Experimental Design:

Input Device Protocols and Collaborative Learning

Joanna McGrenere
Kori Inkpen
Kellogg Booth
Maria Klawe

June 1996
The University of British Columbia

Acknowledgments

Our thanks go to Dr. Shelley Hymel and Dr. Rena Upitis for their helpful comments on this report. This research was supported in part by the Natural Science and Engineering Research Council of Canada (through a postgraduate scholarship and a Cooperative R&D project grant), TeleLearning Network of Centres of Excellence (through Projects 3.2 and 4.4), BC Advanced Systems Institute, Electronic Arts, and the Media and Graphics Interdisciplinary Centre (MAGIC) at the University of British Columbia.

Table of Contents

1. INTRODUCTION

1.1 General Introduction

This document outlines an experimental design for a study that investigates peer collaboration in a computer supported learning environment. Sections of this document have been adapted from a CPSC 533b term report by McGrenere et al. (1995) which outlined a similar experimental design. In the proposed study we examine different ways of supporting peer collaboration which, for the purposes of this study, refers to two students working on a single computer playing an electronic game. The standard computer is configured with only one mouse and therefore when two students share a computer they need to share the mouse as well. We want to investigate the impact of adding a second mouse to the configuration such that each child would have their own mouse.

The difficulty with adding a second mouse is that most computer applications or game software only accept input from a single mouse. In order to add a second mouse, additional software is needed. The software developed for this study accepts input from two mice and passes on one stream of mouse input along to the application level software. The issue of determining which of the two mice is active (i.e., the one whose input stream is sent to the software) is resolved through the use of mouse control passing protocols. A protocol is simply a set of rules for interaction. We have developed software to support two separate control passing protocols: Give and Take. These protocols rely on the use of different mouse buttons. In both protocols there is only one mouse that is active within the application software at any given time. It is always the case that the left mouse button of the active mouse is used to perform actions within the application software. The right mouse button passes control between the two mice. In the Give protocol the right mouse button of the active mouse can be used to give control from the active mouse to the non-active mouse. In the Take protocol the right mouse button of the non-active mouse can be used to take control from the active mouse.

Another configuration alternative for using two mice is to actually modify the application level software to accept two streams of mouse input. We call this the Concurrent protocol because both mice are concurrently active. As such, both children can perform actions within the software at the same time.

Some of the questions that we would like to address in this study are: How does the performance of a single child on a computer compare to the performance of two children sharing a computer? What is the nature of collaboration when two children must share a single mouse compared to when they each have their own mouse? When using two mice, how does the interaction among students and their performance differ under the protocols Give, Take, and Concurrent?

We would like to take this one step further. We would like to determine if peer collaboration using computers affects learning and, if so, which computer configuration best supports collaboration and learning. A more precise problem statement for this experiment can be found below.

1.2 Problem Statement

Does peer collaboration increase a student's ability to perform problem-solving tasks in a computer supported learning environment and if so how can it best be supported?
Our study examines how children learning in pairs (co-discovery) differs from children learning alone (self-discovery). For children learning in pairs, we investigate what type of computer configuration is most effective in co-discovery learning: one mouse setup vs. two mouse setup as well as what type of mouse control protocol (Give, Take, or Concurrent) is most effective. Hence, the focus of this research is the various ways in which children interact with computers in a learning environment.

1.3 Importance/Benefits of this Study

This study is an attempt to build on the previous studies conducted by Inkpen et al. (1994,1995). These studies did not assess whether subjects learned concepts and strategies while they performed an experimental task. In the proposed study we address this issue by having all the subjects play *individually* after they have successfully completed a first round of game playing. This second round of play allows us to test for individual learning. This study also introduces the Concurrent dyad condition which we feel is an important sharing modality that must be investigated.

On a larger scale, this research has implications for the design and installation of computer-based instructional software for children. As Inkpen et al. (1994) note, there is a growing emphasis on cooperative learning in schools. We believe that research on computer-supported collaborative learning (CSCL) will help to aid designers of instructional software as well as educators who must create the collaborative learning environments in schools. Finally, the study of CSCL can possibly benefit research in the larger domain of computer-supported cooperative work (CSCW).

1.4 Organization of the Report

The remainder of this report will be organized as follows: Section 2 presents background information on both the system under consideration and previous studies, Section 3 discusses the sampling method, Section 4 discusses the experimental design, Section 5 discusses the data collection techniques considered in the study, Section 6 discusses issues of reliability and validity, and Section 7 discusses the evaluation and data analysis.

2. BACKGROUND

2.1 Description of System

2.1.1 Software and Hardware Used

The computer game chosen for our experiment is a puzzle-solving game, called "The Incredible Machine" (TIM), created by Sierra On-Line, Inc., (Coarsegold, CA 93614) in 1993. This software runs on any IBM-compatible computer, as well as on Macintosh Apple computers. For the study being documented we intend to use IBM-compatible PCs. This choice of platform was made based on the differences in the TIM user interface for the two different platforms. Previous research found that the user interface for the PC version of TIM is more intuitive for children than that of the Macintosh version [Inkpen et al, 1996].

The Macintosh version has some inconsistencies in the interface which makes it more difficult to use than the PC version.

The dyad conditions Give, Take, and Concurrent require the use of a second mouse. For these conditions there will be two mice hooked into a single PC and a second mouse driver will be needed. Software that implements the control passing protocols will be used for the Give and Take conditions. For the Concurrent condition, the source code of TIM will be modified to accommodate two inputs.

2.1.2 Playing the Game

Puzzles are presented to the player in the form of an unfinished "machine" that the user has to complete, using a wide variety of simulated tools, so that the completed machine, when run, performs a predefined action. This action can be, for example, to shoot a basketball into a hoop. The tools resemble those used in every day life, and include gears, pulleys, ropes, ramps and levers, and also balls, scissors and trampolines. The game features a total of 45 animated parts that can be used to create working machines and over 75 levels of puzzles. The player also has the option to change air pressure and gravity and use the free-form mode to invent his/her own fantasy machines, but this function is not used in our study.

When the game is started, the player is first presented with a control panel, where he or she can select a puzzle to solve (see Figure 1) . The listed puzzles are of approximately increasing difficulty. On the control panel, the user can also see the initial (incomplete) machine, and below, a short textual description of the problem, indicating the goal of the completed machine. The next screen contains three main areas: the playing area, which is the largest; on the right side, a toolbox containing parts which can be used to complete the machine; and in the top right corner, a "run the machine" icon, which, if clicked-on (pressing the mouse button when the cursor is over the icon), starts the machine (see Figure 2). To stop the machine, the player clicks anywhere on the puzzle screen. The process of placing and/or attaching the provided parts and running the machine can be repeated until a correct solution is found, at which point a message box will pop up on the screen, informing the user that the puzzle has been solved.

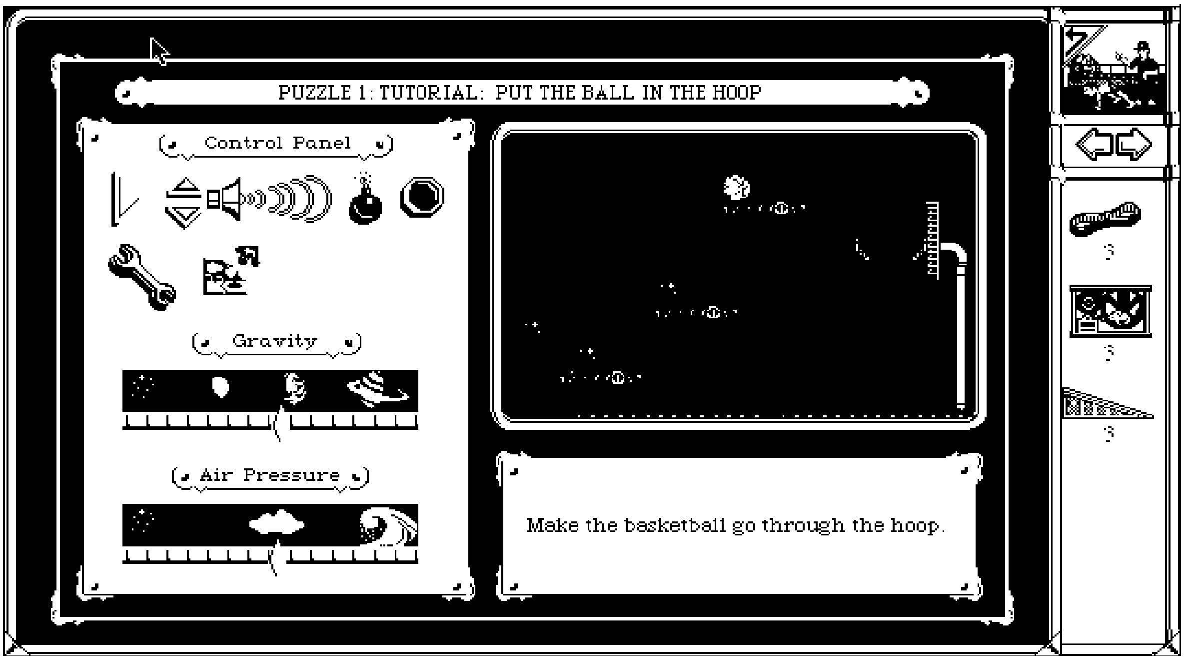


Figure 1: control panel in The Incredible Machine

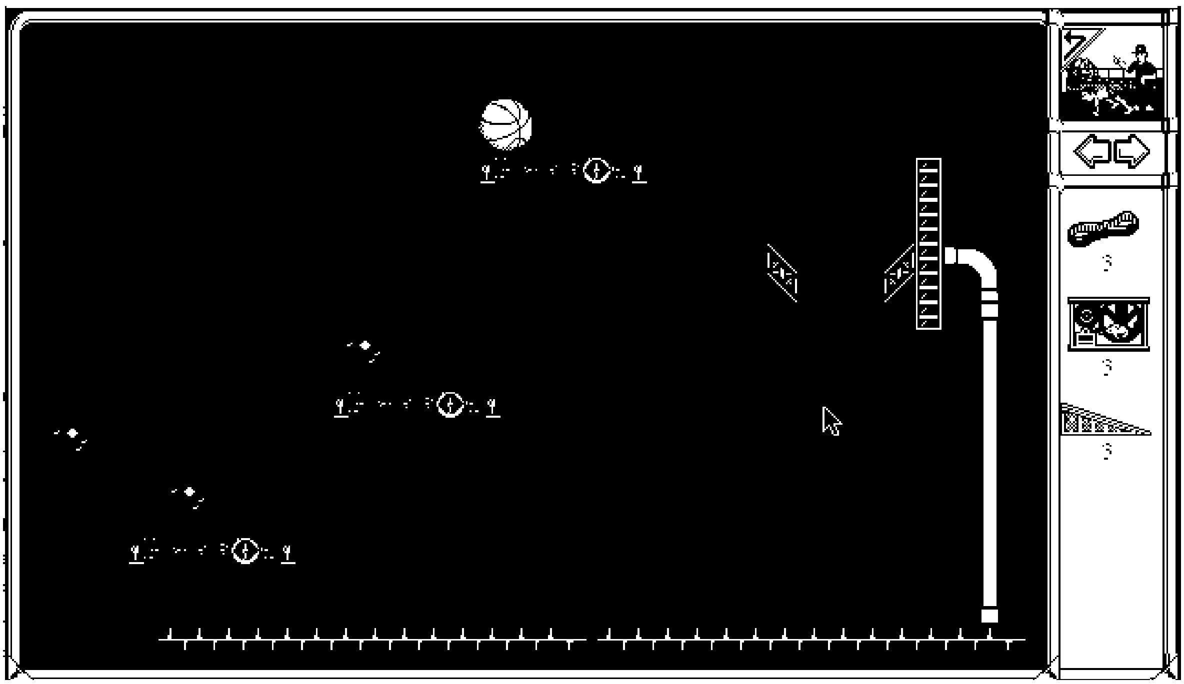


Figure 2: play screen for the first puzzle in The Incredible Machine

Many of the puzzles have more than one solution. The manipulation of the parts is not always straightforward and usually requires some experimentation. Some of the parts, such as scissors, may need to be "flipped" (i.e., turned to face in the opposite direction) in order for the sharp blades to be used in the workings of the machine. In many cases, the player has to think of novel uses of the available parts to complete the machine.

2.2 Previous Studies

Inkpen et al. (1995) conducted a study, “Give and Take,” in which dyads of same-gender children (ages 9 to 13) were asked to solve TIM puzzles on a computer equipped with either one or two mice. In the two-mice setup the left mouse button was used to play the game and the right button passed control between the two mice (note that there was just one cursor on the screen). In the Give protocol, pressing the right button gave control to the other player, while in the Take protocol, pressing the right button took control from the other player. In addition to the Give and Take experimental conditions, there was also a Solo condition in which a child played TIM by himself/herself. The results suggested that having two mice instead of a single shared mouse positively affected the performance of a pair of children playing on a shared computer. The results were also gender dependent. Girls solved statistically more puzzles in the Give condition than in the Solo condition. Boys showed a trend for solving more puzzles in the Take condition but the sample size was not large enough to find statistical significance.

An earlier study, "Cooperative Learning in the Classroom" conducted by Inkpen et al. (1994), examined the cooperative behavior of children playing The Incredible Machine computer game. Each of the children was placed in one of three conditions: Solo, Parallel, or Shared. In the Solo condition a child played on the computer by himself/herself. In the Parallel condition two children were placed side by side each with their own computer. In the Shared condition two children worked together on one computer and had to share the mouse. The combination of gender, as well as whether one or two computers were used (note that only one mouse per computer was used as input in this study), affected the number of puzzles completed in the game. The comparisons analyzed were: Solo play vs. group play (Parallel and Shared), Shared vs. Parallel, as well as gender pairings.

Female/Female pairs sharing one computer, on average, completed more puzzles than females in either Solo or Parallel conditions. This difference was statistically significant between the Shared and Parallel conditions. The male/male comparisons did not reveal any statistically significant differences, probably due to sample size.

Both studies addressed achievement and they indicated some gender differences, but our ability to extrapolate valuable information from their results is somewhat limited. The dependent variable, number of puzzles completed, is of limited utility because fewer puzzles solved does not necessarily reflect less learning or even less fun. The studies would have had to include a test for individual learning (subjects were not retested individually after playing in groups) and/or a retention test in order to provide more valuable information.

2.3 General Background

2.3.1 Cooperative Learning in Schools

There is vast research literature on cooperative learning in schools. Inkpen et al. (1994) provide a summary of some of the findings related to the benefits of cooperative learning in schools. These findings include a positive increase in student achievement and positively affected attitudes toward school and classmates (Hymel et al., 1993; Johnson et al., 1981). Many teachers today are incorporating cooperative learning methods in their classrooms. The curriculum development for elementary schools also tends to incorporate group interaction (Inkpen et al., 1994, 1995).

There have been several studies which have demonstrated the benefits of cooperative learning in various settings. Lim, Ward, and Benbasat (1994), for example, conducted a study which showed that co-discovery groups (pairs working together) outperformed self-discovery subjects when they learned an electronic messaging system. Results in this study showed that co-discovery subjects formed better mental models of the task than self-discovery subjects.

However, there is a lack of research related to the design of computers and their interfaces to facilitate cooperative learning in the context of the traditional classroom (Inkpen et al., 1994). Our study looks at these HCI issues.

2.3.2 Electronic Games and Collaborative Play

There are many electronic games that promote collaborative play. Children, when playing electronic games, very often exhibit collaboration, since they are highly motivated to solve problems and improve their skills during play. Electronic games are based on challenge, fantasy and curiosity, thus combining intrinsic motivations for learning with interpersonal motivations, such as cooperation and competition. Computer-based issues in education have not yet been explored to the point of suggesting ways to design computers and their interfaces so as to maximize collaborative learning in the classroom.

2.3.3 E-GEMS (Electronic Games for Education in Math and Science)

This experiment will be executed within the framework of the E-GEMS research group, thus benefiting from the relationships that E-GEMS has already established with several public schools in Vancouver.

General Goals. The E-GEMS project is a collaborative effort among scientists, educators and professional game developers, aimed at increasing the proportion of children in Grades 4-8 who enjoy learning and exploring concepts in math and science. The research focuses on exploring and developing teaching materials that will integrate both video and computer games with existing classroom practices.

Projects. The E-GEMS research team focuses on human-computer interaction issues related to learning. Specific projects include research on existing computer games as well as designing prototype games that incorporate specific mathematical or scientific concepts. The aim is to find game formats that are attractive to students and, at the same time, are suitable for learning particular concepts. Effects of collaborative play with respect to computer games and ways to integrate educational electronic games into the school curriculum are also studied.

3. SAMPLING METHOD

3.1 Problems in Previous Studies by Inkpen et al.

As mentioned in Section 2.2, Inkpen et al.(1994, 1995) performed two independent studies of collaborative uses of computers, the "Cooperative Learning" and the "Give and Take" studies. The populations for the two studies differed slightly. The "Cooperative Learning" study was performed in a Vancouver public school in an upper-middle class neighbourhood in January of 1994. 52 girls and 52 boys in the age range of 9-12 participated. The study looked at same-gender dyads as well as mixed-gender dyad groupings, and addressed both solo play in addition to collaborative play. The second study, "Give and Take", which was performed at Science World B.C. in August 1994, included 66 girls and 66 boys in the age range of 9-13. This study only looked at collaborative learning of same-gender dyads; there were no mixed dyads and no solo play conditions.

There were two problems with the populations chosen in these studies and some of the analysis that was performed.

1. Self-Selection Bias and Generalizability

The "Give and Take" study used a self-selected population (i.e., children visiting Science World). The subjects chose, by themselves, to participate or were encouraged by their parents to participate. One would suspect that the population of children visiting Science World would have a keen interest for science and therefore wouldn't reflect the general population of children. The fact that many non-science camps bring their children to Science World over the summer season does, however, somewhat mitigate the argument that only children interested in science go to Science World. Overall, however, the degree to which this self-selected population generalizes to children in the age range of 9-13 is highly suspect. The issue of generalizability also comes into play with regards to the sample of the students for the "Cooperative Learning" study. Since the subjects were from upper-middle class neighbourhoods, the generalizability of the results could be limited. We question whether these results can be generalized to children of all economic backgrounds.

2. Statistical Significance

The results of the studies showed trends but statistical significance was only found in a few areas. We suspect that larger samples would reveal that some of the trends are actually statistically significant differences.

The approximate number of observations per condition is recorded below for each of the two studies. Note that in the Parallel, Integrated, Shared, Give, and Take conditions, two subjects are required for each observation.

Collaborative

	Solo	Parallel	Integrated
F/F	8	8	8
M/M	8	8	8
M/F		8	8

Give vs. Take

	Shared	Give	Take
F/F	10	10	10
M/M	10	10	10

These studies showed that there is a medium size effect (approximately 0.6). For adequate power of 75% or greater with this effect size, we need 17 or more observations per condition. (See Appendix 1 for more information on power analysis.)

3.2 Target Population

The target population for this experiment is middle class Canadian school children. There are a number of variables that should be controlled to reflect this population. These variables include: age, economic class, cultural background, and setting. We will assign the following values to these variables:

ages:	10 - 13
economic:	middle class
cultural:	Canadian
setting:	public school

The justification for the age range is to replicate and extend the studies conducted by Inkpen et al. and hopefully find statistical significance where they did not. There were two main reasons for choosing this age range in the original studies: (1) the level of difficulty of the game TIM is appropriate for children of this age, and (2) this is the E-GEMS target research population because it is this age at which children tend to lose interest in math and science, particularly females. We chose to examine middle class children in public school because this group, in our opinion, represents the majority of Canadian children. As such, the study would have a higher degree of generalizability than the other Inkpen et al. studies. Clearly we would reach a higher degree of generalizability if we could sample students from varying populations from all across the country. The cost to run this study with such a sample, however, would exceed our resources. A school setting was chosen over a science centre type of setting in order to minimize the effects of a self-selected population.

There are a number of other variables that would impact the outcome of the study. Some obvious variables are the children's problem solving skill and the nature of pairings for the pairs play condition. We might expect a child with a high problem solving ability to perform better than a child with a low problem solving ability. And in pairs play, the nature of the interaction between the children and the computer would likely be dependent on whether the pair of children were friends or not and the social skills of the children. Randomization will be used to factor out the impact of the above mentioned variables. Students will be assigned randomly to the various play conditions (individual - single mouse; dyad - single shared mouse; dyad - two mice with Give protocol; dyad - two mice with Take protocol). Given a sufficiently large sample size, randomization mitigates the effects of the many extraneous variables.

3.3 Sample Size

360 subjects are required; 180 girls and 180 boys. This sample size has been determined by power analysis. Given that the effect size is approximately 0.6, this sample size will give a power of about 0.83.

4. EXPERIMENTAL DESIGN

4.1 Experimental Conditions

The five experimental conditions are:

(1) individual - single mouse (Solo)

(2) dyad - single shared mouse (Shared)

(3) dyad - two mice with Give protocol (Give)

(4) dyad - two mice with Take protocol (Take)

(5) dyad - two mouse with Concurrent protocol (Concurrent)

The following table provides the number of observations required for each experimental condition:

Number of Observations

	Solo	Shared	Give	Take	Concurrent
Male	20	20	20	20	20
Female	20	20	20	20	20

There will only be one observation recorded for each dyad, therefore, the following allotment of subjects to experimental condition will provide the required 20 observations per cell.

Subject Requirements

	Solo	Shared	Give	Take	Concurrent
Male	20	40	40	40	40
Female	20	40	40	40	40

Total Number of Subjects: 360

4.1.1 Reasons for the Choice of Experimental Conditions

We selected these five experimental conditions because they allow us to make comparisons in children's achievement and learning when working alone vs. working in pairs, in a computer-supported learning environment. Further, they enable us to examine what type of input device configuration facilitates learning and achievement in a collaborative environment.

The Inkpen et al. “Cooperative Learning” study also includes a one-computer vs. parallel (two-computer) setup for children working in a collaborative environment. We chose not to include this condition in the present study because the results showed that pairs working on the same computer solved more puzzles than pairs on side-by-side (parallel) computers (Inkpen et al., 1994).

4.1.2 Comparisons

Our primary interest is to determine which condition is more conducive to an individual's learning, and if and how this relates to gender. This is measured by counting the number of puzzles completed within a fixed time period and then retesting all subjects individually and comparing the number of puzzles solved. Collaboration is assessed by observing the nature of mouse exchanges and communication as well as by logging the number of mouse exchanges. Attitude information is gathered using questionnaires and direct observation.

Subjects will be retested with puzzles that are conceptually similar to the puzzles that are used in the first round of play but are structurally different. This will allow us to control for memorization effects. For example, if Puzzle #1 in the first round of game playing consisted of three mouse cages, three conveyor belts, a bowling ball, a basket ball and a basketball hoop and the goal of the puzzle was to get the basketball through the hoop. A conceptually similar yet different puzzle would use the same objects and would have the same goal but would require the objects to be placed in locations different than in the first puzzle in order to be solved. Thus a student who had simply memorized the placement of the objects in the solution to Puzzle #1, would not be able to solve the conceptually similar puzzle simply by placing the objects as they were in Puzzle #1. The student would need to re-solve the puzzle.

4.2 Experimental Setting

The experiment will take place in six middle-class schools (to be selected at a later time) in Vancouver, British Columbia, Canada, from March to June, 1996. A room will be set-up to run three to four sessions in parallel. The Solo, and Shared conditions will utilize a standard IBM compatible PC with one mouse. The Give, and Take conditions will also use a PC computer equipped with two mice and a program to transfer control between the two mice depending on the type of protocol being tested. The Concurrent condition will use a PC equipped with two mice and a modified version of TIM that permits concurrent play.

The room will also contain camcorders and a VCR equipped with a VGA to Video converter box. The camcorders will be used to record the children's interactions in the Concurrent and Shared conditions. The VCR will be used in the Concurrent condition to record the screen output to create an archived copy of the children's achievement and interaction within the game[1].

4.3 Experiment Schedule

The average school has seven 40-minute periods per day. There are four scheduled periods in the morning and three in the afternoon. In order to minimize the total length of the study, it is essential that we utilize as many of these periods as possible. Below is a possible scenario for the two rounds of play that would enable us to use the periods most effectively.

ROUND 1	
Activity	Minutes
arrive	5
general intro	2
interface training	6
intro to playing	2
game playing	25
TOTAL	40

ROUND 2	
Activity	Minutes
arrive	5
general intro	2
game playing	25
questionnaire	8
TOTAL	40

[1] In the Concurrent Play condition there will be two different colour cursors on the screen. Thus in the archived screen output it will be possible to distinguish the manipulations of each subject.

Before Round 1 there will be a general introduction given to the whole class. Any information that the students should know about the experiment will be discussed at this time. Students will also have an opportunity to ask questions about the study. Having a collective introduction enables the general introduction scheduled in Round 1 to be as short as possible (only a reminder of the task will be required) so that the majority of the period can be used for playing the game.

Directly following the collective introduction, each subject will individually complete a questionnaire (see Appendix 2). This questionnaire will help the experimenters to ascertain the general computer and electronic game experience of the subjects as well as the subjects' attitudes towards electronic games. By administering the questionnaire collectively rather than in Round 1, more time will be devoted to game playing in Round 1.

During Round 1, but before playing TIM, all children will participate in an interface training session. First, the control panel screen will be described to the children. They will then be shown the box that describes the goal of the puzzle and will be told to click on the puzzle screen to start playing the puzzle. Once inside the game, the three parts of the screen will be described to the children: the playing screen, the toolbox, and the "start the machine" icon. The children will then be shown how to drag tools from the toolbox onto the playing screen and how to attach elastics and ropes. Next the children will be shown how to flip and resize objects. Finally, the children will be shown how to start and stop a machine that they have built. Lastly, the children will be given a picture of puzzle pieces on a screen which they will be asked to replicate (see Appendix 8). In order to complete this task, the children will have to perform all of the manipulations previously shown to them (dragging, attaching, flipping and resizing). Each child is required to duplicate this picture to demonstrate that they know how to operate the interface. At this stage the observer will recognize if any of the children are having difficulty with a particular part of the interface and will then assist that child with the item of difficulty[2].

---

[2] This interface testing appeared to be a success in a pilot study. Very few children had difficulty with the interface compared to children in previous studies.

ROUND 1 (assessment of achievement)

Arrive (5 min)	Allow the children time to move from their classroom into the experiment room.
General Introduction (2 min)	Subjects are welcomed by the experimenters and are briefly reminded about the research being done. Subjects are put into their predetermined experimental condition. Subjects in a dyad condition are told that they will be working with a partner. Subjects in the Solo condition are told that they will be working alone.
Interface Training (6 min)	Subjects perform a few tasks that will help to familiarize them with the interface of TIM.
Introduction to Playing (2 min)	Subjects are told to solve as many puzzles as they can, and that they can use the game manual if they so desire.
Play (25 min)	Subjects play TIM starting with Puzzle #1. They aren't given any help from the experimenters.

ROUND 2 (assessing individuals' learning)

Arrive (5 min)	Allow the children time to move from their classroom into the experiment room.
General Introduction (2 min)	Subjects are welcomed by the experimenters and are told that they will be playing the game again but this time they will each be playing by themselves.
Play (25 min)	Subjects play the puzzles that are conceptually similar to the puzzles in the first round but are structurally different.
Questionnaire (8 min)	Subjects will complete the second questionnaire and answer three pinwheel questions. (See Appendix 4 for a description of the format of these questions.)

Round 2 is always played in the Solo condition. In order to process the maximum number of children it is best to use the last four periods for Round 2. Given that there are four sessions running concurrently this will allow sixteen subjects to complete Round 2 per day. Thus we can process up to a maximum of sixteen subjects for Round 1 which would be run in the first three periods of the school day. Processing 360 subjects at a maximum of sixteen per day requires a minimum of twenty-three days to run the study which is approximately five weeks. Given that problems will arise an estimate of seven weeks is probably more realistic.

Another factor that will contribute to a lengthier study than otherwise might be anticipated is that children who have previously played or been exposed to TIM must be given an opportunity to play during the study even though their data will not be used[3]. At a minimum, these children must be given a chance to play Round 1. Thus seven weeks is probably an accurate estimate.

This time estimate for running the study could possibly be reduced if the schools have a PC lab. In this case, Round 1 could be run in one day for all subjects in a class and Round 2 could be run in only a half day.

4.4 Anticipated Problems

Some possible problems we might encounter are our relationship with the class teacher, undisciplined children, and unforseeable interruptions. The class teacher has to be instructed not to influence the children’s attitude and/or motivation toward play. Children may also be sharing information about the individual puzzles during the time in between Round 1 and Round 2. See Section 6.3 for a discussion of some problems related to internal validity.

5. DATA COLLECTION TECHNIQUES

5.1 Description and Justification of Data Collection Methodology

A number of data collection techniques will be used in this experiment. They include questionnaires, direct observation, video, computer logging, and screen logging. The use of multiple techniques is important. Some techniques are used to “triangulate” with one another

[3] Based on a pilot study run in December of 1995, approximately 25% of the students will have been exposed to TIM.

in order to ensure that the “correct” data is being collected. For example, video will be used, in part, to confirm that data collected through direct observation is accurate.

Each of the data collection techniques that will be used in this experiment is discussed below. The following details are covered: when a technique will be used in the experiment cycle; the purpose of its use; the subjects on which it will be used; the data that will be gained; and the pros and cons of using this technique in our particular experiment.

5.1.1 Questionnaires

See Appendices 2 and 3 for samples of the questionnaires.

***who*:** All children.

***when*:** There will be two slightly different questionnaires. The first will be given before Round 1 of game playing and the second will be given at the end of Round 2.

***why*:** The first will determine the general computer and video game experience of each child and his/her basic attitude towards playing video games alone and collaboratively. The questionnaire will address the following: familiarity with computers; whether a child likes video games or not; whether a child plays video games frequently and, if so, whether he/she plays on a computer or on a video game (Sega, Nintendo) platform; whether a child prefers playing video games alone or with a friend, and whether or not a child has played TIM before. The second questionnaire will be similar to the first except that it will not include general computer and video game experience. It will only address the children’s attitude towards playing video games alone and collaboratively. Comparing the attitude portion of the two questionnaires will enable the experimenters to determine whether the children’s attitudes change once they have played in a pair during the experiment.

***pros*:** Quick and easy to analyze since, in general, a range or selection of answers will be provided for each question and the students only need to circle the most appropriate answer. This enables us to perform statistical analysis on the questionnaires.

***cons*:** Doesn't provide a deep exploration of the child's attitude towards collaborative play.

5.1.2 Direct Observation

See Appendices 5 and 6 for samples.

***who*:** A subset of the children. An experimenter can only effectively observe one dyad or solo subject at a given time. There will be two experimenters at all times and so 2-4 subjects will be observed during each session.

***when*:** While children play the computer game.

***why*:** To obtain qualitative data on the nature of the children's interaction with each other and the computer. We want to discover unexpected behaviour. For example, two children operating a single mouse each with one hand on the mouse (as reported in one of the Inkpen et al. studies). Things that will be looked at are the nature of mouse transfers, verbal communication, pointing to the screen, aggressive or disinterested behaviour, and general difficulties with the interface.

***pros*:** Observation can be done in real time.

***cons*:** We could miss something while recording. The observer is often selective in what gets recorded.

5.1.3 Video

***who*:** Subjects in the Shared and Concurrent conditions only.

***when*:** While children play the video game.

***how*:** One video that captures the children interacting with the mouse/mice.

***why*:** To obtain data on mouse sharing and mouse exchanges. This data will be used to compare with similar data obtained from logging in the Give and Take conditions.

***pros*:** Provides a permanent record of the required data.

***cons*:** Similar to direct observation, children could perform differently if they know they are being videotaped. Video generates a huge amount of data that cannot be analyzed in real time.

5.1.4 Computer Logging

See Appendix 7 for a sample computer log.

***who*:** Subjects in Give and Take conditions.

***when*:** While children play the video game in Round 1.

***what*:** The number of exchanges of the mouse and how long (in seconds) each child has control of the mouse.

***why*:** Could provide insight into the performance of a child in Round 2. For example, if a child solves significantly more puzzles in Round 1 while playing collaboratively than in Round 2 while playing individually, it might be explained by the lack of control the child had over the mouse during Round 1.

***pros*:** Easy to do and is completely non-obtrusive.

***cons*:** Data lacks contextual richness.

5.1.5 Screen Logging

***who*:** Subjects in the Concurrent condition.

***when*:** While children play the computer game in Round 1.

***what*:** Screen output which shows the activity of each subject within the game.

***why*:** Similar to computer logging, this could provide insight into the performance of a child in Round 2. For example, if a child solves significantly more puzzles in Round 1 while playing collaboratively than in Round 2 while playing individually, it might be explained by the lack of control the child had over the mouse during Round 1.

***pros*:** Easy to do and is completely non-obtrusive.

***cons*:** Additional data to analyze.

6. RELIABILITY AND VALIDITY

6.1 Reliability

A measure is reliable to the extent that it supplies consistent results. More reliable results are obtained from interpretive measures (coding videotapes and direct observation) when more than one observer or coder is used. The extent of the agreement between the coders is an indication of how reliable the measures are. We feel that coders should preferably be naive about both the experimental design and the hypothesis to be tested -- hence they should not be the researchers. Due to our budget for running this study, we will be unable to have naive observers. Instead, the two main experimenters will be performing the direct observation. This we feel is justifiable because the data from the direct observation will only be used for qualitative analysis. The data from the videotapes, on the other hand, will be used for the quantitative analysis of mouse sharing and exchanges. We will have a single person unfamiliar with this study doing all of the coding of videotapes.

6.2 External Validity

External validity refers to the ability to generalize results across persons, settings, and times. We would like to generalize our experimental findings to children between the ages of 10 and 13. However, because of our limited sample (we are suggesting the use of six middle class public schools), and because of the artificiality of controlled laboratory environments, we realize that our results may not necessarily be generalizable to other environments. We will take some steps, however, to improve the external validity of our study. We have chosen a public school setting from which to sample the children. This is a more generalizable population than children at Science World, because public school children are more representative of children across Canada given that the Science World environment may have a self-selected bias. In addition, we have reduced the obtrusiveness of our study whenever possible, encouraging the children to act as naturally as possible.

6.3 Internal Validity

Since we are constructing an experimental study, internal validity issues are important. Internal validity refers to the extent to which the conclusions we draw about an observed experimental relationship truly imply cause. There are many possible threats to internal validity:

- Children could communicate with one another and talk about the games, thereby affecting the results. The experiment is set up such that all subjects perform both Round

1 and Round 2 in the same day. The majority of subjects play Round 1 in the morning and then play Round 2 in the afternoon and hence have a lunch period in between the two rounds. Communication could take place at this time.

- There could be problems with the selection and randomization procedures to assign subjects to the various experimental conditions. Extra care must be taken to ensure that children are assigned in a non-biased, random way. A teacher, for example, must not select which students can participate in our studies. One method which is often used to perform randomized assignment is the use of random number tables.
- Attrition in the experimental group can sometimes happen. With each dropout, the makeup of the group changes, and the results that are obtained can be skewed. Care must be taken to ensure that children participate fully in the experiment to the very end. Based on previous studies by Inkpen et al., only a few dropouts are expected. The data from these subjects will be discounted.
- There are many other threats to internal validity. These could be present even if the experimenter does take extra care to randomize subjects' assignment to treatment groups. For example, subjects in the control group (in our case, the Solo condition) could be aware that they are in a control group and will, as a result of competitive pressure, work harder to perform. If possible, subjects should not know what the other subjects are doing and what kind of treatments they are getting.

6.4 Triangulation of Different Data Collection Methods

We will try, whenever possible, to triangulate the results of the different data collection methods to obtain increased validity. For example, we could find out whether there are dyads in which one member is more dominant than the other through software logging. This would give us a quantitative measure of the number of mouse exchanges as well as the amount of time each subject is in control of the mouse. Such a quantitative measure could be compared with direct observation of the subjects.

7. Evaluation and Data Analysis

Controlled variables, such as age range, were described previously in Section 3.2. This section will briefly introduce the independent and dependent variables before moving on to a discussion of the data analysis.

7.1 Independent Variables

7.1.1 Experimental Treatments

discovery mode

People learn in many distinct ways. These ways include learning by themselves (self-discovery), or learning by working together with someone else on a problem (co-discovery). For our particular study, we will look at both self-discovery and variations within a co-discovery mode. Note that we are using the term co-discovery somewhat loosely in this study because the mouse sharing protocols do not ensure co-discovery.

Possible values: {self-discovery; co-discovery}

input device

At each computer station there will be one or two mice. Subjects in a co-discovery mode may be sharing a single mouse or may each have their own mouse. Subjects in the self-discovery mode will (obviously) only have one mouse.

Possible values: {one mouse; two mice}

protocol

When subjects have two mice, they utilize either the Give protocol, the Take protocol or the Concurrent protocol.

Possible values: {give; take; concurrent}

treatment condition

This is a composite variable of the three independent variables: discovery mode, input device, and protocol. In our study, it can take on five possible values.

Possible values: {single subject - Solo;
two subjects, one mouse - Shared;
two subjects, two mice, - Give;
two subjects, two mice, - Take;
two subjects, two mice, - Concurrent }

7.1.2 Gender

gender

In the self-discovery mode, both individual males and individual females are examined. In the co-discovery mode, two pairs are possible: male/male and female/female. (There will be no male/female pairs in this experiment[4].)

Possible values: {male only; female only}

7.2 Dependent Variables

7.2.1 Achievement

Achievement is measured strictly as the number of puzzles solved. We do not differentiate between skill in using the interface and skill in solving puzzles (conceptual understanding of the puzzles). The overall time it takes to complete a puzzle will probably be a function of both of these components. We do, however, have an interface training period which should minimize the effects of achievement due to skill in using the interface. An extension to this study might be to separate out these two effects.

[4] Previous research has shown that in mixed-gender pairs there may be an interaction between gender and other variables that produces a more complicated result (Lockheed & Hall, 1976).

Round 1 scores

Round 1 scores represent the number of puzzles completed by pairs utilizing co-discovery, and the number of puzzles completed by individuals in the self-discovery mode during the first round of play.

7.2.2 Learning

Round 2 scores

All subjects are subsequently tested individually for learning by having them participate in a second round of game playing. We will use the scores achieved in this second round as a measure of learning.

We are using the school paradigm for assessment of learning. In the same way that a school gives students individual tests and uses the test scores as an assessment of learning, we will use the scores from the second round of play, which represent individual performance, as an assessment of learning.

7.2.3 Attitude towards Collaboration

questionnaire and direct observation

We will collect information about the subjects' attitudes towards collaboration using direct observation. This information includes the verbal communication between the subjects, the attitude of the subjects towards the game and their fellow subjects, and group dynamics (including issues such as whether one child in the dyad dominates the game). Children's attitudes towards collaborative play will be measured in the questionnaire.

7.2.4 Nature of Collaboration

We examine the nature of collaboration as an attempt to understand *why* variations in learning between experimental treatments take place. Collaboration is measured by two variables provided to us through software logging for the Give and Take conditions and through video coding for the Shared and Concurrent conditions. The two variables are the number of mouse transfers and the percentage of mouse time. We believe that there may be interaction effects between these two dependent variables.

mouse transfers

The number of times control of the mouse was transferred between the two subjects.

percentage of mouse time

The percentage of time for which a subject had control of the mouse.

The following table summarizes possible hypotheses about the roles of these variables.

	low number of transfers	high number of transfers
even time distribution	• Taking turns playing games.	• Fighting for control. • Taking turns placing objects.
lopsided time distribution	• One subject dominates the play. • One subject controls mouse while other directs.	• Fighting for control • One subject dominates the play.

7.3 Data Analysis and Interpretation

7.3.1 Statistical Analysis

There are many different ways of analyzing a set of data. This means that we can use more than one statistic to test for an effect. The effects that we are interested in testing can be found below. For completeness, we list a number of possible statistics that we could use for testing each of these effects. When we run the actual analysis, we expect to use only a single statistic per effect.

Effect of Treatment Condition

The main effects of interest are to test whether there are differences in *Round 1 achievement* and in *learning* among the five experimental conditions.

Round 1 achievement is somewhat more straightforward so we will discuss that first. There are three possible ways that we can analyze this data. They are listed below in increasing order of power.

1. χ^2 (chi-squared)[5]
2. 2 x 5 factorial ANOVA
3. pre-planned comparisons

The χ^2 and the ANOVA are less powerful than pre-planned comparisons. These two weaker statistics can only tell us if an effect exists. They cannot tell us which treatment conditions are better than others or which is the best overall condition. With the ANOVA, however, we can run post-hoc comparisons on all possible pairs using the Tukey test.

The pre-planned comparisons are the most powerful statistic we can use. They will enable us to determine where, if at all, the effect lies. Given that we have five treatment conditions, we are able to perform four pre-planned comparisons without violating Type I error. Here are the four comparisons that we expect to run:

1. *self-discovery vs. co-discovery:*
   Solo condition vs. Shared condition for both boys and girls
2. *Give vs. Take*
3. *shared-mouse vs. two-mice sequential*:
   Shared vs. Give for girls
   Shared vs. Take for boys
4. *two-mice sequential vs. Concurrent*:
   Give vs. Concurrent for girls
   Take vs. Concurrent for boys

[5] A chi-squared statistic will be the only possible statistic we can use if the data is found to be non-parametric.

We are able to test different two-mice sequential conditions for boys and girls in comparisons #3 and #4 because previous studies by Inkpen et al. showed that girls performed better in the Give condition and boys performed better in the Take condition.

We next discuss Round 2 *learning*. There are a number of ways that we could operationalize *learning*. The analysis we perform will depend on how it is operationalized. Using T_1 and T_2 to denote *Round 1* and *Round 2 achievement* respectively we have three possibilities for representing *learning*:

1. T_2
2. $T_2 - T_1$
3. T_1 and T_2 change

In the cases of options #1 and #2 we would have the same three statistic possibilities that were discussed for *Round 1 achievement*, namely, χ^2, ANOVA, and the same four pre-planned comparisons. In option #3 we are treating T_1 and T_2 as two different times at which the achievement of subjects is tested. In this case a repeated measures design would be required.

Effect of Mouse Control

To test the effect of mouse control we need to look at the dependent variables *mouse transfers* and *percentage of mouse time* and test for the following correlations:

1. mouse transfers vs. Round 1 achievement
2. mouse transfers vs. learning
3. percentage of mouse time vs. learning

Here we will only be looking at the data for those subjects who played in a dyad in Round 1 because the variables *mouse transfers* and *percentage of mouse time* are only valid for Round 1 dyad play.

We must be careful when doing these correlations because some of the data involved represent dyad scores and others represent individual scores. The first correlation is straightforward since both *mouse transfers* and Round 1 *achievement* are dyad scores. In the case of the second and third correlation, however, we are correlating dyad scores with individual scores. Further, *learning* is an individual score and so a dependency issue arises.

Statistically speaking, observations must be independent. It is not possible, therefore, to include in the analysis two individual *learning* scores if they are dependent on each other in some way. Because the two *learning* scores for each dyad are dependent on their play together in Round 1, they cannot both be used. To circumvent this dependency problem, we must randomly select one score from each of the dyad *learning* scores and only use these randomly selected scores in the analysis. Dependency is an issue for the third correlation as well. Here again we will randomly select one subject from each dyad and then only use the *percentage mouse time* and *learning* scores for these selected subjects.

We hope to be able to determine what effect, if any, having dominant control of the mouse in Round 1 has on learning. Similarly we want to know if the number of mouse transfers in Round 1 has an effect on learning.

Effect of Attitude

A factor analysis will be performed on the questionnaire. This analysis is used to determine whether the questions included reflect one or many distinct aspects of one's attitude towards collaboration. (e.g., some of the twelve items included in the questionnaire address preference for playing alone vs. playing with others, while other questions reflect preference for being watched while playing electronic games). These questions can then be used to understand attitude.

7.3.2 What Affects Learning: Possible Scenarios

Why might there be differences? One hypothesis is that one child of a dyad might dominate the play, and consequently perform better in the individual round of play. Alternatively, the subject who dominates the mouse usage may spend more time manipulating the mouse, rather than thinking, and consequently perform comparatively poorly in the individual play. A low number of mouse transfers, and an unbalanced distribution of mouse time may be indicative of this case. "High" and "low" numbers will be determined once sufficient data has been acquired.

A low number of mouse control transfers would occur if both subjects are content for one subject to control the mouse, or if one subject dominated the play completely.

Observation will be needed to distinguish between these cases. We expect that learning will differ in these cases, even though software logging output will be similar. In the first case, both may learn to play the game equally well. It is not clear whether one needs to control the mouse to learn how the interface works. In the case where one subject dominates, it is possible that only that subject learns.

If two children spend most of their time fighting over the mouse, they might both perform poorly on the individual play. Excessive mouse control transfers might be indicative of this case.

If the subjects take turns controlling the mouse in alternate games, we would expect a fairly equal distribution of mouse time, and a low number of transfers.

7.3.3 Possible Questions to Examine

The "Give and Take" study by Inkpen et al. (1995) emphasized the differences due to gender, but did not attempt to provide any underlying explanation of the interactions taking place. Here, we will attempt to understand, at a finer level, the actions involved. These are some of the possible questions to address:

Mouse Control:

- Does the number of mouse transfers, by itself, affect learning?
- Does controlling the mouse facilitate learning?

Gender Differences:

- Are there significant differences in gender among the five experimental conditions for learning, achievement, and collaboration?
- Are the number of mouse control transfers related to gender?
- Is the amount of time spent controlling the mouse related to gender?

Attitude/Enjoyment:

- Does the experimental treatment to which a subject belongs affect their attitude towards the game?

- Does the number of mouse transfers affect a subject’s attitude or enjoyment?
- Do subjects enjoy playing more when they control the mouse?

8. Appendix 1: Power Analysis: Determining an Appropriate Sample Size

Whenever one conducts an experiment, it is important to be concerned with statistical ***power***, or the ability to reject the null hypothesis. Stated another way, if you are looking at different experimental groups, power is your ability to say that they differ (on some dependent variable) when in fact they do. Statistical power is based on three factors (Stevens, 1992):

- the α level set by the experimenter (typically 0.01 or 0.05)--this is the level at which you accept or reject the null hypothesis. A significance level of 0.05 was used in the Inkpen et al. studies.
- sample size
- effect size--how much of a difference the treatments make (the extent to which the groups differ on the dependent variable).

To better understand the notion of power, we consider an example: If you have two groups where the true differences are very large, even small sample sizes should be able to detect them. For instance, imagine two groups of people: college professors and elementary school students where height differences are being measured. On the other hand, imagine two similar groups (grade 11 students and grade 12 students). Here, you would need a larger sample size to detect height differences. One important consideration, then, in sample size determinations is that you want a good chance of detecting differences among your experimental groups. The power statistic (denoted by β) provides a means to determine an adequate sample size given the α level chosen, and the effect size. The power statistic represents the probability that you will detect a difference when it does in fact exist. Tables typically provide sample sizes for powers of 0.70, 0.80, and 0.90.

Coven (1977) and many others have noted that the small and medium effect sizes are very common in social science research. We believe that this may be a concern for our study and may be a key reason why Inkpen et al. failed to obtain more significant results. When sample size is a real constraint, Stevens offers these recommendations:

- Adopt a more lenient α level--instead of using 0.05 as the basis to find significance, 0.10 or even 0.15 might be used. Of course, this may lead you to find statistical significance where there is none (a Type I error).
- Consider ways of reducing *within-group* variability, so that you have a more sensitive design. By reducing within-group variability, you are increasing your effect size since effect size is given by **$(\mu_1 - \mu_2) / SD$**, where $\mu_1 - \mu_2$ is the difference between the two groups and SD is the standard deviation (variability) of the groups. On this count, we recommend that as homogeneous a group of children as possible be chosen for this study. The choice of a middle class public school is likely a better choice than Science World (where children may come from a variety of backgrounds). In addition, if low effect size is still problematic, we could consider controlling our population even more stringently. For example, we could control for intelligence based on some aptitude test which the children have taken--selecting only children in the middle academic range. By doing this, however, we would threaten external validity, or the ability to generalize our results to the larger population of children.

9. Appendix 2: E-GEMS QUESTIONNAIRE Page 1

First Name ________________________________

Last Name ________________________________

Grade _______ **Age** ____

Birthday __________________ **Girl / Boy**

Circle all the things you have at home:

Computer	Nintendo	Sega Genesis	Sega CD
Super Nintendo	Sega Saturn	Game Boy	Game Gear
Virtual Boy	Other: ____________________________		

How often do you play video or computer games:

never	a few times a year	a few times a month
a few times a week	almost every day	every single day

How do you like to play computer and video games:

by myself with my friends it doesn't matter

Who would you rather play computer or video games with:

girls boys both girls & boys it doesn't matter

I only like to play alone

Have you ever played "The Incredible Machine", "The Even More Incredible Machine" or "The Even More Incredible Machine 2?"

never a little lots of times

E-GEMS QUESTIONNAIRE Page 2

Note: Electronic Games means both video and computer games

1.	I like playing electronic games with my friends	NO	NO	MAYBE	YES	YES
2.	I like two player games better than one player games	NO	NO	MAYBE	YES	YES
3.	Having people around while I play electronic games makes me nervous	NO	NO	MAYBE	YES	YES
4.	I like talking while I play electronic games	NO	NO	MAYBE	YES	YES
5.	Electronic games are easier to play with a friend	NO	NO	MAYBE	YES	YES
6.	I would rather play electronic games by myself	NO	NO	MAYBE	YES	YES
7.	I don't like having to share my electronic games with my friends	NO	NO	MAYBE	YES	YES
8.	Electronic games are boring when you play by yourself	NO	NO	MAYBE	YES	YES
9.	I like having a friend around when I play electronic games to help me get through the hard parts	NO	NO	MAYBE	YES	YES
10.	I think you can learn more in an electronic game when you play by yourself	NO	NO	MAYBE	YES	YES
11.	Two player games are no fun because you have to sit and wait for your turn	NO	NO	MAYBE	YES	YES
12.	I like people watching me play electronic games	NO	NO	MAYBE	YES	YES

10. Appendix 3: E-GEMS 2nd QUESTIONNAIRE Page 1

First Name ________________________________

Last Name ________________________________

Grade _______ **Age** ____

Birthday ___________________ **Girl / Boy**

How often do you play video or computer games:

never a few times a year a few times a month

a few times a week almost every day every single day

How do you like to play computer and video games:

by myself with my friends it doesn't matter

Who would you rather play computer or video games with:

girls boys both girls & boys it doesn't matter

I only like to play alone

E-GEMS 2nd QUESTIONNAIRE Page 2

Note: Electronic Games means both video and computer games

1.	I like playing electronic games with my friends	NO	NO	MAYBE	YES	YES
2.	I like two player games better than one player games	NO	NO	MAYBE	YES	YES
3.	Having people around while I play electronic games makes me nervous	NO	NO	MAYBE	YES	YES
4.	I like talking while I play electronic games	NO	NO	MAYBE	YES	YES
5.	Electronic games are easier to play with a friend	NO	NO	MAYBE	YES	YES
6.	I would rather play electronic games by myself	NO	NO	MAYBE	YES	YES
7.	I don't like having to share my electronic games with my friends	NO	NO	MAYBE	YES	YES
8.	Electronic games are boring when you play by yourself	NO	NO	MAYBE	YES	YES
9.	I like having a friend around when I play electronic games to help me get through the hard parts	NO	NO	MAYBE	YES	YES
10.	I think you can learn more in an electronic game when you play by yourself	NO	NO	MAYBE	YES	YES
11.	Two player games are no fun because you have to sit and wait for your turn	NO	NO	MAYBE	YES	YES
12.	I like people watching me play electronic games	NO	NO	MAYBE	YES	YES

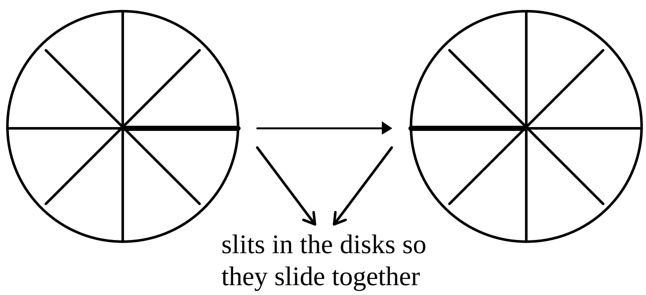
11. Appendix 4: Visual/Pinwheel Questions

At the end of the second round of play the subjects are each asked the following questions:

1. Would you rather play computer of video games alone or with a friend?
2. Who would you rather play computer or video games with, boys or girls?
3. Would you like to play this game again, yes or no?

For each question the subject is able to quantify their response through the use of a multi-coloured pinwheel type of device. The pinwheel is made up of two differently coloured circular pieces of cardboard, in our case blue and red. By twisting the wheel in one direction, more of the blue disk becomes visible until only blue can be seen. By twisting in the opposite direction, the red portion gets larger. The two different colours represent the two different answers for a particular question. So the subject is told to visually display their answer by twisting the pinwheel. There is a different pinwheel for each of the three questions.

Take, for example, the first question. The red disk represents the response “alone” and the blue, “with a friend”. (The word “friend” is written on the blue disk and “alone” on the red disk.) If the wheel is twisted such that the resulting disk is 3/4s blue and 1/4 red then we know that 75% of time the subject likes to play with a friend but does also, 25% of the time, like to play alone.



12. Appendix 5: Direct Observation Coding Sheets (Pairs Play)

Date:____________________. Time: ________________.
Observer: ________________.
- Subject 1: _____________. # - Subject 2: ___________.
give / take / shared / concur

start video	
start logging	
give manual	

1. mouse exchanges:

REGULARITY OF EXCHANGES	✓
subjects establish a pattern for exchanges	
no pattern to exchanges	

NATURE of MOUSE CONTROL	S1	S2
controls mouse while other subject directs		
directs while other subject controls		
dominates the play		
ignores the other subject's suggestions		
fights for control		
willingly gives control away		
willingly takes control		
indifferent to control		

VERBALIZATION	S1	S2
subject with control requests control switch		
subject without control requests a control switch		
denies request for control		

2. Difficulties with interface:

DIFFICULTIES	S1	S2
moving objects		
flipping objects		
resizing objects		
using elastic bands		
using ropes		

3. General attitude:

ATTITUDE	S1	S2
frustrated		
bored		
happy		
indifferent		

4. task oriented:

	S1	S2
subject is task oriented		

5. Communication between subjects:

COMMUNICATION	S1	S2
task related		
not task related		
explanatory		
exploratory		
minimal communication		
medium communication		
constant communication		
friendly		
agitated		
hostile		
frustrated		

6. Other observed behaviours:

OTHER BEHAVIOURS	✓
pointing to the screen	
two subjects controlling mouse at same time	

7. Number of puzzles solved: 1 2 3 4 5 6 7 8 9 10

8. Puzzle times:

PUZZLE	1	2	3	4	5	6	7	8	9	10
start time										
end time										

9. Average time per puzzle: __________.

10. By the end of the round:

	mastered interface	has minor difficulties with interface	has considerable difficulty with interface
Subject 1			
Subject 2			

	clear plan of attack to solve puzzle	basic plan but still relies heavily on trial and error	random trial and error
Subject 1			
Subject 2			

11. Did subject(s) leave before end of the round? yes / no

12. Should video be reviewed for this round? yes / no

13. Appendix 6: Direct Observation - Coding Sheet (Solo Play)

Date: ___________ Time: ___________ Name - Observer: ____________________.
- Subject: ________.

Round: 1 / 2

1. Subject is task oriented: yes / no

2. Difficulties with interface:

DIFFICULTIES	✓
moving objects	
flipping objects	
resizing objects	
using elastic bands	
using ropes	

3. Attitude:

ATTITUDE	✓
frustrated	
bored	
happy	
indifferent	

4. Number of puzzles solved: 1 2 3 4 5 6 7 8 9 10

5. Puzzle times:

PUZZLE	1	2	3	4	5	6	7	8	9	10
start time										
end time										

6. Average time per puzzle: __________.

7. By the end of the round:

	mastered interface	has minor difficulties with interface	has considerable difficulty with interface
Subject			

	clear plan of attack to solve puzzle	basic plan but still relies heavily on trial and error	random trial and error
Subject			

7. Did subject leave before end of round? yes / no

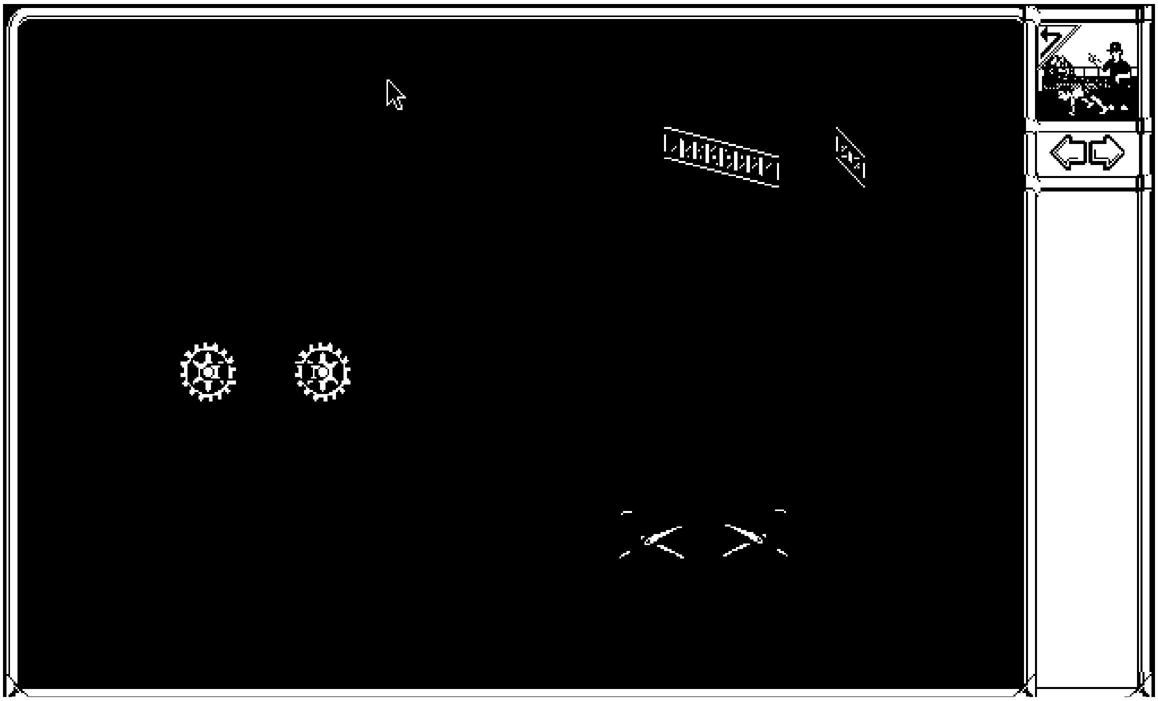
8. Should video be reviewed for this round? yes / no

14. Appendix 7: Sample Session Log

```
In session!
*** New Session Started ***
Right Mouse: Give selected (57 s)
Left Mouse: Give selected (84 s)
Right Mouse: Give selected (85 s)
Left Mouse: Give selected (86 s)
Right Mouse: Give selected (139 s)
Left Mouse: Give selected (207 s)
Right Mouse: Give selected (224 s)
Left Mouse: Give selected (225 s)
Right Mouse: Give selected (237 s)
Left Mouse: Give selected (239 s)
Right Mouse: Give selected (252 s)
Left Mouse: Give selected (298 s)
Right Mouse: Give selected (337 s)
Left Mouse: Give selected (372 s)
Right Mouse: Give selected (430 s)
Left Mouse: Give selected (493 s)
Right Mouse: Give selected (495 s)
Left Mouse: Give selected (497 s)
Right Mouse: Give selected (507 s)
Left Mouse: Give selected (575 s)

*** Game 1 took 9m 59s finished at 9m 59s
Right Mouse: Give selected (622 s)
Left Mouse: Give selected (736 s)
Right Mouse: Give selected (763 s)
Left Mouse: Give selected (799 s)
Right Mouse: Give selected (829 s)
Left Mouse: Give selected (831 s)
Right Mouse: Give selected (832 s)
Left Mouse: Give selected (833 s)

*** Game 2 took 5m 4s finished at 15m 3s
Right Mouse: Give selected (980 s)

*** Game 3 took 1m 21s finished at 16m 24s
Left Mouse: Give selected (1092 s)
Right Mouse: Give selected (1134 s)

*** Game 4 took 3m 4s finished at 19m 28s
Left Mouse: Give selected (1207 s)
Right Mouse: Give selected (1210 s)
Left Mouse: Give selected (1271 s)
Right Mouse: Give selected (1346 s)
Left Mouse: Give selected (1444 s)
Right Mouse: Give selected (1461 s)
```

```
Left Mouse: Give selected (1515 s)
Right Mouse: Give selected (1560 s)


STATISTICS
----------------------------------------------------------
Total session duration:                  28m 50s
Number of puzzles solved:                4
Average time per puzzle:                 7m 12s
Number of mouse exchanges:               39
Time left player had control:            17m 14s
Time right player had control:           11m 36s
Average time per control period:         0m 43s
----------------------------------------------------------
```

15. Appendix 8: Interface Training



16. References

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press.

Ericsson, K.A. and Simon, H.A. (1993). *Protocol Analysis: Verbal Reports as Data.* Cambridge, MA: The MIT Press.

Hymel S., Zinck, B., and Ditner, E. (1993). Cooperation versus competition in the classroom. *Exceptionality Education Canada 3,* (1 & 2), 103-128.

Inkpen, K., Booth, K.S., and Klawe, M. (1994). Cooperative learning in the classroom: The importance of a collaborative environment for computer-based education. Technical Report 94-5, Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, 1994.

Inkpen, K., Gribble, S.D., Booth, K.S., and Klawe, M. (1995). Give and take: Children collaborating on one computer. In *CHI '95: Human Factors in Computing Systems, Denver, CO, May 1995. Conference Companion.*

Inkpen, K., Booth, K.S., and Klawe, M. (1996). Interaction styles for educational computer environments: A comparison of drag-and-drop vs. point-and-click. Working Draft.

Johnson, D.W., Maruyama, G., Johnson, R., Nelson, D., and Sken, L. (1981). Effects of cooperative, competitive & individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin 89,* 47-62.

Lim, K.H., Ward, L.M., and Benbasat, I. (1994). An empirical study of mental models and computer system learning. Working Draft.

Lockheed, M., and Hall, K. (1976). Conceptualizing sex as a status characteristic: applications to leadership training strategies. *Journal of Social Issues 32* (3), 111-124.

Mayer, R.E. (1992). *Thinking, Problem Solving, Cognition.* New York: W.H. Freeman and Company.

McGrenere, J., Nakatsu, R., Neumann, M., and Summers, V. (1995) *Input Device Protocols and Collaborative Learning*. Course report.

Stevens, J. (1992). *Applied Multivariate Statistics for the Social Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.