Adaptive and Interactive Methods for Gathering User Preferences in Educational Games

by

James Gauthier

B.A., The University of Regina, 1994
B.Sc., The University of Regina, 2000

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Department of Computer Science)

We accept this thesis as conforming
to the required standard

. . . . . .

. . . . . .

THE UNIVERSITY OF BRITISH COLUMBIA

August 23, 2003

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature) _____

Department of Computer Science

The University Of British Columbia
Vancouver, Canada

Date _____

Abstract

Web-based learning environments often use games and simulations to enrich the learning process. Understanding the response of learners to these non-text-based environments is usually accomplished through laboratory testing or field surveys. One common method of gathering user feedback is the "click-to-submit" web-form questionnaire. However, these questionnaires are plagued by low response rates and inconsistent results. Research into survey design reveals many potential problems in the construction and administration of questionnaires — problems that are exacerbated by the nature of communication over the web.

With its interactive NFBkids website, the National Film Board of Canada uses web-form questionnaires to gather user feedback. Unfortunately, the web-based questionnaires at NFBkids fail to obtain useful feedback from those who visit the website. Moreover, the managers of NFBkids feel the website fails to create a sense of "community" among the site's users — one of the design goals of NFBkids.

By applying existing research into cognitive theory and psycholinguistics to the design of interactive methods of gathering user feedback, this paper investigates an alternative methodology for the collection of user preferences. In addition to the familiar "clickable" web-form, the research described in this paper explores interactive tools that complement web-forms as a method for gathering feedback from users. Real-time, peer-to-peer assessment techniques are introduced as a way to collect valuable user feedback and to foster a sense of peer presence among a community of web users.

Contents

List of Tables

List of Figures

Acknowledgements

Support for this research was given in part by the Natural Sciences and Engineering Research Council of Canada.[1] Special thanks go to Ines Hardtke, Susan Nosov and Svend-Erik Eriksen at the National Film Board of Canada, and to Val Day and Grant Miller at the Vancouver School Board.

Also, I owe many thanks to Maria Klawe and Brian Fisher. Their guidance, patience and insight enabled me to complete this thesis. I remember many great meetings with Brian and Maria — jovial meetings that always seemed to enlighten and inspire. Thank you.

[1] NSERC provided support through a Discovery Grant to Brian Fisher — under the title "Design of Information Systems for Human Cognition and Communication".

Dedication

This thesis is dedicated to the Electronic Games for Education in Math and Science (E-GEMS) laboratory at the University of British Columbia.

E-GEMS is an interdisciplinary team that researches children's interactions with computers. Beginning in the early 1990s, E-GEMS research has investigated curriculum, behaviour and psychological issues involved in the design and use of educational computer games. Recently, several E-GEMS projects have focused on the use of multi-player games to encourage collaborative learning.

The E-GEMS group has a proud history of excellent work. One of E-GEMS's strengths has been its ability to bring together computer scientists, game developers, artists, writers, teachers, education researchers and students.

Changes in the UBC Department of Computer Science along with the departure of E-GEMS's founder mean the immediate future of E-GEMS is uncertain. It is my sincere hope that E-GEMS may continue to play an important role within the Department of Computer Science — so the long-term future of E-GEMS may be as bright as its past.

The voice of the intellect is a soft one, but it does not rest until it has gained a hearing. Ultimately, after endless rebuffs, it succeeds. This is one of the few points in which one may be optimistic about the future of humanity.

— Sigmund Freud, *The Future of an Illusion*

Chapter 1

Introduction

> *Video games, being the first example of computer technology applied to toy making, have nonetheless been the entryway for children into the world of computers. These toys, by empowering children to test out ideas about working within prefixed rules and structures in a way few other toys are capable of doing, have proved capable of teaching students about the possibilities and drawbacks of a newly presented system in ways many adults should envy.*
>
> *Video games teach children what computers are beginning to teach adults — that some forms of learning are fast-paced, immensely compelling and rewarding.*
>
> — Seymour Papert, *The Children's Machine.* [16]

This project grew out of discussions with two groups within the National Film Board of Canada — the Kids Destination Project and the Animation, Children, Interactive (ACI) Project. In early 2001, the National Film Board (NFB) launched the NFBkids website as a "creative space" where children might explore interactive and informative activities ranging from word puzzles and science games to historical video clips and "do-it-yourself" documentaries [14].[1] In order for the NFBkids website to be successful, the managers of Kids Destination and ACI wanted a user-centred, interactive website. Specifically, they felt that feedback from those who visit NFBkids should be incorporated into the design of future versions of the site. The goal was to begin a dialogue with the children who use the website, so that NFBkids might "evolve and grow" based on users' comments and suggestions [13].

With the National Film Board's plan to have feedback from current users affect the NFBkids experience of future users, one of the big challenges for the

[1] The NFBkids website is available in English at *http://www.nfbkids.ca/* and in French at *http://www.onfjeunesse.ca/*

website was to solicit and gather user feedback. Accurate and reliable responses from those who visit NFBkids would be crucial in order to foster a meaningful "dialogue" between NFBkids and its users. Initially, user feedback at NFBkids consisted of web-log data and the results from online questionnaires (Figure 1.1). In an interview with the managers of Kids Destination and ACI in November 2001, it became obvious that these methods had not proven successful in obtaining useful information that could be passed on to the site designers [13]. Furthermore, Kids Destination and ACI felt that NFBkids had failed to create a sense of "community" for the children who visit the site.

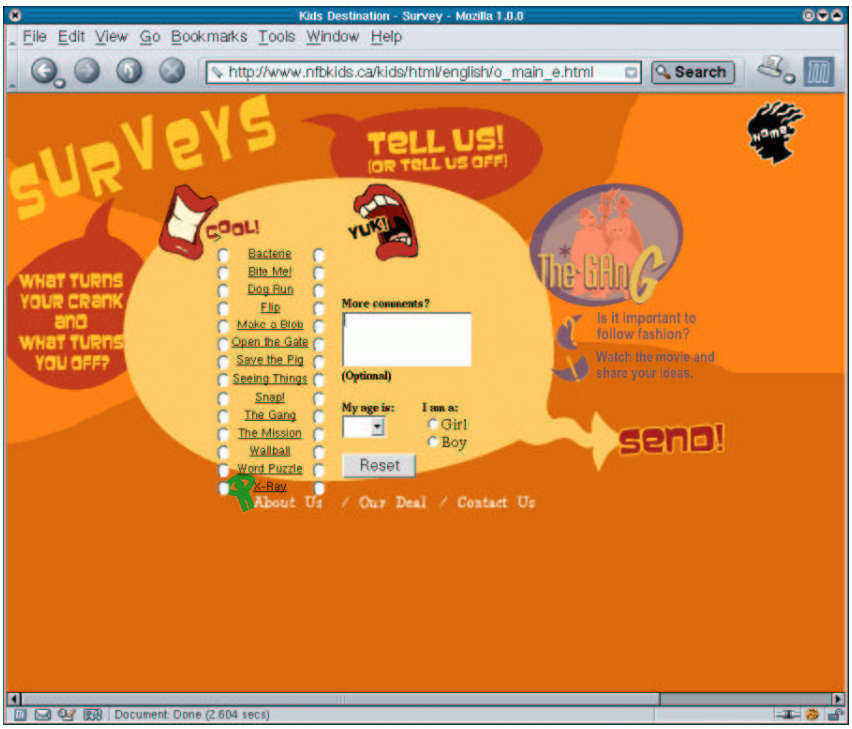


Figure 1.1: The click-to-submit Surveys page at the NFBkids site.

After further consultation with the National Film Board, the research described in this paper began with two goals: (1) to examine methods that would promote a sense of "community presence" among users of NFBkids and (2) to implement a system that would succeed in gathering valuable user feedback. A preliminary review of cognitive theory and psycholinguistics research in early 2002 provided some clues to solving the problems at the NFBkids website and hinted at alterna-

tive methods for gathering user preferences — methods that also might be able to foster a sense of community among web users.

The psycholinguistics research that seemed most relevant in analyzing the shortcomings of the NFB's user-designer "dialogue" at NFBkids focuses on speakers' intentions in language use and how this may be applied to survey design. As Herbert Clark and Michael Schober point out, surveys and questionnaires often fail to accurately convey the intentions of the questionnaire designer and those of the respondent [3]. Clark and Schober outline several intentions-related problems that include (1) recognizing common ground between the author of the questionnaire and the respondent, (2) perspective, (3) common purpose, (4) grounding, (5) pressure to respond and (6) consistency. These problems appear central to an understanding of "what went wrong" with the web-form questionnaires at the NFBkids website.

Closely related to the intentions-related pitfalls presented by Clark and Schober is the way in which a computer-assisted data collection system affects what is sometimes called the respondents' conceptual model of the survey. Researchers like Donald Norman argue that a well-designed system must make design intentions clear and consistent so that users form the correct mental model of the system [11]. Such design considerations represent another potential source of concern with the original NFBkids surveys.

In order to apply existing research in survey design to the problems of obtaining user preferences and promoting "community presence" at the NFBkids website, the next step was to examine the particular strengths and limitations of a site like NFBkids. Most researchers believe the World Wide Web to be an interactive system that allows for efficient communication, however, some authors (for example, Lee Sproull and Sara Kiesler) add that electronic forms of communication are hampered because they lack various contextual clues associated with most forms of human communication [22].

Accordingly, research into psycholinguistics, user-centred design and electronic communication formed the basis of the inquiry into the questions and concerns associated with the NFBkids web-form questionnaire. Furthermore, this background research provided a theoretical framework for the project and led to the conclusion that a "solution" to the problems with the NFBkids survey might take the form of an alternative methodology for gathering user preferences. Instead of working to "improve" the NFBkids web-form, it was felt that an alternative approach (*e.g.,* one that moved beyond the familiar click-to-submit questionnaire) might be able to deal with some of the effects outlined by Clark and Schober and also succeed in promoting a sense of community presence.

Figure 1.2: A snapshot of the Popcorn Meter.

Therefore, the project became an exploration of alternative methods for soliciting preferences from users. The design that emerged aims to attenuate the problems known to affect online surveys and to simultaneously promote a sense of presence in an interactive, virtual "space" populated by web users. The result is a system that incorporates real-time, peer-to-peer assessment into an on-line game-playing environment so that users no longer see "rating a game" as separate from the game activities.

Along with the exploration of real-time, peer-to-peer assessment as a method of gathering user preferences, came the hypothesis that as feedback is integrated with gameplay, users may become more inclined to provide comments and ratings. Consequently, if it is possible to shift game-assessment so that comments and ratings are intended for users instead of *"for the designers"*, this may both contribute to a sense of community and improve the overall quality of feedback. One of the main components of the new system is the Popcorn Meter (Figure 1.2 and Figure 1.3), a peer-to-peer assessment tool that allows users to send game-ratings to each other in real-time. It is introduced as a first step in testing this hypothesis.

The Popcorn Meter is one component of a system designed to provide a context for users in which rating games may be collaborative and entertaining. The background research provides the framework for a design that is peer-centred, aims to encourage interaction and collaboration among users, attempts to promote a sense of peer-presence and also acts as the catalyst for a dialogue on the game activities.

The remaining chapters in this thesis describe the research undertaken in an attempt to explore real-time, peer-to-peer assessment as a viable method of gathering user preferences. Chapter 2 details the NFBkids website and the goals behind NFBkids, along with the concerns and frustrations of the NFBkids managers.

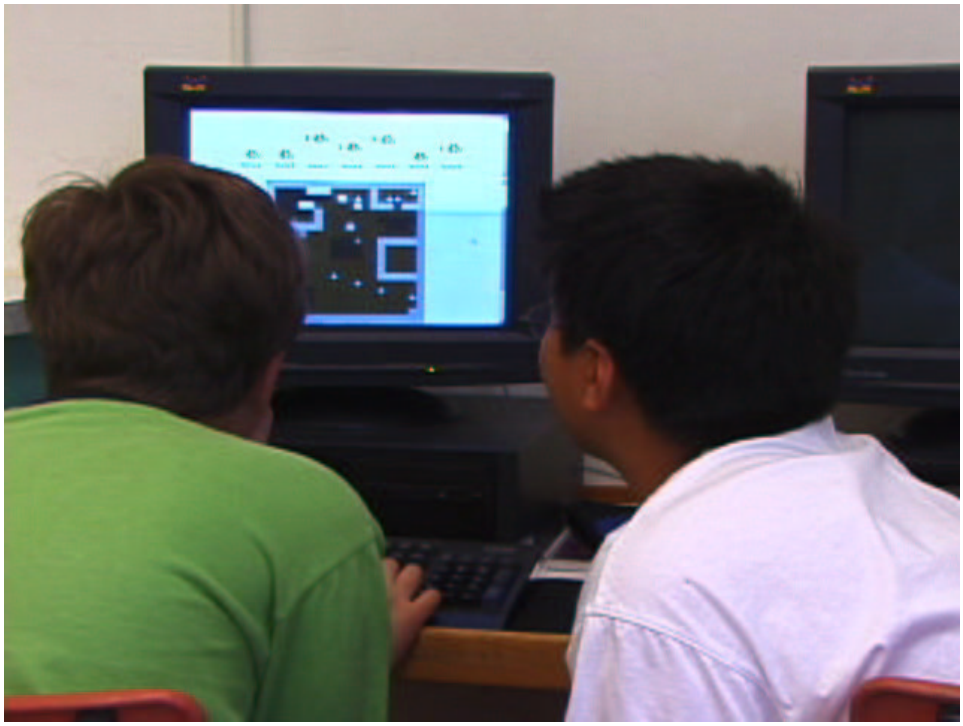


Figure 1.3: Grade Six students playing with the Popcorn Meter-enabled system.

Chapter 3 summarizes background research from a variety of disciplines that impacts on the analysis of NFBkids and the search for alternative methods of soliciting user feedback on the World Wide Web. Chapter 4 outlines the design of a prototype that incorporates real-time, peer-to-peer assessment as a way of gathering user preferences and promoting a sense of community among users. Chapter 5 describes two user studies conducted in order to introduce the prototype to Grade Six children. Chapter 6 discusses the results of the user studies, and Chapter 7 concludes the thesis with an agenda for research, along with ideas for future work into real-time, peer-to-peer assessment.

Chapter 2

The NFBkids Website

> *The NFB brand is well known around the world and is a continuing source of pride in Canada. For the past few years this brand has been too solidly anchored to its past, and not purposeful enough for its future. We are renewing the NFB again today through our recommitment to art for a social purpose.*
>
> *The same NFB going forward, but exploring new media, creating a new film vocabulary, working with new diverse communities all in a two-way dialogue for the public good.*
>
> — Jacques Bensimon, Government Film Commissioner and Chairperson of the NFB, November 9, 2002. [1]

2.1 The Goals of NFBkids

The National Film Board created the NFBkids website as a place where children might "see things in a whole new way, get creative, play games and test their skills" [15]. In several press releases from 2001, the Film Board explained how "creativity" is a key component of the NFB's vision for its children's site.

> This year the NFB will also be offering a greatly enhanced Internet experience on its website, *http://www.nfb.ca/*. The redesigned NFB site will feature a new destination for kids, providing a safe place on the web where young people will be able to explore and develop their creativity, as well as interactive, informative and visually rich resources on auteur animation and point-of-view documentaries [14].

Along with the "creative" nature of the NFBkids website was the National Film Board's desire to allow users the ability to shape the future direction of

NFBkids. Accordingly, the "two-way dialogue" to which Jacques Bensimon refers was also a design goal for the NFBkids website. The NFB hoped to enter into a dialogue with the children who used the site, so that the website might evolve and grow based on the input of those who use it. According to the Kids Destination Project manager in 2001, the first phase of NFBkids was designed "for the kids" (by the web design team at the Film Board), but future versions of the site would be developed, in part, "by the kids" who visit the website — based on comments and suggestions supplied by the site's users [12].

2.2 NFBkids Version 1.0

Although the National Film Board had big plans for NFBkids, even as late as Fall 2002 the site was still "under construction". However, even though the NFBkids site has always been a work-in-progress, the activities at NFBkids have been interesting, varied and, for the most part, fun and entertaining [20]. In November 2001, when the idea for this thesis was first discussed at meetings with representatives of the Film Board, the NFBkids site consisted of three main sections, divided by age group. There was *Hide and Seek*, an area for children ages 3-8, *Fun and Games* for 9-12 year-olds and *Critical Thinking* for ages 15-18.

All three sections of NFBkids were "under construction" in 2001, but *Fun and Games* was by far the most "complete" section at the website. As the intended age group for *Fun and Games* matched well with the ages of Vancouver School Board children involved with E-GEMS, it was decided that this thesis would concentrate on the *Fun and Games* section and attempt to deal with the shortcomings associated with *Fun and Games*, as outlined by the managers of Kids Destination and ACI. The plan was that as the research for the thesis progressed, prototypes and other versions of *Fun and Games* could be demoed to Vancouver school children through E-GEMS, shown to kids involved with the NFB's *Reel To Real* program or tested at *http://www.nfbkids.ca/* by linking new activities to the site.

Fun and Games

Inside *Fun and Games* were fourteen game activities, a collection of Quicktime and Shockwave movies (known as *Pix 'n Flix*), and a feedback section called *SURVEYS*. The "dialogue" between the site designers and the NFBkids audience consisted of web logs and the submissions from the *SURVEYS* section. Most of

Figure 2.1: Main Page of *Fun and Games* (June 2002).

the games had a didactic quality, ranging from activities associated with general science and arithmetic to vocabulary and spelling.

The *Fun and Games* home page (Figure 2.1) notified users of the still-under-construction state of the website and also prompted users to visit the *SURVEYS* section in order to register their comments regarding NFBkids.

> *We're still building the site. What works for YOU? Look around and then go to "Surveys" and tell us what you think!*

Rather than soliciting feedback on a finished product, the designers of NFBkids were experimenting with several different activities and concepts — then looking to users for opinions and suggestions. The Film Board, in an effort to start the aforementioned dialogue between itself and its audience, was using the *SURVEYS* page (Figure 2.2) as the first step in building this dialogue. It was hoped that user

feedback obtained through web logs and *SURVEYS* would help shape the design and implementation of NFBkids version 1.1.

The web-form at the *SURVEYS* page allowed players to rate games as *COOL!* or *YUK!*, provided them with a text box for "More comments" and asked them for their age and gender. An animated GIF "mouth" was associated with each of the *COOL!* and *YUK!* columns. Whenever an activity was marked *COOL!*, the mouth above the *COOL!* column would smile; a click on the *YUK!* column would cause the *YUK!* mouth to stick out its tongue.

The central column that listed the 14 games inside *Fun and Games* was also a set of links to the web pages for each activity. Along with a *SEND!* button (not visible in Figure 2.2; see Figure 1.1), the page also contained links to the information pages *About Us*, *Our Deal* and *Contact Us*.

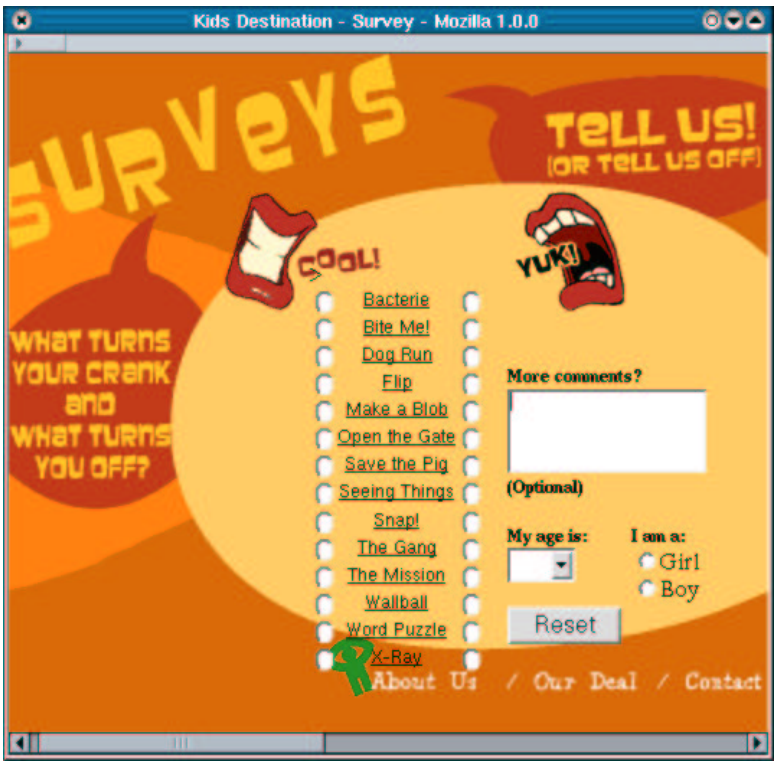


Figure 2.2: A close-up view of the *SURVEYS* page.

The game activities inside *Fun and Games* were "powered" mostly by Java, Macromedia Flash and Shockwave. These games had been put together by the

web designers at the Film Board, known for skilled and creative work. Some of the activities were simple diversions that might hold a player's interest for two or three minutes; others were addictive adventure-type games that could take much, much longer to play. One of the most involved games at NFBkids was a Shockwave-based adventure called *La Mission* — initially only available in French, but an English version was added in 2002. *La Mission* tested (and taught) general science knowledge by sending players on animated *missions* that included building a fire, "catching" lightning with a kite and surfing the periodic table in order to collect elements and build molecules. Figure 2.3 shows a screenshot from *Mission 321...Fire!*, a game that required players to start a fire by combining the correct combination and placement of fuel (twigs, sticks and logs), oxygen (provided by the elephant's breath) and spark (the proto-human banging together two rocks).

Figure 2.3: *Mission 321...Fire!* An activity inside the *Fun and Games* section of NFBkids.

Failure to build the fire in under two minutes would result in the "freezing" of the unfortunate proto-human (note thermometer on the top-right part of the screen). Prior to this mission, players received a lesson in combustion that ex-

plained the role of fuel, oxygen and heat source. After *Mission 321...Fire!* a quiz tested their knowledge of these concepts.

The 14 game activities that made up *Fun and Games* were in various states of development — some were fully functioning games, others were nearly complete but missing levels, a few were not ready at all and were simply listed as "Coming Soon…". Here is an overview of the games listed on the *SURVEYS* page:

Bacterie

This game never became part of NFBkids. Players who clicked on the *Bacterie* link at the *SURVEYS* page would be taken to a brightly coloured "Coming Soon…" page.

Bite Me!

A less-than-inspiring Javascript game where the mouse controlled a mouth that would "bite" on every mouse click. The idea was to bite a roaming hot dog and hamburger as many times as possible in one minute. A "stomach" on the left side of the screen kept score by recording the number of successful "bites" (and also by becoming enlarged). All in all, *Bite Me!* was a clickfest of a game and it lacked the engaging graphics and sound of most other NFBkids activities. Still, watching the stomach "fill up" was entertaining and would have appealed to some players.

Dog Run

The *Dog Run* set of activities included several Flash movies and "Music Videos" along with a "Dog Quiz". For example, the "Dirt on Dogs" movie ran about two minutes in length and provided a fun and wacky look at 12 problems people might face with their dogs. The movie benefited from simple and effective animation, a lively soundtrack with good music and excellent sound effects. Children "of all ages" would have liked this.

The three music videos inside *Dog Run* were also done in Macromedia Flash. They were 2-4 minutes in length and all had a hip-hop feel to them. "Unleashed" (Figure 2.4) was one of these Flash videos — it had smooth animation, fun music and many pop-culture references (some of which might have been missed by *Dog Run*'s target audience). The two other videos were entitled "Fetch Me" and "Wag Your Tail Tonight". These music videos also had intriguing characters, vocals, music and animation.

The Dog Quiz (Figure 2.5) was an interactive Flash presentation. The quiz had nice background music, kooky voices for the "Quiz Masters" and some interesting facts and trivia on dogs. This could have been a dull, lifeless activity, but clever design and creative use of Flash animation and sound effects made this a fun game. As with the other *Dog Run* presentations, the NFB team demonstrated a talent for putting together entertaining activities like the Dog Quiz.

Figure 2.4: Screenshot of the "Unleashed" video from the *Dog Run* activities.

Flip

A 40-second Quicktime movie named *Flip* showed a cartoon of a backyard barbecue chef cooking – and then flipping – burger patties on to hamburger buns. As the BBQ chef bit into the first burger, he realized something was wrong. In an odd twist, it turned out that a small alien spaceship was hidden inside the burger. A couple of strange and comical scenes followed before the Quicktime movie

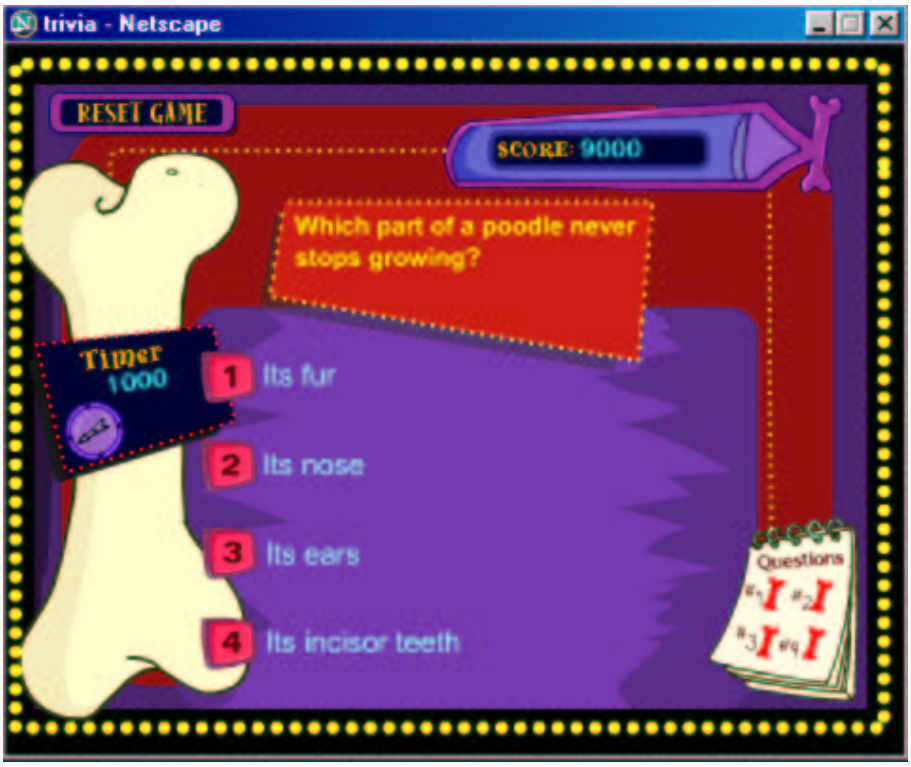


Figure 2.5: A page from the Flash-based game “Dog Quiz”.

seamlessly looped back to the beginning of the clip. Total playing time was approximately 50 seconds per loop.

Make A Blob

A simple, build a *Blob* activity using GIFs and Javascript. At this page, kids could have “creative fun” assembling a bizarre-looking creature by selecting and combining a *Blob’s* constituent parts: a head, a body and various accessories.

Open The Gate

Open The Gate was a Java-based game that tested a player’s skill with the mouse and his or her knowledge of simple arithmetic. In order to open the gate of a stone castle, the player had to position the mouse (very carefully!) on a catapult and

launch stones at banners hanging over the castle walls. There were three sets of banners: five numbers on the left, four arithmetic operators in the centre and five numbers on the right (Figure 2.6).

Figure 2.6: *Open The Gate*.

The player was presented with a number that would "open the gate". The trick was to produce this number using a combination of banners — one number from the banners on the left, one operator among the banners in the middle, and one number from the banners on the right. Hitting the appropriate banners with stones from the catapult would produce a simple equation. By selecting banners (using accurate shots from the catapult) the player could put together an equation that would give the desired result: the number that would open the gate.

At times, the interface for *Open The Gate* was incredibly frustrating. Hitting banners was very difficult — especially for novice players. However, this seemed to add to its addictiveness. A satisfying "boing" reinforced the good feeling of

successfully hitting a banner and a simple tune congratulated a player upon opening the gate.

Save The Pig

In *Save The Pig* (Figure 2.7) players saw a Shockwave-powered game with smooth animation and good sound effects that placed a pig at the top of the screen — falling, ever-so-slowly, towards the bottom of the screen (into water, ice or the Earth's atmosphere, depending on the level).

Figure 2.7: Screenshot of *Save The Pig*.

It took the pig about 30 seconds to fall and the only way to "save" the pig was to type in 7-10 words (depending on the level of difficulty) taken from the letters of a keyword shown in the top-left corner of the Shockwave canvas. In Figure 2.7 the player has successfully entered seven words using combinations of letters

from the keyword “spring”. This has caused a log to appear at the bottom of the screen — fortunately, the pig is in for a soft landing.

Seeing Things

The *Seeing Things* page included a few activities based on extreme-close-up photography (see Figures 2.8 and 2.9). One of the games inside *Seeing Things* was called *Eye Spy*, a Flash-based presentation with a good use of sound, animated text and a few visual effects. A game of *Eye Spy* began with these lines...

- Look closely.
- What do you see?
- If you see it, you will know the answers!

Eye Spy would then show an extreme-close-up photograph of an object. It would ask a series of yes/no questions — some of these were leading questions and others were “misleading”. After 6-12 questions, the object would be revealed and *Eye Spy* would provide a more conventional photograph of the “mystery” object.

Snap!

Another game that never made it beyond the “under construction” stage. The *Snap!* page simply read “Coming Soon...”

The Gang

The Gang opened with a Flash movie called “The New Shoes”. It portrayed a group of adolescent-esque characters making fun of one character’s heavy work-boots. The mocking tone of the group changes after they see a teen idol wearing the same boots. Following the movie was a page entitled “It’s Important to Follow Fashion” where two written articles from “experts” (one a fashion designer, the other a principal at a high school where uniforms are mandatory) outlined issues related to peer-pressure and fashion.

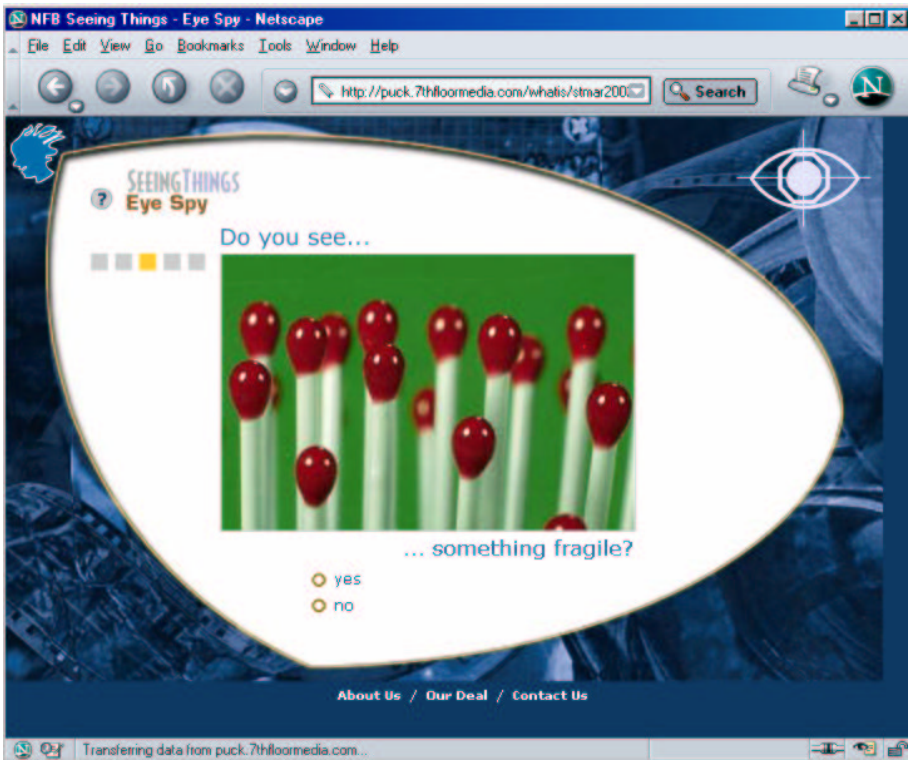


Figure 2.8: Screenshot of a *Seeing Things* activity.

Wallball

This was a Javascript game similar to *Pong* where a single player could score points by positioning a paddle in order to "bounce" a dot into certain regions of the screen. *Wallball* was much less sophisticated than most of the other activities at *Fun and Games*. Some players might have found the repetition of bouncing the ball monotonous (rarely would the slow-moving ball get past the player's paddle), others may have enjoyed running up the score.

Word Puzzle

Word Puzzle was a Javascript game that placed several "jumbled" letters on the screen. Up to three "hints" would be given so that the player might rearrange the letters to form the "hidden phrase". A surprisingly entertaining game.

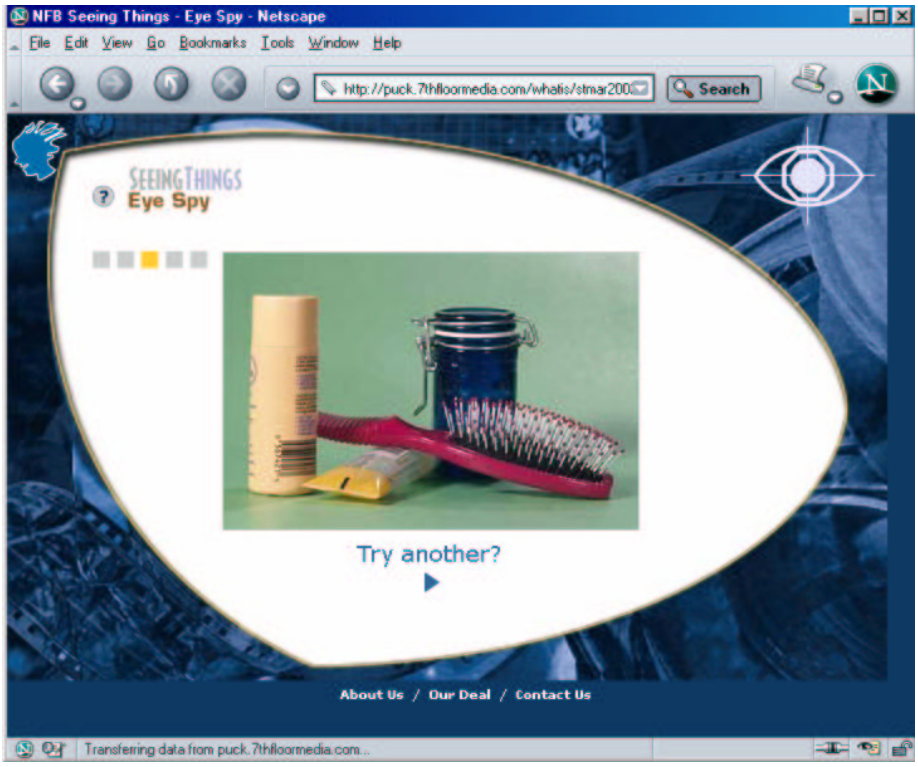


Figure 2.9: A mystery object is revealed at *Seeing Things*.

X-Ray

Also implemented in Javascript, the *X-Ray* activity displayed a "family photo" of a normal-looking cartoon family. The mouse acted as a rectangular-shaped *X-Ray* device. The player could use the *X-Ray* box (by "hovering" the enlarged mouse cursor over different sections of the family photo) to provide a penetrating look at the family. The *X-Ray* revealed the true nature of these characters: hideous, ghastly creatures.

2.3 Initial Survey Results

With NFBkids 1.0, the Film Board had succeeded in putting together a diverse collection of activities that seemed likely to appeal to a wide audience of young web users. As the NFB had hoped, at *http://www.nfbkids.ca/* children would find

a “safe place on the web” with many fun and creative activities. Furthermore, the Film Board could be proud that its children’s site was meeting certain aspects of the NFB mandate as many of the NFBkids activities were fundamentally different from the games and activities seen at most commercial web sites designed for kids.

However, despite the apparent success of the web-design team, after a few months on the web, Kids Destination and ACI were unhappy with the overall performance of the NFBkids site. Their disappointment with NFBkids centred on the feedback users were providing through the *SURVEYS* page. Intended as the springboard to a dialogue with the site’s users, the web-forms were failing to bring in useful feedback from those who visited NFBkids. By November 2001, fewer than half of those who had visited the site had submitted any information at all and those who did send comments rarely provided the site’s administrators with what might be considered “interesting and useful comments” [13]. Also, very few NFBkids users were answering the age and gender questions — fewer than 20 percent [13].

According to the Kids Destination project manager, the vast majority of submissions were short “not-so-meaningful” comments, exemplified by submissions like these:

- “*this sucks*”
- “*way too slow*”
- “*ABSOLUTELY GREAT - THANKS*”
- “*how does it work?*”
- “*SSSLLLLLOOOOOOWWWW*”
- “*What?!?*”
- “*stoopid*”
- “*Fun, I guess.*”
- “*i like it*”
- “*boring!*”
- “*Good.*”

- "*Very fun. Thanks.*"

At a meeting in November 2001, after reviewing several pages of a printout of comments gathered at the *SURVEYS* page, the manager of ACI admitted, "these were not the sort of comments we were hoping for" [13].

Even the simple rating scheme (*COOL!* versus *YUK!*) was not collecting useful feedback for the web-design team. Web-logs showed that many users would rate games they did not play and often they would not provide rating for games they did play. All of this feedback left the administrators of the site with their own "comments":

- What part of it sucks?
- What does "too slow" mean? Does a particular game take too long to download? Does it not run at the appropriate speed on this user's machine? Is the game action not fast enough?
- Exactly what about the activities do they like? What about the site do they not like? What do they mean by "stoopid", "fun" or "GREAT"?
- Have these users put any thought into their ratings or their comments?
- Why are they rating games they have not played? Why are they not rating games they have played?
- Who are these users?

Although dismayed by the quantity and quality of the feedback from the *SURVEYS* page, the managers of ACI and Kids Destination still believed in the central idea of moving NFBkids from a "for the kids" website to one that might be considered "by the kids". In November 2001, it was decided that the NFB strategy would not be abandoned — NFBkids would continue to give its audience an "experimental" site with a wide variety of activities, and then look to the audience to shape the direction of the site by allowing users to rate the games and submit comments. However, everyone involved could see that NFBkids needed something better than the *SURVEYS* page.

Not knowing exactly how to proceed, it was proposed that a research project and/or Master's thesis would examine the "feedback problem" at NFBkids and outline possible solutions. These solutions could then be explored at the NFBkids

site or by showing prototypes of new *SURVEYS* to children involved with the NFB's *Reel To Real* program. Also, it was proposed that Vancouver school children might play a role in the development of a new system, as prototypes could be shown to Grade Six students through E-GEMS.

A Classic HCI Problem

The National Film Board had designed and built NFBkids as a website that could make the transition from a "for the kids" site to a "by the kids" site. However, by the end of November 2001, the managers of ACI and Kids Destination had determined that NFBkids was not performing as they had hoped. The feedback collected at the *SURVEYS* page did not appear to provide any direction for the future development of NFBkids. In fact, the user data collected through the *SURVEYS* compromised the user-centred design intentions of the NFBkids team as "bad" user data breaks the process of user-centred design. Furthermore, those responsible for the site felt that one of the initial design goals had been forgotten — NFBkids was to be an interactive, "community" site where users might feel "connected" to each other [12].

Faced with a "classic HCI problem" at the NFBkids website (where the system had been designed to meet certain goals, but was returning disappointing results), the next step was to research the process of soliciting information on the World Wide Web. It seemed as though the web designers at the Film Board had done nice work as far as the games and activities at NFBkids were concerned, but that they had paid little attention to the task of gathering user feedback. Even at first glance, there appeared to be several problems with the *SURVEYS* page:

- There was no measure of the degree to which a user liked or disliked a particular activity — was it really *COOL* or was it very *YUK*?
- The sterile text box seemed to provide little incentive for players to enter comments. Also, the radio buttons for *COOL / YUK* and *GIRL / BOY* along with the drop-down menu for *MY AGE IS* seemed unlikely to inspire players to provide feedback. In fact, it was possible the entire *SURVEYS* page was perceived as more of nuisance to users than anything else.
- Unless the respondent provided such information in the text box, there was no way of knowing why some activities were *COOL* and others were *YUK*. For example, was a player making a general comment on the entire site with a page of *YUK*s? (A situation that occurred surprisingly frequently)

- If players were indifferent to an activity, what was expected of them? Would they leave it unchecked or would it receive an arbitrary rating of *COOL* or *YUK*? Could this be a possible explanation for the high number of played-but-unrated games?

- Did the terms *COOL* and *YUK* prejudice the results? Also, did the animated images above the *COOL* and *YUK* columns bias or alter the results? Some users might have been interested in seeing the *COOL* mouth smile — or in watching the *YUK* mouth stick out its tongue. Was this responsible for the high number of rated-but-not-played games?

These questions and concerns with the *SURVEYS* page, led those involved with the project to conclude that NFBkids deserved, at the very least, a "better" online questionnaire. Research into questionnaire design (described in more detail in the following chapter) would reveal many additional shortcomings with the *SURVEYS* page.

Moreover, the questionnaire design research would suggest that it might be possible to move beyond simple online questionnaires. A fast, interactive system like the World Wide Web could take advantage of other methods of gathering user feedback. Therefore, in examining the "feedback problem" at NFBkids, the research project would not only look at a "better" questionnaire for the website, but also explore an alternative methodology for gathering user feedback on the WWW.

In preparing the design for a new method of soliciting user preferences, the emphasis would be on creativity, interactivity and collaboration — precisely the goals the Film Board had in mind when designing its children's site. It was hoped that this novel approach to gathering user feedback could also address the NFB's concerns that NFBkids was not fostering a sense of community among its users. Accordingly, the new project would attempt solve both of the HCI-related problems at NFBkids: it would attempt to promote a sense of "community presence" and successfully gather user preferences — allowing NFBkids the chance to become more of a "by the kids" website.

Chapter 3

Background Research

> *Survey questionnaires are useful measurement instruments to the extent that questions convey to the respondent the desired intent of the researcher. For much of survey research this had been assumed to be an unproblematic part of the measurement process. Yet throughout the history of survey research, there has been a stream of methodological research, now grown large with the application of concepts from cognitive psychology to surveys, which has challenged that assumption.*
>
> — Robert Groves, Nancy Fultz and Elizabeth Martin, "Direct Questioning About Comprehension in a Survey Setting". [6]

As seen in the previous chapter, the managers of Kids Destination and ACI were not getting the results they wanted from the NFBkids website. However, despite a few obvious problems with the click-to-submit *SURVEYS* page, it was not entirely clear why the web-forms and web-logs were performing so badly. As a method of gathering user feedback the web-forms were a conventional approach to user-centred design, but could the HCI shortcomings of the web-forms be "responsible" for the terrible performance of the *SURVEYS* page? January 2002 began with a survey of cognitive psychology and questionnaire design research, in an attempt to find out where in the process of gathering feedback, the *SURVEYS* were "going wrong".

Initially, the research questions dealt with determining the validity of the NFBkids questionnaire and then searching for empirical ways to test and improve the *SURVEYS* page. Concerns included "what makes a good questionnaire?" and "how might traditional questionnaire research be applied to web-based questionnaire design?" However, by the end of January 2002, the focus of the research had shifted from self-administered, web-based questionnaires, to an exploration of alternative methods of gathering user feedback on the web. Existing research in

psycholinguistics and cognitive psychology hinted that a "solution" to the NFB's HCI problem might not necessarily lie in an improved web-form, but in exploring different methods of obtaining user preferences on the World Wide Web.

3.1 The Process of Asking and Answering Questions

The process of soliciting information through a survey or questionnaire might seem straightforward. Participants receive a series of questions from an interviewer or from a written questionnaire and the participants' answers to these questions are recorded. If the authors of the survey feel they have "properly designed" the questionnaire, the information gathered is assumed to reflect the opinions and concerns of the participants. However, in the words of Herbert Clark and Michael Schober,

> How a question is worded makes a difference, but so do many other factors — how the question is introduced, what questions come before and after, what answers are allowed and much, much more. The factors are so diverse that they may seem impossible to account for. Even wording is mystifying. [3]

A 1974 study by Seymour Sudman and Norman Bradburn examined these "many other factors" that may influence participants' responses to survey questions [23]. Using the term *response effect* to refer to the amount of error in the response to a question associated with a particular factor, Sudman and Bradburn outlined forty-six independent variables as potential sources of response effects. Ellen Wentland and Kent Smith discuss these findings:

> These [forty-six independent variables] included variables associated with the interviewer role, such as the interviewer's social background characteristics; variables associated with the respondent's role, including characteristics and respondent role motivation; and variables associated with the task, such as the perceived social desirability of the answer. [...] Sudman and Bradburn concluded that task variables seemed to be the most important factors related to response error overall. [28]

The web-form at NFBkids was not a traditional paper and pencil questionnaire, however, it would seem that *respondent role motivation* is an important consideration for a web-form like the NFBkids *SURVEYS* page. In a sense, as far as users are concerned, when it comes to providing comments and ratings through a click-to-submit form, "what's in it for them?"

Furthermore, other questions relating to interviewer and respondent *roles* seem instructive in analyzing the "feedback problem" at NFBkids. Although the slogan *TELL US OR TELL US OFF* was created to offer players the incentive to send feedback, could it have had the opposite effect by setting up specific "roles" for the respondent and the designer? For example, were those who visited NFBkids reluctant to share their comments and preferences with the "site designers"? Most users probably visited NFBkids in order to experiment and play with the games. If they saw their "role" as game players, the *TUOTUO* page might have appeared as an oddity — the website was a place to have fun, yet filling out the *SURVEYS* page may have been seen as distinctly not-fun. *TELL US OR TELL US OFF* may not have prepared users for this change of "role", nor have conveyed to them the importance of this new role.

In terms of the *task variables* at the *SURVEYS* page, users might have been frustrated with the yet-another-web-form format. Not only did the "task" at *TUOTUO* differ from those found at the other NFBkids pages (provide feedback versus play games) the *TELL US OR TELL US OFF* page was relatively bland, flat and "boring" — in stark contrast to the game activities at NFBkids that benefited from the use sound and animation. An honest reaction to the *SURVEYS* page might have been "get me back to the action — the challenging and rewarding games".

Also, although the one use of animation at the *SURVEYS* page (the animated GIF mouths) was intended to make rating the games more "fun", this animation might have corrupted the rating process. As already mentioned, users might have clicked on the *COOL* and *YUK* columns simply to see one mouth smile or another stick out its tongue. These animated GIFs could have enticed players to "rate" the games, but the *task* of clicking may have had little to do with a player's thoughts on a game; it simply may have reflected the player's desire to "play" with the mouths.

From the speculation and discussion above, research into traditional survey design appears to hold interesting lessons and observations for the analysis of "what is wrong" with the NFBkids *SURVEYS* page. Although the comments from Clark and Schober, along with those from Wentland and Smith, seem to indicate that a diverse and almost "impossible to account for" number of variables affect survey design, an examination of language and human communication reveals that

many of the pitfalls associated with survey design may be avoided.

The "answer"? A growing number of theorists claim that good survey design benefits from paying close attention to the linguistic and social rules associated with human communication. An analysis of language and communication can help to "refine" the questions and speculation above in order to build a more complete analysis of the NFBkids "feedback problem".

Questions, Language and Communication

Some researchers see the process of soliciting information as a "social encounter". Although a survey may take many forms (*e.g.*, written questionnaire, interview, click-to-submit web-form), it is still governed by the linguistic and social rules associated with "ordinary" human communication. A survey may be considered to be a very "special" type of conversation in that it is usually a voluntary social encounter between strangers, but regardless of its unique qualities, a survey remains a conversation — one bound by language and social convention. Theorists use this concept of survey-as-conversation to suggest that an understanding of the survey process should rely on an understanding of the rules that guide human conversations and social encounters in general. In the words of Sudman, Bradburn and Schwarz:

> The vehicle through which survey questionnaires are delivered is language. Understanding how people comprehend speech and written material deepens our understanding of the ways in which questionnaire design affects people's answers. [...] Trying to understand how respondents comprehend survey questions leads inevitably to a more basic search for information about how people understand the world around them and how they communicate with one another. [24]

Therefore, understanding how humans interact and communicate (with a particular emphasis on language) may provide insight into the cognitive processes that respondents go through when completing a questionnaire. Accordingly, an application of cognitive psychology to survey design will aid in understanding how survey design can impact respondents' answers.

In terms of the NFBkids project, cognitive psychology has a role to play by looking at survey participants and how they deal with questionnaires. Exploring the social and cognitive complexities of the response process may lead to a better understanding of what is "wrong" with the NFBkids site. It might be possible to

uncover the "feedback problem" by seeing key elements of the site (for example, the *SURVEYS*) from the point of view of the kids — or, more precisely, from the point of view of the kids' brains.

A self-administered, online questionnaire must (1) motivate the respondents to answer the questions, (2) make sure the respondents understand the questions, (3) provide respondents with an incentive to answer the questions truthfully and (4) accurately collect the data. Sudman, Bradburn and Schwarz would suggest than an examination of how such a questionnaire might succeed would start by looking at how the participants will process the various stimuli they will encounter at the website (*e.g.*, written material, visual and auditory feedback) and how they might respond to the tasks with which they are presented. This approach to survey design is not foolproof, nor does it guarantee a survey that performs as intended. However, it does offer a set of guidelines and it suggests how one might proceed in addressing a "feedback problem" like the one at NFBkids.

3.2 The Psychology of Survey Response

Many theories deal what is sometimes called the "response process". In their 2000 study of the psychology of survey response, Roger Tourangeau, Lance Rips and Kenneth Rasinski outline many approaches to survey response, including

- Psychometric theories
- Process Theory
- High Road/Low Road theories
- Satisficing models
- Two-Track Theories [26]

As mentioned in the previous section, in terms of the problems at the NFB-kids website, one approach that seems particularly promising is the application of cognitive processes to survey methodology — conducted by several researchers including Seymour Sudman, Norman Bradburn and Norbert Schwarz, as well as Herbert Clark and Michael Schober. These authors attempt to improve survey design by examining what is often termed "the problem of meaning".

Judith Tanur explains that "Meaning [reflects] the view that a respondent must share the meaning of a question intended by the survey researcher if he or she is to respond usefully" [25]. Accordingly, understanding the rules for "normal" or "ordinary" human communication may provide insight into the study (and possibly the reduction of) response effects in a survey interview or on a questionnaire. Sudman, Bradburn and Schwarz put it this way:

> Language comprehension is based not only on formal structures like syntax, but more importantly, on pragmatic factors that may deeply affect meaning. A study of the effects of question wording has to focus on the factors that influence respondents' comprehension and on how the inferred meaning may differ from the researcher's intended meaning. Variations in contexts may result in different respondents answering, in effect, different questions (in such cases, the answers cannot be compared) or even questions that the researcher never intended to ask. [...] Cognitive psychology has a rich research literature on memory and information processing, processes that are fundamental to understanding how respondents answer survey questions. [24]

Robert Peterson agrees that respondents must be able to interpret the meaning of the researchers and he issues similar warnings on the dangers of misinterpretation and miscommunication. Peterson uses a slightly different terminology by claiming that several "communication-related activities must be carried out" [17].

> A researcher must first encode the research question so that all study participants interpret it in the same way. Study participants must decode the research question in the way that the researcher intended and encode answers that contain the requested information. If study participants decode the research question differently, they are, in effect answering a different question. Last, the researcher must decode the answers provided by study participants in a way they intended the answers be decoded. [17]

Clark and Schober examine the problem of meaning more rigorously by applying the psycholinguistic concept of *intentions* to survey design. Their research into the cognitive processes that underlie the procedure of asking and answering questions reveals many pitfalls in questionnaire design and administration. They apply psycholinguistics, the study of the mental faculties involved in the perception, production and acquisition of language, to survey design in order to show that

survey responses are linked to the intentions of the interviewers and participants. Therefore, understanding "what goes wrong" in a survey or interview requires an investigation into the intentions of the speaker (or author) and how the participants interpret the speaker's intentions. According to Clark and Schober, "once we understand the role of speakers' intentions in language use, we will find many of the problems of survey design more tractable" [3].

Speaker's Meaning

> *The idea is that language use, whether it is in conversations, interviews, debates, or writing, is built on what people intend by what they say and do. An essential part of these intentions is captured in this principle:*
>
> ***Principle of speaker's meaning:***
> *Speakers and their addressees take it for granted that the addressees are to recognize what the speakers mean by what they say and do.*
>
> — Herbert Clark and Michael Schober [3]

In laying out their argument for a psycholinguistic analysis of survey design, Clark and Schober suggest that *meaning* is ambiguous when it comes to almost every aspect of human communication — words often have many alternative definitions; most sentences have several interpretations. They claim that human communication is possible, in large part, because meaning is *inferred* from experience and context. They use examples from Herbert Paul Grice and Charles Perfetti to illustrate this:

> When Ann utters "Sit Down" to Ben, she *means something.* She intends Ben to recognize that she wants him to sit down — and that she has this particular intention (Grice, 1957); that is, she is trying to make certain of her intentions public – open, accessible, shared, mutually known – between Ben and her [...]
>
> What counts, then, is not the meanings of words *per se,* but what speakers mean by using them. The point is so obvious that we rarely

give it a second thought. Take these actual newspaper headlines (Perfetti *et al.,* 1987)

Girl, 13, Turns in Parents for Marijuana, Cocaine

Toronto Law to Protect Squirrels Hit by Mayor

Deer Kill 130,000 [3]

Each of the headlines above has more than one interpretation, but Clark and Schober claim most readers infer the intended reading quickly, unconsciously and "without apparent effort". When the meaning of a sentence or passage is ambiguous, the mind of the reader uncovers the intentions of the author by relying on several cues, including common ground, perspective, common purpose and grounding.

Common Ground

According to Clark and Schober, mutual knowledge, mutual beliefs and mutual suppositions form the common ground between speaker and addressee. They note that common ground may not necessarily exist between the speaker and the addressee, but that at some point during the process of communication these actors "believe they share" at least some of this information in common [3]. This comes into play even before the speaker puts together the first gestures or utterances — as speakers choose their words and behaviour according to a principle of "utterance design".

> ***Principle of utterance design:***
> *Speakers try to design each utterance so that their addressees can figure out what they mean by considering the utterance against their current common ground.* [3]

Although there are many reasons to communicate (*e.g.,* flirtation, comraderie, experimentation), generally, people want to say things their addressees will understand. In order to do this, speakers "need to root what they say in information they believe they share with [their addressee(s)] — their common ground" [3].

Clark and Schober divide common ground into *cultural common ground* and *personal common ground*. Whereas cultural common ground is linked to the information that is common to cultural groups, personal common ground draws on "joint personal experiences as viewed against [the speakers' and the addresses'] cultural common ground" [3]. The following paragraph illustrates how people establish common ground in order to foster efficient, intimate, pleasurable conversation. (This quotation is from a "scholarly" work. Slight elitist overtones aside, it provides a good illustration of the importance of common ground in human communication.)

> When Veronica and John meet at a party, and as they establish the cultural groups they have in common, they can each assume an enormous body of cultural common ground. Once they realize they are both university graduates, for example, they can assume as common ground all those facts and beliefs they assume university-educated people take for granted. These range from theories of gravity, light, and biological systems to the basic facts of geography, history, and social organization. Or as two speakers of western American English, they can assume as common ground the phonology, syntax and vocabulary of that dialect of English. As two baseball fans, they can assume as common ground the names of the major players, their statistics, and such jargon as *rbi* and *era*. [3]

Cultural common ground may provide the context for the conversation described above, yet it is personal common ground that plays the major role in making *party-talk* (and, in fact, much of human communication) rewarding. If one assumes that human beings are social animals, communication has many purposes. Obviously, a main function of human communication is the sharing of ideas, but people also "get something" out of the act of conversation (and other forms of communication).

Generally speaking, human nature entices people to communicate. From an evolutionary perspective, although there are many factors to consider, human beings have a better chance of survival if they are able to converse and share ideas — even with strangers. A first encounter at a social gathering makes an interesting example as an exploration into what might be called the establishment of personal common ground.

> Personal common ground is established from joint perceptual and linguistic experiences interpreted against cultural common ground.

> But how does this work? In language use, common ground accumulates in a highly systematic way, as expressed in this principle:
>
> ***Principle of accumulation:***
> *In a conversation the participants add to their common ground each time they contribute to it successfully.* [3]

Clark and Schober continue with the idea that common ground is cumulative in human discourse by claiming that every time a participant contributes "successfully" to a conversation, she or he adds to the common ground between the participants. In terms of Veronica and John,

> When Veronica speaks, John interprets her utterance against their initial common ground, and then they both add the content of what she says to that common ground. Then when John speaks, Veronica interprets him against their updated common ground once more. And so it goes. [...]
>
> When Veronica says, "The guy next door just bought a motorcycle," she *presupposes* there is a man John can readily identify in their common ground as "a guy next door". She treats this as *given information*. What she and John are to add to their common ground is her belief that the person so identified just bought a motorcycle. She treats this as *new information*. But listeners often have to draw *bridging inferences* to get the speaker's presuppositions to fit into their current common ground. If John goes on, "And how bad is the noise?" he presupposes that there is a uniquely identifiable "noise" in common ground. Since there has been no explicit mention of any noise, Veronica has to draw the bridging inference that the motorcycle makes noise, and that is the noise John is referring to. [3]

The bridging inference in Clark and Schober's example (the motorcycle is responsible for the noise) is a simple inference. However, other bridging inferences may be much more sophisticated. According to Clark and Schober, inferences of this type are not the exception: "bridging inferences are ubiquitous in discourse".

Perspective

Within the boundaries of typical human communication, the *perspective* of the speaker can heavily influence the direction of the dialogue. With their discussion of common ground, Clark and Schober have shown that presuppositions and inferences are universal components of conversation. Therefore, there is a great deal of room for the speaker to influence the addressee's interpretation. Often the speaker may be conscious of the influence of his or her perspective and intend to convey facts in a "leading" manner. However, the effect of perspective on the addressee's interpretation may also be unintentional.

> Speakers ordinarily expect their addressees to accept their perspectives. When Veronica says "the guy next door" or "that awful pest you met," she takes it for granted that she and John will now view the man this way. That is the way presuppositions work. If John doesn't object, he implies that he accepts her perspective, at least for the moment. [...]
>
> Perspectives are easy to plant in the common ground of a discourse. There are many perspectives that speakers can take on a situation, and for addressees it often matters little which one is selected. It is also polite for addressees to accept speakers' perspectives. After all, speakers choose perspectives they judge their addressees will accept, so to object is to question their judgement. [3]

Of course, speakers often frame what they say in language that will persuade their audience to accept a particular perspective. This framing of facts may be subtle or obvious, depending on the skill and intent of the speaker (the attention and perspective of the audience is important, too). An accomplished orator is able to "take advantage" of the power of language and carry the audience through presuppositions and bridging inferences to an apparently obvious conclusion. However, in terms of survey design, if the goal is to solicit accurate feedback, the survey designers must be aware that the language and format of a survey can influence the perspective of the respondents.

Common Purpose

One aspect of dialogue that becomes crucial in the establishment of common ground is the *purpose* of the dialogue itself.

> Whenever we take part in a discourse, we do so purposefully. Some of our goals are private, even embarrassing if they were to come out. But others become public, a shared, mutually recognized part of the discourse. When you plan a party with a friend, your primary goal is to arrive at that plan, and it must be mutually recognized as such if you and your friend are to progress. [3]

Therefore, to a certain extent, conversation is a rational and co-operative effort. What emerges (again, to certain degree) is a common purpose or set of purposes. As the participants build the dialogue, they recognize this common purpose and guide the conversation in a "mutually accepted direction" [3]. Clark and Schober quote the sociolinguist Herbert Paul Grice[1] for a principle of discourse that captures the co-operative aspect of conversation:

> ***Co-operative Principle:***
> *Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.* [3]

Clark and Schober point out that the co-operative principle (if it is observed) means the mutually accepted purpose of an "exchange" is essential for determining the speaker's meaning. An example to illustrate the importance of this principle in terms of the speaker's intention...

> Suppose you run a restaurant, and one day a woman with a pleasant voice telephones, says "Hello," and asks one of three questions:
>
> 1. Do you accept American Express cards?
> 2. Do you accept credit cards?
> 3. Do you accept any kinds of credit cards?
>
> With each question she asks whether you accept something, but what that something is varies from "American Express cards" to "credit

[1] In their discussion of Grice's co-operative principle, Clark and Schober also draw on other work by Grice: his conception of conversational implicature and his 4 maxims of conversation (quantity, quality, relation and manner). Herbert Paul Grice's own description of these concepts may be found in [5].

> cards" to "any kinds of credit cards." As it happens, you can answer yes to each one, so if you take her literally, you should respond simply, "Yes, we do." But would you? [3]

Clark and Schober answer this question with "No, you wouldn't." To back up their claim, they discuss the responses elicited from 150 restaurants (50 per question) in San Francisco in 1979 (also, see Table 3.1):

> Restaurateurs interpreted the caller by inferring her purpose. They supposed that she wanted to patronize their restaurant and pay with a credit card, so she wanted to know if they accepted a card she owned. She signalled the cards she was interested in by her question. She specified an American Express card in question 1 but credit cards in general in questions 2 and 3. By mentioning "any kinds" of cards in question 3, she showed an interest in the *particular* cards they accepted, and most restaurateurs told her about them. So restaurateurs looked for the caller's mutually recognizable purpose, and their responses were attempts to satisfy it.
>
> This example brings out an important point: Questions can themselves serve many different purposes. We usually think of questions as *information questions*, as asking for information the questioner doesn't or couldn't possibly know. In our example, "Do you accept American Express cards?" is taken as an information question. But there are also *exam questions*, which ask for information the questioner already could or does know. [3]

Along with information questions and exam questions, Clark and Schober identify another set of questions they call *presequences* — questions that test a precondition for an action the questioner would like to take. "Do you accept any kinds of credit cards?" is a presequence element known as a *prequestion*[2] [3]. When a speaker asks a prequestion, he or she is not really interested in the direct answer to that question, but in the request it *prefigures*. Therefore, a prequestion is tool that allows the speaker to ask a question without really asking it. Such a question is effective in everyday conversation because respondents are familiar with them — *i.e.*, respondents are able to interpret the speaker's purpose with a prequestion based on the common ground and common purpose they already share with the speaker.

[2] Along with prequestions, Clark and Schober also describe preinvitations (*e.g.*, "What are you doing tonight?") and preannouncements (*e.g.*, "Do you know what just happened to me?")

	Question		
Response Examples	1	2	3
Yes, we do	100	44	10
Yes, we do. We accept American Express and Visa	0	38	56
We accept American Express and Visa	0	16	34
Other	0	2	0
Total	100	100	100

Table 3.1: Clark's 50-restaurant survey. San Francisco, 1979. Results shown are percentages of responses elicited from 50 restaurants per question. [3]

Grounding

Another important principle to come out of Clark and Schober's work is the principle of grounding. When two people build a conversation, they speak in order to express what they mean (in *most* circumstances). According to the principle of accumulation, speakers have to make sure their utterances become part of the mutual common ground they share with their respondents. This implies they must "ground" what they say:

> ***Principle of grounding:***
> *For each contribution to discourse, the participants try to reach the mutual belief that the addressees have understood what the speaker meant to a criterion sufficient for current purposes.*

> Grounding is ordinarily achieved through collaboration, through joint actions. When Veronica speaks, she looks for evidence from John that he has understood her. John, in turn, tries to provide that evidence by saying "uh huh," nodding his head, or taking the relevant next turn. If he hasn't understood her, he will ask her to repeat, confirm, or paraphrase what she said. [...]
>
> What Veronica and John accomplish in this process is a shared construal of what Veronica meant. [3]

The five principles described above form the basis of Clark and Schober's theory of *speaker's intentions* in language use. The *principle of speaker's meaning*

claims that people mean what they say and are able to recognize the meaning in what others say. According to the *principle of utterance design*, speakers prepare their words based on the common ground they currently share with their respondents — so that respondents may easily interpret the meaning of the speakers' words. As discourse proceeds, the participants accumulate common ground, according to the *principle of accumulation*. Participants collaborate in building a shared understanding of the speakers' intended meaning (*principle of grounding*) and both speakers and addressees have "mutually recognizable purposes" as outlined by the *co-operative principle* [3].

According to Clark and Schober, an understanding of these principles allows for a deeper analysis of language use — which, in turn, greatly benefits survey design. However, although an analysis of the principles of "social encounters" deepens an understanding of the dialogue between interviewer (or questionnaire) and respondent, these principles do not tell the whole story. There are other issues and concerns that may not play a major role in an analysis of everyday discourse, but are significant with respect to surveys. Some of these survey-specific concerns that seem particularly important for a web-form questionnaire like NFBkids include *Pressure to Respond* and *Pressure for Consistency*.

Pressure to Respond

In a conversation, certain rules govern the conduct of the speaker and the addressee. Such rules dictate, for example, that in most circumstances it is rude to interrupt the speaker. Another rule that guides conversation: when people ask a question, they expect an answer. The respondent may take a few seconds to respond, but a delay of more than several seconds is usually not tolerated. Once a question has been asked and the speaker has finished talking, it is the respondent's *turn* to speak — and there is pressure on the respondent to, at the very least, do *something*...

> Questions and answers are subject to the conventions governing turn taking in conversation. Veronica, in asking John a question, designates him as the next speaker. He in turn is obligated to begin his turn the moment she completes hers. Ordinarily, he will start in on his answer without a pause. But if he cannot, he must show Veronica that he realizes that it *is* his turn at that moment. He is under pressure to say something — for example, "um" — within about one second. [3]

On the telephone, turn taking is similar. Dead silence on the telephone for any significant period of time often makes people uncomfortable — regardless of whether the call is a chat between friends or a survey interview between strangers. Accordingly, survey interviews that are conducted face-to-face or by telephone share the pressures to respond of everyday conversation.

However, the pressures to respond are different when it comes to a written questionnaire. When answering questions on a self-administered questionnaire, any pressure to respond depends on the respondent's attitude toward the survey. If there are time pressures and the respondent feels his or her answers are important (for example, if the "questionnaire" is a written examination), it is likely the respondent will work efficiently to complete as many questions as possible. This may not be the case if there are no obvious "consequences" to the respondent's answers.

Pressure for Consistency

In everyday conversations, the participants are under pressure to be consistent with what they say. Clark and Schober's principles of utterance design and accumulation tend to indicate that whenever a participant speaks, his or her words are interpreted against what has already been said. There is a bit of room for misinterpretation and error, but for the most part, contradictory statements appear as either dishonest or confused. A survey faces similar pressure for consistency.

> Many surveys contain what are called *knowledge* filters, such as these two questions:
>
> 1. Do you have an opinion about gasoline taxes?
> 2. Have you thought enough about gasoline taxes to have an opinion?
>
> When respondents answer no to a knowledge filter, they should be less willing to offer an answer to the next substantive question, and they are. They are discouraged from doing so apparently because they think they need to know a lot in order to answer the questions. To be consistent, they should refuse to answer them. That, indeed, is the purpose of knowledge filters. [3]

Set 1	Set 2
Up to $\frac{1}{2}$ an hour	Up to $2\frac{1}{2}$ hours
$\frac{1}{2}$ to 1 hour	$2\frac{1}{2}$ hours to 3 hours
1 hour to $1\frac{1}{2}$ hours	3 hours to $3\frac{1}{2}$ hours
$1\frac{1}{2}$ hours to 2 hours	$3\frac{1}{2}$ hours to 4 hours
2 hours to $2\frac{1}{2}$ hours	4 hours to $4\frac{1}{2}$ hours
More than $2\frac{1}{2}$ hours	More than $4\frac{1}{2}$ hours

Table 3.2: Response alternatives for a German study on TV watching. [3]

Knowledge filters demonstrate that survey respondents try to be consistent in what they say — even on a written questionnaire. There is further evidence that pressures to be consistent are at work when people answer surveys, as respondents often use their existing answers as evidence when making judgements. In a 1985 study on television viewing habits, German citizens were asked "How many hours a day do you spend watching TV?"[3] Half of the participants were given the response alternatives in Set 1 and the other half received Set 2 (see Table 3.2).

Set 1 respondents estimated 2.7 hours of TV viewing per day with Set 2 respondents estimating 3.2 hours per day. These results suggest that issues related to perspective and common ground are influencing the respondents to answer the TV-viewing question in terms of what is considered typical or normal in the population — implied by the options provided in each set. However, an even more interesting result follows from a subsequent question in the survey...

> People whose earlier answer led them to believe they watched TV *more* than average judged that TV played a *more* important role in their lives. The difference was induced entirely by the response alternatives they were offered with the first question, set 1 or 2. [3]

[3]This study was conducted by Schwarz, Hippler, Deutch and Strack. It is referenced by [3]

Another example of respondent consistency may be found in a 1987 study dealing with Ohio residents' interest in government and public affairs.[4] The respondents were asked one of these questions:

> 1a. Do you happen to remember anything special that your United States representative has done for your district or for the people in your district while he has been in Congress? (Yes 12%; no 88%)
>
> 1b. Do you happen to know the name of the governor of Ohio? Yes 69%, no 31%)

Then they were asked, either immediately or later in the interview:

> 2. Would you say you follow what's going on in government and public affairs most of the time, some of the time, only now and then, or hardly at all?

> Respondents who said no to whether question 1a or 1b were, in effect, giving the interviewer evidence that they were not really following public affairs. Indeed, once they had said no to either question 1a or 1b, they were less likely to say they followed public affairs "most of the time" for question 2. But since question 1a prompted more no answers than question 1b, it should also keep people from saying "most of the time" more than question 1b would. It did, by a reliable margin of 32% to 26%. Respondents, trying to be consistent, brought their later answers into line with their earlier ones. [3]

A possible explanation for the influence of one survey question on another is *priming* — where the first question "primes" the respondent's memory in such a way that it affects the answer to the second question. An alternative explanation is that survey participants simply pursue consistency as they present themselves in an interview or on a written questionnaire. As they complete a survey, respondents may prefer to appear as though they have a consistent set of beliefs (even if they do not) [3].

[4] Study conducted by G.F. Bishop. Referenced by [3].

Cognitive Burden

Another important consideration in survey design is the so-called cognitive burden it places on participants. For example, Donald Norman argues for design principles that co-operate with users. Usually Norman's principles of design are associated with the design of objects, tools, and "everyday things", but they may be extended to survey design. According to Norman,

> The human mind is exquisitely tailored to make sense of the world. Give it the slightest clue and off it goes, providing explanation, rationalization, understanding. Consider the objects — books, radios, kitchen appliances, office machines, and light switches — that make up our everyday lives. Well-designed objects are easy to interpret and understand. They contain visible clues to their operation. Poorly designed objects can be difficult or frustrating to use. They provide no clues — or sometimes false clues. They trap the user and thwart the normal process of interpretation and understanding. [11]

Surveys fall into the category of Norman's "well-designed objects" that should be easy to interpret and understand. In terms of survey design (particularly the design of written surveys), Tourangeau, Rips and Rasinski summarize Norman's principles of design: in their words, a design should...

- take advantage of knowledge that is available externally, as well as knowledge stored in long-term memory;
- simplify the structure of the tasks;
- make both controls and the actions they perform visible to the user;
- rely on natural mappings between actions and their consequences;
- exploit both physical and cultural constraints;
- allow for errors; and
- rely on standardization when other design principles do not apply. [26]

Another set of design principles that aims to minimize the cognitive burden of a survey may be found in the work of Cleo Jenkins and Don Dillman. Jenkins and Dillman concentrate specifically on written questionnaires and base their findings on research into visual perception. Paying particular attention to patterns of eye-movements during reading, Jenkins and Dillman refer to the research of Patricia Wright and Phil Barnard[5] in arguing the importance of both verbal and non verbal language on a questionnaire:

> Wright and Barnard (1978) write that the problems of completing self-administered questionnaires fall into two classes: problems with the language used and problems arising from the way information is arranged spatially. This statement suggests that the spatial arrangement of information is not "language". However, it is more precise to label both as *graphic language* and to further subdivide them into "verbal" versus "non-verbal" language. [8]

Jenkins and Dillman proceed to develop five principles for the design of self-administered questionnaires that take into account both verbal and non-verbal language.

1. Consistently use graphical elements (*e.g.*, contrast and spacing) to define a clear path through the questionnaire.
2. Use prominent visual guides to redirect respondents when conventions within a questionnaire must change.
3. Place directions where they are easily seen and close to where they are needed.
4. Keep separate pieces of information physically close when they must be connected to be understood.
5. Ask only one question at a time. [26]

[5]The article by Wright and Barnard to which Jenkins and Dillman refer is "Asking Multiple Questions About Several Items: The Design of Matrix Structures on Application Forms", *Applied Ergonomics*, 9 (1978) pp. 7-14. Much of Wright and Barnard's research deals with reading skills. For example, reading for decision-making, reading for action and reading for reference.

The principles described in this section were written to address problems with "traditional" surveys — face-to-face interviews, interviews by telephone and written questionnaires. Although much of this research may be applied directly to web-forms and other computer-assisted questionnaires, there are certain specific concerns when a computer is involved in collecting the data.

3.3 When a Computer Asks the Questions...

Computer-assisted data collection brings a new set of problems, challenges and opportunities to survey design. For example, replacing the interviewer or written questionnaire with a computer seems likely to affect the "nature" of the survey. Many of the issues and concerns associated with the principles of survey design deal with the cognitive and social complexities of the response process — a process that is bound to change when a computer is inserted into the procedure of asking and answering questions.

Donald Norman's work provides a framework by which the role of the computer in a computer-assisted survey might be understood. He focuses on conceptual models as an important concept in design:

> *mental models* [are] the models people have of themselves, others, the environment, and the things with which they interact. People form mental models through experience, training, and instruction. The mental model of a device is formed largely by interpreting its perceived actions and its visible structure. [11]

Mental models are an important consideration in the design of any survey. According to Tourangeau, Rips and Rasinski, "evidence as is available suggests that survey respondents do not always have clear notions about what is expected of them" [26]. With a non-computer-assisted survey, the respondents base their conceptual model of the survey on their encounter with the interviewer or with the written questionnaire. It is best if the respondents' mental model of the survey matches closely with that of the survey's designers. In the words of Donald Norman,

> The *design model* is the designer's conceptual model. The *user's model* is the mental model developed through interaction with the system. The *system image* results from the physical structure that has

> been built (including documentation, instructions, and labels). The designer expects the user's model to be identical to the design model. But the designer doesn't talk directly with the user — all communication takes place through the system image. If the system image does not make the design model clear and consistent, then the user will end up with the wrong mental model. [11]

In the quotation above, Norman deals with the design of "everyday things". However, his comments have clear consequences for the design of surveys and the "system image" that results from the structure of the survey — regardless of whether the survey consists of an interview, a pencil-and-paper questionnaire or a click-to-submit web-form. In light of Norman's comments, the addition of a computer to the data collection process alters the way respondents "interact" with a survey. Ultimately, the machine affects respondents' conceptual model of the survey. Mick Couper has determined several characteristics he believes to be crucial for a computer-assisted data collection system, if it is to promote a mental model for users that matches the mental model of the designers [4]. Couper's thoughts on what he calls CAI (Computer Assisted Interviewing) are summarized by Tourangeau, Rips and Rasinski:

- Functionality (*i.e.*, the system should meet the requirements for carrying out the tasks);
- Consistency (the system's conventions and mappings between actions and consequences should be the same within a questionnaire and, if possible, across questionnaires and other interviewer tools);
- Informative feedback (the system should provide some feedback, such as a confirmation message of movement to the next screen, for every user action);
- Transparency (the system should carry out certain functions — such as checking that the answer entered corresponds to one of the options — without drawing the user's attention to them);
- Explicitness (the system should make it obvious what actions are possible and how they are to be performed);
- Comprehensibility (the system should avoid jargon, abbreviations, and arbitrary conventions);

- Tolerance (the system should allow for errors, incorporating facilities to prevent, detect and correct errors);
- Efficiency (the system should minimize user effort by, for example, simplifying the actions needed to carry out common operations);
- Supportiveness (the system should recognize the cognitive limits of the users and make it unnecessary for them to memorize large numbers of commands, providing ready access to online help instead);
- Optimal complexity (the system should avoid both oversimplification and extreme complexity). [26]

In addition to the design characteristics mentioned above, there are specific concerns with the mental model of a system when the target audience consists of children [7]. As the intended audience of the *Fun and Games* section of the NFBkids website is 9-12 year-olds, it is important to consider the mental models that are formed by children and how they might differ from those of adults.

> One difference between adults and children is the mental models that they form. Mental models are formed by users in order to help guide them through attempts to perform tasks and correct errors while using computer software. One particular design strategy is to encourage the correct mental model of the user. This can be accomplished by having the application use a metaphor for something that the user already knows. [10]

Lee Sproull and Sara Kiesler raise yet another concern for any system that relies on networked, electronic communication:

> Electronic interactions differ significantly from face-to-face exchanges. [...] In an electronic exchange, the social and contextual cues that usually regulate and influence group dynamics are missing or attenuated. [22]

In light of the research seen in the previous section, Sproull and Kiesler's warning seems particularly significant. Written questionnaires and interviews already face the problems of common ground, perspective, common purpose, pressure to respond, *etc*. If one accepts that *meaning* is ambiguous in everyday human conversation (where participants rely on several cues to accurately infer the

speaker's meaning), a method of communication as crude as a web-form seems extremely limited. In order to design a web-questionnaire that "works", the designers must be very careful and pay close attention to the concerns and warnings mentioned above. To simply throw together a web-survey and assume it will accurately collect opinions and responses from users is dangerous. In a sense, a poorly designed survey is worse than no survey at all: a bad survey brings in data the designers might assume to be valid — basing decisions on such data is risky, at best.

The observations, guidelines and theories mentioned in this chapter offer several warnings and suggestions for survey designers. There is a certain amount of overlap in the design principles mentioned above and, although principles like these are "vague enough that applying them is likely to remain more of an art than a science" [26], they do provide a framework for analyzing the NFBkids "feedback problem". Also, these principles and guidelines provide important clues in the search for solutions. Couper points out that web-based surveys are still poorly understood — partly because this method of conducting surveys is in its infancy:

> [...] while Web surveys in general may become increasingly easy to do (both cheaper and quicker), good Web surveys (as measured by accepted indicators of survey quality) may become increasingly hard to carry out. [...] we simply do not yet know how things will turn out for this exciting new method and for the industry as a whole. [4]

The strategy for this research project (*i.e.*, the Master's thesis) is to bring together theories from various areas of research in an attempt to better understand the shortcomings of the NFBkids site, outline possible improvements and develop a prototype. Couper's comments (among others) indicate that it may not be possible to "fix" the problems at NFBkids (*e.g.*, the response process is complex; Web-based surveys are new and exciting, but not well understood). However, the hope is that existing research will point toward alternatives to the NFBkids *SURVEYS* and that these alternatives will improve the feedback performance of NFBkids while promoting a sense of community for NFBkids users. The search for such alternatives leads to a survey of research into social relationships on the internet. If NFBkids is to foster a sense of "belonging" or entitlement among its users, future development of the site may benefit from an examination of the feelings of kinship and "connectedness" many users already find on the World Wide Web.

3.4 Social Relationships on the Internet

The tremendous growth of the internet has allowed for the widespread development of a different type of social interaction for millions of computer users — *online* social relationships. Those who are "connected" have the opportunity to communicate with an ever-increasing number of users from around the globe. This networked audience may choose from a wide variety of internet-based forms of communication, including: electronic forums, instant messaging, Multi-User Dungeons (MUDs) and their variants, electronic mail, Bulletin Board Systems (BBSs), virtual communities and file-sharing networks. People use these systems for many purposes — some of which are decidedly "personal".

> Although the first computer networks and conferencing systems supported communications between people who used them at work, the biggest growth now seems to be for people to use them outside of work — for finding friends and lovers, pursuing hobbies, scavenging for investment tips, political organizing, and so on. [9]

Many authors approach what might be called the *social aspects of computing* with distinct utopian and anti-utopian biases:

> [...] computing has become a kind of projective device for social scientists and other social critics. Often enough, each analyst foresees in the social world to come the consummation of whatever trends he or she finds most heartening or most deplorable in the world as it is. And computing is often pictured as the agency by which these fondest hopes or deepest fears are to be realized. [21]

However, even with this warning in mind, the literature on "social computing" holds important lessons when it comes to the deeply personal experience many authors associate with networked e-communication. For example, this passage from (determined utopian) Howard Rheingold:

> Since the summer of 1985, for an average of two hours a day, seven days a week, I've been plugging my personal computer into my telephone and making contact with the WELL (Whole Earth 'Lectronic Link) — a computer conferencing system that enables people around the world to carry on public conversations and exchange private electronic mail (e-mail). The idea of a community accessible

> only via my computer screen sounded cold to me at first, but I learned quickly that people can feel passionately about e-mail and computer conferences. I've become one of them. I care about these people I met through my computer, and I care deeply about the future of the medium that enables us to assemble.
>
> I'm not alone in this emotional attachment to an apparently bloodless technological ritual. Millions of people on every continent also participate in the computer-mediated social groups known as virtual communities, and this population is growing fast. Finding the WELL was like discovering a cozy little world that had been flourishing without me, hidden within the walls of my house; an entire cast of characters welcomed me to the troupe with great merriment as soon as I found the secret door. Like others who fell into the WELL, I soon discovered that I was audience, performer, and scriptwriter, along with my companions, in an ongoing improvisation. A full-scale subculture was growing on the other side of my telephone jack, and they invited me to help create something new. [19]

The sense of entitlement Rheingold feels for the WELL demonstrates that at least for a certain group of computer users, e-communication may foster a sense of community among those who participate. The interaction Rheingold describes as "emotional attachment" is valuable when considering a redesign of the NFBkids site. The managers of ACI and Kids Destination want a site that conveys a feeling of "connectedness" to users. This (by itself) is an important goal, however, a sense of entitlement among users offers further potential benefits — if users feel an attachment to the system (and to the community of users) they might also take an interest in the future of that system. Rheingold claims that he not only cares about the people he has met through the WELL, but he also cares "deeply about the future of the medium that enables us to assemble". In terms of solutions for NFBkids, *caring* about the future of the system seems likely to be linked to the quality of feedback users might provide. Users who "care" may be more inclined to enter comments and suggestions than users who do not.

One key to building a system that fosters a "caring" community of web users may be found in the amazing variety of video games and other entertainment software that relies on *interaction* among users to promote a sense of community [10]. Many of these games incorporate advanced computer graphics and sound to build a computer-generated avatar for each player — for example, networked versions of ubiquitous first-person-shooter video games. However, there is con-

siderable evidence that even systems "limited" to text-based communication can evoke powerful feelings of entitlement among users. According to Sherry Turkle, this is often the case with text-only chat rooms and MUDs:

> Some MUDs use screen graphics or icons to communicate place, character, and action. The MUDs I am writing about here do not. They rely entirely on plain text. All users are browsing and manipulating the same database. They can encounter other users or players as well as objects that have been built for the virtual environment. MUD players can also communicate with each other directly in real time, by typing messages that are seen by their players. [...]
>
> The term "virtual reality" is often used to denote metaphorical spaces that arise only through interaction with the computer, which people navigate by using special hardware — specially designed helmet, body units, goggles, and data gloves. [...] In MUDs, instead of using computer hardware to immerse themselves in a vivid world of sensation, uses immerse themselves in a world of words. MUDs are a text-based, social virtual reality. [27]

Therefore, the strength of the text-based MUDs lies not in advanced graphics or sound, but in the *interaction* and imagination of the participants.

The challenge for the redesign of NFBkids is to consider the "feedback problem" in light of the survey design theories and e-communication research mentioned in this chapter. The new feedback-gathering system will attempt to avoid the pitfalls of questionnaire design mentioned by those like Clark and Schober, observe the principles of design put forward by Norman and others, and deal with the shortcomings of communication on the WWW outlined by Sproull and Kiesler. Also, the system will try to take advantage of the speed and connectivity of the internet, and keep the best parts of the original NFBkids — a site that makes excellent use of colour, sound, animation and game design. The new design task is to focus on collaboration and interaction while paying attention to the obstacles of pressure to respond, consistency, functionality, *etc.* that may be hampering the NFBkids *SURVEYS* page.

Chapter 4

System Design: An Interactive, Peer-Centred Approach to Obtaining User Feedback

> *Because people have a strong positive bias towards social relationships and predictable environments, the more a media technology is consistent with social and physical rules, the more enjoyable the technology will be to use. Conforming to human expectations means that there is instant expertise just because we're human, and there are positive responses, including feelings of accomplishment, competence and empowerment.*
>
> — Byron Reeves and Clifford Nass, *The Media Equation.* [18]

At the first meeting with the National Film Board in November 2001, the general feeling with the NFBkids site was "what can be done to bring in *better* and *more accurate* feedback?" [13]. The consensus at that meeting: the web-form data from the *SURVEYS* were disappointing, yet it was unclear how the system might be "fixed".

The background research that now makes up Chapter 3 of this thesis shows the quest for *better* and *more accurate* feedback to be more than simply a shot in the dark, but a certain amount of speculation and conjecture was necessary in order to isolate potential problems with the *TUOTUO* page and to brainstorm suitable solutions. The early work consisted of putting together the "facts" on the NFBkids *SURVEYS*:

- The activities at NFBkids were interesting and fun, but the *SURVEYS* page was a boring instrument — especially in contrast to the game activities.

- The NFBkids designers had added an ornament to the *Tell Us Or Tell Us Off* page (the *COOL!* and *YUK!* mouths), but the gameplay characteristics of this ornament impeded or distracted from the purpose of the interface (at best). In fact, it was likely that *COOL!* and *YUK!* were adding artifacts to the web-form data by encouraging users to click on these columns in order to animate the mouths — this data was then submitted as their "ratings" for the games.

- In general, NFBkids users were not inclined to enter text in the *More Comments?* box, nor did they volunteer information regarding their age and gender.

- *Tell Us Or Tell Us Off* was not a motivation for most users. The response rate was low and the few comments that were submitted were deemed "less than meaningful" by the managers of NFBkids.

Next, the NFBkids site was examined in light of the psycholinguistics research. The work of psychologists like Clark and Schober provided insight into what might be "going wrong" as far as the *Tell Us Or Tell Us Off* approach was concerned. *TUOTUO* seemed likely to suffer from problems common to many questionnaires — particularly questionnaires that are web-based. Some of these problems were related to the role of the *intentions* of the respondents and the designers of the *SURVEYS* page. For example, common ground, perspective, common purpose, grounding, pressure to respond and pressure for consistency all appeared to be likely candidates for the disappointing results associated with the *TUOTUO* web-form.

Moreover, the work on mental models revealed that a *Tell Us Or Tell Us Off* page could be inappropriate in terms of soliciting information on a children's website. Children were already facing a website designed "by adults". To a certain extent, NFBkids users would be aware of a dichotomy between themselves and the adult designers — and that the designers were "watching" as the kids completed the *SURVEYS*. The *Tell Us Or Tell Us Off* page was intended to give the kids a chance to voice their concerns, but users might have seen the asymmetrical communication of the web-form as pointless — from a user's perspective there was no chance at "feedback" on any feedback they would provide. Essentially, there was little to motivate users to complete the web-form. This would be a problem on any site, but it becomes more significant when the target audience is a group of 9-12 year-olds. Grounding, common purpose and perspective are

concerns with any group of users, however, a group of kids could find the whole process of *"Telling **us**"* unimportant or meaningless. "This sucks", "Good!" and "stoopid" seem like likely responses given the context of the web-form. In retrospect, it is not surprising that responses like these made up the vast majority of *TUOTUO* submissions.

Furthermore, the design of the *SURVEYS* page left users with a task that was not well integrated with the rest of the site. *TUOTUO* stood in stark contrast to the entertaining activities at NFBkids. The click-to-submit task of submitting feedback was very "un-gamelike" and the appearance of the *SURVEYS* page was likely seen as an interruption of gameplay. Pulling users away from the games so they may answer questions that go to "the designers" raises concerns related to the Clark-Schober effects already mentioned (pressure to respond, common purpose, *etc.*); doing so with an instrument that signals to users "the fun is over, now it's time to Tell Us or Tell Us Off" adds new concerns related to task and design, such as those studied by Norman and others.

Certainly, the background research from Chapter 3 was helpful in identifying potential problems (and plausible causes) associated with the methods of gathering feedback at NFBkids. Moreover, the psycholinguistics research and HCI theory seemed to indicate that any "improvement" in design of NFBkids would involve moving away from characteristics that impede or distract from the purpose of the interface (specifically, the *COOL!* and *YUK!* mouths or, more generally, the *SURVEYS* page itself) and attempt to align feedback activities with the goals of the interface. In a sense, a new interface would aim to align the goals and *intentions* of users and designers.

The next step was to apply the theoretical framework of survey design to web-based techniques for gathering user input. This would mean a shift to a new model of soliciting user feedback. A design capable of integrating feedback with gameplay would aim to get *better* and *more accurate* feedback from users and, at the same time, attempt to make users feel "part" of the NFBkids community. Therefore, by incorporating the psycholinguistics and Human-Computer Interaction research, a new interface would attempt to meet both of the NFB's demands — to foster a sense of community among NFBkids users and to gather valuable user feedback.

4.1 Moving from "for the kids" to "by the kids"

The suggestions and conclusions presented in Chapter 3 seemed to indicate that in order to make NFBkids more of a "by the kids" site, the methods of gathering feedback would have to be more kid-centred. Based on the background research, one idea was to shift users' mental model away from "adults" and towards "kids". The NFBkids website was something adults had built for children and, in terms of most users' mental model of the site, "adults" would always be present. However, if the focus of the interaction could be changed, perhaps the site's intended users (9-12 year-olds) might find the process of submitting preferences more "fun" and they might also be more inclined to rate games and provide comments.

With the goal of making feedback less adult-centred, it was time to explore alternatives to *Tell Us Or Tell Us Off*. As mentioned above, *TUOTUO* worked to reinforce a mental model where designers ("Us") were separate from game players. Perhaps the users of NFBkids really had little or nothing to say about the activities at the NFBkids site — this is a possible explanation for the low response rate and the not-so-meaningful one-liners. However, it was also possible that the poor performance of the *SURVEYS* was related to the way in which users were being asked to provide ratings and comments.

Perhaps the Clark-Schober effects of common purpose, common ground, pressure to respond and perspective were playing a role in users' reluctance to supply ratings and comments. If a 10-year-old is placed in a situation where she is asked to tell an adult her thoughts on a child's activity, she might be less-than-willing to do so. She may feel she lacks common ground with the adult. She might also feel pressured to provide a response an adult will understand (or approve of). Furthermore, she might (consciously or not) assume the adult's motives in asking the question are suspect or, at least, different from her own. If all of this takes place on a webpage, where no questions may be asked or clarification given, it provides little motivation for the 10-year-old to take the matter seriously.

One alternative to the "comments for designers" approach of *Tell Us Or Tell Us Off* would be to have a computer-generated agent or avatar ask the user for his or her ratings and comments — the designers could then monitor the recorded "conversation" between the computer and the user. If those who visited NFBkids were, in fact, reluctant to provide feedback to an unknown adult designer, perhaps they would be more inclined to share their preferences with some sort of on-screen animated character. An animated character would likely make for a much more engaging scenario than the *SURVEYS* page. The use of an avatar as the basis for interaction between the system and user was considered (briefly) as part of this

study, however it was decided that this strategy might not deal effectively with the Clark-Schober effects nor would it move the system in a kids-centred direction.[1] In the words of Byron Reeves and Clifford Nass:

> Consider this research example, a prototype of our work on the media equation. If either of us called you on the phone to ask how well you liked this book, you would likely be polite and say that it was fine. But if *someone else* were to make the same phone call, the evaluation might be less positive. This is an example of a simple *social* rule: People are polite to those who ask questions about themselves.
>
> We found the same rule applies to media. When asked to evaluate the performance of a *computer*, people gave more positive answers about how well a computer performed if it asked questions about itself than if a different computer asked the same questions. People are polite to computers, too. [18]

Reeves and Nass illustrate the importance of having *someone else* other than the web designers ask the questions about a website's design. However, the social conventions that figure so prominently in the psycholinguistics research also appear in their discussion of human-computer interactions. This begs a larger question on the social dynamics of any "conversation" that involves soliciting preferences from a group of users: if the designers should not be involved in the conversation (asking the questions), who should?

Given that the goals of the NFBkids redesign included (1) to change the mental model of users by making the "adult" a little bit smaller and the "kid" a bit bigger, in the hope of obtaining *better* feedback and (2) to foster a sense of community presence for the children who use the site, the answer seemed simple: in order to make NFBkids more of a "by the kids" website, the online conversation of preferences and ratings should include only kids — to the highest extent possible.

[1] A comparison of several feedback-gathering systems, some using animated characters as the basis for interaction between the system and user, would make for an interesting study. However, it is beyond the scope of this thesis.

4.2 Peer Presence and Real-Time Assessment

The search for an alternative to *Tell Us Or Tell Us Off* began with an attempt to minimize the response effects mentioned in Chapter 3 and to reduce the potential for an us-and-them mental model among those who would visit NFBkids — all the while adhering as closely as possible to the principles of purposeful design outlined by Norman and others. In keeping with the design goals of NFBkids, the new system would stress interaction and collaboration, and it would promote a sense of community among users.

Early in the design process, it was thought that shifting the focus of game-rating away from "Us" (adult web developers) would be a delicate procedure. One additional concern was that the rating system would have to be independent of the NFBkids activities — a dissimilar collection of games that would change over time. Individual games would be added, modified and deleted; the rating scheme could not be closely tied to any one activity in particular. Furthermore, the system would have to encourage users to rate games, but not interfere *too much* with the game activities — *i.e.*, some sort of balance would have to be found where the rating system would be present and "enticing", but not distract players' attention from the games to such an extent where it would impede or obstruct gameplay. To summarize, a successful feedback-gathering system would:

- be able to address the Clark-Schober effects — particularly common ground, common purpose, perspective, grounding and pressure to respond
- reduce the role of "adults" in the feedback-gathering process
- follow HCI principles of low cognitive burden and purposeful design
- function independently of the NFBkids game activities
- promote a sense of community and "entitlement" through user-interaction and user-collaboration
- encourage users to submit feedback, but interfere minimally with the game activities

Based on the background research and on the apparent overlap between some of the goals listed above, a few hypotheses emerged concerning feedback, game-play and "sense of entitlement". First, as feedback became integrated with game-play, perhaps users would become more inclined to provide comments and ratings

(*i.e.*, if the *task* of rating games became more gamelike, players might begin to "enjoy" submitting feedback and do so more freely — or, at least, they might no longer see rating games as a chore; something apart from gameplay). Second, if it were possible to shift game ratings and comments so that, ostensibly, they were intended for users (and aligned with the purposes and intentions of users) instead of "for designers", this might help the system to contribute to a sense of community while simultaneously improving the quantity and quality of feedback.

This line of thinking lead to a central hypothesis on how an interactive system might simultaneously foster community presence and collect accurate user feedback: *perhaps a symbiotic relationship exists between peer presence and interactive, collaborative assessment.* If this were the case, a collaborative feedback-gathering system might be able to successfully meet both of the Film Board's key goals (and shortcomings) with NFBkids — if good feedback is linked to a sense of peer-presence, an interactive peer-centred system should be able to stimulate and encourage both.

Moreover, a tool or set of tools that allows users to submit comments and game-ratings to each other in real-time might further contribute to a sense of peer presence and "system entitlement" — this is the essence of many of the powerful interactions seen in Turkle's studies of social relationships on the internet [27]. Therefore, the new system would incorporate tools based on these hypotheses. It would be the first step in testing this potential symbiotic relationship between peer presence and real-time, collaborative assessment.

4.3 The Popcorn Meter

Peer presence and real-time assessment were two key goals for the new interface. To address the Clark-Schober effects, the new system would allow players to "rate the games" by sending ratings and comments to each other. The game-rating dialogue would take place between users — no longer would comments and ratings be intended for "the designers".

It is not possible, based on the psycholinguistics research, to conclude that feedback intended for other players will be *better* than comments and ratings intended for designers. However, the research does suggest that peer-to-peer feedback will likely be *different* from user-to-designer feedback. In terms of rating games, a player's intentions are likely to be more closely aligned with those of other players than with the perceived intentions he or she might assume designers have for wanting feedback — especially when the players are kids and the design-

ers are adults. Likewise, issues related to pressure to respond, common purpose, common ground and perspective are presumed to be *different* when players rate games for other players than when they "know" their comments are intended for designers. For example, peers often assume similar language, goals and attitudes. This seems likely to affect their approach to game-rating when comments and ratings are exchanged through a peer-to-peer rating system.

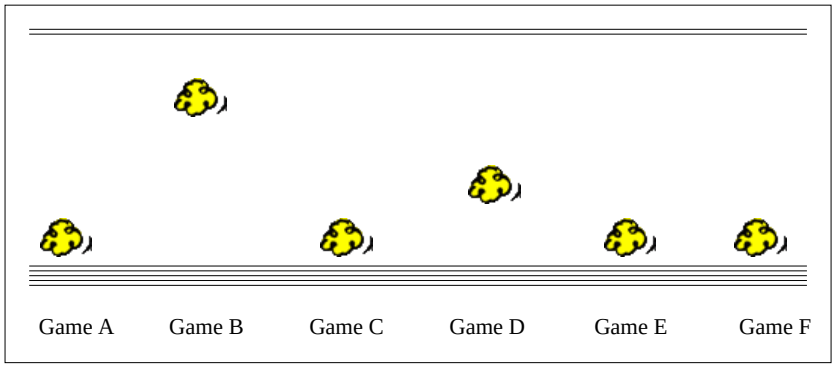


Figure 4.1: Design concept for the Popcorn Meter.

Early in the design process, it became clear that not all of the information (*i.e.*, comments and ratings) could be sent to users in real-time. An instrument capable of conveying real-time ratings and comments from even a modest number of users would demand a significant amount of on-screen attention from every game player. Such a tool could not be a subtle instrument. When on-screen, a comments-and-ratings device would interfere with the playability of any other on-screen game activity. Therefore, it would have to appear on-screen sporadically, lest it become too distracting or risk being ignored.

In order to maintain a sense of peer presence, it was felt that peers (or, at least, peer activity, *e.g.*, their ratings) should always be "present" to all players. Therefore, an elaborate all-in-one device was rejected in favour of a discreet, real-time tool that would be "omnipresent" — present on every player's screen, all the time, so that every player could receive at-a-glance feedback from his or her peers at any moment. The hope was that this omnipresent tool would reinforce the idea that others are "out there" (*i.e*, it would be a constant reminder of other players' activity), but still allow players to focus on the primary game activity. This device evolved into the *Popcorn Meter* (see Figure 4.1 for an early sketch). The Popcorn Meter would always occupy space on-screen and it would convey every player's

ratings to all players, in real-time.

The concept is as follows: when any player rates a game (this takes place after he or she finishes with a game activity; more on this in the next section), the "kernel" on the Popcorn Meter associated with that game "pops" according to the player's rating. The player may assign a game a *kernel rating* of zero, one, two, three or four kernels (this rating scheme offers more granularity than the 2-position *COOL!* and *YUK!* scheme seen on the *SURVEYS* page). In the sketch, *Game B* has received a rating of three kernels and *Game D* has received a rating of one kernel. The kernels hover in their "popped" position for a few seconds before returning to the initial position — essentially the zero kernel position. As mentioned, all Popcorn Meters are networked; a *kernel vote* by a single player affects everyone's Popcorn Meter.

At first glance, one might view the Popcorn Meter as nothing more than an ornament to embellish the rating process. However, if players know that *others* have Popcorn Meters on their screens, the co-operative principle discussed by Clark and Schober [3], along with the maxims of conversational implicature mentioned by Grice [5], imply that the Popcorn Meter has the potential to become an instrument of communication. Therefore, the Popcorn Meter (when combined with the Ratings Screen and the Comments Screen, to be discussed in the next section) provides users with the chance to *contribute* to a dialogue (albeit, somewhat abstract) on the game activities. It provides context for the discussion to take place and informs users that others are "out there", engaged in the conversation.

Furthermore, the observations by Turkle [27] and others suggest that peer-to-peer interaction (like that of the Popcorn Meter, Rating Screen and Comment Screen) has the potential to arouse feelings of entitlement among those who participate. Accordingly, the Popcorn Meter is one component designed to test the central hypothesis presented in the previous section: that peer presence and real-time assessment are linked in a "mutually beneficial" way.

4.4 Ratings and Comments "to other players"

A "Post-Game Screen" (Figure 4.2) was proposed as the method by which players would submit their ratings. This screen would appear when a player completed all levels of a particular game activity or in the event he or she quit the game prematurely. At the Post-Game Screen users would be able to assign a game 1-4 kernels or choose not to enter a rating. This information would then be sent to every player's on-screen Popcorn Meter and to a database. A "Ratings Board"

screen would follow the Post-Game Screen and call upon the database to list the kernel ratings (from all players) for that particular game (Figure 4.3).

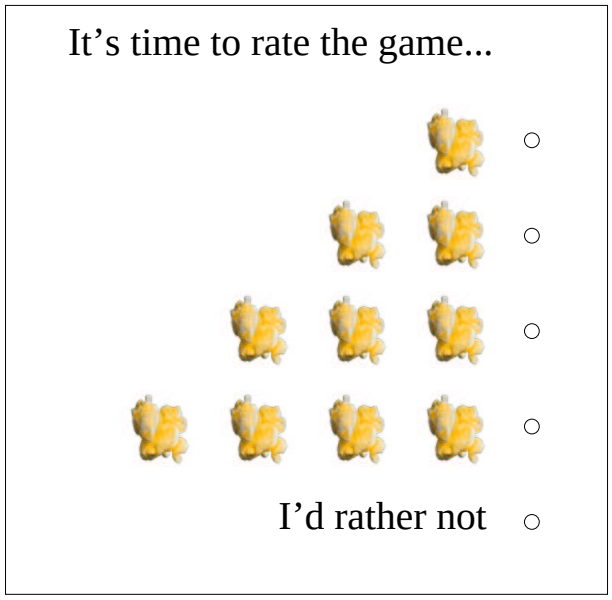


Figure 4.2: Proposal for the post-game Rating Screen.

The Ratings Board would provide feedback to the user, showing him or her that the kernel vote from the previous screen was accepted and registered (his or her kernel vote would be on the board with the other ratings) and it would give the user a chance to see the history of kernel ratings, along with the names of players who submitted these ratings. The Clark-Schober work [3], and the research by Grice [5], suggests that the Ratings Board might also reinforce the concept of a co-operative dialogue. As players see their ratings "on the list" with the names and ratings of others, they have the opportunity to observe a "community" of players voicing differing opinions on the game activities.

Seeing the names (or handles, if users log into the system with false names) of other players may strengthen users' concept of this community. By giving users the chance to observe an *exchange* of ratings among a community of players (to which they belong) the Ratings Board may provide further context for a "conversation" on the game activities. The Popcorn Meter issues immediate, peer-to-peer feedback. The Ratings Board allows users the chance to spend as much time as they like viewing the ratings of their peers. Both function to provide the context,

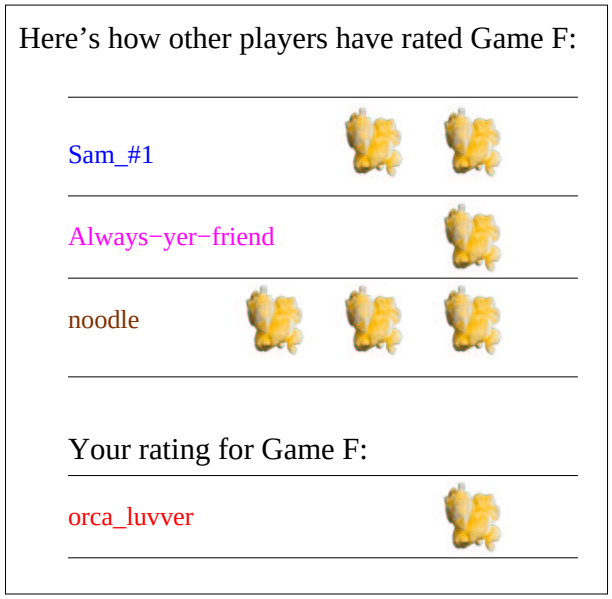


Figure 4.3: Proposal for the Ratings Board — a screen that lists all ratings for a particular game.

the catalyst and the vehicle for a game-rating dialogue among players.

As mentioned above, a real-time tool with a direct chat feature was rejected in favour of the Popcorn Meter because a "comments tool" was thought to be too disruptive and likely to interfere with the playability of the main on-screen activity — the NFBkids game. However, user-comments were deemed crucial to any online "discussion" of the game activities. User-comments would allow for a more thoughtful response from players (this is of particular importance since the designers wanted to know *why* players liked or disliked certain games) and such comments would further support the peer-to-peer dialogue and sense of community that was being designed "into" the system.

Like the Post-Game Screen and Ratings Board, submitting and viewing comments would have to take place after a player was "finished" with a game. The "Add Comments Screen" was proposed as the tool that would allow users to enter comments (Figure 4.4). The system would present players with the Add Comments Screen after they were done viewing the Ratings Board. Any information a user would submit through the Add Comments Screen would be stored in a database and retrieved on the subsequent screen — the "Comments Board" (Fig-

Figure 4.4: Proposal for the Add Comments Screen

ure: 4.5).

As with the Popcorn Meter and kernel rating screens, it was hoped that the Add Comments Screen and Comments Board would function as tools for co-operative discourse and contribute to an environment where a peer-to-peer game-assessment dialogue could take place. Like the Ratings Board, the Comments Board would show users that their comments had been registered by the system and let them see the comments (and handles) of others.

The Popcorn Meter, Ratings Board and Comments Board were designed to provide users with a context in which game-rating could be an entertaining, collaborative activity. The psycholinguistics research became the framework for a design that would be peer-centred, encourage interaction and collaboration among users, promote a sense of peer-presence and act as the catalyst for a dialogue on the game activities. Whether or not users in the target audience (9-14 year-olds) would actually use the system as intended was unknown at this point in the study, but it was felt that the Popcorn Meter, Ratings Board and Comments Board *might* be able to gently push the attention of users in a particular direction — to get them to think about rating the games and communicating their preferences to their peers.

According to the secondary research that was prepared prior to the design of

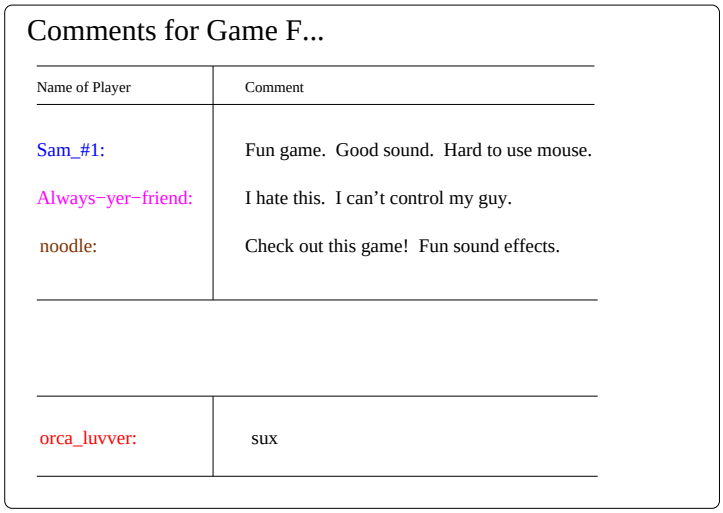


Figure 4.5: Proposal for the Comments Board.

the system, the PM-RB-CB (Popcorn Meter, Ratings Board, Comments Board) design had followed many of the principles of HCI, psycholinguistics and questionnaire design. Here is some of the speculation that surfaced after the the design was completed:

- The flickering of the Popcorn Meter might act as constant (friendly and/or not-too-intrusive?) reminder to users of their presence in an online community of other NFBkids gameplayers — and get them thinking about how much they "like" the games they are playing.
- The Ratings Board and Comments Board might further players' sense of attachment to the community and encourage them to submit their "honest" feelings on the games.
- These tools might shape user interaction so the ratings and comments databases contain useful information for the designers. Information that would help designers incorporate player preferences in the design of the next version of NFBkids — on the road to building a more "by the kids" website.

However, it also seemed possible that children using a PM-RB-CB version of NFBkids would be less than thrilled by the system. In "darker moments" these concerns surfaced:

- Would the Popcorn Meter be an annoyance? Would it detract users' attention from the main game activity, leaving users frustrated and even less interested in game-rating?

- Would the intentions-related research (as interpreted) apply to a click-to-submit web-form? NFBkids Version 1.0 users were not interested in clicking-to-submit. It seemed likely that a question of intentions (theirs versus those of the designers) was part of the problem, but was a web-form simply too "boring" an instrument? Perhaps a PM-RB-CB system would suffer a similar fate.

- Even if the PM-RB-CB design was able to make the adult "smaller" and the kid "larger", in terms of users' mental model of the system, would it be *enough*? Might kids find the rating system contrived instead of "fun"?

Some of the speculation above was valid — based on the theoretical framework behind the project — and some was not — simply the hopes and fears of an inexperienced software designer. The implementation and demonstration of a prototype to children in the target audience would help identify the valid concerns and separate them from naive hopes and fears.

Chapter 5

User Studies: Goals, Objectives and Method

> *The misconception that playing electronic games is an anti-social activity has been dealt with in the recent literature. While some children do prefer to play alone, many children prefer to play while others are present. Girls are in fact more likely to play if there's the possibility of interacting with others while they play [...] And boys do play collaboratively and are often involved in game-playing parties and everyday conversations about games. Both of these findings support the claim that for many children playing electronic games is not an anti-social activity*
>
> — Joanna McGrenere, "Design: Educational Electronic Multi-player Games". [10]

The proposal document was sent to several groups at the National Film Board (including ACI and the Kids Destination Project) in April 2002. The initial response from the managers of ACI and Kids Destination was positive. However, changes at the Film Board meant that NFBkids would no longer be a priority for the NFB. By June it was clear that an investigation into peer-to-peer real-time assessment would become a solo project without significant NFB involvement.

During the summer months of 2002, the decision was made to put together a collection of online game activities similar to what existed at NFBkids. A prototype of the PM-RB-CB system would be integrated into this set of online games, and the system would be shown to Grade Six students in Vancouver schools through E-GEMS.

Accordingly, two versions of a Java- and Javascript-powered website were built: one with a comments-to-designers method of soliciting user feedback, and a second system with a Popcorn Meter, Ratings Board and Comments Board. The

first system would be used to see how the traditional "Tell Us" methods would fare with a Vancouver Grade Six class (in a laboratory setting) and the second system, intended to promote a dialogue on users' opinions of the game activities, would demonstrate peer-to-peer rating to a different set of Grade Six students.

Both web-based systems were designed to operate on a single Linux computer running an Apache web server. This would simplify the user study process, allowing kids in the Vancouver school system to surf their way to the website using networked machines in their school's computer laboratory.

The two systems ran on a modestly-powered computer at the University of British Columbia[1] connected to the University's high-speed network. Both versions of the website ("tell designers" and PM-RB-CB) relied on Java, Javascript and PHP. A mySQL database (integrated with Apache 2.4 through PHP) stored all comments and ratings, and would later help in sorting and analyzing the data.

In mid-October 2002, the comments-for-designers website was shown to students at University Hill Elementary School, near the UBC campus. A few weeks later, the interactive, peer-to-peer system was unveiled to students at Maple Grove Elementary in the upscale Vancouver neighbourhood of Kerrisdale.

5.1 Exploring "traditional" methods: University Hill, October 2002

A collection of ten game activities formed the basis of the *Fun and Games* system that was presented to University Hill (U Hill) students on 18 October 2002. The Grade Six class (ages of students ranged from 10-12 years-old) was given a five-minute presentation on the system and the study, then the students spent 50 minutes in the University Hill computer lab playing with the system. After the 50-minute session in the computer lab, they returned to their classroom and were given 15 minutes to complete a post-game questionnaire.

During the pre-session presentation, the University Hill students were told they were the first group of elementary students to "try out" the new system and that one of the goals of the study was to collect their comments and opinions on the games. Accordingly, after each game activity they would have the chance to assign the game a *happy face* rating. Originally, the rating scheme for the University Hill study was to consist of kernel ratings, but a preliminary study

[1]The machine, known as *faxe*, had a 450 MHz Intel Pentium III processor, 128 MB of RAM, ran Red Hat Linux 7.3 and made its home in the Department of Computer Science.

at UBC (with adults) had participants asking "why are these popcorn kernels?" Without the context of the Popcorn Meter, pilot study participants found popcorn kernels to be an odd choice for a rating scheme. When pilot groups at UBC were shown a comments-to-designers system that rated games with *happy faces*, the participants did not question the rating scheme. On this basis, the University Hill system was designed with happy faces instead of popcorn kernels.

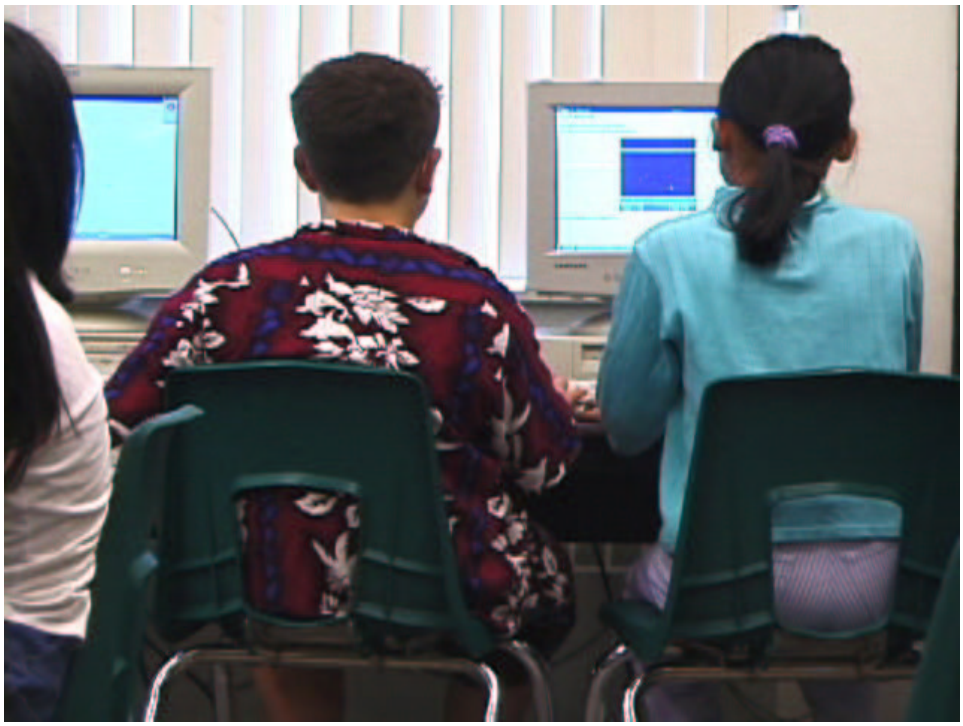


Figure 5.1: A pair of University Hill students playing a 2-player game of *Blister Ball*.

It may seem as though the presence of *faces* instead of *kernels* adds an unnecessary variable to the process of experimenting with a comments-to-designers system and a peer-to-peer rating system. However, the U Hill study was not designed to be a control group for an experiment on comments-to-designers and peer-to-peer rating. At this early stage in the testing and development of a peer-to-peer rating system, the user studies (both U Hill and Maple Grove) served as the point of departure for an analysis of peer-to-peer rating as a method of gathering user feedback. The prototypes shown to Vancouver elementary students represent

an early step in the process of refining a design based on peer-to-peer interaction and collaboration. It was hoped that the University Hill and Maple Grove studies would indicate whether or not students would accept a peer-to-peer rating system like PB-RB-CB. If such a system were shown to be in the "ballpark" in terms of what students would accept, then more comprehensive user testing could follow.

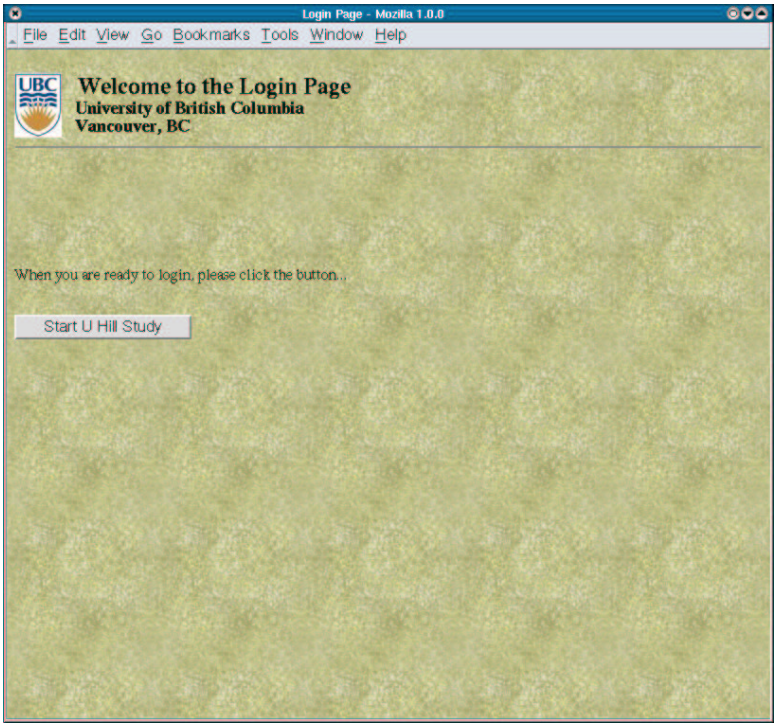


Figure 5.2: University Hill login screen.

In the University Hill computer laboratory, 28 students gathered around 15 personal computers. On 13 of the machines, students worked in pairs[2] — see Figure 5.1. Two students (one a late-comer) had a machine to themselves. Video cameras recorded the actions of two of the paired groups. The Apache server running on *faxe* logged every page request.

[2]The number of available computers made this pairing-up necessary. Other studies by McGrenere [10] and Inkpen [7] indicate that observing students in pairs may be beneficial — as certain interactions occur when students are paired. These interactions may provide insight into students' thoughts, opinions and feelings. Therefore, pairing-up was not seen as a drawback, but as an opportunity.

After the students were seated in front of their computers, they were given a username and password (one username and password per computer) along with the Uniform Resource Locator (URL) of the study homepage. The homepage was a simple screen that asked students to click on a Javascript-powered button and "log in" to the system (Figure 5.2). Clicking on the *Start U Hill Study* button brought up a pop-up window that prompted users for their *Username* and *Password*. The username and password were implemented to control access to the system and to personalize "Confirmation Screens" (screens that would appear after a user submitted ratings or comments).

Once the students had successfully logged into the system, they were presented with a modest "Palette Screen" where web-links to the 10 game activities were arranged in a simple list — a "palette" of game links. The Palette Screen gave no indication of the type of activity to which each link pointed. Games were listed simply as *GAME A*, *GAME B*, *GAME C*, . . . , *GAME J*. Clicking on these links would take users to the following games:

Game A

Game A was a Java applet version of a timeworn video game called *Lunar Lander* that was popular on the Atari 2600 console, and later the Commodore 64 (Figure 5.3). The controls for Game A were simple: the SPACE BAR activated the "Main Thruster" and the LEFT ARROW and RIGHT ARROW keys rotated the lander. With a delicate touch, the player could "land" the lander on several platforms — and earn points by doing so. A not-so-delicate touch would send the lander into the "canyon" walls where the vector graphics components of the lander would disintegrate. Generally, the lander was difficult to control and crashing the lander was a common experience.

In the Game A version, the lander could crash an infinite number of times, but each crash would reset the score to zero. When players were "finished with this game", they could click on the link below the applet canvas. This link would take them to a post-game Rating Screen.

Game B

Billed as a "race against the clock", Game B (*a.k.a.*, *Concentration*) was a simple Javascript-based matching game (Figure 5.4). As with many games of this type, the goal was to uncover pairs of images as quickly as possible. The Javascript button located beneath the game board acted as both a clock and reset button.

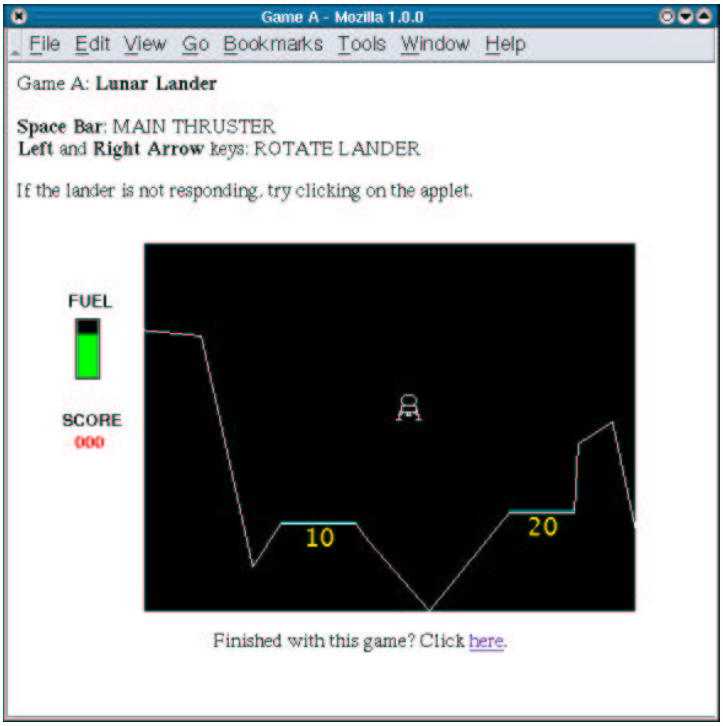


Figure 5.3: Game A — the *Lunar Lander*.

Clicking on the clock would re-jumble the images and set the clock to 0:00. When a player succeeded in matching all pairs of images, the clock would stop. As with the Game A page, players could click on the *Finished with this game? Click here* link at the bottom of the screen in order to move on to the Rating Screen.

Game C

Game C (*Box Builder*) was a single-player strategy game written in Javascript. The player (in red) would take turns with the computer (in blue) to see who could build more "boxes" on an 8 x 8 grid of dots (Figure 5.5). The dimensions of the gameboard meant that a total of 49 boxes were built per game.

Boxes were built one "line" at a time. A game would start with an empty grid of dots (*i.e.*, a grid with no lines) and the human player would take the first turn by placing a vertical or horizontal red line on the board — this was done by clicking on any pair of adjacent dots (radio buttons). The computer "player" would add a

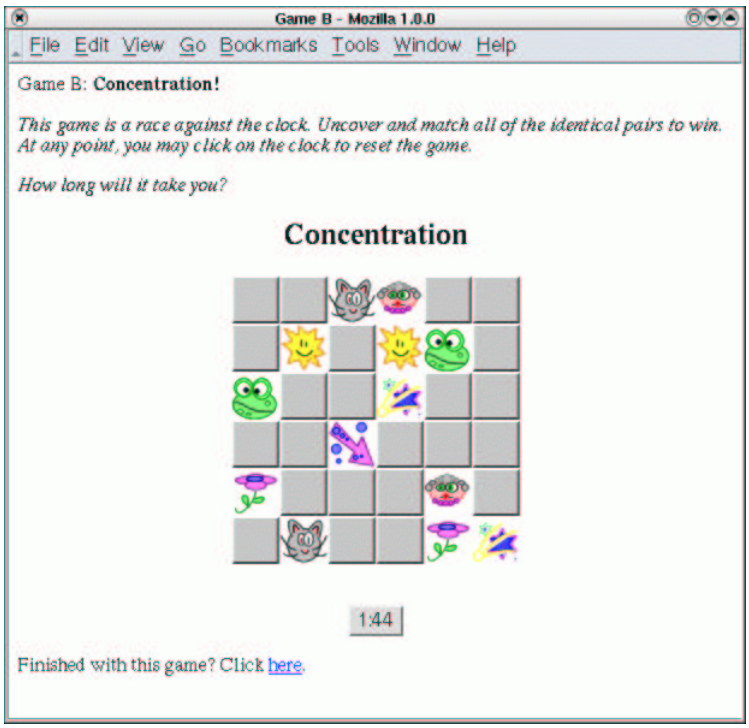


Figure 5.4: *Concentration*. Also known as Game B.

blue line to the board, and then it would once again be the human player's turn. Play would continue in this way until a "box" was formed by one of the players. When either the human or the computer completed a box (*i.e.*, by adding its fourth side), this player would "claim" the box and the square within the lines would be drawn in the that player's colour. Whenever a player claimed a box, that player would get to add another line to the board. The game ended when all lines (112 of them) were drawn.

The computer would easily beat novices. However, with a bit of practice, most players would consistently build more boxes than the *Box Builder* algorithm.

Game D

This game, also known as *COBS*, was a *Tetris* clone played either with the arrow keys or with the mouse (Figure 5.6). The slight twist with *COBS* was that each piece that "fell" to the bottom of the Java canvas was made up of four animated

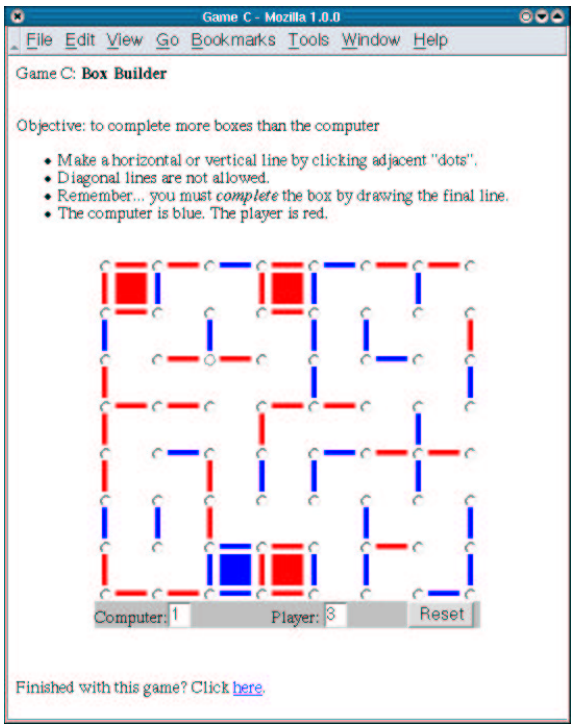


Figure 5.5: The midpoint of a game of *Box Builder.*

"heads". Most of the heads appeared to be angry, although the occasional piece looked "happy".

Game E

Javanoid (*a.k.a.* Game E) was a clone of the popular video game *Arkanoid*, itself a take on other games like *Breakout* or *Fast Ball* (Figure 5.7). Like *Arkanoid*, this game featured "pills" that would fall towards the player's paddle — catching these pills with the paddle would add special powers (*e.g.*, a longer paddle, a "shooting" paddle, an extra paddle, a sticky paddle, or three balls in play at once).

Game F

This single player game was given the name *3-D Vector Air Hockey* (Figure 5.8). The mouse controlled the near paddle and the object was to put enough "english"

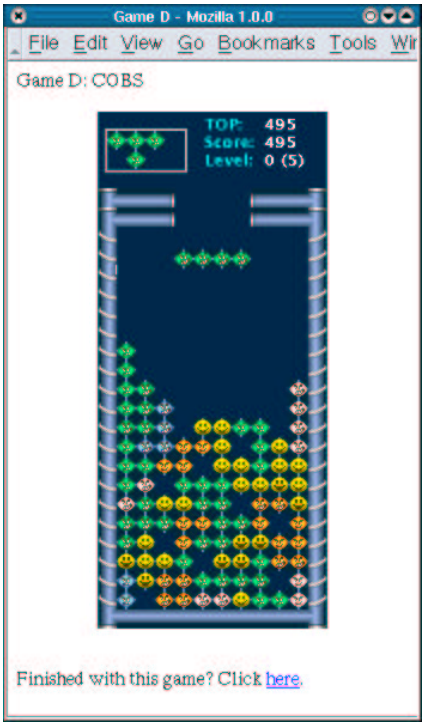


Figure 5.6: *COBS* — like Tetris, but with animated "heads".

on the puck to get it past the computer's paddle. Mouse control for *3-D Vector Air Hockey* was very sensitive (some pilot study participants called it "touchy"). As a result, it was very easy to lose. The computer would play indefinitely — at no point would one player be declared the winner.

Game G

Game G was the *Hoop Shooter* (Figure 5.9). The mouse controlled the placement of the basketball player. To shoot the ball required two mouse clicks – the first mouse click would "lock" the player and stop the progress meter labelled "Forza" (Force), the second click would stop the progress meter labelled "Direzione" (Direction). Most pilot study participants found this to be an addictive game. Even with considerable practice, getting the ball in the hoop was a challenge.

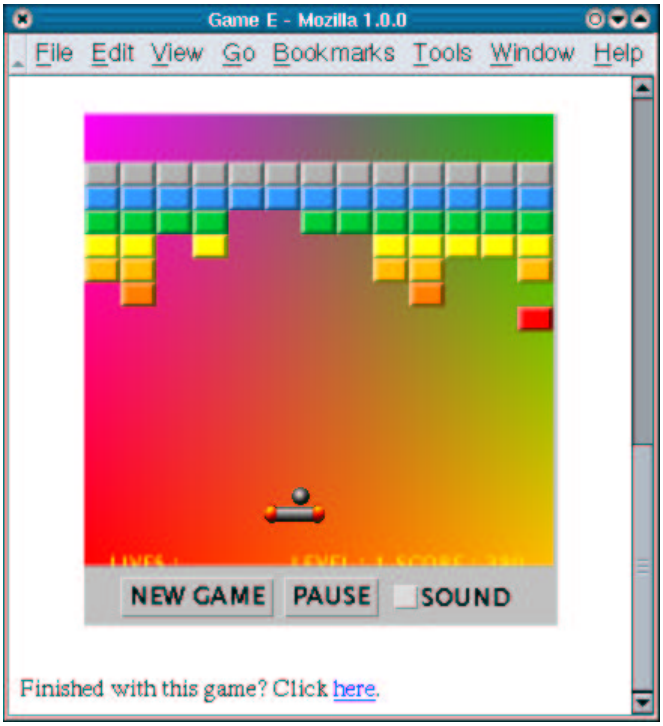


Figure 5.7: *Javanoid.* A Java-powered version of *Arkanoid*.

Game H

The only two-player game in the collection, Game H (*Blister Ball*), was a clone of an old Apple II game (Figure 5.10). In two-player mode, it was a co-operative activity where both players would "shoot" at targets (balls) as these targets bounced around the screen. The players' "bases" (from which they would shoot) could move horizontally at the bottom of the Java canvas. Player 1 used the Z and C keys to move her base and X to shoot, while Player 2 moved his base with 1 and 3 (on the keypad) and used 2 to fire. The game could also be played in one-player mode, in which a single player would use the Player 1 keys to shoot the targets by herself.

A player's base had three "levels". If the base were struck by a ball, it would lose a level. Therefore, after being struck 3 times, the player's base would expire and the game would be over for that player.

One of the more frustrating aspects of *Blister Ball* was that players' "shots"

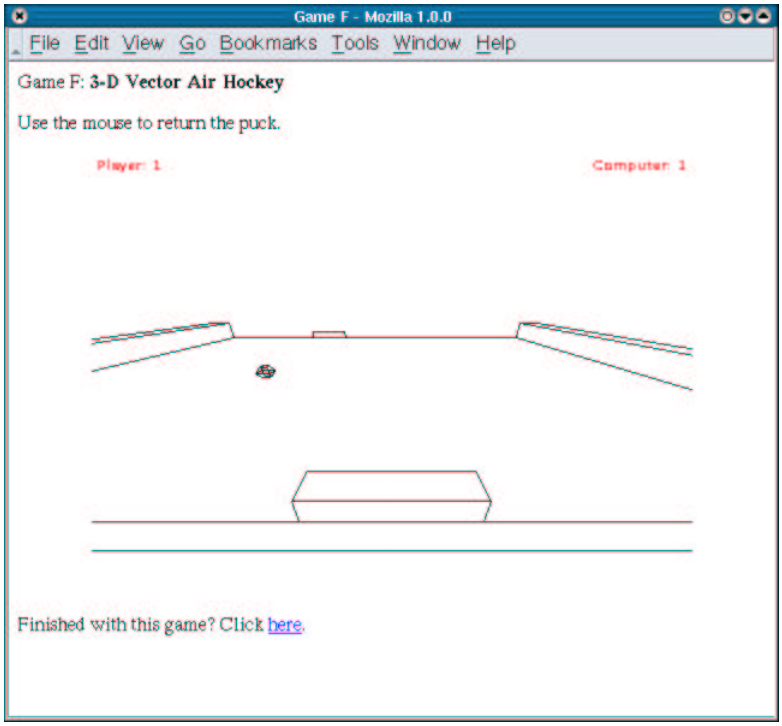


Figure 5.8: The retro-looking screen of *3-D Vector Air Hockey.*

moved very slowly, and only one shot per player was allowed on the screen at any time. When a player missed a target, it would be a second or two (although it might feel like an eternity) before that player could fire another shot.

Game I

Game I, *Boulderdash*, was the most adventure-minded of the games. Level 1 of *Boulderdash* (Figure 5.11) was the least challenging of the game's ten levels, but still difficult for most novice players. The goal was to move the friendly-looking robot around the game-board (using the arrow keys), while avoiding various hazards and collecting enough diamonds to open the portal to the next level.

As the robot moved around the screen, it would create a path that might cause dangerous insects to be released or cause boulders to fall. When an insect touched the robot it would cause instant death (for the robot, that is) and boulders falling from any height were also fatal. Careful positioning of the robot would allow

Figure 5.9: Not as easy as it looks. . . The *Hoop Shooter.*

it to push boulders (in some cases onto insects, killing them) and cause "chain reactions" where many boulders might fall, causing an explosion. This was a skill learned after many failed attempts.

Game J

This was a somewhat difficult turn-taking game called *Bridges* where a single player would attempt to build a "bridge" across the game-board more quickly than the computer. Written in Java, *Bridges* left the player to connect the left and right sides of the Java canvas with a red line, while the computer attempted to connect the top and bottom edges of the "board" with a solid blue line.

The game opened with a Java canvas that looked similar to a checkerboard, but with red, blue and white squares (Figure 5.12). The player would take turns with the computer "filling in" the white squares — a click to a white square would turn it red. The computer would claim a white square by turning it blue. If the

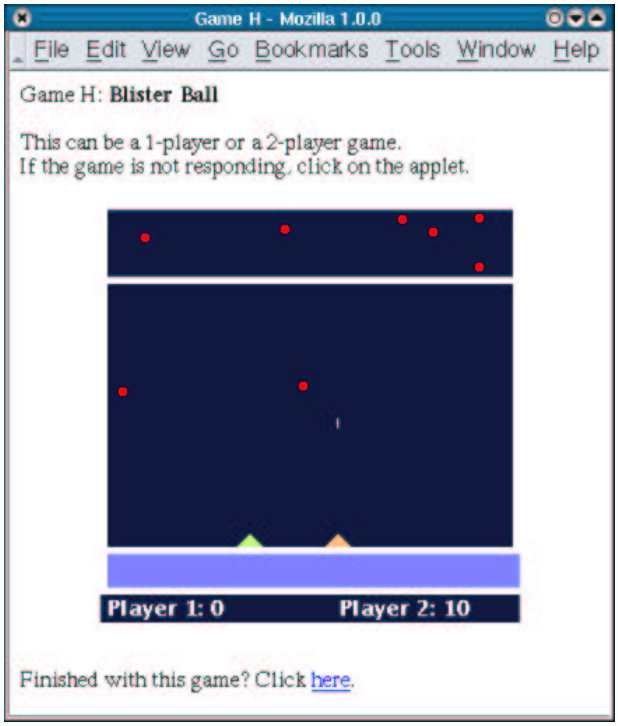


Figure 5.10: The two-player game *Blister Ball*.

player could build a continuous path of red from left to right before the computer could build a blue path from top to bottom, the player would win. Rarely were novices able to beat the computer — it took quite a few tries to find a way to beat the machine.

Sending Ratings and Comments to the Designers

The Rating Screen that followed each game allowed users to assign zero to four *Happy Faces* to the game they had just played (Figure 5.13). Users could also select "I do not want to vote" or simply hit *SUBMIT QUERY* without clicking on any of the radio buttons[3]. The system recognized all of these possible outcomes

[3]Of course, *SUBMIT QUERY* is a poor name for a button that submits a player's vote to the designers — there is no "query". The Rating Screen webpage used PHP to send data to the mySQL database and, as this sort of operation is known as a mysql_query(), "SUBMIT QUERY" is the default tag for this type of button.

Figure 5.11: *Boulderdash.* Avoid the hazards and collect enough diamonds to open the portal to the next level.

and stored them as distinct submissions in the mySQL database.

This implementation (0-4 Happy Faces and "I do not want to vote") differed slightly from the proposal document that offered users 1-4 kernels and the "I'd rather not" option. The change was made to offer players the zero faces/kernels option because it was felt that a zero rating might be an important option for some users.

After hitting the *SUBMIT QUERY* button, players would see a new screen that gave them a polite "feedback message" and the opportunity to send comments to "the web designers" regarding the game they had just played (Figure 5.14). The comment box on this screen was implemented with a 40-column, 8-row PHP text-form that would connect to the mySQL database running on *faxe* and "insert" the contents of the text box into the database.

A click on the *SUBMIT QUERY* button would trigger the PHP-form action and take users to a final "feedback" page. This feedback screen gave the user a

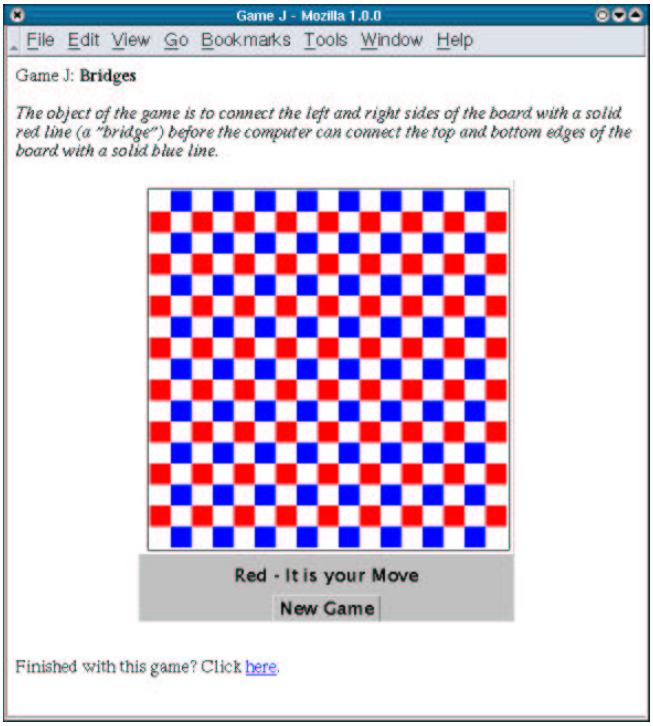


Figure 5.12: A "fresh" *Bridges* game board.

short, personalized "thank you" message and echoed the comments entered on the previous page (Figure 5.15). From here, users could return to the palette screen by clicking on the link at the bottom of the screen.

Post-Session Questionnaire

The two-page written questionnaire given to U Hill students after the 50-minute session in the computer lab (see Appendix A for a sample) was designed to gather students' opinions on "sending comments and ratings to designers". There were six sets of questions on the questionnaire:

1. Of the games you just played, please list your top 3 games.
2. Think about one of your top 3 games. If you had to describe this game to some friends (who had not seen the game) what would you tell them? (Please circle this game in the list above.)

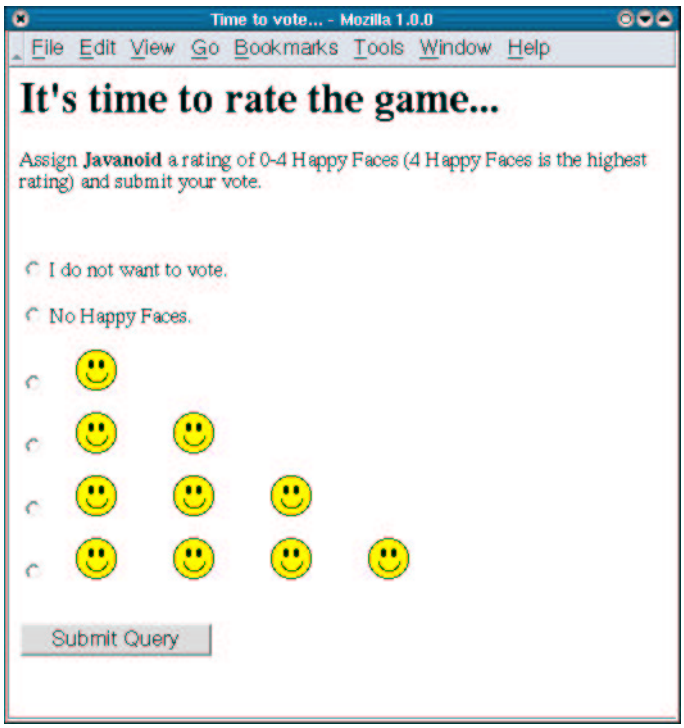


Figure 5.13: The University Hill post-game Rating Screen for *Javanoid*.

3. In this study, after you played a game, you had the chance to send your ideas and opinions to the game designers. Is this something you like? Are you pleased that the game designers will be reading your comments?

4. Do you think the game designers will pay attention to the comments you entered today?

5. Why do you think the game designers want to read your comments and see how you rated the games?

Section 1 of the questionnaire was intended to get students to "think" about the games they had just played — a way to tie the written questionnaire to the experience in the computer lab. Also, the handwritten list of favourite games could be compared to the *Happy Face* ratings and comments recorded by the system.

Section 2 was an attempt to see if students' comments "for friends" (albeit, on a post-game, written questionnaire) might differ from comments intended for

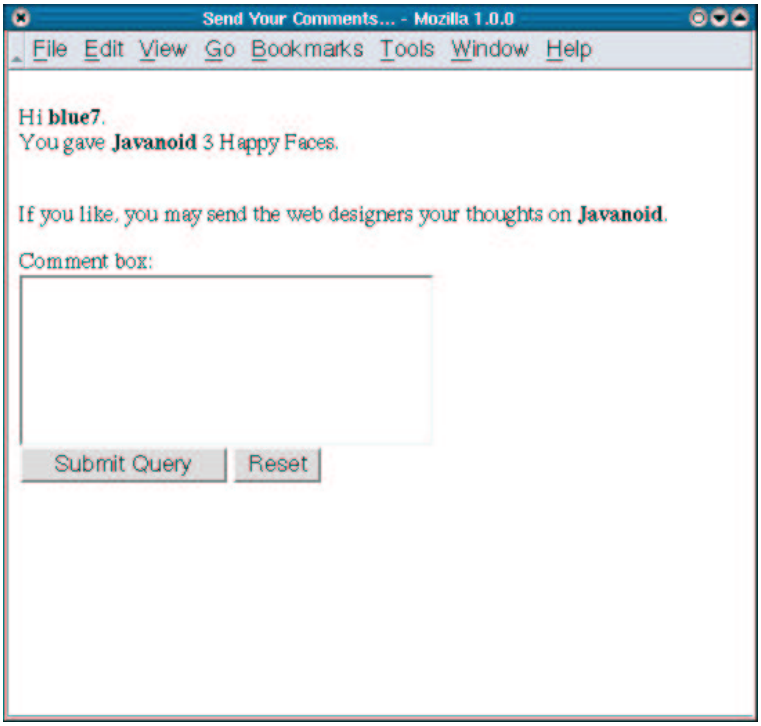


Figure 5.14: The University Hill "Comments to Designers" screen for *Javanoid*. At the previous screen, user *blue7* assigned *Javanoid* a rating of 3 Happy Faces.

designers. Of course, typing comments into a web-form during playtime (in the computer laboratory) is very unlike answering a post-session written questionnaire (in the classroom), but it was felt that the questionnaire should at least try to find out what kids might "tell their friends" about the games.

In Section 3, it was hoped that students might simply state whether or not they liked the idea of sending post-game ratings and comments to designers.

Sections 4 and 5 were aimed at isolating some of the Clark-Schober effects. Would the kids think their comments "matter" to the game designers? Why might students "want" to submit comments and ratings? What did they assume the purpose might be behind the designers' desire to have users (*i.e.*, kids) submit ratings and comments? How did the kids see the process of gathering user preferences? Also, how did the students see the intentions and roles of designers and kids (as they relate to the system)?

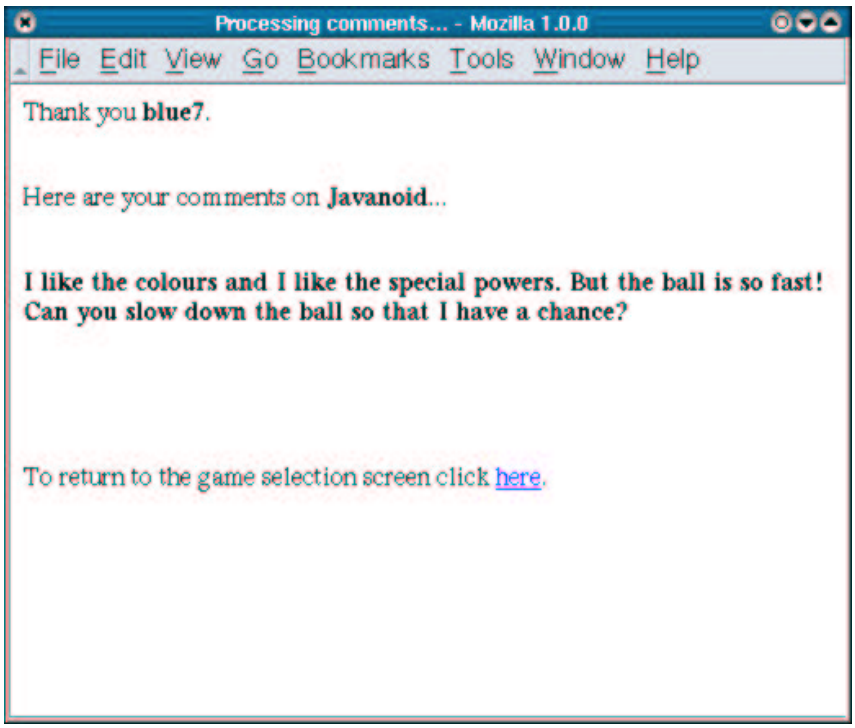


Figure 5.15: The final screen in the University Hill ratings/comments process — a screen that echoes the user's comments and allows the user to return to the palette screen.

After 15 minutes, all students had completed the written questionnaire. The questionnaires were collected, and a short "thank you" concluded the University Hill study.

5.2 An Interactive System: Maple Grove, November 2002

On 7 November 2002, 27 Grade Six students at Maple Grove School took part in the second *Fun and Games* user study. As with the University Hill study, before students received their passwords and usernames, the Maple Grove (M Grove) students were given a short presentation on the system. They were told they would be trying out a collection of games and that, once logged into the system, they

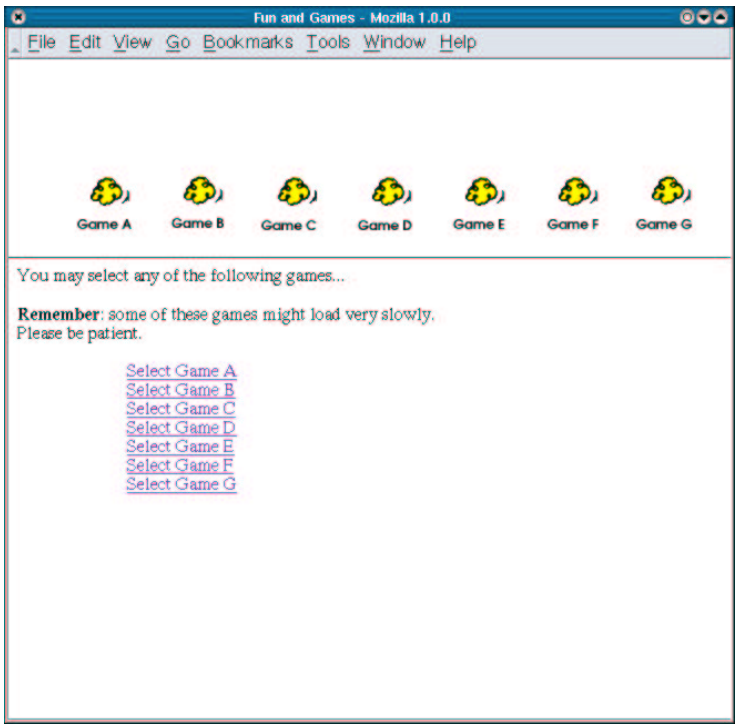


Figure 5.16: The Maple Grove Palette Screen. The Popcorn Meter is "idling" in the top frame of the page.

would be able to signal to other players how much they liked particular games by using a real-time tool (the Popcorn Meter) and by adding their ratings and comments to each game's cumulative post-game Ratings Board and Comments Board. At no point during the presentation were "the designers" mentioned as a target for ratings or comments. The focus was on rating games for "other players".

Before logging into 17 personal computers in the school's computer lab,[4] the students were given instructions on how and when they would be able to rate the games, and told that their kernel ratings would cause everyone's Popcorn Meter to "pop". Also, various screens (including the Palette Screen, the Ratings Board, the Add Comments Screen and the Comments Board) were explained before the kids were allowed to surf their way to the Login Screen.

Accordingly, the Maple Grove students saw a system that was substantially

[4]Twenty students worked in pairs and seven students had machines to themselves.

different from the one shown to University Hill kids three weeks earlier. The most striking difference was the presence of the Java-powered Popcorn Meter that occupied the top frame of every HTML page (for example, see Figure 5.16). Also, the system consisted of seven games instead of ten.[5]

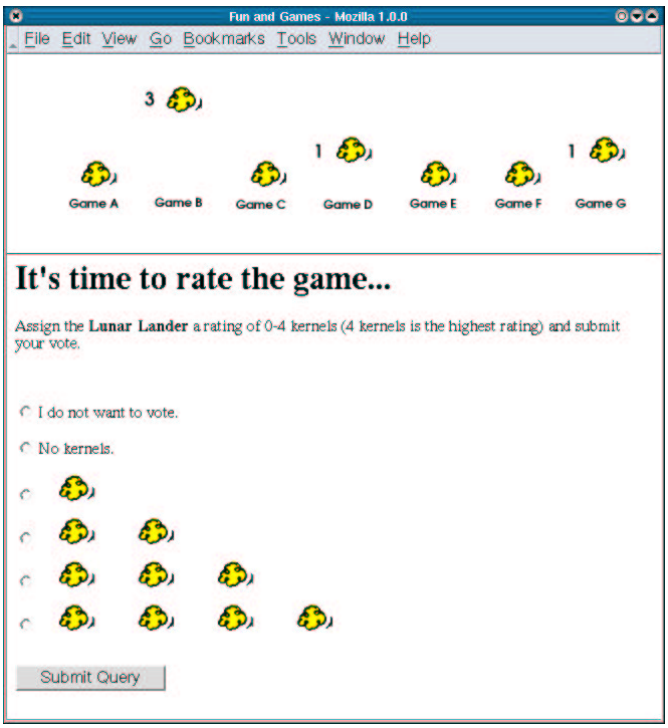


Figure 5.17: The Maple Grove post-game Rating Screen. Note active Popcorn Meter in the top frame.

Consistent with the PM-RB-CB design proposal, the post-game Rating Screen at Maple Grove displayed *Kernels* instead of *Happy Faces* (Figure 5.17). Kernel votes were recorded in the mySQL database and seen on every player's Popcorn Meter. After submitting a kernel rating, a player was shown an Add Comments Screen (Figure 5.18) similar to the one from the U Hill study, however, the M

[5]It was necessary to cut the number of Maple Grove games to seven, because a Popcorn Meter wider than seven kernels would not fit on the screens of the M Grove lab computers (given their size and resolution). *Box Builder*, *Hoop Shooter* and *3-D Vector Air Hockey* were cut from the Maple Grove system.

Grove screen told users: “If you like, you may send other players your thoughts on [the name of the recently played game]”.

Next, M Grove students saw the Ratings Board — a screen that featured a list of all kernel ratings and the usernames of those who had played/rated the game (Figure 5.19). The Comments Board (Figure 5.20) followed the Ratings Board and listed a table of usernames and users’ comments. Hyperlinks placed at the top and bottom of the table (in the event of a long list of comments, two links would reduce the amount of scrolling that might be required in order to find a single link) allowed players to return to the Palette Screen.

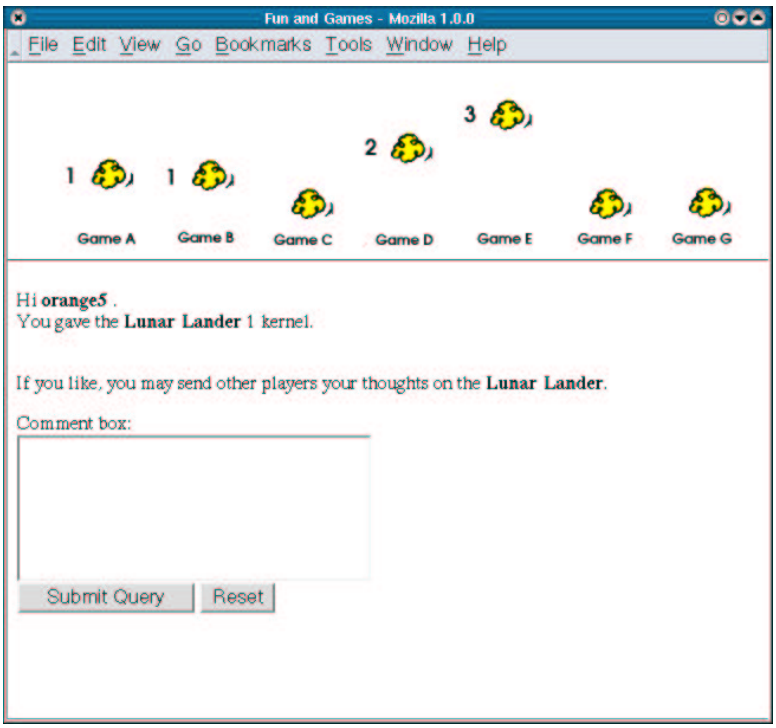


Figure 5.18: The Maple Grove Add Comments Screen.

As with the previous study, the Apache server on *faxe* logged every page request, the mySQL database stored every rating and comment submitted by the M Grove students, and two of the paired groups of students were videotaped while they used the system.

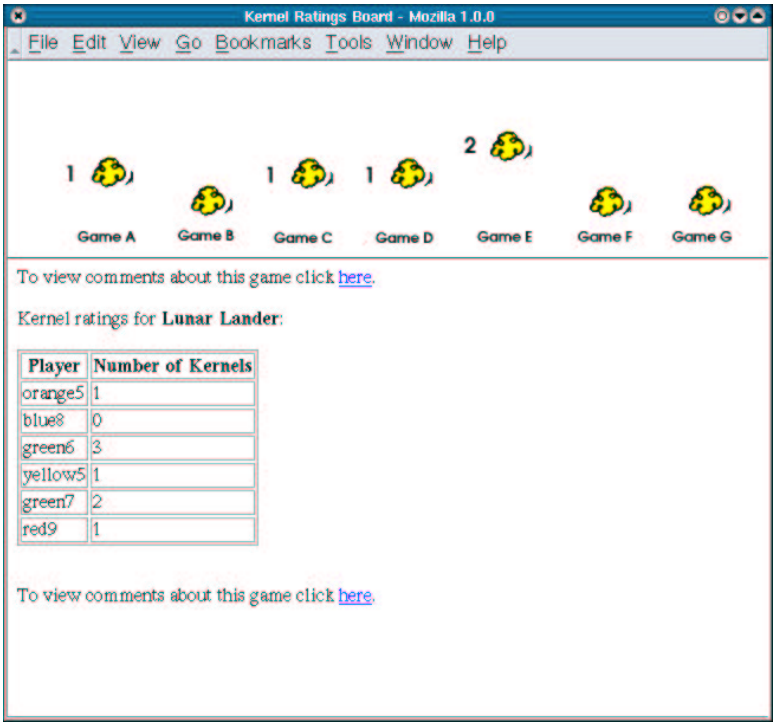


Figure 5.19: The Maple Grove Ratings Board. Ratings are listed beside the username of the player who submitted the rating.

Questionnaire

After 50 minutes in the computer lab, the students returned to their classroom and completed a two-page, written questionnaire (Appendix B contains a sample questionnaire). All students finished the questionnaire in under 15 minutes. There were some similarities to the U Hill questionnaire, but also some important differences. The questionnaire had eight sections of questions:

1. Of the games you just played, please indicate your top 3 games with a “1”, “2” and “3”. [A two-column list of the eight games followed this sentence.]

2. Think of one of your top 3 games. If you had to describe this game to some friends (who had not seen the game), what would you tell them? (Please circle this game in the list above.)

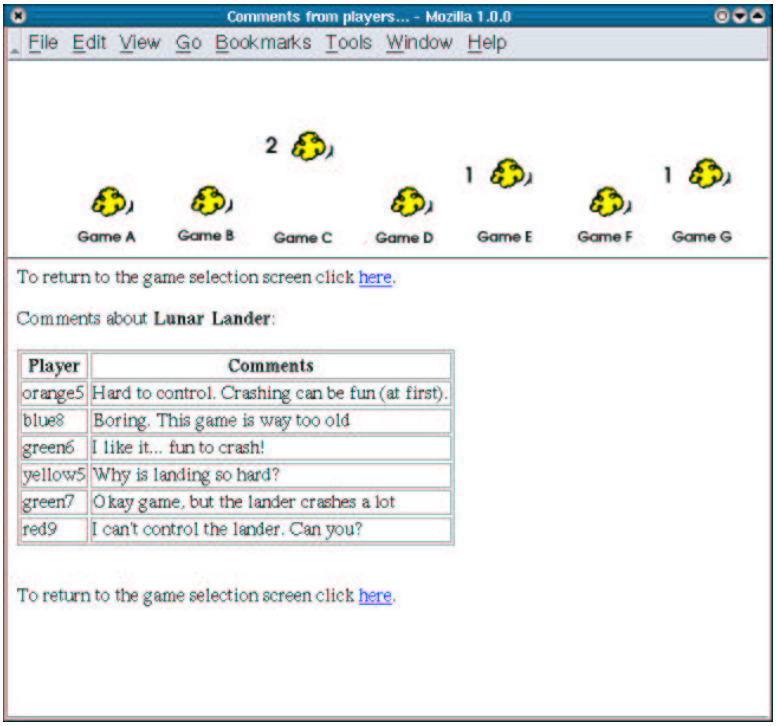


Figure 5.20: The Maple Grove Comments Board.

3. In this study, after you played a game, you had the chance to send ratings and comments to other players. Is this something you like?

4. Did you send many comments to other players? Why or why not?

5. Did you send many kernel ratings to other players? Why / why not?

6. What did you think of the Popcorn Meter?

7. The Web designers would like your thoughts on these games. For each game listed below, did you find the game “Cool” or was it “Yuk”?

8. Do you have and other comments for the Web designers?

As with the University Hill questionnaire, the first section was designed to link the written questionnaire to the activities in the computer lab. It was also an

attempt to see how students would rate games on a post-session written survey (versus an online web-form).

The questionnaire was intended to gather students' impressions of the system — in particular, what did they think of comments and ratings "to other players"? And, did they like the Popcorn Meter? Some of the questions (*e.g.*, sections 2, 7 and 8) were an attempt to solicit user feedback in a different way than what the system offered — a post-session, paper-and-pencil question on what they would "tell friends" about one of their favourite games, along with comments and ratings "for the Web designers".

Fortunately, the overwhelming majority of students (at U Hill and M Grove) took the time to carefully complete the written questionnaires. It should be noted that most of these students had participated in other E-GEMS studies prior to seeing the *Fun and Games* website. Also, for most of the study participants, it seems that 70 minutes with a set of computer games (even when a written questionnaire is involved) is definitely preferable to normal classroom activities. This may explain, in part, why most students indicated a willingness to play, rate and comment on the *Fun and Games* game activities. However, the web-log data, the comments and ratings databases, and the written questionnaires contain some interesting (in some cases, surprising) results. The next chapter outlines and debates some of these findings.

Chapter 6

User Studies: Results and Discussion

> *I sent a few comments because I thought I had to.*
> — blue4
>
> *I check the comments first before playing.*
> — green1
>
> *I didn't send comments because I don't like to waste time.*
> — red4
>
> *You can tell other players about your opinion and read theirs.*
> — green5
>
> *I like people to know my comments, and vote for games I like.*
> — blue5
>
> *I liked it because I can know what my friends think about my favourite game.*
> — blue4
>
> Selected Maple Grove students' answers to section 4 of the written questionnaire.

The user studies at University Hill and Maple Grove were preliminary studies that were, in a sense, "exploratory". The purpose of these studies was to see how Grade Six students (in a laboratory setting) would react to a "for designers" feedback-gathering system and a system built around a peer-to-peer model. These studies did not use a random sample of kids, nor do they allow for the use of inferential statistics to determine if differences in behaviour between the two groups

are statistically significant. There are simply too many discrepancies between the two studies for this sort of analysis.

The goal with these studies was to see the two systems "in action" — *i.e.*, to observe students' behaviour, and to look for tendencies and clusters of data that might suggest evidence of Clark-Schober effects or, in the case of the PM-RB-CB system, indicate that the system is cultivating a peer-to-peer, game-rating dialogue. Ultimately, the data from these studies would help determine whether or not the project's theoretical approach offers a plausible explanation to feedback-related problems at a "for designers" website.

Also, it was hoped that the study results might suggest whether or not a design based on the psycholinguistics research can alter user feedback by changing the model of interaction and shifting the focus of such feedback. In the event of this sort of evidence, the U Hill and M Grove studies would represent a first step in refining the design of the PM-RB-CB system and would "pave the way" for future research.

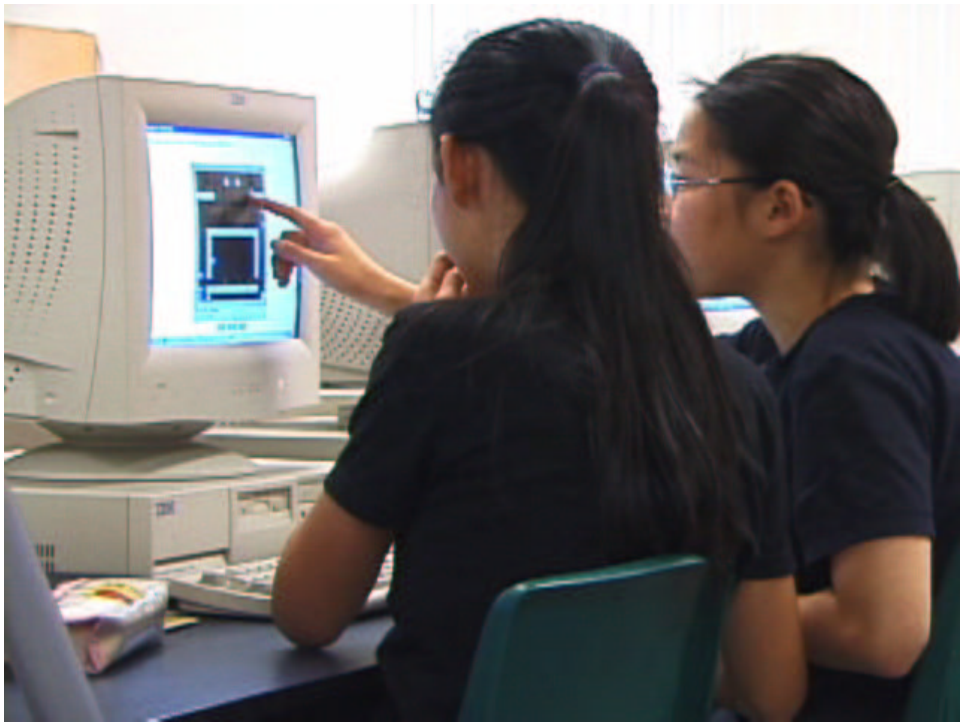


Figure 6.1: Two University Hill students discuss *Boulderdash* strategy.

Furthermore, the study data would be an early exploration of the central hypothesis: that a symbiotic relationship might exist between peer presence and interactive, collaborative assessment. However, before "big" issues like these could be settled, important "simple" questions needed to be answered. For example, would the user studies show that kids are willing to accept a peer-to-peer, online assessment model?

One conclusion that can be drawn (unequivocally) from the user study data is that the students at U Hill and M Grove had fun during the user studies. A great deal of data supports this conclusion:

- Most of the video footage shows kids laughing, pointing at the screen and talking about the games — usually discussing strategy or sharing game experiences. For example, the video stills seen in Figures 6.1 and 6.2.
- Comments by the students (in the mySQL database, on the post-session questionnaire and in verbal comments from kids to adults involved with the study) indicate they enjoyed themselves during their 50-minutes in the computer lab.
- Positive remarks and observations by teachers at U Hill and M Grove on students' enjoyment and their participation in the studies.
- The round of applause received at both schools at the conclusion of each user study.

Other evidence suggests the students enjoyed their user-study experience: the Apache server logs. The server running on *faxe* was shut down roughly one hour after the completion of each user study. The study at University Hill was conducted in the morning of October 18, 2002, but the Maple Grove study was held in the afternoon of November 7 — the written questionnaire was given to students during the first 15 minutes of their final period of the day. Although the Apache server would only grant access to computers in the *.ubc.ca* and *.vsb.bc.ca* domains, the server logs show that several Vancouver-area computers attempted to connect to the system after 3:15 PM on November 7. The URL for the M Grove study (*http://faxe.cs.ubc.ca:2000*) was unusual and used a non-standard port number (2000). It seems unlikely that anyone other than participants (who were given the URL prior to logging in) would have requested this address.[1] In fact, it is probable that students surfed to the Login Page when they got home from school.

[1]Web-bots are a possible explanation for the Login Page requests, however this is not likely. The system was up for only a few hours at a time and a URL with a port number of 2000 is unlikely

Figure 6.2: *Concentration* at University Hill.

6.1 University Hill — Ratings and Comments

Generally, the results at University Hill correspond to the findings at NFBkids. The U Hill students submitted 137 ratings and 131 comments (*i.e.,* collectively they clicked on *SUBMIT QUERY* 137 times at the Rating Screen and 131 times at the "Comments for Designers" Screen.). As far as the *Happy Face* ratings are concerned, 30 of the "submissions" were not submissions at all — users simply clicked *SUBMIT QUERY* without clicking on any of the radio buttons. This means roughly 23% of the time, students skipped the rating process.

A larger percentage of students skipped the "add comments" process, as 63 of the comments were non-submissions. Therefore, roughly 46% of the Comment Boxes were empty when U Hill players clicked to go on to the next screen. Of the 68 Comment Box submissions that did contain data (*i.e., any* sort of text at

to attract web-bots. Also, it is improbable that all web-bot requests would be from machines located in Vancouver.

all), 16 were classified as gibberish — leaving 52 of 131 (approximately 40%) of players' comments to be catagorized as "valid" comments.

The U Hill students did have a higher response rate on comments and ratings than NFBkids users, yet rating and commenting on games was a larger part of the U Hill system (*e.g.*, every game had its own Submit Rating page and Add Comments page, and all submissions were echoed to the user). Moreover, unlike NFBkids users, the University Hill students knew they were part of a study and received a short presentation telling them the Web designers wanted their comments and opinions on the games.

The usage of ratings and comments by University Hill students may be broken down so that "clusters" of users appear — *i.e.*, so that groups of students with similar behaviours are isolated from the data. In Table 6.1 the "use of ratings" is seen in terms of the percentage of the time a registered username would be associated with a rating. For example, seven "users" always rated a game after they played it[2] — *i.e.*, 100% of the time they clicked on *SUBMIT QUERY*, they had also selected a *Happy Face* rating or had selected "I do not want to vote". A single username entered a rating 91.7% of the time after playing games, and two usernames entered ratings between 80% and 90% of the time: one username had a valid rating response rate of 81.8% and the other entered valid ratings at a rate of 80.0%.

The largest "cluster" in the Table 6.1 data is at 100%. This cluster may be extended to include 12 usernames (80% of the usernames in the U Hill computer lab) by saying that 12 usernames rated games at a rate of 75% or higher. This represents what might be considered a high response rate when it comes to *Happy Face* ratings. The response rate is not so high for the Comment Box on the Comments To Designers page (Table 6.2).

In terms of University Hill students' use of the Comment Box, there is still a group of students (four of them) at 100% — they entered text into the Comment Box after every game they played. However, the other clusters are at much lower response rates with five usernames clustered between response rates of 45% and 67%, two users at 23.1% and four usernames at 0% (these four usernames supplied no comments whatsoever during the 50-minute lab session.).

[2] "Users" appears in quotation marks because most usernames were shared by a pair of U Hill students.

Valid Rating Response Rate			Number of *usernames*
100%			7
90%	0.917		1
80%	0.818	0.800	2
70%	0.769	0.750	2
60%			0
50%	0.500		1
40%	0.464		1
30%			0
20%			0
10%			0
0%			1

Table 6.1: University Hill students' use of ratings listed in terms of the percentage of the time a *username* submitted a valid rating after playing a game. (N=15)

Comment Box Response Rate				Number of *usernames*
100%				4
90%				0
80%				0
70%				0
60%	0.667			1
50%	0.500			1
40%	0.455	0.429	0.400	3
30%				0
20%	0.231	0.231		2
10%				0
0%				4

Table 6.2: University Hill students' use of the Comment Box (including gibberish) in terms of the percentage of the time a *username* would submit a comment after playing a game. (N=15)

Comments for Section 3	Number of Students
Only Positive Comments:	18
Positive and Negative Comments:	6
Only Negative Comments:	4

Table 6.3: University Hill students' reaction to sending comments and ratings to designers — based on their answers to Section 3 of the written questionnaire. (N=28)

6.2 University Hill — Written Questionnaire

The results from the written questionnaire hint at why some U Hill students might have not "bothered" to submit post-game ratings and comments. Of course, in a time-limited study, it is understandable that students would rather play games than type text into Comment Boxes. However, the questionnaire results suggest there may be additional reasons for the poor performance of the Comment Boxes.

For the most part, the University Hill students had positive things to say about rating games and sending comments to designers, but an important segment of the U Hill Grade Six class indicated a few misgivings when it came to the process of rating games and sending comments to the Web designers.

Sections 3-6 of the questionnaire were open-ended (an attempt to avoid prejudicing the results), so categorizing responses required a somewhat broad set of criteria. Accordingly, based on their answers to Section 3, students' reactions to "sending comments and ratings to designers" were classified into three categories, shown Table 6.3. Four students openly disliked the process of rating and commenting on games, six were ambivalent, and 18 basically said "Yes" for Section 3.

Answers to Section 4 (see Table 6.4) offer several explanations for students' like or dislike of the rating process. Fifteen students said "yes" to the question in Section 4 ("Do you think the game designers will pay attention to the comments you entered today?"), but the other 13 students were unsure that their comments mattered to the designers (and, in some cases, they were convinced the designers would not pay attention to their comments). It is interesting to examine students' answers to Section 4 in combination with their answers to Section 3. Representa-

tive examples follow:

- username: *blue3*
  Section 4 answer: *Yes, I think they will pay atention to them because they want to be good designers, Don't they?* [sic]
  Section 3 answer: *Yes, I'm pleased, but also down because I told the truth of how I feel and some were pretty borring* [sic]

- username: *orange3*
  Section 4 answer: *No. I don't think so!!!*
  Section 3 answer: *No!!! I don't like sending comment!!!* [sic]

- username: *orange2*
  Section 4 answer: *Yes because I said good things about there games* [sic]
  Section 3 answer: *I think this was fun for research.*

- username: *red2*
  Section 4 answer: *Yes, because they might want to see how the game did.*
  Section 3 answer: *Yes, because it's good to feel good about what you design.*

- username: *red5*
  Section 4 answer: *They might.*
  Section 3 answer: *Yes, because then they can improve the game, if needed.*

- username: *blue2*
  Section 4 answer: *No, because there were too many people that entered today, and the game designers have no time for that.*
  Section 3 answer: *yes*

- username: *yellow3*
  Section 4 answer: *maybe. It's not in my control, so I not sure* [sic]
  Section 3 answer: *yes, because they know how i feel about the game* [sic]

- username: *orange4*
  Section 4 answer: *Yes and No. There can be a lot of game comments. Maybe he/she will read it to develope games* [sic]
  Section 3 answer: *Yes, I didn't get to write it but I think it will help game to be better* [sic].

- username: *blue2*
  Section 4 answer: *No I don't think they will pay attention.*
  Section 3 answer: *Yes.*

- username: *yellow1*
  Section 4 answer: *Hopefully, but I don't expect them to change the whole game. It's okay as it is.*
  Section 3 answer: *I didn't really want to give comments but I thought of a few things they could might change. I would mind if the designers read some of them* [sic]

- username: *orange5*
  Section 4 answer: *Maybe, but not very much. Maybe they need feedback.*
  Section 3 answer: *Yes. because the good games I wanted the designers to know that they create great games for kids to sample!* [sic]

- username: *orange5*
  Section 4 answer: *they Better pay attention!*
  Section 3 answer: *yes. caus then the designers can make the game better what good for them and better for use that we playing and enjoying from it* [sic]

The suspicious nature of many students' answers to Section 4 may be an indication that the system is "suffering" from some of the intentions-related pitfalls associated with questionnaire design discussed by Clark and Schober [3], Sudman, Bradburn and Schwarz [24], and Tourangeau, Rips and Rasinski [26]. Certainly, some of the comments above raise questions related to common ground, perspective, common purpose and pressure to respond. For example, user *blue3*

Section 4 Answers	Number of Students
Yes:	15
Maybe / Yes & No / Not Much:	3
No:	6
Don't Know:	4

Table 6.4: University Hill students' answers to Section 4 of the questionnaire — "Do you think the game designers will pay attention to the comments you entered today?" (N=28)

openly voices concerns that harsh user comments might offend the designers — the "I told the truth [...] and some were pretty borring" line on *blue3*'s written questionnaire and this comment submitted by *blue3* during the lab session: "hi im sorry to say this but i thought this game was boring and pointless :(" [sic].

Just as *blue3* writes "im sorry to say this" when criticizing a game, other U Hill participants appear to suggest a relationship between the tone of their comments and how "happy" the designers will be to receive them (*e.g., orange2*'s answer to Section 4: "Yes because I said good things about there games", *red2*'s insinuation with "it's good to feel good about what you design" or *orange5's* focus on "great games"). This appears to be evidence that the way in which users perceive the intentions (and feelings) of designers affects the comments users put into the Comment Box.

When asked why game designers want to read user comments and see how users rate games (Section 6), many U Hill students had similar thoughts — and their remarks centred on the concept of improving games based on user feedback:

- *So they can improve the games and see which ones were the most popular.* (blue2)
- *They want to see what kind of game is wanted these day.* [sic] (yellow1)
- *Because it give the designers some ideas and make the game easier. And funner.* [sic] (yellow1)
- *I think that the game designers will whant to read my comments about the*

game so they can improve it amd make it better so the kids we can enjoy the game. [sic] (orange5)

- *So that they can make the game more intersting and cool and fun* [sic] (blue5)
- *so maybe they can improve the game according to our comments.* (yellow3)

However, even though most students feel it is a "good thing" designers want their feedback and they see a role for users in the process of improving games, the answers to Section 4 of the written questionnaire indicate that nearly half of U Hill participants wonder if the designers will "pay attention" to their comments. As seen in the answers above, their reasons for feeling this way range from a belief that "too many" user comments mean designers cannot pay attention to all submissions (*e.g., blue2* and *orange4*) to an agnostic approach to understanding designer motivations (*yellow3*).

Ultimately, the comments of the University Hill students do provide "designers" with feedback that could be used to shape the future design of the system. There are clusters of students who prefer certain games and even the very-limited Comment Box submissions contain information on what the kids like and dislike about some of the games. There is evidence of Clark-Schober effects in the way U Hill students appear to approach the process of rating and commenting, however, it could be argued that intentions-related problems are common to most questionnaires, particularly those that are web-based.

The question that remains: is it possible to shift the focus of game-rating with a peer-to-peer method of game assessment so that users' mental model of the system (especially as far as rating games is concerned) is more user-centred.[3] Furthermore, will a user-centred and/or kid-centred method of game-rating result in *different* feedback — *i.e.,* will comments and ratings intended for users differ from those intended for designers?

The theory behind the design of the PM-RB-CB system (as interpreted) suggests that intentions-related pressures are different for user-to-user (peer-to-peer) communication than for user-to-designer communication. Therefore, will the PM-RB-CB system succeed in providing the context for a peer-to-peer dialogue to take place? And, will such a dialogue (if it occurs) result in a fundamentally *different* set of ratings and Comment Box submissions?

[3] And, in terms of this system, more kid-centred.

Figure 6.3: A pair of Maple Grove students look at the post-game Rating Screen for *Javanoid*.

6.3 Maple Grove — Ratings and Comments

All in all, the Maple Grove students enjoyed rating games and sending comments (see Figures 6.3, 6.4 and 6.5 for video stills of M Grove students in front of the rating-related and comment-related screens). In just under 50 minutes, they sent 154 ratings and submitted 147 comments.[4] In terms of Kernel Ratings, only 16 were "non-ratings". This means 138 Kernel Votes were cast during the M Grove user study — which translates to a surprisingly high response rate of 90% for submissions on the post-game Rating Screen. If one user, *orange2*, a rare case who submitted Kernel Ratings infrequently, is removed from the data, the Rating Screen response rate rises to 134/141, or 95%.

[4] As with the University Hill study, these figures include "empty" submissions. Collectively, Maple Grove students clicked on *SUBMIT QUERY* 154 times on the post-game Rating Screen and 147 times on the Add Comments Screen.

Among the comments submitted by M Grove students, 55 were empty non-submissions. Therefore, roughly 63% of Maple Grove "comments to other players" contained text. Removing the eight submissions considered to be gibberish leaves 84 "valid" comments for 147 clicks on the *SUBMIT QUERY* button — this means just slightly better than 57% of submissions were classified as "valid comments". If the "rare case" *orange2* user is removed from the data (this user did not like to submit comments, either) the text-to-submissions ratio rises to 90/133 (68%) and the valid-text-to-submissions ratio becomes 83/133 or 62.4%.

Figure 6.4: Maple Grove students look over the *Javanoid* Ratings Board.

When the Kernel Rating data is broken down into the "percentage of the time a username submits a rating" analysis used to study the University Hill *Happy Face* ratings, one major cluster appears (Table 6.5). Twelve usernames (out of 17 — roughly 71%) submitted ratings on every game they played (*i.e.,* a radio button was clicked before *SUBMIT QUERY* was clicked on every Add Comments page). If the three usernames who sent ratings between 80% and 95% of the time are grouped with the twelve who always rated games, the response rate for

Kernel Ratings at Maple Grove may stated as follows: 88% of M Grove usernames submitted valid ratings at least 83% of the time (after playing a game).

Figure 6.5: Two Maple Grove students examine a Comments Board.

The response rate for sending comments was also high (Table 6.6). There are two clusters in the data: one group of seven who sent comments 100% of the time after they had played a game, and a second group of three usernames who submitted comments between 14% and 17% of the time. The rest of the usernames are distributed between a response rate of 38.5% and 92.3%. One noteworthy result: there are no users with a comment response rate less than 14.2%. This is surprising — students who refuse to submit comments were expected. For example, at University Hill, four usernames (27% of usernames) had a comment response rate of 0% — they submitted no comments during the entire study.

Valid Rating Response Rate			Number of *usernames*
100%			12
90%	0.910	0.900	2
80%	0.833		1
70%			0
60%	0.625		1
50%			0
40%			0
30%	0.310		1
20%			0
10%			0
0%			0

Table 6.5: Maple Grove students' use of ratings listed in terms of the percentage of the time a *username* submitted a valid rating after playing a game. (N=17)

Comment Box Response Rate				Number of *usernames*
100%				7
90%	0.923			1
80%	0.875			1
70%	0.750			1
60%	0.690			1
50%	0.500			1
40%	0.455			1
30%	0.385			1
20%				0
10%	0.167	0.154	0.142	3
0%				0

Table 6.6: Maple Grove students' use of the Comment Box (including gibberish) in terms of the percentage of the time a *username* would submit a comment after playing a game. (N=17)

6.4 Maple Grove — Written Questionnaire

The results from the post-session questionnaire at Maple Grove show that students enjoyed their time with the system. Generally, students' answers to Section 2 were enthusiastic and informative — most students found a game (or games) they liked and wrote impassioned comments about "what made the game fun". Also, many students wrote enthusiastic comments when it came to Sections 3-6 on the Popcorn Meter and sending comments to other players.

As with the U Hill survey, most sections on the M Grove questionnaire consisted of open-ended questions so that students' answers might be "in their own words" as much as possible. In order to categorize responses to questions like *"What did you think of the Popcorn Meter?"*, broad criteria were sometimes necessary. The goal was to capture common ideas and group them. Accordingly, reaction to the Popcorn Meter was classified into the following groups:

- I liked it / I chose games based on it / I shared ideas with it
- Fun / Cute / Funny / Cool
- I did not like it / I found it annoying / it got in the way
- I did not see it
- "no comment"

Table 6.7 lists the grouped results for Section 6 of the Maple Grove questionnaire. Twenty students (74% of students in the M Grove study) wrote positive comments when asked about the Popcorn Meter and five students (18.5%) wrote negative statements. One student wrote "no comment" and another wrote "I didn't see it". Here is a sample of users' comments on the Popcorn Meter:

- *Cool. It was actually funny.* (red5)
- *Got in the way of the games.* (blue3)
- *I think it's interesting and cool.* (yellow3)
- *It was a smart idea. I like it a lot.* (blue4)
- *I think the popcorn meter was very good for rating.* (green3)

Comments for Section 6	Number of Students
I liked it / chose games based on it / shared ideas with it:	11
Fun / cute / funny / cool:	9
Did not like it / found it annoying:	5
Did not see it:	1
"No comment":	1

Table 6.7: Maple Grove students' reaction to the Popcorn Meter — based on their answers to Section 6 of the written questionnaire. (N=27)

- *It was pretty accurate. Like I submitted 1 kernel and 1 kernel instantly appeared (from 0 kernels). I looked at the meter to choose games.* (blue5)
- *The popcorn metre was a bit of a guide to me.* [sic] (green5)
- *Oh... it is really cute.* (orange1)
- *It okay. It sort of persuade me to choose the game that had more points.* [sic] (green4)
- *I don't like it. It annoyed me.* (blue2)
- *Sometimes I don't like it, because they're moving. I can't play the game.* (green4)

It was anticipated that some students might find the Popcorn Meter "annoying". The Popcorn Meter is a large, active Java applet that occupies a significant amount of "real estate" on the screen.[5] What is surprising is the relatively low

[5]The computers at Maple Grove were equipped with 17-inch monitors, but at a resolution of 800 x 600 several game screen webpages were "too tall" for the M Grove monitors. This required students to scroll the main frame (the frame in which the game activity was located) slightly in order to be able to display the entire game — game instructions often occupied the top portion of this frame. It was thought this might be a major "annoyance" for M Grove kids. However, during the study, students scrolled up and down to read game instructions, *etc.* and they did not seem to be bothered by scrolling.

percentage of users who did not like the Popcorn Meter (again, 18.5%) and the considerable number of students who used language like "cool" or "cute" to describe the Popcorn Meter: 9 students (33%).

The fact that the rating tool itself could become a *cool* part of the system was somewhat unexpected. Of course, one of the design goals was to build a rating system that kids would accept and use, but frequent comments by students referring to the rating instrument as "cute" and "fun" came as a surprise. It seems possible that the language used by many students to describe their feelings on the Popcorn Meter may reflect their enthusiasm not just for the Popcorn Meter, but for the entire rating system. At any rate, their comments focus on the Popcorn Meter (and not, for example, the Comments Board) and it seems as though the Popcorn Meter was a bit of a "hit".[6]

Section 3 of the Maple Grove questionnaire asked students for their thoughts on sending "ratings and comments to other players". The students' answers in this section were varied — many students simply said "yes" they "liked sending comments" because they "wanted to" or found it "fun". Others wrote they "did not like" sending comments, often adding there was "not enough time" to send comments or that they "would rather play games". And, there was a group of students who wrote about the *value* of sending comments — that it allowed them to tell others "what I think" about games. Therefore, the answers to Section 3 fell into three categories:

- Students who wrote they could "express their feelings", "share thoughts" or "give preferences" by sending comments to others.
- Students who simply wrote they "liked" sending comments.
- Students who wrote they "did not like" sending comments.

The Section 3 results are listed in Table 6.8. Nineteen students (70% of the M Grove participants) wrote positive statements regarding comments to other players, and eight (30%) wrote negative statements. Eight of the students who "liked" the ability to send comments added that submitting comments allowed them to communicate their thoughts to other players — and this was a reason they enjoyed submitting comments.

[6]For example, the most flattering comment to describe the PM-RB-CB system without referring to the Popcorn Meter is this written statement by *green1*: "the comment system is very neat and well-organized". Only references to the Popcorn Meter contain colourful adjectives like "fun", "cute" or "cool".

Comments for Section 3	Number of Students
I like it and I can express my feelings / share my thoughts:	8
I like it:	11
I did not like it:	8

Table 6.8: Maple Grove students' reaction to "sending comments to other players" — based on their answers to Section 3 of the written questionnaire. (N=27)

Here are some of the students' responses to Sections 3 and 4 of the Maple Grove written questionnaire. Section 3 simply asked students if they liked having the opportunity to send ratings and comments to other players, and Section 4 asked "Did you send many comments to other players? Why or why not?"

- username: *green5*
  Section 3 answer: *yes you can tell other players about your opinion and read theirs* [sic]
  Section 4 answer: *I did, because I wanted to see if anybody else liked the game* [sic]

- username: *red4*
  Section 3 answer: *Yyep! It gives you a chance to rate the game* [sic]
  Section 4 answer: *I sent them alot of comments* [sic]

- username: *red2*
  Section 3 answer: *Yes. Some of the comments were funny.*
  Section 4 answer: *no. I sent one because I really liked the games.*

- username: *blue3*
  Section 3 answer: *Yes, you can express your feelings.*
  Section 4 answer: *Yes, because to say what I think.* [sic]

- username: *blue2*
  Section 3 answer: *Yes.*
  Section 4 answer: *No because I was to focused on the game.* [sic]

- username: *blue5*
  Section 3 answer: *Yes. So you can easily choose what games to play first.*
  Section 4 answer: *Yes, I like sending messages.*

- username: *green1*
  Section 3 answer: *Yes. I check the comments first before playing.*
  Section 4 answer: *Yes, I did, it is because I want them to know what I think of the game so they could rate it well* [sic]

- username: *green1*
  Section 3 answer: *The rating system is good but the popcorn thing is too unsophisticated. The comment system is very neat and well-organized.*
  Section 4 answer: *I sent 4 comments, 3 of them jumbled letters. I send comments when have tons of time, not when have less than 1 hour.* [sic]

- username: *green4*
  Section 3 answer: *Yes. I could express how I feel about the game. I also like to see what other felt about it too.* [sic]
  Section 4 answer: *Not really, there wasn't much to say.*

Not surprisingly, the students at Maple Grove found rating games to be more "fun" than typing comments into Comment Boxes. One reason for this seems to be the amount of time and effort required to rate a game versus the time and effort associated with the Comment Box.[7] Also, kernel ratings offered real-time feed-

[7]This was a concern when designing the system — using the keyboard to enter text can be a slow process, especially for Grade Six kids. Also, coming up with "something to say" takes time. Prior to the user studies it was felt that many users would probably not be willing to spend the time to "think about and type" comments. As seen in their questionnaire responses, many of the kids did find the process of entering comments time consuming. Surprisingly, only a minority of Maple Grove students felt this way.

back through the Popcorn Meter — feedback from Comment Box submissions was not as immediate.

As far as the amount of time required to submit comments and ratings is concerned, the submission of a kernel rating took seconds and required only a couple of mouse clicks. Web logs show that students who entered text into comment boxes spent between 10 and 100 seconds at the Add Comments screen before clicking on *SUBMIT QUERY*. Therefore, in order to enter comments into the Comment Box, students' desire to "express their feelings" and contribute to the discussion on the Comments Board had to outweigh the costs associated with spending time away from the games. An interesting result is that 10 usernames out of 17 (59%) took the time to enter text into the Comment Box on at least 75% of the games they played.

The post-session questionnaire sheds light on the clusters seen in Tables 6.5 and 6.6. The positive statements made by students in Section 6 ("What do you think of the Popcorn Meter?") echo the results seen in Table 6.5: the overwhelming majority of Maple Grove students liked the Popcorn Meter and used it (regularly) as a game-rating tool.

Students' responses to Section 3 and Section 4 of the questionnaire show that roughly 70% of participants claimed they "liked" to submit comments and enjoyed reading the Comments Board. During the study, these students experimented with the Comment Box to varying degrees — with most students sending comments after they played a game.

Those who did not use the Comment Box complained that comments simply "took too long", ate up time they would rather spend with the games, or that they simply did not have anything to say. This information may help to explain the clusters in Table 6.6 as a small minority of users (3 of 17 usernames) used the Comment Box less than 20% of the time. Two of these users (*red5 and orange1*) submitted kernel ratings frequently and one username (the aforementioned *orange2*) seemed uninterested in sending ratings or comments.[8] Here are some of the questionnaire results for students associated with these usernames:

- username: *orange1*

  Section 3 (Do you like sending ratings and comments to other players): *Yes!! because you get to tell how good the game was and see which game was the best.* [sic]

[8] *red5* faced the kernel-voting screen 12 times and submitted 10 ratings (for a kernel voting ratio of 83%). *orange1* saw the post-game Rating Screen six times and submitted six kernel votes (100%). *orange2* visited the Rating Screen 13 times and submitted four kernel votes (31%).

Section 4 (Did you send many comments to other players?): *No, because I just want to keep playing then sent comments, I cannot stop!!* [sic]

Section 5 (Did you send many kernel ratings to other players?): *Yes to see if I could see which game is the best.* [sic]

Section 6 (What did you think of the Popcorn Meter?): *Oh... it is really cute.*

- username: *red5*

  Section 3: *Yes. Because I'll know if everyone have same thoughts.* [sic]

  Section 4: *No. because I was lazy.* [sic]

  Section 5: *No. I was lazy.*

  Section 6: *Cool. It was actually funny.*

- username: *orange2*

  Section 3: *Yes, because I can saw the other studens' comments and I can sent my comments.* [sic]

  Section 4: *A few, because I didn't have time.*

  Section 5: *No, because I didn't have time*

  Section 6: *I hate that tings.* [sic]

A few of the questionnaire responses (*e.g.*, user *green1* who wrote "I check the comments first before playing") suggest the design of the system might be improved. The system shown to Maple Grove students made sending and reading comments strictly a post-game activity. Perhaps the design of the Comments Board and the Add Comments Screen could be changed — for students like *green1* who want to see the comments before they play a particular game or before they enter their comments.

Allowing players to read the Comments Board (along with the Ratings Board) before, after or during a game seems to be a good idea and it is supported by the psycholinguistics research. If the interactions among users of the Ratings and Comments Boards may be considered to be a dialogue or conversation, then it is perfectly "natural", as far as human conversation is concerned, that some participants would prefer to respond rather than initiate the conversation. Moreover, a

"flashier" set of Ratings and Comments Boards (perhaps a Java applet that expands when players want to submit or read comments, and shrinks when they want to focus on the primary game activity) might further entice players to read and send comments. Also, it might make the process of sending comments more of a "dialogue". This type of "chat" tool could be used instead of the Comments Board or in tandem with the existing Comment Box and Comments Board. At any rate, a system that allows players to bring up the "Boards" at any time would make for an interesting comparison with the Maple Grove PM-RB-CB system.

6.5 U Hill and M Grove — General Discussion

Based on their comments from the post-session questionnaire, it seems as though the students at University Hill "*know what's going on*" as far as the process of rating games is concerned. The majority of University Hill participants assumed the web designers want user feedback in order to "find out what kids like" and "improve the games according to our comments".[9]

However, even though most U Hill participants seem to understand their role in the process of improving the *Fun and Games* website (to provide the designers with valuable feedback), many of them were reluctant to submit comments at the post-game Add Comments screen. Clearly, these students were not motivated to provide meaningful comments — for them, sending comments is "boring", it takes "too long" and/or they would "rather play games". Other reasons behind students' unwillingness to enter comments seem to be associated with the intentions-related issues discussed by Clark and Schober: for example, the fact many U Hill students were unsure whether or not the designers would pay attention to their comments.

Of course, the PM-RB-CB system shown to Maple Grove students was designed to encourage users to submit comments and ratings, and to change the focus of the interaction from user-to-designer to peer-to-peer. However, PM-RB-CB was not simply about *more* submissions from users nor was it strictly concerned with promoting a happy community of kids — if the goals had been only to foster dialogue and a sense of community among users, then a flashy, attractive chat feature could have been added to the system. The fear was that such a chat tool (on its own without the Popcorn Meter to frame the discussion) might motivate students to chat, but it would not focus the dialogue on the games. With a "fun" chat tool, users might have a great deal to say, but the topic of conversation could stray

[9]This observation was made by University Hill user *yellow3*.

unpredictably. Ann Brown *et al.* discuss a similar problem in their analysis of the "ethos of the classroom" by examining the role an experienced teacher plays in shaping classroom interactions [2]. According to the situated cognition work of Brown *et al.*, an expert teacher does not let students discuss anything they want. Instead, the teacher narrows classroom discussion. Similarly, one of the goals of the PM-RB-CB system was to act as a bridge between users in order to foster a sense of community, but also to limit user interaction so the focus of Comment Box submissions would be on the game activities.

There is evidence that PM-RB-CB successfully motivated Maple Grove students to rate and comment on the games. Generally, M Grove participants said and wrote "good things" about the Popcorn Meter and the process of rating and commenting on game activities. Also, a large cluster of M Grove students regularly submitted comments and ratings. Furthermore, the presence of the Popcorn Meter seems to have *"planted a seed"* in the minds of many Maple Grove participants — it got them thinking about rating games even before they had encountered their first game activity. Instead of a strictly post-game rating system, like the traditional web-form at University Hill or the NFBkids *SURVEYS*, the Popcorn Meter may have made game-rating "easier" and "more thoughtful" by priming users — inducing them to consider the merits and drawbacks of activities before and during gameplay. A future study could examine this possibility more closely.

One of the promising results from the Maple Grove study is the large percentage of M Grove students who wrote that they enjoyed sending kernel ratings and comments because it allowed them to *communicate* with other players. Based on questionnaire responses from kids who described the Popcorn Meter, Ratings Board and Comments Board as a way to "express my feelings", "show others that I like a game", "see how others like my favourite game", or as a guide "to choose games", it seems as though the PM-RB-CM system did provide a platform for peer-to-peer interaction and it did shift the focus of rating games. According to Clark and Schober (and others), this change in focus is likely to result in *different* user feedback than the standard "for designers" techniques, as the context for a peer-to-peer dialogue is associated with different pressures than user-to-designer communication.

Moreover, there is evidence that the PM-RB-CB design is a step towards a system that meets two of the NFB's goals with NFBkids: a website and survey method that foster community presence and gather "meaningful and useful" user feedback. More study is necessary, but the Maple Grove data suggests that a symbiotic relationship may exist between peer presence and interactive, collaborative assessment. Students who felt a *connection* with their peers through the Popcorn

Meter or the Ratings and Comments Boards were more inclined to rate games and submit comments.

Certainly, the vast majority of students at Maple Grove were not "against" the Popcorn Meter. Early concerns that users might not accept peer-to-peer rating were not confirmed by the M Grove data. In fact, the evidence indicates that Maple Grove users enjoyed the process of sending, receiving and viewing kernel ratings and comments. Unexpectedly, many of the Maple Grove students indicated a special fondness for the Popcorn Meter — the M Grove study uncovered the fact that the rating instrument itself could become "cool" or "cute", thereby engendering a sense of entitlement because of its "coolness".

Overall, the preliminary user studies at University Hill and Maple Grove suggest that the psycholinguistics research of Clark and Schober, Tourangeau, Rips and Rasinski (and others) holds valuable lessons for the design and implementation of web-based methods that aim to solicit user feedback. In a sense, the work of Clark and Schober "predicted" the results at NFBkids, University Hill and Maple Grove. The standard techniques (*e.g.*, the *Tell Us Or Tell Us Off* web-form) often fail because of the effects outlined by Clark and Schober. Adding tricks or animations that are orthogonal to the purpose of the interaction (the animated GIF "mouths") does not improve the performance of such methods.

To the extent that PM-RB-CB *"succeeds"* as an alternative technique for gathering user feedback, it relies on research into psycholinguistics and Human-Computer Interaction — in order to align tasks and goals ("popping" kernels and rating games) and to align motivations and intentions (the interests of users in rating games for each other, along with those of the designers). As mentioned, most of the Maple Grove students "liked" sending ratings and comments to each other, and many of them found the Popcorn Meter to be "cool". A more comprehensive set of user studies will help identify exactly why the Popcorn Meter is *cool* and how PM-RB-CB might be improved, but the data from the U Hill and M Grove studies indicate that PM-RB-CB may have successfully integrated important elements of theory into its design. Future work along with an agenda for research will determine whether or not PM-RB-CB is a successful design, but at this point, the system (especially the Popcorn Meter) appears to be a promising first step.

Chapter 7

Future Work / Research Agenda

> *I think it's weird, I suggest use the heads in COBS.* [sic]
> — red4
>
> *PRETTY! ... and Kewl! But make them bunnies...*
> — green5
>
> *I think there should be more numbers.*
> — red2
>
> *Kernel ratings are unsophisticated compared to star ones, so I rated only 2-3.*
> — green1
>
> *Um, Why is it popcorn? Can't they make a meter?*
> — green2
>
> Students at Maple Grove offer a few suggestions on the Popcorn Meter.

The PM-RB-CB project is a preliminary step in exploring real-time, peer-to-peer assessment techniques as a way of gathering user feedback and encouraging a sense of peer presence among a community of web users. With promising results from the user studies at University Hill and Maple Grove, the next step is to show the system to more kids, and observe.

Specifically, what is needed is a set of more comprehensive user studies that use random samples of participants so that inferential statistics may be used to investigate and analyze observed differences. For example, Maple Grove kids liked the Popcorn Meter and they sent more comments and ratings than University Hill

students. Does this mean the PM-RB-CB system "outperformed" the comments-to-designers system seen at U Hill? Or, were M Grove kids simply "chattier" than U Hill students on a particular day? There are too many differences between the two studies in order to answer such questions. However, a more thorough experiment would be able to address questions like these.

Also, the results from University Hill and Maple Grove have uncovered some interesting questions for future study. One interesting result from the questionnaires at Maple Grove involves the comments of several students when asked *"What do you think of the Popcorn Meter?"* A few students answered this question by saying they like the Popcorn Meter, but that they want their "own kernel". Although the meaning of such a statement is open to interpretation, it could be that these comments are similar to the suggestions (quotations) that begin this chapter. These students might be asking for a system that allows them to *individualize* or *personalize* the Popcorn Meter.

A future experiment could test the viability of a personalized rating scheme where every username is associated with a unique icon on the real-time meter. One way of doing this would be to use coloured popcorn on the existing Popcorn Meter. For example, when *red2* rates a game, the kernel associated with that game could turn red (and be branded with a number *"2"*) to indicate which user submitted the rating. Another implementation might allow students to select from a wide variety of icons — bunnies, stars, heads, *etc.*

The work by Clark and Schober supports a personalized, real-time meter. By "individualizing" the kernels, the Popcorn Meter could further become an instrument of grounding. In a sense, Grade Six students cannot discuss the topic of grounding on a post-session questionnaire, but they can write that they want their "own kernel".

Another interesting result that deserves further study is the discrepancy between U Hill and M Grove students use of the "I do not want to vote" option on the post-game Rating Screen. Unlike the kids at University Hill who often clicked on *SUBMIT QUERY* without submitting a *Happy Face* vote, Maple Grove students went out of their way to click on "I do not want to vote" before clicking their way to the next page. Are kids at M Grove more "polite" or more thorough than the participants at U Hill? Or, is this difference in user behaviour related to the systems the kids were using? A study based on a random sample of users could look at these questions.

Other tools and tricks to test on future studies include a rating scheme that allows for a "dud" rating on the Popcorn Meter. For the PM-RB-CB system shown to Maple Grove students, the idea was that more activity on a particular kernel

would translate into a "hotter" game. Adding duds to the meter would change this, but the results might be interesting.

As mentioned in the previous chapter, an alternative way of sending and reading comments (*e.g.*, an interactive chat tool) should be explored. A Java applet for comments that always remains on the screen might be problematic, but it should be tested. Allowing students to open and close the Comments Board at any time might address the concerns/desires of some M Grove students' who wrote that they wanted to read a game's comments before playing the game, *etc.*

Another interesting variation on the Comments Board would be to seed the board with bogus comments. Brown *et al.* conclude that teachers shape the classroom climate through "demonstration, modeling, and guided feedback" [2]. Perhaps the system could foster a similar climate on the Comments Board. At any rate, it would be interesting to play with the "ethos of the Comments Board". If the Comments Board were seeded with thoughtful comments, would this "demonstrate" to users that "quality" comments are expected? If the CB were seeded with one-liners (*e.g.*, "SUCKS", "Great", "luv it", "hate it") would this have an effect on users' behaviour, on their attitude to rating and commenting, or on the feedback they provide? Yet another way of *modeling* the CB climate would be for the system to "reach out" to users who do not send ratings and comments. If a user were to consistently submit short messages (or none at all), the system could ask, prod or prompt the user to submit more detailed messages — like Brown *et al.*'s responsible teacher. This could introduce some very odd effects, but might also make for fascinating research.

There are, potentially, many "other uses" for peer-to-peer, real-time assessment. Adults play games; would they respond differently than 9-12 year-olds to an instrument like the Popcorn Meter? Also, many agencies solicit feedback from "users" — often these organizations use networked computers to gather such information. Peer-to-peer, real-time techniques could be tried on a variety of platforms, from information kiosks to corporate websites.

The strategy with any future studies should be to incorporate background research that includes cognitive psychology, Human-Computer Interaction, the design of everyday things, social relationships on the internet and the Media Equation. Such research helps to explain why standard techniques of soliciting user feedback (for example, those used for the NFBkids survey) are often ineffective. Knowing why and how the standard techniques fail is crucial for any design methodology that aims to be "user-centred". One of the problems with the NFBkids website was that their user-centred design process relied on the data from the survey — disappointing results from the *SURVEYS* halted the user-centred de-

sign process. Essentially, user-centred design "breaks" when the survey instrument fails to bring in usable data.

As far as PM-RB-CB is concerned, the project began with a real-world problem, followed by an attempt to find a theoretical framework within which the problem might be studied, understood and tackled. Of course, as an alternative method of gathering user preferences, PM-RB-CB seeks to attenuate the problems known to affect online surveys. The preliminary user studies suggest that real-time, peer-to-peer tools might be an effective method of soliciting user feedback and of fostering a sense of peer presence. Future research should build on this work through careful experimentation and by exploring the concepts mentioned above more thoroughly.

The user studies at University Hill and Maple Grove, along with the background research, suggest several "starting points" for future work. Although the eventual direction of future work is unknown, the hope is that the Popcorn Meter, Ratings Board and Comments Board — and the research behind them — form an introduction to peer-to-peer assessment techniques as a method to collect user feedback and to foster a sense of peer presence among a community of web users.

Bibliography

[1] J. Bensimon. National Film Board of Canada: A Hammer or a Mirror? Speech delivered at The Chicago Humanities Festival. Chicago, Illinois. November 2002.

[2] A. Brown, D. Ash, M. Rutherford, K. Nakagawa, A. Gordon, and J. Compione. Distributed expertise in the classroom. In G. Salomon, editor, *Distributed Cognitions — Psychological and educational considerations*. Cambridge University Press, New York, 1993.

[3] H. Clark and M. Schober. Asking questions and influencing answers. In J. Tanur, editor, *Questions About Questions*. Russel Sage Foundation, New York, 1992.

[4] M. Couper. Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64:464 – 494, 2000.

[5] H. Grice. Logic and conversation. In P. Cole and J. Morgan, editors, *Speech Acts*, volume 3 of *Syntax and Semantics*. Academic Press, New York, 1975.

[6] R. Groves, N. Fultz, and E. Martin. Direct questioning about comprehension in a survey setting. In J. M. Tanur, editor, *Questions About Questions*. Russel Sage Foundation, New York, 1992.

[7] K. Inkpen. *Supporting Collaborative Learning Environments For Children*. PhD thesis, University of British Columbia, 1997.

[8] C. Jenkins and D. Dillman. Towards a theory of self-administered questionnaire design. In L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz, and D. Trewin, editors, *Survey Management and Process Quality*. Wiley-Interscience, New York, 1997.

[9] R. Kling. Social relationships in electronic forums: Hangouts, salons, workplaces and communities. In R. Kling, editor, *Computerization And Contro-*

versy — Value Conflicts and Social Choices. Academic Press, San Diego, second edition, 1996.

[10] J. McGrenere. Design: Educational electronic multi-player games – a literature review. Master's thesis, University of British Columbia, 1996.

[11] D. Norman. The psychopathology of everyday things. In R. M. Baecker, J.Grudin, W. A. S. Buxton, and S. Greenberg, editors, *Readings in Human-Computer Interaction – Toward the Year 2000*. Morgan Kaufmann, San Fransico, second edition, 1995.

[12] National Film Board of Canada. Interview with Susan Nosov, Kids Destination Web Manager and Animation, Children, Interactive (ACI) National Product Manager, and Svend-Erik Eriksen, Kids Destination Project Leader and Executive Producer of ACI West. November 2001.

[13] National Film Board of Canada. Interview with the managers of the Kids Destination Project and the Animation, Children, Interactive Project, November 2001.

[14] National Film Board of Canada. *National Film Board's English Program Launches New Season With "Preview 2001"*. Press Release, May 2001.

[15] National Film Board of Canada. Original home page of NFBkids website, 2001. http://www.nfbkids.ca/.

[16] S. Papert. *The Children's Machine: Rethinking School in the Age of the Computer*. Basic Books, New York, 1993.

[17] R. Peterson. *Constructing Effective Questionnaires*. Sage Publications, Thousand Oaks, CA, 2000.

[18] B. Reeves and C. Nass. *The Media Equation*. Center for the Study of Language and Information. University of Cambridge, CSLI Publications, New York, 1996.

[19] H. Rheingold. *The Virtual Community: Homesteading on the Electronic Frontier*. MIT Press, Reading, Massachusetts, 1993.

[20] L. Rice-Barker. Playback News, 11 June 2001 – National Film Board Preview: New docs, animation and a kids' site. *Playback Magazine*, June 2001. http://www.playbackmag.com/articles/magazine/20010611/nfb.html.

[21] J. Rule and P. Attewell. What Do Computers Do? *Social Problems*, 36(3):225 – 241, June 1989.

[22] L. Sproull and S. Kiesler. Computers, networks and work. In R. M. Baecker, J.Grudin, W. A. S. Buxton, and S. Greenberg, editors, *Readings in Human-Computer Interaction – Toward the Year 2000*. Morgan Kaufmann, San Fransico, second edition, 1995.

[23] S. Sudman and N. Bradburn. *Response Effects in Surveys: a review and synthesis*, volume 16 of *National Opinion Research Center – Monographs in social research*. Aldine Publishing Co., Chicago, 1974.

[24] S. Sudman, N. Bradburn, and N. Schwarz. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass Publishers, San Fransico, 1995.

[25] J. Tanur. Cognitive aspects of surveys and this volume. In J. Tanur, editor, *Questions About Questions*. Russell Sage Foundation, New York, 1992.

[26] R. Tourangeau, L. Rips, and K. Rasinski. *The Psychology of Survey Response*. Cambridge University Press, New York, 2000.

[27] S. Turkle. *Life On The Screen: Identity in the Age of the Internet*. Simon and Schuster, New York, 1995.

[28] E. J. Wentland and K. W. Smith. *Survey Responses: an evaluation of their validity*. Academic Press, Inc., San Diego, 1993.

Appendix A

First Appendix

This appendix contains the entire list of ratings and comments submitted by University Hill students during the user study on October 18, 2002. It also contains a sample web-log for University Hill user *orange3* (this sort of web-log was prepared for each University Hill username) and scanned images of the two-page questionnaire completed by user *orange3*. The final section of this appendix contains a few captured images from the two video cameras present during the user study.

The database tables that store rating information code web-form submissions in the following manner:

Web-form action	DB entry
Zero *Happy Faces*	0
One *Happy Face*	1
Two *Happy Faces*	2
Three *Happy Faces*	3
Four *Happy Faces*	4
I do not want to vote	9
No submission	-

The University Hill games are coded as follows:

Game A	Lunar Lander
Game B	Concentration
Game C	Box Builder
Game D	COBS
Game E	Javanoid
Game F	3-D Vector Air Hockey
Game G	Hoop Shooter
Game H	Blister Ball
Game I	Boulderdash
Game J	Bridges

A.1 Ratings for Designers: University Hill

Game	User	Vote
A	blue2	9
A	red2	2
A	red1	0
B	red3	1
A	blue3	3
A	yellow2	1
A	red4	1
F	blue5	4
A	blue5	3
D	red5	3
H	orange5	2
B	red4	4
J	red5	9
C	red4	0
H	blue5	2
B	blue5	9
B	red2	1
J	blue3	3
B	red1	9
A	yellow3	4
A	orange5	9
F	yellow2	4
J	red2	0
J	blue5	0
H	red3	2
A	yellow1	1
C	orange2	9

Game	User	Vote
D	red4	4
B	blue4	2
B	blue2	2
E	blue3	4
F	red5	3
F	red4	4
F	yellow1	9
J	orange5	1
E	red3	4
J	red3	9
D	red1	0
F	red3	2
A	red5	1
G	blue3	2
E	red1	0
B	orange5	3
E	orange3	4
C	blue2	3
F	blue3	3
E	red4	4
I	red3	3
C	red3	9
E	red3	-
J	blue3	-
I	orange2	3
J	orange2	1
D	blue2	1
D	blue5	4
H	orange2	4
C	blue3	3

Game	User	Vote
B	orange3	-
E	orange2	3
F	red2	4
G	red3	1
J	blue5	-
E	yellow1	4
I	red3	-
E	blue2	9
H	red3	-
A	orange2	4
D	red2	1
C	blue4	9
C	red2	0
G	red4	4
G	red4	-
H	red4	0
E	yellow3	4
G	orange3	-
E	red2	4
B	yellow1	9
B	orange2	0
I	orange5	4
A	red3	2
D	red3	-
E	blue5	4
H	red3	-
F	blue2	1
D	blue3	4
I	red4	0

Game	User	Vote
E	red3	-
A	orange3	2
E	red3	-
I	red5	2
E	red3	-
J	red3	0
I	red3	-
G	blue5	3
F	red3	-
G	red2	4
F	red4	-
G	blue2	9
G	red3	3
I	red4	-
C	red3	0
H	blue3	2
H	red2	0
B	red3	0
A	red3	-
H	red5	0
G	yellow3	9
H	red3	-
J	red3	-
H	orange3	2
G	yellow1	4
H	blue2	0
I	red2	2
F	red4	0
J	yellow3	-

Game	User	Vote
C	blue5	4
B	blue3	3
D	red1	9
I	yellow1	9
I	yellow1	9
C	yellow3	9
I	blue2	0
F	red1	0
C	orange5	4
E	red5	-
E	red5	-
I	orange3	4
E	red3	-
H	blue5	-
H	red3	-
I	blue3	4
D	orange3	-
J	blue2	1
C	orange3	-
A	red2	4
D	orange5	4

A.2 Comments to Designers: University Hill

Game	User	Comments
A	blue3	hello i found this game pretty amusing but i didnt get the point!
A	red4	Hi that game is good game.
F	blue5	EXALENT
A	blue5	the turning is good
B	red3	This game is sort of dull.
A	red2	

A	blue2	Your game is not interesting.
D	red5	
B	red4	This game is good
J	red5	
C	red4	HI
H	blue5	
B	blue5	
B	red2	
H	orange5	this was a cool game but really too hard. it was hard to control and shoot the blister balls. In the end: "it wasn’t so good it can be better!"
B	red1	
J	blue3	hello this game was fun and exiting but it is hard to beat!
A	yellow3	A really awsome old Antari copied game.
J	red2	
J	blue5	SUCKS
A	orange5	It’s too complicated to control this game. booooring.
F	yellow2	It’s fun
H	red3	O.k. but too hard
C	orange2	
D	red4	t
A	yellow1	this game is very interesting but the instructions is very confusing and hard to understand we think you should add something to make the spaceship go down quicker.
F	red5	good!
F	red4	h fjkashcb
B	blue2	it is fun. But it’s little bit hard
E	blue3	this is one of my favorite games d i think that it will be populer
F	yellow1	very hard to understand what were we supposed to use to move the

```
                thingy
E       red3    Fun and Funny.
J       orange5 It’s a hard game.
J       red3
D       red1
F       red3    Hard.
A       red5    Great game
E       red1
B       orange5 cool.but sort of boring
G       blue3   hi
                im sorry to say this but i
                thought this game was boring
                and pointless:(
E       orange3 I really like this game challenging
C       blue2   we lose but it’s fun...
E       red4    adfghgfuyi
F       blue3   we thought this game was up beat
                and veryb easy
I       red3    Funny.
C       red3
E       red3
J       blue3
I       orange2
J       orange2
J       orange2
D       blue2   WHY WE CAN’T CHANGE THE SHAPE????
                ?????????
H       orange2
D       blue5   i LOVE TETRIS!!!!
B       orange3
E       orange2
F       red2
G       red3
J       blue5
C       blue3    we think that it was boring
                 and fun!
I       red3
A       orange2
```

```
D       red2
E       blue2     .......
C       red2
G       red4      -[[b-opkj][p[’
G       red4      [p[-p[
E       yellow1   this game is very fun.but the
                  ball moves toooooo fast. you
                  shouldn’t make the pills so hard
                  to get people will get angry and
                  lose their balls.
H       red4
G       orange3
E       red2
B       yellow1   no comments
B       orange2
E       yellow3   nice game!
A       red3
I       orange5   Cool the best game. this was really
                  cool its kinda tricky
D       red3
E       blue5     i LIKE THIS GAME
H       red3
F       blue2     ......????????????????
D       blue3     that is my favorite game
I       red4      k;sdhjaiyfweigfjhaw
E       red3
A       orange3   ok
I       red5
E       red3
J       red3
I       red3
F       red3
G       blue5
G       red2
F       red4      jmkjghiv
G       red3      Good
G       blue2     ?????????????????????/
I       red4      hhg
```

```
C       red3
H       red2
B       red3
H       blue3     this game is BORING
A       red3
H       red3
H       red5      ok bit boring!
J       red3
G       yellow3
H       orange3   not bad
I       red2
H       blue2     ??????????
F       red4      lp;,l;
G       yellow1   THIS GAME IS REALLY REALLY
                  REALLY REALLY REALLY REALLY
                  REALLY FUN!!!!!!!IT ROCKS!!!
                  !!!!!!!!!!!MY FRIEND AND I
                  LOVE IT!!!!!!!!!!!
C       blue5
J       yellow3
B       blue3     this game was ok
B       blue3     this game was ok
D       red1
C       yellow3
I       blue2     WEIRD...
I       yellow1   NO COMMENT
F       red1
F       red1
E       red5
C       orange5   The game  is really really
                  good (hard and tricky)
E       red3
H       blue5
H       red3
I       orange3   this is a very fun game!!!:)
I       blue3     this game was SOOOOOOOOOOOO
                  FUN
D       orange3
```

```
J       blue2    your game is BORING
C       orange3
A       red2
D       orange5  Good!
```

A.3 Sample Web Log: University Hill

```
University Hill School
October 18, 2002

User: orange3
Database Log
```

Game	User	Vote
E	orange3	4
B	orange3	-
G	orange3	-
A	orange3	2
H	orange3	2
I	orange3	4
D	orange3	-
C	orange3	-

Game	User	Comments
E	orange3	I really like this game challenging
B	orange3	
G	orange3	
A	orange3	ok
H	orange3	not bad
I	orange3	this is a very fun game!!!:)
D	orange3	
C	orange3	

Web Log

```
09:03:18 gameD
09:04:38 gameD

09:07:33 palette

09:07:46 gameF
09:08:28 gameF

09:10:59 gameI
09:11:17 gameI

09:11:46 gameA
09:12:17 gameA

09:13:34 gameE
09:22:28 gameE/process_vote  (4)
09:23:24 gameE/process_comments  (I really...)

09:27:19 palette

09:27:24 gameB
09:27:35 gameB/process_vote  ( )
09:27:38 gameB/process_comments ( )

09:27:46 gameG
```

```
09:31:01 gameG/process_vote  ( )
09:31:02 gameG/process_comments  ( )

09:31:09 gameA
09:34:53 gameA/process_vote  (2)
09:34:59 gameA/process_comments  (ok)

09:35:27 gameH/
09:39:43 gameH/process_vote  (2)
09:39:49 gameH/process_comments  (not bad)

09:40:09 gameI
09:45:54 gameI/process_vote  (4)
09:46:21 gameI/process_comments  (this is...)

09:46:27 gameD
09:46:37 gameD/process_vote  ( )
09:46:39 gameD/process_comments  ( )

09:46:45 gameC
09:47:12 gameC/process_vote  ( )
09:47:13 gameC/process_comments  ( )   Final Entry
```

A.4 Sample Questionnaire: University Hill

University Hill School
October 18, 2002

USERNAME: Orange

Post-game questionnaire

1. Of the games you just played, please list your top 3 games.

i) boulder dash
ii) Hoop shooter
iii) lunar landing

2. Think about one of your top 3 games. If you had to describe this game to some friends (who had not seen the game), what would you tell them? (Please circle this game in the list above.)

For example, consider these questions:

What parts of the game did you like?
What made this game fun?
What (if anything) was not so great about the game?
Why is it one of your top games?

(You may answer any, all or none of these questions. Also, feel free to provide your own comments on the game.)

This game called boulder dash is a very funny game with lots of fun how you play is get all the diamonds and watch out for the boulders I like how the way that the bugs chased you and it was funny the explosions

3. In this study, after you played a game, you had the chance to send your ideas and opinions to the game designers. Is this something you like? Are you pleased that the game designers will be reading your comments?

(You may answer this question with a simple YES or NO. Add more information, if you like.)

I think this was fun for research My favorite was boulder dash! Yes I think they will pay attention to my comments

4. Do you think the game designers will pay attention to the comments you entered today?

Yes because I said good things about there games

6. Why do you think the game designers want to read your comments and see how you rated the games?

because I sent very nice comments and answers about my favorite game

Thank you very much.

Have a nice day!

Appendix B

Second Appendix

This appendix contains the entire list of ratings and comments submitted by Maple Grove students during the user study on November 7, 2002. It also contains a sample web-log for Maple Grove user *red3* and scanned images of the two-page questionnaire completed by user *red3*. The final section of the appendix contains a few captured images from the two video cameras present during the user study.

The database tables that store rating information code web-form submissions in the following manner:

Web-form action	DB entry
Zero *Popcorn Kernels*	0
One *Popcorn Kernel*	1
Two *Popcorn Kernels*	2
Three *Popcorn Kernels*	3
Four *Popcorn Kernels*	4
I do not want to vote	9
No submission	-

The University Hill games are coded as follows:

Game A	Lunar Lander
Game B	Concentration
Game C	Bridges
Game D	COBS
Game E	Javanoid
Game F	Blister Ball
Game G	Boulderdash

B.1 Ratings for Other Players: Maple Grove

Game	User	Vote
D	orange1	9
E	orange2	9
B	red3	4
C	green3	0
A	orange3	4
C	orange1	4
G	red5	4
B	yellow3	4
A	green3	0
B	red4	0
D	orange2	2
A	blue5	4
E	green3	2
D	red1	2
B	orange2	-
F	blue4	4
A	orange2	9
E	blue3	3
F	orange2	-
E	green2	4
C	green5	3
F	green1	2
E	blue4	4
A	red4	1
A	red4	3
E	green1	4
D	blue4	0
A	orange1	3
G	green3	4
E	blue2	4

Game	User	Vote
G	green3	4
C	red1	9
C	blue4	0
F	green2	3
F	red5	9
D	blue3	0
B	blue4	4
B	red1	2
E	red5	4
G	green2	1
E	red4	9
A	green4	9
D	green1	9
D	red5	4
A	green5	9
D	green2	0
A	red3	1
C	blue3	2
C	red5	9
C	green1	9
B	green3	0
B	green1	9
C	green2	9
B	red5	9
B	green2	9
A	green1	9
A	red5	2
A	green2	0
A	blue4	9
A	green3	9

Game	User	Vote
G	green1	9
A	red1	0
B	blue2	2
B	blue3	0
A	blue3	1
A	blue2	9
B	yellow4	0
F	blue1	2
G	blue1	4
E	blue1	4
D	blue1	2
C	blue1	2
B	blue1	1
A	blue1	1
B	orange1	0
C	orange2	1
C	green4	3
F	green3	2
E	green1	9
D	yellow3	9
D	green4	2
E	red1	2
C	orange2	-
C	orange2	-
D	green3	1
D	green3	1
E	red1	2
B	orange3	9
B	green4	3
D	red5	4

Game	User	Vote
D	green1	9
C	orange3	9
F	red1	3
F	red1	-
D	orange3	9
F	green5	4
F	orange2	-
G	red5	4
E	green4	3
E	orange3	9
G	green2	0
G	blue4	4
E	orange3	9
E	red3	4
F	green5	4
F	orange3	9
E	green2	9
G	orange2	-
G	orange3	9
E	green2	9
F	red2	4
F	blue5	4
G	blue4	4
G	blue4	9
C	blue4	9
C	red3	2
B	green5	2
C	red3	2
D	blue4	9
E	orange1	4

Game	User	Vote
F	green5	-
C	red4	4
D	red3	0
E	red5	9
G	blue4	0
G	green2	1
A	red1	0
A	red1	0
A	orange3	9
A	orange3	9
D	green5	-
F	green4	4
F	red3	0
D	blue4	4
D	red4	0
B	blue5	1
G	green3	4
F	red1	1
E	green2	4
F	orange2	-
A	green3	0
A	orange2	-
B	orange2	-
F	red4	9
F	orange1	4
G	green4	4
G	red3	0
G	green1	-
E	green5	-
G	red5	-

Game	User	Vote
G	red5	-
G	blue2	4
E	green3	3
G	red1	9

B.2 Comments to Other Players: Maple Grove

Game	User	Comments
A	green2	this is noy fun except when it crashes
A	blue4	
A	green3	
G	green1	IYRTYUEtu
B	blue2	bbbbbbbbbbooooooooooorrrrrr rrrrrriiiiiiiiinnnnnnnnggg ggggggg
B	blue3	BOring
B	blue3	BOring
A	blue3	BOring
A	blue2	boring
B	yellow4	Not much action. I don’t recommend this.
F	blue1	At first, you might hate this. 2-player is not that bad!
G	blue1	Action and strategy. Took me a while to learn, but a lot of fun.
E	blue1	Great fun -- you must try this! The gun/shooter is so cool.
A	green5	No comment.
D	green1	
D	blue1	I like this. Cool little "guys".

```
C       blue1   Hard to beat computer--a good
                strategy game.
B       blue1   So much clicking...
A       blue1   Hard to control!
C       blue3   MAtt:Computer is too good
                Sam:BBBBBOOOOOORRRRRIIIIINNNN
                NNGGGGGGG!!!
D       red5
D       green2  gerljgn
C       red5
C       green1
B       green3
C       green2  THIS IS NOT FUN!!!!!!!!!!
B       green1
B       red5
A       green1  I dont know
B       green2
A       red5
A       red1    cause:it’s not easy to play!!!
                !!!!!
A       red3    its too hard to controll!!!!!!
                and it’s weird but it is
                interesting in a weird way...
                ...-_-zzzz
E       red4    it was cool but classic
E       red4    it was cool but classic
G       green2  nothing speciaL
A       green4  No commet.
E       red5    cool
B       blue4
B       red1    i can make me happy!!!!!!!!!!
                !!!!!!!!!!!!!!!
D       blue3   MAtt:this is tetris.  BORING!!!
                plageurism.!!!!
F       red5
F       green2  best here
C       blue4   boring!
C       red1    i dont’t know!!!!!!!!
```

E	blue2	fun, soso, hard to win
G	green3	very cooll..... SQUISH lllllll llooooooooooooollllllllll
A	orange1	
D	blue4	
E	green1	AWESOME!!!!!!!!!!!!!!!!
A	red4	ummmmmmm
A	red4	ummmmmm it wuz B0ring and lame sort of i didnt get how u points worked
F	green1	Hard
E	blue4	This game is bad but it’s the best out of them all...
C	green4	No commet
C	green5	1.Pretty Good! But it is too hard to win! 2.Very hard, but I’ll win SOMEday
F	green3	Ivan: ummmmmm stupid apples Anson: I hate apples grrrrrrrr rrrrrrrr LOL
C	orange2	This game is worst
B	orange1	
A	orange2	
F	blue4	not bad..
E	blue3	GOOD BUt copied
B	orange2	
B	orange2	
D	red1	cause i play that game about 111000000000 000000 times !!!!!!!!!!!!!!!!!!!
E	green3	ivan: eeeeeewwwwwww metal balls!
D	orange2	...
A	blue5	it was fun i guess
B	red4	it is fun (crossing my hand)
A	blue5	it was fun i guess
A	green3	ahhhhhhhhh shut up copy righted
G	red5	so cool

```
B       yellow3 cool games
C       orange1
D       orange1
D       red5
E       orange2
C       green3  IVAN: HOW DO U PLAY!!!!!!!!
                !!!!!!!!!!!!!!!!!!!!!!!!!!
                !!!!!!!!!!!!!!!!!!!!!!!!!!
                !!!!!!!!!!!!!!!!!!
                ANSON:AHHHHYAHHHHHHHHH
B       green4  Sort of fun
                Soso
B       red3    you should make a varaity of
                drawings and make the icons
                click faster!!!!!!
E       red1    cause it’s very boring!!!!!
                !!!!!!!!!!!!!!!!!!!!!!!!!!
                !!!!!!!!!!!!!!!!!!!!!!!!!!
                !!!!!!!!!!!!!!!!!!!!!!!!!!
                !!!!!!!!!!!!!!!!!!!!!!!!!!
                !!!!!!!!!!!!!!!!!!!!!!!!!!
                !!!!!!!!!!
C       orange2
E       red1
D       green4  So boring
D       green3  ANSON20:GOOD GRAPHICS BUT
                BAD GAME I HATE TETRIS
                IVAN: OKKKKKKKKKKK
E       green1  akjfasdfjasdl;kf=
F       orange2
F       orange2
E       green2  SWEEEEEEEEEEEEEET!!!!!!!!
                !!!!!!SSSSSSWWWEEEEEEEEEEE
                ET!!!!!!!!
D       green1
F       red1
F       red1
F       green5  Very exciting!
```

```
                This is the best game we've
                played before.
F       orange2
G       red5
G       red5
E       green4  Fun
G       blue4   so good man!
G       green2  dasf,gmknfv hewnkgmn;k.qhfl
                jkegtrqk3hwebqwghfbewjf hil
                qewkfhejlnf;kewjfpiewhfngow
                4l;kmnfqwuiorjf[we4olthgqiu
                ejktm;4wtj=04wo[th[3oq2ktjm
                jkq4l2
G       green2  dasf,gmknfv hewnkgmn;k.qhfl
                jkegtrqk3hwebqwghfbewjf hil
                qewkfhejlnf;kewjfpiewhfngow
                4l;kmnfqwuiorjf[we4olthgqiu
                ejktm;4wtj=04wo[th[3oq2ktjm
                jkq4l2
E       red3    cool...!!!-__-
F       green5  Hahahahahahahhahahahaaha
                hahh
E       green2
G       orange2
E       green2
F       red2    this game is hard,
                but it is fun!!!!!!!
F       blue5   Best game 4 ever!
G       blue4   best out of them all
                i guess...
C       blue4
G       blue4
D       blue4
E       orange1 wow, i got 2 in a row!!
C       red3    ^__^xxxxzzzzxxxzzxzxzxhey
B       green5  BBBBBBBBBBOOOOOOOOOOORRRRR
                RRRRRIIIIIIIINNNNNNNNNNG
                GGGGGGGGGGGG!!!!!!!!!!!
```

```
F       green5  LOLOLOL
D       red3    make it better seriously...
D       red3    make it better seriously...
C       red4    c|o.Op monkey says: oh ha
                aha ha ah ah ah ah
                ( cool game )
E       red5
C       red4    c|o.Op monkey says: oh ha
                aha ha ah ah ah ah
                ( cool game )
E       red5
G       blue4
G       green2
A       red1
D       green5
F       green4  Fun, very exiciting
F       red3    BORING......-__-ZZZZZZZZZZ
D       blue4
D       red4    Boring!!! wuts up wit da
                uglie faces
F       red1
G       green3  tHE BEST OF ALL BEST DUDE
                TRY IT!!!!!!!!!!!!!!!!!!!!
                !!!!!!!!!!!!!!!!!!!!!!!!!!
                !!!!!!!!!!!!!!!!!!!!!
B       blue5   The boringest game ever
                invented, i think>
F       orange2
E       green2  bjmnv dmsvjsbcljsncjklasb
                dkmns fbam fc;ljkasfnj
                sbjahkbcjlsacbjlkasbncjkl;
                sabljasvckaskljascb
                ljsadcnljascnb;klsac
A       green3  THE WORST OF ALL
A       orange2
B       orange2
F       red4    i dont no
F       orange1
```

```
G       red3
G       green4  okok
G       green1
E       green5
G       red5
G       red1
G       blue2   excellent
E       green3  FHUAERIN VSEDHRZKL89AZSE
                .DT790 H5
```

B.3 Sample Database Log and Sample Web Log: Maple Grove

```
Maple Grove School
November 7, 2002

User: red3
Database Log
```

Game	User	Vote
B	red3	4
A	red3	1
E	red3	4
C	red3	2
C	red3	2
D	red3	0
F	red3	0
G	red3	0

Game	User	Comments
A	red3	its too hard to controll!!!!!! and it's weird but it is interesting in a weird way.....-_ -zzzz
B	red3	you should make a varaity of drawings and make the icons click faster!!!!!!
E	red3	cool...!!!-__-
C	red3	^__ ^xxxxzzzzxxxzzxzxzxhey
D	red3	make it better seriously...
D	red3	make it better seriously...
F	red3	BORING.....-__-ZZZZZZZZZ
G	red3	

Web Log

```
13:44:06 gameA
13:46:17 gameA/process_vote.php  (1)
13:47:02 gameA/process_comments  (its too hard...)
13:47:07 gameA/recent_votes
13:47:21 gameA/see_comments

13:47:30 gameB
13:59:55 gameB/process_vote  (4)
14:00:36 gameB/process_comments  (you should...)
14:00:46 gameB/recent_votes
14:00:48 gameB/see_comments

14:01:00 gameE
14:10:02 gameE/process_vote  (4)
```

```
14:10:15 gameE/process_comments.php  (cool...)
14:10:22 gameE/recent_votes
14:10:24 gameE/see_comments

14:10:33 gameC
14:13:12 gameC/process_vote  (2)
14:13:14 gameC/process_vote  (2)
14:13:33 gameC/process_comments  (^__^...)
14:13:36 gameC/recent_votes
14:13:38 gameC/see_comments

14:13:41 gameD
14:15:22 gameD/process_vote  (0)
14:15:38 gameD/process_comments  (make it...)
14:15:39 gameD/process_comments  (make it...)
14:15:43 gameD/recent_votes
14:15:45 gameD/see_comments

14:15:48 gameF
14:18:18 gameF/process_vote  (0)
14:18:33 gameF/process_comments  (BORING...)
14:18:35 gameF/recent_votes
14:18:37 gameF/see_comments

14:18:40 gameG
14:22:39 gameG/process_vote (0)
14:22:42 gameG/process_comments  ( )
14:22:43 gameG/recent_votes
14:22:45 gameG/see_comments  (Final Entry)
```

B.4 Sample Questionnaire: Maple Grove

Maple Grove School
November 7, 2002

USERNAME: red3

Post-game questionnaire

1. Of the games you just played, please indicate your top 3 games with a "1", "2" and "3".

Game A - Lunar Lander
Game B - Javanoid 1
2 Game B - Concentration
Game F - Blister Ball
3 Game C - Bridges
Game G - Boulderdash
Game D - Cobs

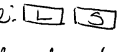
2. Think of one of your top 3 games. If you had to describe this game to some friends (who had not seen the game), what would you tell them?

(Please circle this game in the list above.)

For example, consider the following questions:

What parts of the game did you like?
What made this game fun?
What (if anything) was wrong with the game?
Why is it one of your top games?

(Also, feel free to add your own comments.)

It was very exciting and make you jump out of your seat! This game was fun because it is suspensful and hard because of those icons like:  etc.  But it was really hard to move the bed.. this part.

3. In this study, after you played a game, you had the chance to send ratings and comments to other players. Is this something you like?

Yup!

If so, what do you like about it? If you do not like it, why not?

I liked it because I like to decide things like wether I like it or not. Plus it was entertaining!

4. Did you send many comments to other players? Why or why not?

No, because I usually get carried away with comments..

5. Did you send many kernel ratings to other players? Why / why not?

Yes I sent many kernel ratings but it depend on if I liked the game or not and my friend. I sent it because I wanted to tell them that this game was good or that game was bad. So that everybody can play the fun games but someone could've liked a game that I hate.

6. What did you think of the Popcorn Meter?

It was a good idea and I liked the drawings...

7. The Web designers would like your thoughts on these games. For each game listed below, did you find the game "Cool" or was it "Yuk"?

COOL!	Game	YUK!
✓	Lunar Lander	✓
✓	Concentration	
✓	Bridges	
	Cobs	✓
✓	Javanoid	
✓	Blister Ball	✓
✓	Boulderdash	✓

8. Do you have any other comments for the Web designers?

no comment. --zzz ^^ hihihi

THANK YOU!

Appendix C

Video Stills from University Hill and Maple Grove

Figure C.1: University Hill students see the Palette Screen for the first time.

Figure C.2: The University Hill *Happy Faces* Rating Screen.

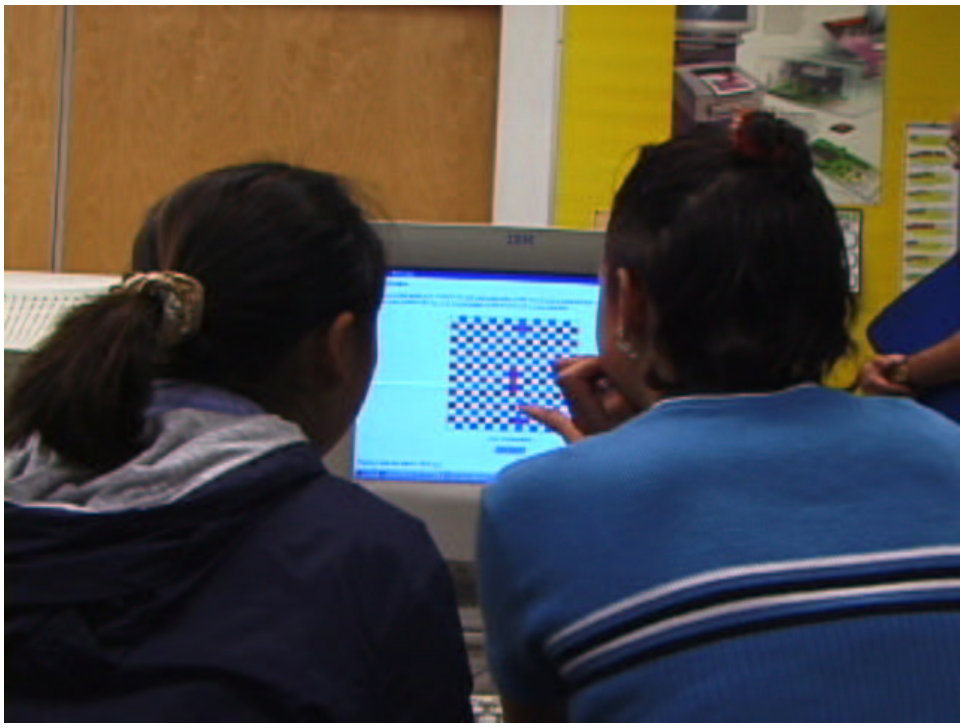


Figure C.3: *Bridges* at University Hill.

Figure C.4: The *Javanoid* game screen.

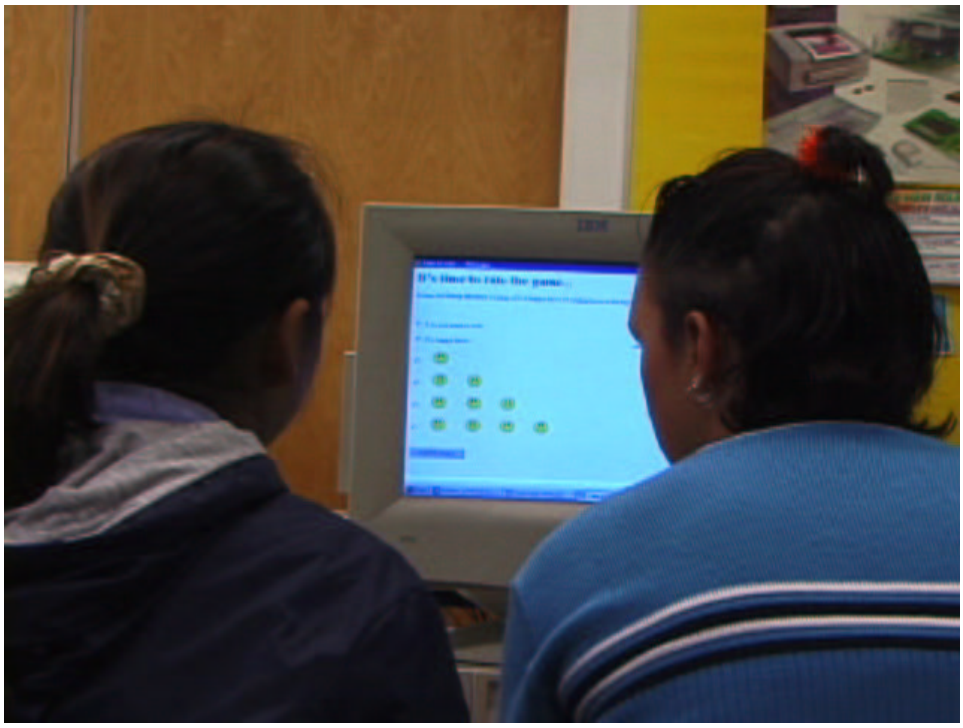


Figure C.5: Two U Hill students in front of the *Hoop Shooter* Rating Screen.

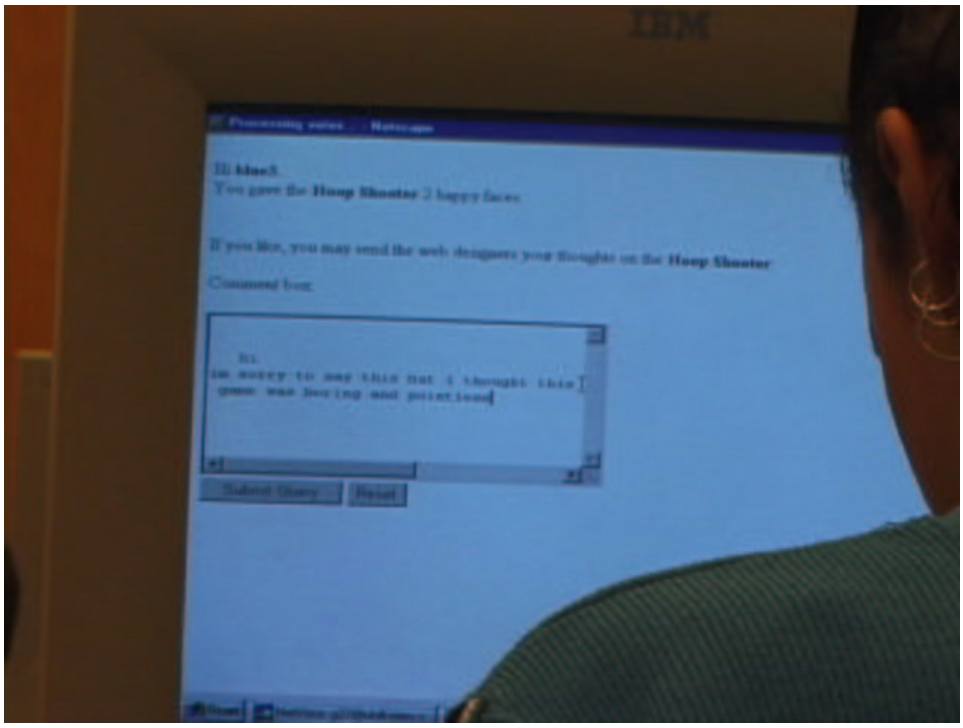


Figure C.6: Under the username *blue3*, a University Hill student enters the comment: “hi im sorry to say this but i thought this game was boring and pointless :(”

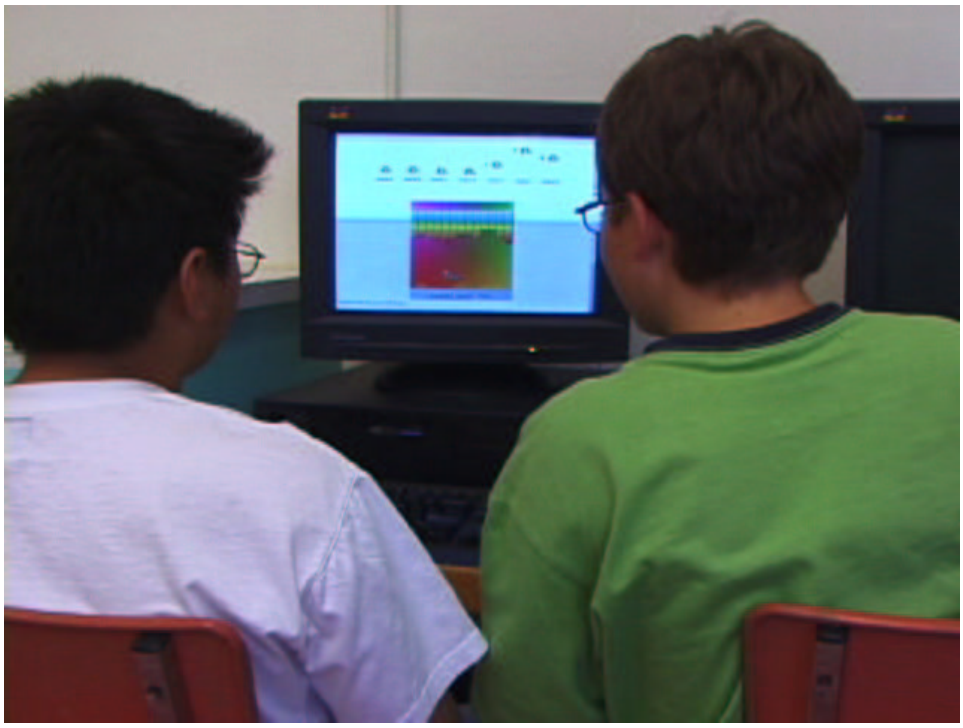


Figure C.7: Two Maple Grove students at the *Javanoid* screen.

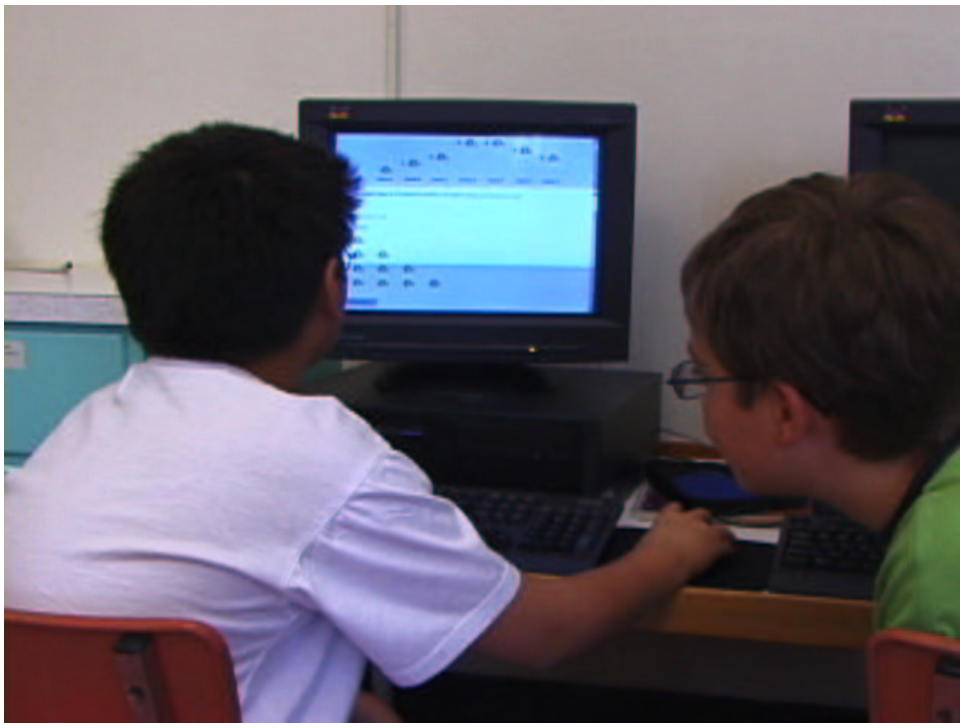


Figure C.8: A post-game Rating Screen at Maple Grove.

Figure C.9: The Maple Grove computer lab.

Figure C.10: Maple Grove students playing *COBS* (the near screen) and *Boulderdash.*

Figure C.11: *Boulderdash* at Maple Grove.

Figure C.12: The Maple Grove *Javanoid* game screen.