OK	Good to Good	Alright to OK	OK to OK	Good to OK
Boys	2	1	1		1		4	1
Girls				1	1	1	5	

Table 7: Self-assessment of ability at science before and after using Astroversity

This change in opinion is not necessarily reflected in the justification. B3a, for example, believes that he has gone from OK to good and his reason shifts from: "I listen to things" to "I have a good understanding of scientific theory". While B1b believes he remains OK but his reasoning shifts from: "I work hard, but dont share ideas with the class" to "I revise". While B1c sounds as if the self-assessment ought to improve as although he reports listening both times by the end of the third session in addition he believes he has a scientific brain. However, analysing the reasons given in Table 8 shows that for most there is a consistency between their own perception of what it is to be good at science and their self-assessments. Moreover, the different reasons given in within some members of the groups show that the students tended to work independently when answering.

	To be good at science you must...	I am...	at science because...	I am...	at science because...
B1a	have a scientific brain	good	I have a scientific brain	good	I have a scientific brain
B1b	study and work hard	ok	I work hard, but don't share ideas with the class	ok	I revise
B1c	listen and have a scientific brain	ok	I listen to things	ok	I listen and have a scientific brain
G2a	listen	ok	I enjoy doing experiments	ok	I like working in groups
G2b	understand it	not good	I don't understand many things	ok	I understand some things and I can work in a group sometimes!
B3a	be able to understand the things surrounding you	ok	I listen to things	good	I have a good understanding of scientific theory
B3b	be able to understand the subject of each type well	ok	I listen to the topic and undertand well	good	I discover new things and like doing lots of pratical work
B3c	don't know	bad	-	ok	but find it hard to remember things
G4a	listen + learn	good	I enjoy it, I listen and I cooperate	good	I enjoy it and I listen and learn from others

G4b	listen and learn	ok	I will listen and try to understand different problems	ok	I enjoy it and find it interesting
G4c	pay attention	ok	I like doing group experiments	ok	I enjoy doing experiments and working in groups
G4d	understand it	alright	some things I understand and other things I don't	ok	I'm not really bad and not really good
B5a	listen in class	good	I like practical work	ok	I like experiments
B5b	have a scientific mind	alright	I have been doing it for 9 years...	good	I have been studying it for 9 years
B5c	-	-	-	good	I find it interesting
G6a	listen, have good understanding I enjoy it!	ok	I concentrate, listen and try my best at working to the best of my ability	ok	it's enjoyable and I work to the best of my ability
G6b	Try to understand, listen	ok	I listen carefully and understand what I'm being taught	ok	I listen carefully and am willing to learn
B6c	have good logic skills	ok	I use logic skills	ok	I think things through and discuss

Table 8: Changes in justification of ability

Perception of strategies for Astroversity
After the third session using Astroversity the students were asked to individually complete the statement "My strategy for doing the rescue is to...". Their answers are given in Table 9.

Group	My strategy for doing the rescue is to...
B1a	finding safe rotes for all sensors
B1b	talk to others
B1c	use different sensors
G2a	work as a team and to listen to each other's ideas
G2b	use everyone's notes to find a safe route
G2c	listen to everyone in the team + combine our results to work out a route safe for all of us
B3a	plot a point to the casualty and then slowly plot my way from this route to the exit
B3b	work together and work out a strategic route with my team mates
B3c	listen to the other team members and follow their instructions
G4a	to see the goods and bad for each organ, in the body, and we used different colours
G4b	find random safe spots then put it together with my group
G4c	plot the best rescue

G4d	combine everybody's safe routes - use everybody's notes and ideas
B5a	plot the best route
B5b	search all of the squares for any danger and find the safest route
B5c	Find the causalty. Find a safe route. Begin the rescue
G6a	use good navigation techniques listening to each other & writing/marking it down
G6b	Split the room into 3. Each of us search a different part. Put our results together and work out a route
B6c	find the student and then the quickest and clearest route to the exit

Table 9: Recorded rescue strategies

Table 9 shows that the groups generally agree about their shared strategy. However, the sentences are short and generally at a high level rather than a discussion of methods employed. There is a focus on teamwork in that they should listen to their teammates and plot safe routes. The short statements can be partially explained by the fact that many students dislike writing and wrote the minimal amount rather than their actual approach.

There is limited evidence that the advice of the online tutor influences the data-logging strategy. When asked at the end to describe the strategy used to plot a path four students, all who changed strategy, gave a method that could be directly related to tutor advice:

- to see the goods and bad for each organ, in the body, and we used different colours (Group 1 - girl)
- find the student and then the quikest and clearest route to the exit (Group 3 - boy)
- split the room into three. Each of us search a different part. Put our results together and work out a route (Group 6 - girl)
- search all of the squares for any danger and find the safest route (Group 5 - boy).

Observations of students using the system show that tutor selection appeared to be based on the cartoon features of the character, in particular Captain Bolt was repeatedly chosen despite his guidance being: "I think you should be the leader". In the sessions observed and recorded the time the advice was displayed on screen was insufficient for it to be read. However, there were examples of students using strategies suggested, as shown in Transcript 6.

G8a: *Do we need to record more data? We all recorded lots of data but some of it wasn't used*
G8c: *Next time I should review all the data at the same time. How about dividing the area into three and each of us explore one?*
G8a: *OK*
G8a: *Em...* [Confidently] *yes. The person who finds the location of casualty should tell all of us as soon as possible. This could save time!*
G8b: [Smiling and looking to G8a] *yeah, do we play again?*
G8c: *I forgot to say that I will explore the left part of this area from 'H-L' and G8b, middle 'M-Q', G8a, Right 'R-U'. Is this all right?*
G8a: *OK*
G8c: *Let's try*

Transcript 6: Group of girls doing Level 1

G8b recorded a strategy of: "listen to everyone in the team + combine our results to work out a route safe for all of us" - this does not closely match the strategy observed suggesting that there is a discrepancy in relating actual and recalled behaviour.

Strategies for recording data
Analysing the maps collected in the training level 11 students just shaded the cell asked to record where the casualty lay - O15, four have O15 (casualty) and Q6 (which is out of permitted area), and three just have Q6. Only two have marked more points - Q16, S14, H11 and M8 - which are grid references from second training exercise! This suggests that before starting the mission students have not yet reflected upon how to record data from the sensors.

Once the collaborative stage of the game had been reached four main strategies for recording data were recorded. All of them involved marking the casualty, in Figure 13a, b and c this is I19. Then the simplest, shown in Figure 13a is marking cells that have significance, either an exceptionally high or low amount of toxin in that square. These appear to be as an after thought rather than integral to the search strategy. Only boys in this study appeared to use this method to record data.

The second type is marking a safe route. This is usually a line between crosses or shaded squares as shown in Figure 13b.

The third type is incrementally identifying a route, that is, marking safe cells up until a high toxic area and then going back to the last junction - see Figure 13c.

The most sophisticated method, employed predominantly by girls, is general exploration, ie, marking safe and dangerous areas with a key - which progressed to multiple colours when the number of toxins increased. In addition two girl groups who appeared to be working independently used a key. In Level 1 where there is only one toxin 'S' marks safe cells on the map, and 'X' dangerous cells. Every group member shared the key. By the third level where the number of toxins has increased the girls introduce colour for each gas while still using 'S' and 'X'. An example is shown in Figure 13d. This group also demonstrated some division of labour, eg: "Shall I go near the exit?... Is anyone else near the exit? I'm going near the exit."

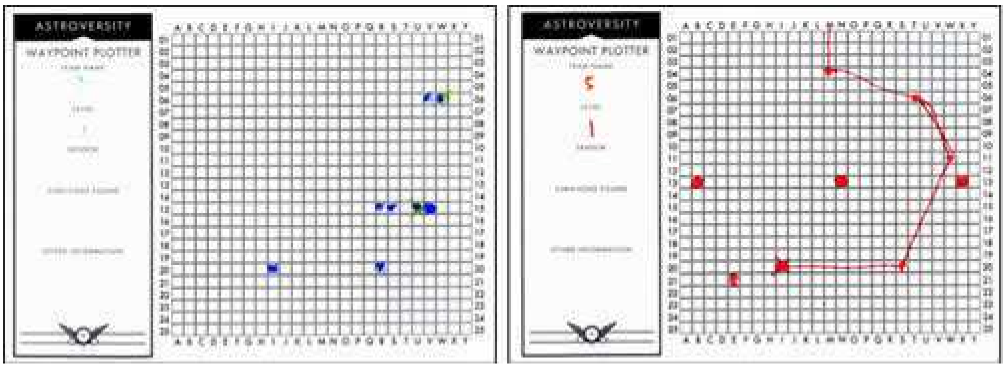

Figure 13: a) Marking points; b) Marking single path

c) Incrementally identifying a route; d) General exploration - colour coded

The data collected on strategy was collated and eleven of the 21 students where more than one paper representation was collected showed a changed strategy over the three sessions. Of the seven groups (Groups 3 and 5 did not return their maps after the first session and Group 7 stayed in the same group for the last two session) only one group showed a shift of all group members to another strategy, with another three groups having two students employ a different approach. Girls tended to record more data throughout and were more likely to use multiple colours to distinguish between toxins and distinguish between safe and dangerous areas. The first group of boys actually appeared to use less effective strategies in later levels while the remaining nine students progressed. This is shown in Table 10.

	Level 1			**Level 2**			**Level 3**		
Group 1 - boys	Po	MP	MP				Po	Pa	Po
Group 2 - girls	MP	MP	MP	MP	MP	MP	MP in colour	MP	MP
Group 3 - boys				Po	Po	MP	MP	MP	GE - colour coded
Group 4 - girls	GE	GE	GE	GE - colour coded	GE	GE	GE - colour coded	GE - colour coded	GE - colour coded
Group 5 - boys				N	N	Pa	N	N	Pa
Group 6 - mixed	Po	Po	Po	Pa	Pa	Pa	Pa	Pa	Pa
Group 7 - boys				Po	Po	Po	Pa	Pa	1 dot

Key to table: N=None Po=Points Pa=Path MP=Multiple GE=General paths exploration - marking safe and dangerous areas

Table 10: Strategies employed for data recording

The impact of paper recording on strategy
The paper representation of data collected was used by some of the groups. Group 3, including with the boy with learning difficulties used the paper effectively:

B3b: *I think I've found a route*
B3c: *I've found a route as well*
B3b: *That's my route* [passes over the paper]
B3c: [Leans over and puts both routes in front of B3b] *I think we have to go round here and up to here, that way cos everything round here is just so bad*
B3b: *We could try that and then we could try that* [points each time]
B3c: *If we start there and came back that way, and then came across that way*

Transcript 7: Discussion of deciding a route in Group 3

The girls used their paper maps extensively, and all contribute to the discussion, as shown in:

G4b: *Yeah but if we use that* [leans over and points to 4's paper] *it goes to... and if we do that [points] we can go down there*
G4c: *Will it work? If we go down there* [points]
G4b: *Yeah* [looks at map]
G4c: *To I11*
G4a: *Is are quite good actually*
G4b: *Yeah*
G4c: *Hmmm all the Is are good* [2 leaning right over] *hmmm 10* [comparing the 2 maps]
G4b: *It's either that or urm... 9, 8, 7 then it goes*
G4d: *That is the exit, in there... so that's the exit in there* [assume pointing at paper but 3 writing]
G4a: [Can't hear]
G4d: *The green one is the exit*
G4c: *How do you get there from /*
G4b: *{H5, 4, 3, 2}*
G4c: *{That one's dodgy a bit}*
G4b: *G4c's got urm 6, 5, 4, 3, 2 [looking at map] 6, 5, 4, 3, 2*
G4d: *That's brilliant - them co-ordinates* [stands up looking over]
G4c: *That's 6, 5, 4, 3, 2 [marks on second map] - so she's got N4*
G4b: *Then 6*
G4c: *I got that one was iffy though*
G4b: *What M?*
G4c: *No that one* [4 has walked round and is looking over]
G4b: [speaks to 4]
G4c: *No iffy for the lungs... it's all right for the brain but it's iffy for the lungs right... right I got a route* [picks up map, girls all stare] - *you go up four more squares, so you come from there, up to there, along to there, up to there and along like that, and up*

Transcript 8: Discussion of route by Group 4

This is an interesting transcript as it shows that there's an appreciation that they need information from all the sensors that means that they have to have data from all the students. It also shows the instinctive use of grid references. There is discussion and some simultaneous speaking but that appears to be as they are engaged, they praise each other, and ask questions.

In particular Groups 1, 5 and 7, all boy groups, tended to recall data rather than use the maps that they had created.

6.1.2 Discussion on scientific enquiry

Triangulating the evidence shows some discrepancy in self-assessment of ability and strategies observed. For example, the strategies in Group 1 appear to be less systematic, yet the boys record no change in perception of ability. In Group 2 a girl appears to believe they improve yet the strategy remains the same. However, Group 3 all believe they improve and their strategies advance over the sessions. There is a slight improvement of strategy in Group 4 but no change in opinion. While in Group 5 one boy thinks he got worse, but there was no change in strategy - although this group displayed the least effective strategy. Finally, in Group 6 there was no change in perception, but there was an improvement in strategy. This questioning technique is not necessarily reliable, as just as asking about group skills, highlighting behaviour before an exercise can cause reflection that changes later opinions - though this data suggests that is not always the case. Alternatively, it could be that the triangulation question was too broad, although this could also imply that students did not link Astroversity with their self-perception of ability or that it had no impact.

The latter reason could explain the low numbers that change strategy; nine out of 21 used more efficient strategies. In the majority of cases without support Astroversity does not automatically lead to an improvement in scientific enquiry skills. In particular, there is little understanding of the role of tutors; the students do not respond positively to anthropomorphic agents given the context. That is, they have little impact on behaviour and are not recalled.

6.2 Group skills

6.2.1 Data on group skill development

The following data was collected about group skills.

Self reports for ability at working in groups
The students were asked to fill in the gaps of the following statement before and after using Astroversity:

"I am................ at working in groups because................ "

Half the girls believed they improved after using Astroversity. However, only one girl believed initially she was poor at working in groups and by the end all felt they were at least OK. Boys tended to have a lower self-perception of ability at working in groups. Only two initially felt they were good. One fifth of boys felt they improved, while one fifth felt they got worse, one progressed from OK to No good. This is shown in Table 11.

	OK to Good	Not very good to OK	Good to Good	OK to OK	Alright to Alright	Not very good to Not very good	Good to OK	OK to No good
Boys	2		2	2	1	1	1	1
Girls	3	1	3	1				

Table 11: Self assessment of ability at group work before and after using Astroversity

Analysing those that had a change in opinion there does not seem to be a change in reasoning between the sessions; for example, B1b went from OK to good, yet his reason for justification decreased, initially it was "share ideas but let others share their ideas as well", in Session 3 it was "I share other information" - with no mention of listening to others. Whereas the girl who went from Good to OK had a shift in reasoning from "I can communicate well" to "I listen to other people" - which does not explain the change in perception. However, the students do relate their ability to their definition of working in a group. This is shown in Table 12.

	To be good at working in a group you must...	**I am...**	**at working in a group because...**	**I am...**	**at working in a group because...**
B1a	work together	good	-	good	-

B1b	communicate	ok	share ideas but let others share their ideas as well	good	I share other information
B1c	communicate	ok	I talk to people	ok	I talk to people
G2a	be able to work as a team	ok	I make sure everyone gets involved	good	I make sure everyone gets involved
G2b	listen to each other, and contribute, help each other	not very good	I don't contribute enough or I'm too bossy!	ok	I give my ideas and listen but sometimes I talk too much or I try to take control
B3a	be able to express your veiws freely	alright	I am very open	alright	I can explain my ideas
B3b	be able to communicate with your friends	ok	I talk, but I don't get really involved	good	I find it easy to communicate and share different ideas
B3c	don't know	ok	-	not good	I don't talk to other team members. They are not my friends
G4a	co-operate	good	I know the people well	good	I can give my ideas out and listen to others as well
G4b	be prepared to listen to others	good	I find it better to to talk about a problem in a group	good	I listen to what other people say and find it easy to work problems out together
G4c	include the rest of the group	ok	I include everyone	good	I include everyone and share my ideas
G4d	listen to each other, help each other	ok	I am quite quiet so I don't give all my ideas, and I let other people take the lead	ok	although I am quite quiet I give my ideas
B5a	be able to participate and listen to other members of the group	good	I contribute and listen	good	I contribute and listen
B5b	have a mature, open mind...	not very good	I work better on my own	not very good	I think differently
B5c	have good communication skills	ok	I sometimes have good ideas but don't communicate them well	ok	I have some good ideas
G6a	cooperative, willing to share ideas, good in teams	good	I cooperate well!	good	I understand, listen and use each others ideas to collaborate a

					good choice
G6b	be able to accept other people's opinions, able to speak out/tell your ideas, able to work with people, not on your own	ok	I listen to others opinion and how they feel, I'll try to make sure everyone's happy with the final decision. Most of the time I prefer to be a leader	good	I like to consider everyone's ideas and make the final idea, agreed
B6c	have communication skills	good	aI can communicate well	ok	I listen to other people

Table 12: Changes in reasoning and self-assessments

This self-perception data is not always accurate when compared to the data from the dialogues recorded or observed and the behaviour displayed. For example, B3a felt he was "alright" at group work as he shared his ideas and he improved at science as he listened and had a good understand of scientific theory. While doing Level 1 in the first session where the only toxic material is bloppo he instructed:

B3a: *I've chosen bloppo, you choose moob*

He then went on to physically select the sensor for his fellow student. There was no evidence that he listened to his team mates.

As stated, B3c has dyslexia and a classroom support assistant. Although he believes his ability at working in a group has got worse, a belief contradicted by his teacher and the level of participation in the Astroversity activity compared to a science class, he has begun to reflect on his behaviour. Initially he did not bother answering, by the end he gave a reason for his belief. However, some changes in opinion seem reasonable. G2b supports her change in opinion by her reflection on bossiness, initially she admits to being bossy, but she puts her improvement down the fact she realised she talks too much or tries to take control.

These inferences need to be viewed with caution, as discussed in Section 6.2.2.

Attitudes towards prompts on team skills and group strategies
The students did not pick up on the prompts designed to encourage reflection or the instructions from the robot tutors. The one example of discussion recorded in the video and in the field notes came from Group 4 being questioned on how well they had used the data collected after having looked at the graph:

G4c: *Yes but not as much as we could*
G4a: *What shall I put?*
G4c: *"Yes, but not as much as I could". Cos we all did some but we didn't do loads*

Transcript 9: Discussion recorded responding to prompt about use of data collected

The failure to pause at the prompts is not unique to the self-reflection task. The students would often rather refer to the visual than written instructions. For example, the selection of the sensors in Level 1, the written text says only bloppo is present but there is an assumption that as there are three sensors each student should take a different one.

Conversation within a group
Groups were not always focused on the task, most often during the first exploration exercise, especially amongst the boys who were interested in the Astroversity environment. Conversations focused on identifying players. Thus a group would establish who had which colour probe, eg "I'm still yellow". Once this was done they explored the capabilities of the probe, although sometimes erroneously. Examples are shown in the next transcripts:

B1c: *Play follow my leader... go*
B1b: *Stay still* [boy 3] *and I'll push you*
B1c: *Where... wait a second*
B1b: *Yeah* [giggles]*... picking on you my son* [B1a also laughing, perhaps pushing as well?]
B1b: *There's one bad thing about chasing somebody - they always know when you're going to run into them... or slide*
B1c: [B1b] *You've got to try and get me - I'm green I'm green* [giggles]

Transcript 10: Group 1, Session 1, planning follow my leader

B1b: *I'm bored now, I'm going to start bashing into [B1c]... [B1c] you have to chase me*

Transcript 11: Group 1, Session 2, interacting through chasing

B7b: *I've shot you* [the others ignore B7a who follows them into exit room - no comment on the RV, they whiz out by him laughing]
B7b: *He's shot now*
B7a: *How do you shoot?*
B7b: *He shot back* [seems to be disregarding B7a's question]
B7a: *Who's red?*
B7b: *[B7c]*
B7c: *What are we suppose to do?*
B7b: *Did you see me then?* [his green swerved in front of B7a's yellow]*... chase him down like a... mine's way faster than* [B7a]

Transcript 12: Group 7, Session 1, shooting in Astroversity

B1b: *I wish this sounded like guns, pheeew pheeeeeew* [more laughing by all 3]*... I knew what you were going to do then* [talking to 3]*... I got clipped there by B1c... you've got 10 more seconds till you have to chase me... 3, 2, 1 now you have to chase me... come on now... ahhha*
B1c: *Oh he went flying past*

Transcript 13: Group 1, Session 1, identifying probe capabilities

An earlier transcript, Transcript 8, shows a group of girls discussing the task, but in addition to the scientific enquiry skills mentioned this highlights the collaborative aspect. The girls praise each other, take turns, and listen. However, these students believed they were OK or good at group work prior to playing Astroversity.

6.2.2 Discussion on group skills

The transcripts were analysed to see if there was a shift from a cumulative turn taking approach seen in the pilots to what Mercer (1995) calls exploratory talk, ie conversations incorporating not only turn taking but justification and questioning. Although exploratory talk was observed there was no increase in the later sessions. Groups that listened, took turns, and provided justification did so throughout. The only change was in Group 3. They spoke more in later sessions than initially.

There is some evidence that the self-reports reflect actual behaviour. The girls talk more and usually consider themselves better at group work. However, B3c believed he got worse, yet as mentioned, his input and the comments of the teacher suggest that he improved over the course of the sessions.

However, these reports are not enough from which to draw conclusions. To ask a question about group skills before and after a group activity ought to imply that the person asked changes their opinions as they have had the opportunity to reflect and an opportunity to observe their own behaviour. Another reason for not drawing conclusions is that the groups were not recorded over extended periods of time so the volume of data is limited and therefore insufficient. Alternatively, students do not spend long on such questions as they are not directly related to the task. There may also be an issue as to what is understood by the terminology to describe group skills.

6.3 Attitudes towards software

6.3.1 Data about attitudes

Attitudes towards the usage of Astroversity

At the end of the study the students were asked for up to three reasons as to the purpose of the software. There were 48 comments made, which are summarised in Table 13.

Activity	No of comments	Example	Comments
Develop team and group skills	12	"make you think like a team/group"	
General educational	7	"help learn"	
Enjoyment	6	"make learning fun"	
Improve communication	5	"help people to communicate, listen and talk to each other"	
Improve thinking skills	4	"help people to learn from their mistakes and rethink their tracks"	
"worth continuing"	3		same group
Practice and understand co-ordinates	3	"learn grid references"	two groups
Make science fun	2	"help with our science"	
So that you can help people in space	2	"teach you how to rescue someone on a spaceship"	two groups
Develop computer skills	2	"use new technology to	two groups

		network games and experience IT"	
Improve the ability to collect data	1	"help you collect data and understand graphs"	
Improve concentration	1	"to make you consate"	

Table 13: Summary of reasons given to play Astroversity

This shows that a quarter of all reasons are involved with developing team skills, in addition, another five of the 48 comments think that it improves communication. There is less appreciation of it as a means to help science, despite occurring in a science lesson. Only two reasons relate to general science and only one about the data collection skills.

The students appeared to work individually when thinking about reasons, for example, all the groups had at least one student that said it supported teamwork, while only two of the six groups had all three students think teamwork was a reason for use - and they did not use the same wording.

The idea that it supports mathematics was mentioned in the interviews. The reason stated was because of the graph and grid references. This implies that the surface features of software are taken as key to its assumed relevance.

The comments about the software were all complimentary. The students liked the look of the software and accepted the narrative despite obvious flaws, for example, why should they manage the rescue when there were robotic tutors that ought to be able to do it more efficiently? And how can they repeatedly perform the rescue of the same person if this is not a simulation? And if the computers were capable of plotting such a simulation why could it not plot a safe route itself?

Finally, the teacher was disappointed that Astroversity was not ready for the start of the autumn term. She had wanted to use it as an introductory lesson to give her new students the opportunity of working together - she appreciated the speaking and listening aspect of the task. Furthermore, the activity would act as a good introduction to the systematic work required in the GCSE course.

Interface issues

The software itself is not always easy to understand for new users. In the training when doing the search instead of pressing F1 for help they would ask their peers. In this situation the students were observed to come round and actually complete the task or make statements like: "it's the last one on your left", rather than providing an explanation. At least two students could not find the high and low sensor readings as they remained in the transporter room, and several did not pick up the casualty in the training level. Regardless of performance at this level the students proceed to the game, even if they have not demonstrated they know how to pick up a casualty or have been shown a meaningful graph. This led to the following exchange:

B1a: *We need something to do... what do we do* [he looks around]

B1b: *You know we need to find... we need to find spots... see the little thing in the bottom left... you have to find...*

Transcript 14: Group 1, Session 1 - Level 1, confusion in purpose

Eventually they summoned the teacher.

When interviewed after the sessions one of the girls admitted that she had thought Astroversity was a waste of time after the training, "The first time was really rubbish 'cos we didn't really know what to do".

The students were also confused as they could not identify what team they belonged to. The three players in Group 1 managed to join two different teams, and this was not confirmed until the exploration stage ended at different times, although the boys had tried to find their teammates during the exercise. The other team would have been impacted when they exited from the software to try again, as that would have caused these machines to hang when it came to information sharing. Another group were confused over what sensors they had chosen in Level 3. This was problematic as their scheme to record data was colour coded.

Engagement with task

The task appears engaging; the students are caught up in the task and are immersed, as illustrated in Transcript 15.

B7a: *I think our survivor's dead - with brain damage*
B7b: *Well why rescue him then?*
B7c: *That's not very nice, what about his great great great great great grandson?*

Transcript 15: Group 7, session 2, engagement in task

In particular the feedback from the rescue stage was motivating. The students related strongly to the casualty:

G4c: *He's going to get picked up... we've just got to watch the brain*
G4b: *It's going to get mushed*
G4c: *No... it's not going to be the brain getting mushed is it? It will be the lungs or the heart. Just watch how perfect they are and watch them fade away slowly*
G4a: *The heart*
G4d: *The heart's getting bigger* [all lean forward]
G4a: *That's not good*
G4c: *Oh yeah, that sensor was for the heart*
G4a: *We're going to die*
G4b: *Well the lungs are doing quite well*
G4d: *Shit our heart's going to explode*
G4a: *Ohhhhh*
G4d: *I didn't do the heart, I just did the brain bit*
G4b: *Shit it's exploded*
G4a: *Sorry sorry*
G4c: *It's no one's fault, we didn't know... we've just got to use the heart sensor... oh and the lungs have gone* [points at screen]
G4a: *The lungs are at a quarter*
G4b: *That's crap*
G4a: *The lungs have gone*
G4d: *Who's idea for a route was that?*
G4b: *The lungs have disappeared*
G4c: *Well it was perfect for the brain... perfect*
G4b: *It's crap - look at it*

G4d: *We have to find out where*

G4a: *We're dead*

Transcript 16: Group 4, Session 3, the rescue

Another positive feature of the trials were the numbers attending. More students attended the final session than the first. The six students who came for the last session came in partly because they had been told that the activity was fun, and all the students were disappointed that they were doing a science task rather than completing Astroversity in the final visit by the researcher.

6.3.2 Discussion about attitude to software

The beliefs about Astroversity in Table 13 reflect the two goals of the software. Although their self-assessment of collaborative skills may not change two thirds of students believe that the software assists the development of teamwork. If thinking skills incorporate scientific enquiry then 22% of the participants recognise that is a goal.

Moreover, the fact that students came in because they had heard it was fun indicates that it was spoken about outside the class and considered engaging. Despite the students displaying no noticeable improvement in task performance. This is positive as it shows that the students are engaged in a classroom activity.

The engagement aspect would explain the focus on the game rather than the reflection - this reflection is not integrated authentically into the activity, ie the rescue can be performed without consideration of the amount of listening or volume of data recorded. It is also an unusual activity to be asked to do. Traditional games do not stop you to ask about strategy efficiency, rather the reflection time is built into an activity - the character dies and you repeat the process.

The comments suggest that students often focus on the surface appearance of the software. For example, the suggestion that it benefits maths or that it enables students to learn how to rescue casualties in space (see Table 13). This implies the students would have difficulty transferring skills as there is little reflection on the underlying principles. Although engaging, this suggests that more discussion is needed around the software to draw out the principles if it is to be used in formal learning settings.

Finally, although not expressed by the student with dyslexia Astroversity appears to have potential for those who have poor reading skills. It provides an environment in which all three students are encouraged to contribute but in which the main task does not require reading or writing. This may be because the task has three elements that are essential to a successful group activity: the students have a shared responsibility in the outcome, they must communicate internally and externally, and to successfully complete the activity reflection is required. There is a degree of '"wicked problem"ness' about the task.

7. KEY FINDINGS

The following summarises the findings across all the iterations of this piece of software. It draws on the research questions outlined in Section 2 above.

Determine if it is possible to create a game that achieves the same level of engagement as mainstream games, but which enables children to learn about and practise collaboration and the processes of generating and testing hypotheses.

These areas are interlinked, engagement is closely related to strategies and collaboration, hence the combined list below.

- The concept appeared to be engaging - however, it was only used in a classroom setting, which means that despite the fact students said that they enjoyed Astroversity a direct comparison cannot be made with mainstream games. Possibly the study should have been revised to incorporate out of school activities, but given the restricted software this was impossible. However, the study did show that the software was engaging and encouraged students to attend a lesson despite the proximity to the summer holiday.
- The findings show that strategies are not necessarily improved, as in Table 10, only 11 of the 21 students changed strategies and of these two actually used less sophisticated strategies towards the end.
- The studies show it is possible to create a game that enabled students to practise collaboration and to a certain extent give them the opportunity to generate and test hypotheses.
- However, Astroversity shows that structured reflection has no impact when it is not embedded within the authentic tasks required of the game, the students do not respond well to the prompts, the time they are displayed is minimal and no discussion was recorded at the prompt, though some verified that they were putting in the 'correct' answer.
- The studies show that the least engaging aspect is the reading of instructions that are frequently ignored. The provision of information in the form of robot tutors does not impact the behaviour. The graphic display of exploding brains, hearts and lungs is the most engaging aspect.

What does this prototype tell us about the best ways of designing digital resources for learning?

This can be subdivided into two areas: interface and development process issues.

Interface

- As researchers we need to take context into account. For example, students that worked with researcher were more reflective than when the same questions were given by computer prompts. Perhaps this means that Astroversity should be provided with external documentation emphasising the need for external mediated discussion?
- Modes of collaboration - shift between working each other in the planning stage and on-screen, almost two sorts of relationships going on (with probe and with each other). This is engaging.
- The importance of an avatar that students can associate with - the students focused on identifying which colour they were, who had the greater speed, and the skill displayed manipulating the probe.
- That text instruction is alone insufficient; students often skip these and become confused.
- The prototype confirms the importance of the look and feel of the software. If the cartoon characters look suitable for a younger audience, regardless of the complexity of the task, the students will say it is too juvenile for them. There is clearly a cut off point for this. The cartoon tutors engaged adults who use the software.

Development process

- If the study was to be run again longer should be spent on the two-dimensional version. It was useful to 'practise' the questions to be asked and scaffolding the type of response. Further work is needed to develop prompts to assist the interpretation of the graph for example. The students did begin to realise there was a relationship between the toxins but they did not use this information to identify safe routes.

- The development of this prototype further emphasises the benefits of an iterative development process. The various stages allowed the researchers to confirm that the goal was valid: it confirmed students were poor at systematic thinking and did not collaborate without support. The discussions led to revisions in the software, for example, the difficulty the participants had in interpreting the graph led to the revision.

What does this prototype tell us about how informal learning processes can be transformed through use of these tools?

- Astroversity can act as a positive 'speaking and listening exercise' - which is applicable for English as well as science. This is because a shared task and representation (including socialising, eg racing) supports a group communicate.
- Astroversity highlights the role of reflection in learning - the need to scaffold and provide tools for the process. As discussed, the prompts were ineffective but the task itself was engaging and provided an opportunity for the students to repeat exploration and refine strategies as a consequence. More research is needed on tools to mediate this. The structure of the task is clearly one, and the fact that the goal is shared with all students having equal responsibility towards a successful rescue.
- Working with paper is feasible but only in a classroom situation and then it is often used ineffectively.

How does this prototype help us understand the potential of next generation technologies to create intrinsically motivating and engaging learning experiences?

- The engagement is clear due to the involvement with the task. The goal is defined but there is no right answer (see existing games), one can solve by trial and error but it is better to reflect on the information available. It is a 'wicked problem'.
- Context is important. Without support of an external mediator explicit reflection does not occur. The evidence for this is observing the response to the same question being asked by researcher and computer. Students need to be supported in the idea that it is necessary and beneficial.

8. RECOMMENDATIONS FOR NEXT STAGE DEVELOPMENT

This section looks at possible future directions for Astroversity based on the studies so far. It also reflects upon research questions that could be addressed.

8.1 Increasing complexity

If the puzzle is to be commercialised the content needs to be increased. The strength of Astroversity is the engagement with the rescue task, yet there is only one activity. However, before extra complexity is added it needs to be confirmed that the students are: developing strategies for plotting data, switching between moving probe, observing readings, recording data on paper, and then collaborating to find a safe route without giving up in a three-dimensional environment. The current findings indicate this potential but are insufficient to base future claims on, for example, the lack of students changing data recording strategy.

Future versions of Astroversity therefore need more than the incorporation of tasks used in the original VMULE. These tasks, such as standing on a stone to release a bridge which your partner could cross so that they could stand on another stone while you crossed, requires co-operation, but not necessarily collaboration. If the focus is still on developing scientific enquiry then the task should require the students to gather data - the question is whether this can be recorded on-line rather than on paper.

One method of increasing complexity without extending the task is the introduction of mentors

- can we design Astroversity so that other students pollute the space station, ie become the aliens and decide the layout of toxins? This could benefit both game players and mentors; game players would have different problems each time, while mentors would have to understand how gases interacted, the natures of the diffusion of heavy gases, the likely strategies tried by the rescuers etc.

8.2 Immediate work

The simplest, and probably most effective, addition to Astroversity is documentation. Teachers should be able to read the goals of the software, possible introductions and discussion questions. The latter is required as studies so far highlight that explicit reflection on group skills and strategy is poor; students need external encouragement to reflect in order to improve. Such documentation should also include the technical requirements, copies of the paper maps, and the arrangement required for computers, ie students will need to gather around one machine.

Minor changes to the software are also required. To assist interpretation of the graph it would make more sense to label 'likelihood of survival' rather than 'damage'. This would mean rescues are about probabilities and these are easier to relate to.

The narrative could be improved by another minor scripting change. If the professors were not advisors on board the Astroversity but in mission control, possibly trying to deal with another scenario, then it would make sense that the students would have to report back or be asked questions about interpretation of data.

8.3 Interaction and interface issues

There is a need for a stronger narrative. In a school environment the task is completed as compulsory. If doing outside then there are discrepancies in the back-story. At the end of the training is the infiltration of alien toxins; this is shown dramatically with the environment becoming distorted. It then makes sense to form groups to participate in the rescue. If choosing to be a returning student you are asked to perform a rescue immediately. Why would they let you dock onto the Astroversity if contaminated? Moreover, why are the robot tutors not doing the rescue when they have an understanding of what is required? (This would not be an issue if they were transferred to mission control and given other tasks.)

The students did not appear concerned about switching between mouse and keyboard for navigating through the game. However, they are not always systematic, the maps indicate a haphazard approach and very little methodical working. This suggests investigating what support would be helpful. Should the teachers discuss data recording methods prior to the activity, or should the robot tutors offer more advice. The former implies it is educational rather than a game, given current attitudes towards the tutors the latter may not get read.

The third area for improvement is the student feedback. Many students do not read the text - thus can we signify their progress in other formats? This is problematic as there is little method of assessing progress without access to their discussions and paper representations - self-reflection alone is not successful with this representation. Possibly corroborating information could be incorporated, eg the time to respond to the reflection prompts. If an individual takes a long time it doesn't follow they are thinking about the questions, but answering "absolutely brilliant" all the time immediately does imply a lack of thought. The robot tutors could suggest that they are not completing the performance review with as much accuracy as the Astroversity requires and delay continuation of the game or reduce time for exploration. Another suggestion would be to relate self-assessment on strategy to performance, so if a team felt they had a good strategy but inflicted lots of damage then the tutor could question their analytical skills. Again, this is fundamentally written feedback and may need to be supported by a delay before exploration to encourage discussion, or if the team is poor a demonstration of a possible approach.

8.4 Competing representational systems

The least successful area in terms of making Astroversity a mainstream game appears to be the requirement to use paper to record online observations. In a classroom setting this is not problematic; the students are forced to use the paper, but it is not instinctive. This is especially true for the boys who are more experienced game players. One solution is to incorporate a logging system so that the students can record online - they already have a map on screen and if they had the facility to mark areas of interest perhaps they would. This leads to the question of how could this best be represented. An advantage of Astroversity is the flexibility in choosing what to record and how. Giving the same choices online would be technically challenging. And the screens where the group come together would have to be redesigned so that all three maps and the way plotting screen were visible.

A second area for development is the graph interpretation. Students focus on the damage 'score'. Not the meaning behind. How can this be made more explicit without having more text? The icons are effective, but thought is needed as to how the relationship between the toxins can be highlighted if the students do not pick up on the graph lines. Does it require human mediation or do repeated plots lead to the software pausing the game or highlighting that area of the graph? These questions need further work.

8.5 Preferred scenario

Futurelab recommend that the next stage of development should consist of:

1) The development of Astroversity as a free online resource. This is because:

- there is insufficient evidence of learning gains for it to be publisised as a mainstream commercial game that supports the development of scientific enquiry and collaboration skills
- teachers and students could benefit from using it in its current form
- this type of resource would raise the profile of ICDC.

2) This requires:

- the development of teaching material
- adapting the interface as specified
- testing to see whether it can be downloaded and work on a variety of networks.

3) Futurelab to liaise with ICDC to find resources for funding this stage.

9. REFERENCES

Mercer, N (1994). The quality of talk in children's joint activity at the computer. Journal of Computer Assisted Learning, Vol 10, No 1, pp24-32

Millar, R and Osborne, JF (Eds) (1998). Beyond 2000: Science Education for the Future. London: King's College London

Wegerif, R (1996b). Using computers to help coach exploratory talk across the curriculum. Computers & Education: An International Journal, Vol 26, No 1-3, pp51-60