Prime Climb: An Analysis of Attention to Student-Adaptive Hints in an Educational Game

by

Mary Anne Midori Muir

B.Sc., The University of British Columbia, 1995
B.Ed. (Sec), The University of British Columbia, 1998
B.C.S., The University of British Columbia, 2008

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

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2012

© Mary Anne Midori Muir, 2012
Abstract

Prime Climb is an educational game that provides individual support for learning number factorization skills in the form of hints based on a model of student learning. Previous studies with Prime Climb indicated that students may not always be paying attention to the hints, even when they are justified (i.e. based on a student model’s assessment). In this thesis we will discuss the test-bed game, Prime Climb, and our re-implementation of the game which allowed us to modify the game dynamically and will allow for more rapid prototyping in the future. To assist students as they play the game, Prime Climb includes a pedagogical agent which provides individualized support by providing user-adaptive hints. We then move into our work with the eye-tracker to better understand if and how students process the agent’s personalized hints. We will conclude with a user study in which we use eye-tracking data to capture user attention patterns as impacted by factors related to existing user knowledge, hint types, and attitude towards getting help in general. We plan to leverage these results in the future to make hint delivery more effective.
Preface

This work is based on the work conducted in UBC’s Intelligent User Interface laboratory of Dr. Cristina Conati. An overview of the existing Prime Climb game is discussed in Chapter 3. The re-implementation of the Prime Climb game (as discussed in Chapter 4) was a collaboration between Peter Rau, an undergraduate student and I. Peter and I worked together to develop the specifications. Peter wrote the majority of the code while I provided support when he had any questions. In addition, I re-implemented the user’s pedagogical agent and hinting algorithm in the game as well as continued with bug fixes and improvements after he left the project.

The user study conducted in Chapters 5 and 6 was done under the approval of UBC Behavioral Research Ethics Board (BREB) certificate H04-80496. I was the main investigator for this study and managed all the recruitment and conducted the pre/post-test and questionnaire with the subjects. I was assisted by Alireza Davoodi who acted as a second player in some of the games as well as assisted in observing the player’s eye-gaze data as they played when I was playing the game. In addition Samad Kardan helped observe subjects when Alireza was not available.


In both papers, I was responsible for conducting the user study, processing all data and generating all figures and tables.
Table of Contents

Abstract ................................................................................................................................. ii
Preface ................................................................................................................................... iii
Table of Contents ..................................................................................................................... v
List of Tables .......................................................................................................................... ix
List of Figures ........................................................................................................................ x
Acknowledgements .............................................................................................................. xiii

Chapter 1: Introduction ....................................................................................................... 1

1.1 Thesis Goal ..................................................................................................................... 2
1.2 Contributions of this Research .................................................................................... 2
1.3 Methodology .................................................................................................................. 3
1.4 Outline .......................................................................................................................... 4

Chapter 2: Related Work ................................................................................................... 5

2.1 Educational Games .................................................................................................... 5
2.2 User-Adaptive Edu-games ......................................................................................... 5
2.3 Intelligent Tutoring Systems ...................................................................................... 6
2.4 Eye-Tracking .............................................................................................................. 7

Chapter 3: Prime Climb Game .......................................................................................... 10

3.1 Description of Prime Climb ....................................................................................... 10
3.2 Tools .......................................................................................................................... 12
3.3 Pedagogical Agent ..................................................................................................... 13
3.3.1 Hinting Strategy ................................................................................................... 13
Chapter 4: Reimplementation of Prime Climb ................................................. 18

4.1 Issues Faced with the Original Prime Climb Game ............................................. 18
4.2 Goals of the New Prime Climb Game ................................................................. 20
4.3 Game Architecture ............................................................................................. 21
  4.3.1 Server ............................................................................................................ 23
  4.3.2 Client(s) ....................................................................................................... 24
  4.3.3 Database .................................................................................................... 26
    4.3.3.1 Database Tables .................................................................................... 27
4.4 Benefits of Re-implementation and my Contribution ......................................... 30

Chapter 5: Preliminary Analysis of Eye-Tracking Data ........................................ 32

5.1 Data Collection .................................................................................................. 33
  5.1.1 Arrangements to Minimize Eye-Gaze Data Loss ............................................. 34
5.2 Assessment Tools ............................................................................................... 36
  5.2.1 Pre-test and Post-test on Factors and Common Factors ..................................... 36
  5.2.2 Student’s Subjective Assessment of the Game ............................................... 37
5.3 Eye-Tracker Data ................................................................................................. 37
  5.3.1 Data Collection ............................................................................................. 38
  5.3.2 Defining Relevant Eye-Data ........................................................................... 38
  5.3.3 Areas of Interest ......................................................................................... 39
  5.3.4 Eye-Tracking Measures .............................................................................. 40
  5.3.5 Data Validation ........................................................................................... 42
  5.3.6 Processing of Eye-Tracker Data ................................................................... 44
5.4 Knowledge Test Results ........................................................................................................45
5.5 Initial Comparison between Fixation Time and Hint Length ........................................46
5.6 Patterns of Attention to Hints ............................................................................................47
  5.6.1 Subject 1 .........................................................................................................................47
  5.6.2 Subject 2 .........................................................................................................................53
5.7 Discussion ..........................................................................................................................56

Chapter 6: In Depth Analysis of Eye-Tracking Data .................................................................58
6.1 Experimental Design & Data Collection ..............................................................................58
6.2 Descriptive Statistics on Game Play ...................................................................................58
  6.2.1 Pre- and Post-test Scores ...............................................................................................58
  6.2.2 Time Spent on the Game, Number and Type of Hints ................................................59
  6.2.3 Moves ............................................................................................................................60
  6.2.4 Magnifying Glass Usage ...............................................................................................61
6.3 Factors that Affect Attention .............................................................................................61
  6.3.1 Impact of Model Accuracy ............................................................................................63
  6.3.2 Additional Factors which Affected Game Play as Measured by Hint Fixation
      Time .....................................................................................................................................64
      6.3.2.1 Results ..................................................................................................................64
  6.3.3 Additional Factors which Affected Game Play as Measured by Fixations per
      Word ..................................................................................................................................66
      6.3.3.1 Results ..................................................................................................................66
  6.3.4 Discussion ......................................................................................................................69
6.4 How Attention Affected Game Play ....................................................................................70
6.4.1 Move Correctness after Hints ................................................................. 70

6.4.2 Magnifying Glass Usage ................................................................. 71

Chapter 7: Conclusion and Future Work .................................................. 72

Bibliography ............................................................................................. 76

Appendices ............................................................................................... 80

Appendix A Permission Forms .............................................................. 80

Appendix B Recruitment ........................................................................... 83

Appendix C Pre-test ................................................................................. 84

Appendix D Questionnaire ...................................................................... 86

Appendix E Server States ........................................................................ 91

Appendix F Study Raw Data ................................................................. 93

F.1 Play Time, Hint Number and Time between Hints .......................... 93

F.2 Pre and Post-test Score and Proportional Learning Gain ............... 94

F.3 Moves .............................................................................................. 95

F.4 Magnifying Glass Usage .................................................................... 96

Appendix G More Details on Specificity and Sensitivity ....................... 97
List of Tables

Table 3.1: Examples of common factor and factorization hints. ........................................ 14

Table 5.1: Statistics on the mean fixations/word for each hint type for the two subjects. ..... 52

Table 6.1: Logistic regression results for Move Correctness after Hint. ............................. 71

Table A.1 Playtime, total hints and average time (SD) between hints for each player. ........ 93

Table A.2 Pre- and post-test scores as well as proportional learning for each player. ........ 94

Table A.3 Moves made by subject including percent of wrong moves and number of wrong/correct moves made after a hint ................................................................. 95

Table A.4 Magnifying glass usage for each subject ........................................................... 96
List of Figures

Figure 3.1: The Prime Climb game................................................................. 10

Figure 3.2: Player swinging over three highlighted numbers after having fallen due to choosing a number that shared a common factor with the partner (red player). ............... 11

Figure 3.3: Magnifying glass tool showing the factorization of 27. The left button lets the player select the number they want displayed in this tool. The right “?” button accesses a help menu.................................................................................................................. 12

Figure 4.1: Original Prime Climb architecture. A server communicates with two clients – the wizard and client. The wizard is the investigators version of the client that allows them to observe when the client receives a hint. The Prime Climb client contains a pedagogical agent which provides user-adaptive hints as they play................................................................. 19

Figure 4.2: Game architecture for the new Prime Climb implementation. The server, database and admin application reside on the same machine while the clients access the server through the internet. ........................................................................................................ 22

Figure 4.3: Log in screen for the Prime Climb game which allows for users to either register or log in with an existing username................................................................. 25

Figure 4.4 Lobby screen of Prime Climb, showing two players sitting in the first game room ready to start a game. ............................................................................................... 25

Figure 4.5: Game screen of Prime Climb showing a definition hint in the upper left part of the screen as well as the effect of using the magnifying glass to show the factorization of 40 in the upper right corner................................................................. 26

Figure 4.6: Database structure for the Prime Climb game................................................................. 27
Figure 5.1: Pre-calibration screen which allows subjects to become more aware of the capabilities of the eye-tracker.................................................................35

Figure 5.2: Prime Climb screen showing the four areas of interest: Hint, Hint Close, Mountain and Tool.........................................................................................................................40

Figure 5.3 Percentage of segments discarded for different threshold values. ......................43

Figure 5.4: Comparison between the mean fixation time on the three types of hints and the expected reading time.................................................................46

Figure 5.5: Total fixation time spent by subject 1 on a) Definition, b) Bottom-Out, and c) Tool hints. Dotted line shows the expected time an average-speed reader would take to read that hint. Dot/dash line shows the median reading time for S1. ..................................................48

Figure 5.6: S1’s scan path for Definition hint #15 showing a more scattered pattern of attention (left). Enlargement of hint region of this hint (right)...........................................50

Figure 5.7: S1’s scan path for Definition hint #9 showing a more regular reading pattern (left). Enlargement of hint region of this hint (right)..................................................50

Figure 5.8: S1’s scan path for Bottom-Out hint #7 (left). Enlargement of hint region of this hint (right)..............................................................................................................51

Figure 5.9: Total fixation time spent by subject 2 on a) Definition, b) Bottom-Out, and c) Tool hints. Dotted line shows the expected time an average-speed reader would take to read that hint. Dot/dash line shows the median reading time for S2. ..................................................54

Figure 6.1 Pre and post-test scores for the 12 subjects given as percentages. .........................59

Figure 6.2: Average time (and SD) between hints for each of the participants.........................60

Figure 6.3: Wrong moves made by the subjects including the proportion of wrong moves that were made after receiving a hint (solid blue portion) which average 37.7% (SD = 11.4).....60
Figure 6.4: Percent of Tool hints that were followed by the use of the magnifying glass..... 61

Figure 6.5: Interaction effects between (Left) Time of Hint and Attitude; (Middle) Time of Hint and Hint Type; (Right) Move Correctness and Hint Type. ......................................................... 65

Figure 6.6 Main effects of: (Left) Hint Type; (Right) Attitude. ......................................................... 67

Figure 6.7: Main effect of Pre-test Score. ........................................................................................................ 67

Figure 6.8: Interaction effects between (Left) Move Correctness and Hint Type; (Right) Move Correctness and Time of Hint. ........................................................................................................ 68

Figure A.1: Sensitivity and specificity of the model for each student. ......................................................... 97
Acknowledgements

I would like to thank my supervisor, Cristina Conati for her support and guidance through this project. Her assistance is greatly appreciated.

I would like to thank those members in the lab who have helped me through my time here at UBC. I especially would like to thank Samad Kardan and Alireza Davoodi who were always available to help with the user studies even on short notice and provided me with feedback when needed. Also to the other graduate students who helped me test the Prime Climb game as we were rewriting it, their feedback and bug detection was most welcome.

Lastly I would like to thank my parents, who have always encouraged and supported me in my endeavors.
Chapter 1: Introduction

As computers and technology have become more commonplace in our society, we are finding that children are spending more time than ever playing computer games. In the latest survey on Canadian youth by the Public Health Agency of Canada, they report that about one-half of boys and one-third of girls play video games for two hours or more per day on average [8]. Children play computer games because they are entertaining but it is also possible to make games educational by designing games that teach academic concepts. These educational games (edu-games from now on) are a promising media for the development of innovative computer-based pedagogy, however while there is ample evidence that edu-games are highly engaging, there is less direct support for their educational effectiveness [12, 32].

We believe that edu-game effectiveness can be improved by making them more adaptive to the specific needs of individual students. We do this by devising intelligent pedagogical agents that can provide individualized support to student learning during game play. It is challenging to provide this support because it requires a trade-off between encouraging learning and maintaining engagement. Another difficulty with edu-games is the fast paced environment which can cause players to rush to complete levels by using heuristics and/or guessing to progress the game rather than engaging in proactive reasoning about the underlying concepts.

One of the most common forms of adaptive interventions is to provide hints designed to gradually help students through specific educational activities when they have difficulties proceeding on their own [34]. Despite the wide spread usage of adaptive hints, there is a growing body of research showing their possible limitations. These limitations range from students “gaming the system”, i.e. using the hints to get quick answers to solve the problems
[4] to help avoidance, i.e. players not using hints altogether [1,29]. Our research is looking at this second issue, specifically we wish to gain a better understanding of which factors may affect a player’s tendency to attend to adaptive hints which the student did not explicitly request.

We use Prime Climb, an educational computer game, as the test-bed for our research. Prime Climb was originally developed in the 1990’s by the Electronic Games for Science Education (EGEMS) group at the University of British Columbia to teach number factorization to children. Since then, extensive research has been devoted to add to the game, the ability to provide individualized support by providing user-adaptive hints [9, 10].

In previous studies [10], attention to hints has been measured by hint display time (the time a hint is open on the screen) but we believe we can get a more accurate measure of attention by using eye-tracker data. Therefore in this paper, we will explore the factors which impact a player’s attention to hints while using eye-tracking data to measure attention.

1.1 Thesis Goal

The goal of this thesis is to explore how user attention patterns to user adaptive hints are impacted by factors relating to user’s existing knowledge, hint type, hint timing, and attitude towards getting help in general. We are also interested in exploring how attention to hints affects subsequent game play.

1.2 Contributions of this Research

In this thesis we contribute to the fields of user-adaptive intervention and intelligent tutoring by using gaze information to understand if and how users attend to an edu-game’s adaptive interventions. There are three main contributions of this work.
• While previous work on help avoidance focused on capturing and responding to a student’s tendency to avoid requesting hints [1, 29], we investigate how students react to unsolicited hints.

• We are investigating attention to adaptive hints during the interaction with an edu-game whereas most previous work on student usage or misusage of hints has been in the context of more structured problem solving activities. Providing didactic hints in edu-games is challenging because it requires a trade-off between fostering learning and maintaining engagement.

• We use eye-tracking data to study user attention patterns to adaptive hints, an approach not previously investigated in hint-related research.

1.3 Methodology

We chose to use Prime Climb, an educational computer game, as an experimental platform to evaluate aspects of the game’s adaptive nature. Prime Climb is a game aimed at teaching number factorization skills. In order to support our evaluation plans, we first needed to re-implement the original version of the Prime Climb game because the games suffered by run-time stability issues. We also wanted to make the game more modular and flexible to facilitate rapid prototyping of alternate game versions. This would help us evaluate the game’s adaptive nature better as we could modify parameters and hinting strategies as well as test multiple variations of the game at the same time. After re-implementing the game, we used the new Prime Climb to determine if and how students were processing the agents’ personalized hints.
1.4 Outline

In Chapter 2, we discuss related work in the areas of educational games, intelligent tutoring systems and use of the eye-tracking technology in Intelligent Tutoring Systems (ITS) and Human Computer Interaction (HCI). In Chapter 3, we introduce the edu-game we used, Prime Climb, while in Chapter 4, we discuss the work we did in re-implementing the game to provide a more flexible test-bed for research. In Chapter 5, we discuss our initial work in examining the eye-tracking data generated by two subjects. We then expand this work by exploring patterns of attention with a larger data set of subjects in Chapter 6. Finally we review the achievement of our thesis goal and provide suggestions for future work.
Chapter 2: Related Work

2.1 Educational Games

Edu-games are a promising form of computer-based education; however while there is ample evidence that they are highly engaging; there is less support for their educational effectiveness. In the paper by de Castell and Jenson [12], they suggest that we have to rethink what game-based education should look like and that test scores may not be the most appropriate measure of the educational value of games. Van Eck [32] provides an overview of the pedagogical potential of edu-games but the results are limited. Linehan et al. [20] suggest that a lack of theoretical and methodological rigor in the design of edu-games is the reason for the current lack of success. They propose the use of Applied Behavior Analysis (ABA), a personalized method of teaching developed by behavioral psychologists, as a medium that could take advantage of the benefits of computer games.

2.2 User-Adaptive Edu-games

User-adaptive edu-games have received increased attention as a way to improve edu-game effectiveness. However, most of the existing work has not been formally evaluated in terms of how adaptive edu-game components affect edu-game effectiveness. Peirce et al [27] discuss the adaptive system (ALIGN) used in the Elektra game. ALIGN uses rule based and other probabilistic methods to provide adaptations in the form of feedback and hinting. They found significant results in the evaluation of the Elektra game and positive though not statistically significant results for the learning outcome and flow experience. Easterday et al [14] compared the effects of providing assistance (in the form of a tutor or game-based approach) on learning and interest of Policy World, an edu-game which teaches policy argument skills. Assistance included situational feedback, error flagging and penalties for making errors while
tutoring provides knowledge based feedback and made students correct the errors immediately. They found that providing assistance increased competence which helps with learning and interest. However adding tutoring increased their interest but did not affect their learning.

Conati and Manske [10] compared a version of Prime Clime with no agent with two versions with agents that differed in their accuracy. While they found no significant difference in learning between the three conditions, they did observe that students paid more attention (based on estimating the time the hints were open) to the hints provided by the less accurate agent. They attribute this to the less accurate agent providing fewer and simpler hints (and therefore fewer interruptions). In our work, we extend the work by using a more accurate measure of attention, eye-tracking data, to better understand if and how users attend to Prime Climb’s adaptive interventions.

2.3 Intelligent Tutoring Systems

Intelligent tutoring systems (ITS) are systems that provide customized feedback or instruction to users while they perform the required task. One common way this feedback is given is by using adaptive incremental hints, but their effectiveness is in question because of extensive evidence that students can misuse them. There are two main categories of help misusage that have been investigated so far in the context of ITS for problem solving.

The first is gaming the system where we see students repeatedly asking for help or purposely entering wrong answers to get to bottom-out hints which explicitly tell the student how to perform a problem-solving step and move on. Baker et al. [4] describe this behavior and compare 6 systems (described in more detail in [1, 3, 5, 6, 17, 33]) that can accurately detect in real-time when a student is exhibiting this behavior. These systems were also able to intervene to reduce this behavior. They found that there were two distinct categories of this
gaming behavior (harmful and non-harmful) based on the learning gain seen by the students. They found that the behavior of those participating in harmful gaming (as they had low/no learning gain) was identifiably different from those that exhibited non-harmful gaming (as they showed learning gains even though they gamed).

The second type of hint misusage uncovered in the literature is help avoidance, where students avoid asking for help even when it is needed. Aleven et al. [1] presented a model that could detect both gaming the system and help avoidance. They also found that help abuse correlated negatively with learning while help avoidance did not correlate with learning. In Roll et al. [29] this model is used to generate hints designed to improve students’ help seeking behavior in addition to hints that help with the target problem solving activities. They showed that their Help Tutor was able to improve hint usage behavior and reduce help abuse but they found this positive effect was a result of student following its advice rather than the students learning the principles of good help seeking behavior. Not much work has been done on understanding if and how students process adaptive hints that they have not elicited but Roll suggests that students often ignore these hints. A similar hypothesis was brought forward by Conati and Manske [10] based on preliminary results on student attention to hints in Prime Climb; however, attention was not measured using an eye-tracker but was calculated based on the time the hint was open on the screen.

2.4 Eye-Tracking

There has been a rising interest in using eye-tracking to gain insights on the cognitive and affective processes underlying a user’s performance with an interactive system. Conati and Merten [11] explored using eye-tracking data for the real-time modeling of user self-exploration and effective exploration. They found that adding gaze shifts to the model im-
proved the assessment of self-explanation and exploration but the use of pupil dilation was not a reliable predictor of self-explanation. Muldner et al. [26] used eye-tracking data, specifically pupil size, to inform a user model on the users’ affect and reasoning. They found that users had significantly larger pupil size when expressing positive affect. In addition they found that subject’s reasoning impacted pupil response but may need additional information to distinguish between self-explanation and analogical reasoning.

Bee et al. [7] track a user’s eye gaze to measure their level of attention during the interaction with the agent. They were able to identify two groups of individuals when they were interacting with the virtual character: First “starers” are users who continuously focus on the virtual character’s face; and second “non-starer” which do not stare and have a more natural behavior towards the character. Those in the “starer” group experienced reduced feelings of social presence while those in the “non-starer” group showed significant variations in their gaze behavior and behaved more naturally with the virtual character. They also found that the “non-starer” group showed a different eye gaze behavior depending on whether their interaction mode was interactive or not.

In Prasov and Chai [28], they use gaze information to improve their application’s ability to identify specific entities (in this case items in a bedroom scene) referred to by linguistic expressions. They found that the use of the eye gaze could improve performance especially in cases where there was little domain information.

Loboda and Brusilovsky [21] used the cWADEIn application which supports students in learning expression evaluation in the C programing language through explanatory visualization. Their user-adaptive version adapts the speed of animations based on the progress that student has done so far (i.e. more progress results in a faster pace of animations). They found
the user-adaptive version allowed students to explore expressions at a higher rate. An exploratory eye movement analysis showed the adaptive version engaged students more and attracted their attention more.

In this work, we extend the use of gaze information to understand if and how users attend to an educational game’s adaptive interventions.
Chapter 3: Prime Climb Game

This chapter describes Prime Climb, the game we are using as a test bed for our research on user adaptive game. It was designed to teach number factorization to students in grades 5 to 7. It was originally developed by the Electronic Games for Science Education (EGEMS) group at the University of British Columbia in the 1990’s. Since then, extensive research has been devoted to add to the game, the ability to provide individualized support by providing user-adaptive hints [9, 10]. In this section, we will provide an overview of the existing game which includes the rules for playing the game as well as the tools the player has while they are playing the game.

Figure 3.1: The Prime Climb game.

3.1 Description of Prime Climb

Prime Climb (as seen in Figure 3.1) is a number factorization game where two players work cooperatively to climb a series of mountains. Each mountain is divided into num-
bered hexagons which the players can click on to make their way up the mountain. On some mountains there are obstacles represented by rocks, trees, bees, or buffalos which players cannot select and must go around. Both players are connected to each other with a “rope” that restricts them from moving more than two hexagons away from their partner and acts as a “safety line” when a player makes a mistake. To help the players, the game highlights all hexagons that a player can reach in green (i.e. hexagons that are adjacent to their current position and are not further than two hexagons away from their partner and don’t contain an obstacle). To successfully move up the mountain, players need to select a green highlighted numbered hexagon which does not share a common factor with the number their partner is on. If the player chooses a hexagon that shares a common factor with the hexagon their partner is on, the player will fall (as seen in Figure 3.2). They then will have the option to choose one of the three hexagons they are swinging over. If the player is less than two levels from the bottom, the player when they fall will fall off the mountain and both players will restart at the bottom of the mountain. In addition, the game does not require the players to take turns but allows the players to move in any order. Lastly when either player reaches the top of the mountain, both players move on to the next mountain.

Figure 3.2: Player swinging over three highlighted numbers after having fallen due to choosing a number that shared a common factor with the partner (red player).
The mountains (currently 12) are designed to provide the player with increasing challenge as they play by increasing both the height of the mountain (and therefore the number of hexagons on the mountain) and the complexity of the numbers on the mountain. The height of the mountain starts at 5 hexagons in height and then increases by 1 hexagon each level. On the seventh mountain, the mountain height resets to 5 hexagons in height and then continues to increase by 1 as before. This reset in height also corresponds to when the complexity of the numbers on the mountains change as the first six mountains contain numbers under 100 while on the later mountains, they start to include three digit numbers. The largest number found on any mountain is 387.

![Figure 3.3: Magnifying glass tool showing the factorization of 27. The left button lets the player select the number they want displayed in this tool. The right “?” button accesses a help menu.](image)

### 3.2 Tools

The Magnifying Glass tool, found in the upper right corner of the game, allows a player to view the factor tree of any number on the mountain. The player accesses the tool by clicking on the magnifying glass symbol on the device (left button in Figure 3.3) and then clicking on the number the player wants to view. A factor tree is one way to visualize the factors of a number in a way that shows its prime factors and the intermediate non-prime numbers that result from the recursive decomposition of a number and its factors. The factor tree is generated by decomposing the number of interest, $z$, into two numbers, $x$ and $y$ such
that \( z = x \times y \). This is repeated recursively on \( x \) and \( y \) until the prime factors is reached. Figure 3.3 shows the factorization of 27. As there can be more than one possible way to decompose a number, the magnifying glass choses the decomposition that resulted in the most balanced factor tree to simplify its display on the device.

### 3.3 Pedagogical Agent

Each player of Prime Climb is provided with a pedagogical agent known as agent Monkey (found in the upper left corner of the game screen as seen in Figure 3.1). The agent provides unsolicited individualized support to the player as she plays the game and also provides help on demand by letting the subject ask for more hints after a hint is given. The hints are displayed in a speech bubble, containing two buttons which allow the player to choose to either receive another hint or to resume playing the game.

These hints are given after both correct and wrong moves based on a model of student learning which will be described in a following sub-section. The hints are designed to provide help when the model predicts that the subject does not know the factorization of one of the numbers involved in the move (correct or incorrect) or does not understand the concept of common factors. In addition to deciding whether to provide a hint, the agent also makes a decision on whether to provide a factorization or common factor hint. The hinting strategy will be discussed in the next sub-section.

#### 3.3.1 Hinting Strategy

The hinting strategy for Prime Climb can be divided into two categories: hints on factorization and hints on common factor concepts. The agent decides which category of hint to provide (common factor or factorization) based on an algorithm [22] which assesses the probability that the subject understands the factorization of each number involved in the last
move as well as the probability that the subject understands the concept of common factors with preset thresholds. If the probability falls below the threshold, a hint is provided. More information on the finer details of this topic can be found in previous work [22]. In practice, most players only see the hints on factorization as it often takes a long time for the assessment of the player’s knowledge of common factors to fall low enough to get below the threshold (i.e. the probability the player does not understand common factors is below 0.5 if the move was wrong or 0.1 if the move was correct) to trigger a hint.

The hints (from both hinting categories) can be grouped into four hint types: Focus, Tool, Definition, and Bottom-out. Examples of each type of hint can be seen in Table 3.1.

Table 3.1: Examples of common factor and factorization hints.

<table>
<thead>
<tr>
<th>Hint</th>
<th>Common Factor</th>
<th>Factorization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus</td>
<td>“You cannot click on a number which shares a common factor with your partner’s number”</td>
<td>“Think carefully how to factorize the number you clicked on / your partner is on.”</td>
</tr>
<tr>
<td>Definition 1</td>
<td>“A common factor is a number that divides into both numbers without a remainder. Here is an example…”</td>
<td>“Factors are numbers that divide evenly into the number. Here’s an example…”</td>
</tr>
<tr>
<td>Definition 2</td>
<td>“A common factor is a factor of both numbers. Read this example…”</td>
<td>“Factors are numbers that multiply to give the number. Look at this example…”</td>
</tr>
<tr>
<td>Tool</td>
<td>n/a</td>
<td>“You can use the magnifying glass to see the factors of the number you clicked on / your partner’s number / any number on the mountain.”</td>
</tr>
<tr>
<td>Bottom-out</td>
<td>“You are right because x and y share no common factors” or “You fell because x and y share z as a common factor. x can be factorized as $x_1^{<em>}x_2^{</em>}…x_n$, y can be factorized as $y_1^{<em>}y_2^{</em>}…y_n$.”</td>
<td>“You are right because x and y share no common factors” or “You fell because x and y share z as a common factor. x can be factorized as $x_1^{<em>}x_2^{</em>}…x_n$, y can be factorized as $y_1^{<em>}y_2^{</em>}…y_n$.”</td>
</tr>
</tbody>
</table>
These hints are designed to provide progressive help by starting with the most general hint (*Focus* hints) and get progressively more detailed. Below we describe each hint in more detail:

- *Focus* hints are designed to channel the player’s attention on the skill that requires help.
- *Tool* hints are hints designed to encourage the use of the magnifying glass tool by reminding the player that they can use the magnifying glass to view the factorization of any number on the mountain.
- *Definition* hints are designed to re-teach the concepts of factors and common factors via explanations and generic examples. There are two different definitions for both the factorization and common factor definition hints, which represents two different ways the concepts can be explained (Examples can be seen in Table 3.1). The game alternates between the two hints each time it provides a definition hint. The examples that accompany the definition hint change for every hint, and are designed to help illustrate the given definitions while still leaving it up to the student to find the factorization of the numbers relevant to the performed move. In addition to the generic examples, it is also possible for the system to provide a more relevant example when a player made a wrong move. These specific examples would use the two numbers from the previous move. These specific examples were added to the game based on comments found on the pilot testing questionnaires in that subjects thought that the hints would be more useful if they provided feedback that was relevant to what they were doing at the time.
- *Bottom-out* hints give the factorizations of the two numbers involved in the move (e.g., “You fell because 84 and 99 share 3 as a common factor. 84 can be factorized as...” or “You are right because 3 and 8 share no common factors.”). It should be noted that the
bottom-out hints provided by Prime Climb differ from bottom-out hints from other systems as they do not provide the player with explicit information on the move they should make next but focus on making the subject understand her previous move in terms of factorization knowledge.

3.3.2 Structure of the Student Model

Prime Climb includes a student model which is able to infer a student’s factorization knowledge by using the students’ actions while they are playing the game. One difficulty with using these actions is that they are not an unambiguous evidence of knowledge. For example, if students make a correct move, it could be evidence of their knowledge or it could be a guess. In the case of a wrong move, the move could be evidence that they don’t know the factorization of one or more numbers involved, or they don’t know the common factors between the two numbers or that they had the underlying knowledge but made an error due to being distracted.

To handle this uncertainty, Prime Climb uses a Dynamic Bayesian Network to model the students’ knowledge. The student model contains nodes for each number on the mountain as well as that number’s factors. In addition, it contains a node representing the student’s understanding of common factors. With each move, a node is created which models the correctness of the student’s latest move. With this information, the model is able to assess the probability that the student knows the factorization of a number on the mountain as well as the probability the student understands the concept of common factors. More details about the design and development of the student model can be found in the work of Dr. Conati’s lab [9, 10, 22].
In the Prime Climb game, the student model requires prior probabilities for some numbers when the model is created. Prime Climb can use one of three possible choices for prior probabilities:

- **Generic**: All prior probabilities are set to 0.5 which correspond to the model having no knowledge of whether the player knows the factorization or not.

- **Population**: Prior probabilities corresponding to numbers found on the pre-test are based on data collected from previous studies.

- **User-specific**: For each student, prior probabilities corresponding to numbers found on the pre-test are derived from that student’s pre-test answers. Priors are set to 0.9 and 0.1 for nodes corresponding to correct and incorrect pre-test answers respectively.
Chapter 4: Reimplementation of Prime Climb

As mentioned in the introduction, the first step in this work was the re-implementation of the Prime Climb game. This was done to improve its suitability as a test bed for our research. We will start by discussing some issues we had with the original Prime Climb game then proceed to discuss the design of the new implementation as well as my contribution to this work.

4.1 Issues Faced with the Original Prime Climb Game

The original implementation of the Prime Climb game consists of three components (as seen in Figure 4.1): the server which manages the communication between the two players; the Prime Climb Wizard which provides the graphical interface for the investigator which allows them to play the game as well as displays a message indicating the type of hint the other player received; and the Prime Climb client which allows the subjects to play the game and receive hints on the game. The original game was run on Windows 98/XP/7 machines which have Microsoft Agent components installed. It was implemented using C++ and Windows API.
Figure 4.1: Original Prime Climb architecture. A server communicates with two clients – the wizard and client. The wizard is the investigators version of the client that allows them to observe when the client receives a hint. The Prime Climb client contains a pedagogical agent which provides user-adaptive hints as they play.

There were several issues with this system that prompted our decisions to re-implement the Prime Climb game which are:

- **Instability**: The Prime Climb game would crash frequently both when the game was first started as well as during the game itself. This is at least partially related to using Windows 7 as the operating system that resulted in underlying changes in the operating system that interfere with the communication between machines. Not only did the crashes interrupt the game for the student but can cause frustration and require the students to repeat portions of the game as there are not any options to start the game at a later level.

- **Inefficient network communication**: Prime Climb uses broadcast network communication to send messages between the two clients. This is an inefficient method as it sends all messages to every machine connected on the network. This also prevents the server from handling more than one game at a time. Playing more than one game on a network re-
quired the recompilation of all three components of the game to use a different network port so that the two games would not interfere with each other.

- **Inflexible game features**: All the features in the game (for example: the text of the hints, parameters for the user model, and the threshold values for providing hints) are hard coded in the game and any changes require recompilation and then the new executable would have to be distributed to any computer running the game. This makes prototyping of alternate versions more difficult as extensive coding was needed to make changes.

- **Planned Obsolescence of Game Components**: The graphical face of the user agent is provided by Merlin, an animated wizard who gives the adaptive hints to the students as they play. Merlin is implemented using Microsoft Agent; however; with the introduction of Windows 7 operating system, Microsoft announced that it would no longer include Microsoft Agent components in the Windows operating system and discontinued support for it. They later provided an installation package for Windows 7 due to customer feedback but also announced that Microsoft Agent would not be supported in any future version. As a result, the original game would only work so long as we use Windows 7 operating system.

- **Restrictive computer requirements**: The game is limited to run on Windows machines which have Microsoft Agent installed. This means that when we want to run a user study in a school, we are limited to running the game on our own machines as many schools run Apple computers or the computers they do have lack Microsoft Agent installed.

### 4.2 Goals of the New Prime Climb Game

The following are the goals we had when we re-implemented the Prime Climb game. First we wanted to replicate the features present in the original game but also make it more
modular and flexible to facilitate rapid prototyping of alternate game versions as testable research hypothesis. Second, we chose to move to a web platform as it allows us to distribute the game more easily as the player only needs to access the webpage. Third, the game needed to be able to handle multiple game sessions at a time and allow researchers to be able to modify relevant game components, such as game levels, interactive hints, and character behaviors on the fly from an administrator interface. Lastly, we wanted these changes to be made without requiring the game to be recompiled or for the researchers to need to be familiar with the code itself.

4.3 Game Architecture

The new implementation consists of three parts: the server, the client(s), and the database. A fourth component, an administrative application was planned but has not been implemented at this point in time. The code for all components resides on the same machine (the server) but the client is accessed by users in the form of a web page. Figure 4.2 illustrates the interactions between the various components and also describes the technologies and/or languages used to build these components. In the next sections we will describe each component in more detail.
Figure 4.2: Game architecture for the new Prime Climb implementation. The server, database and admin application reside on the same machine while the clients access the server through the internet.
4.3.1 Server

This is the backbone of the entire Prime Climb game. It consists of two layers:

1. Communication Layer: This layer manages the communication between the components of the game.
2. Application Layer: This layer handles the processing of game play, state, logging and user models.

The Communication Layer manages all the communication from the server to the other components (clients, database and all other future components). We used Windows Communication Foundation and C# to implement the communication features. We also used Entity Framework components to simplify the communication between the server and the database as it allows us to use existing APIs to access the database. The communication layer takes messages coming in from the clients and passes the information to the application layer to be processed and acted upon as needed.

The Application Layer handles the processing of game play, action logging, and user models. This section was written in C# and used Entity Framework to communicate with the database. The application layer also implements the pedagogical agent using the SMILE reason engine for graphical probabilistic models, developed by the Decision Systems Laboratory of the University of Pittsburgh. SMILE (Structural Modeling, Inference, and Learning Engine) is a library of classes that implements graphical probabilistic models like the Dynamic Bayesian Network we use to model user knowledge. When a correct or incorrect move is made, the application layer will update the game state for the effect of that move. It will also send an update to the user model and make a decision on whether to provide a hint.

We chose to use this technology for several reasons; first, these components are designed to work together. Windows Communication Foundation supports the use of Microsoft
Silverlight which we chose to use to implement the client application. The technology chosen also allow us to run on any computer with the Silverlight plug-in which means it will work with both Windows and Apple computers. This was a bonus as it provides us with more flexibility in the computers we can use to run experiments. In addition, it was compatible with the software library which we used to implement the Dynamic Bayesian network. Lastly one member of the research group during the initial design phase was familiar with programming with these technologies and advocated that it would meet our needs for the re-implementation.

4.3.2 Client(s)

Users of the Prime Climb game directly interact with the client application as it is the graphical front end of the system. While users access the client through a web-based interface, the client itself is implemented using C# with Microsoft Silverlight components. Silverlight was chosen as it is a technology that allows rich graphical applications. The client consists of 3 different screens:

1. Log in / Registration Screen (Figure 4.3): This display gives the user an opportunity to log in with a pre-existing user name or to create a new user by filling in the required information (user name, name, grade, and gender).
Figure 4.3: Log in screen for the Prime Climb game which allows for users to either register or log in with an existing username.

2. Lobby Screen (Figure 4.4): This display provides information on the player (name, grade and gender), other players logged into the game and a mechanism for players to join an available game session.

Figure 4.4 Lobby screen of Prime Climb, showing two players sitting in the first game room ready to start a game.
3. Game Screen (Figure 4.5): This displays the prime climb game. User actions are sent back to the server where they are used to advance game play and get recorded into log files for future analysis:

![Game Screen](image)

**Figure 4.5:** Game screen of Prime Climb showing a definition hint in the upper left part of the screen as well as the effect of using the magnifying glass to show the factorization of 40 in the upper right corner.

### 4.3.3 Database

The database is used to store action logs from the clients, user model data, and authentication data. It also stores the information used by the client to build the mountains, the text of the hints which are provided to the player and various parameters which can be modified by researchers to change the way the user model works and hints are provided. Storing this information in a database rather than the server code allows researchers to modify the parameters dynamically and removes the need to recompile the code when a change is made as you only need to restart the server for these changes to become active.

We chose to use Microsoft SQL as it provides simple integration with the server technology through the use of Entity Framework. This framework reduces the need to write code for the interactions between the server and the database.
The database itself consists of 16 tables that work together to store the relevant data for the Prime Climb game. The tables with their relationships are shown in Figure 4.6. In the next part, I will describe each table and its function in relation to the game.

### 4.3.3.1 Database Tables

- **Config**: This table allows us to store various parameters of the game which we may want to modify on the fly. We use the config table to direct the game on how many game instances are available as well as how many game versions are represented. We want to be able to run more than one version of Prime Climb as it allows us to run comparative studies on various elements of the Prime Climb game. The config table sets the threshold values for when to provide hints as well as the parameters for initialization of the Dynamic Bayesian Network which implements the pedagogical agent. We are also able to control
what values are used to initialize the prior knowledge of the pedagogical agent. We have the choice of 3 types of priors: Generic (which sets all prior knowledge to 0.5); Specific (which uses the results of the pre-test to set the prior probabilities for the numbers involved in the pre-test); and Population (which uses prior study data collected to set the prior probabilities for the numbers involved in the pre-test). Lastly there are settings which allow us to control whether both players receive hints. This setting allows us to turn off the hints to the investigator when an investigator is playing with a subject. This reduces the amount of distraction faced by the subject as they will not be interrupted when the other player would have received a hint.

- **PriorKnowledge**: This table stores the prior values used to initializes the prior knowledge nodes of the pedagogical agent. There are three possible settings: generic priors which are set to 0.5; population priors based on data we have collected from previous studies; or user specific priors which we can set based on the factorization test administered to the individual player.

- **LogAgentModels**: This table records each move as it records the numbers each player is on, which player made the move, the time they made the move, and whether the move was correct. It also records the pedagogical agent’s assessment of the factorization knowledge that the current player has of (i) the number she is on; and (ii) the partner’s number.

- **LogFactors**: This table stores the last assessment made by the pedagogical agent on the factorization knowledge of a number.

- **LogModels**: This table stores the last assessment made by the pedagogical agent on the player’s probability of understanding the concept of common factors.
- **Mountain**: Each entry in this table provides the information needed to build a single mountain. In addition to the height of the mountain, we store as an array of integers, the numbers that go in each hexagon ordered from the bottom left corner and going across each row before moving to the next layer. We also store the id of the picture which should be used as the background for that mountain as well as the x, y coordinates of where the mountain should be located on the screen.

- **Hints**: This table stores the text of the various hints, whether the hint needs to substitute the numbers related to the previous move and whether it requires the calculation of factors (which have to be calculated based on the numbers related to the previous move). We are also able to make a hint inactive (i.e. not given to the players) which allows us to control which hints a player may encounter.

- **HintType**: This table is currently being used to identify whether a hint (whose text is stored in the Hints table) corresponds to a hint which is given after a correct or incorrect move.

- **HintCategory**: This table refers to the category that the hint falls into. Currently there are the following categories of hints: Focus, two types of Definition (Factorization or Common Factor), Tool and Bottom-Out. (Note: in hindsight, this would have been better named Hint Type) ¹

- **Users**: This table includes the demographic information on the subject (grade, gender) as well as the players name and a username which allows them to log into the game server.

---

¹ This can be confusing as we use Hint Type in our study to describe the hints found in HintCategory. This is an artifact of a misuse of terminology made during the development process.
• **LogGames:** This table ties the many components of the game together. It stores the user id of the two players playing the game as well as the time the game started and the hinting strategy being used for that game. The game id is used by other tables to identify the unique game associated with that information.

• **LogHints:** This table keeps track of the hints that the player receives while they play the game. It records which hint it was, the time the hint was first displayed as well as the time it was closed. It also keeps track whether the user requested another hint or not.

• **LogMovesTypes:** This table stores the type of move that a player has made. The available move types include *Restart* (when they start or restart a mountain), *Position* (used at the start of a mountain to show where each player is at), *Move_Success* (when the move they make is valid and occurs), *Move_Fail* (when the player tries to move somewhere that is invalid like an obstacle and as a result fails to make a move), *Fall* (when a user makes a wrong move and falls, it records where the user ends up) and *Mag_Used* (when a user uses the magnifying glass).

• **LogMoves:** This table records all the moves made by both players.

• **Debug:** This table was used for implementation purposes to help debug the program.

### 4.4 Benefits of Re-implementation and my Contribution

There are several benefits to the re-implementation which are:

• The game is more stable and allows multiple games to be played at the same time.

• The game is more flexible as it allows the investigator to easily modify parameters and does not require recompilation of the program to make changes to the game.

• The game allows the investigator to access game logs through the database.
• The game is web based and no longer requires specific software installation to play (though it does require a single browser plug-in but this can be installed automatically when the user first accesses the site if they don’t already have the plug-in installed)

My contribution to the re-implementation of this game is the following:

• Supervised and supported the undergraduate student who implemented the game.
• Contributed bug fixes to the code when bugs were identified.
• Re-implemented the pedagogical agent for the game.
• Re-implemented the adaptive hinting strategy for the game.
Chapter 5: Preliminary Analysis of Eye-Tracking Data

We used the new version of the Prime Climb game to investigate if and how students pay attention to the Prime Climb adaptive hints. In a previous study of Prime Climb by Conati and Manske [10], there were indications that student may be ignoring the agent’s hints even when the hints were well justified (i.e. based on a student model that was evaluated to be rather accurate). We also saw evidence of subjects not attending the hints in our pilot testing. In these cases, attention was estimated by how long subjects had the hint open on the screen (display time). However using display time to measure attention is unreliable because the subject may not be attending a displayed hint when it is open, or they could be fast readers and therefore be processing a hint even when display time is short. Therefore we chose to look at a more accurate measure of attention based on eye-tracking data.

We believe that eye-tracking data will be able to provide us a better measure of attention as we can narrow the measure of attention to what they are looking at rather than what was available to be looked at. In this chapter, we will focus on a preliminary analysis we did on the eye-tracker data from two subjects (S1 and S2 from now on), while our next chapter will focus on a larger study involving 12 subjects.

In the preliminary analysis of eye-tracking data, we start by describing the study design used for data collection including the eye-tracker we used as well as some techniques to deal with data loss. Next we discuss how we process the eye-tracker data. Then we discuss the relevant eye-tracking measures we use and finally we will discuss the results from analyzing the data of these two subjects. We will then conclude with a discussion on what we learned.
5.1 Data Collection

We recruited two grade 6 students from the local UBC community who came to the computer science department to participate in the study. Recruitment was conducted through flyers (see Appendix B) distributed to the local school, youth outreach events held by the Computer Science department as well as at local sports camps. Parents and students completed the permission forms (see Appendix C) which were collected before the experiment started. Each child was compensated with a $10 gift card to a local kid’s bookstore.

The game was run on a Pentium 4, 3.2 GHz machine with 2GB of RAM, with a Tobii T120 eye-tracker acting as the primary screen. The Tobii T120 eye-tracker is a non-invasive desktop-based eye-tracker embedded in a 17” display. It collects binocular eye-tracking data at a rate of 60 Hz (or 120 Hz). In addition to the eye-tracking data, it collects all keystrokes and mouse clicks made. It also collects video data of the user’s face.

Participants started by completing a pre-test which tested their ability to identify the factors of individual numbers (16 numbers tested overall) and identify the common factors between two numbers (5 pairs of numbers tested). After completing the pre-test, they underwent a familiarization and calibration phase with the Tobii eye-tracker (described in more detail below). Next they played Prime Climb with an experimenter as a partner. They played the game until they climbed all mountains which took approximately an hour for the two subjects involved. During this period, we also had a second investigator observing a secondary monitor to detect eye-tracking data collection issues as they occurred. Finally, participants took a post-test identical to the pre-test and completed a questionnaire to gain qualitative feedback on their game experience.
5.1.1 Arrangements to Minimize Eye-Gaze Data Loss

The collection of eye-tracking data is susceptible to error due to excessive head movement of the subject as well as subjects looking away from the screen, blinking, resting their head on their hands and other actions which can cause the eye-tracker to lose track of their eyes. [30] discuss some additional factors that cause eye-gaze data loss in relation to a head mounted eye-tracking system but factors like eye characteristics of individuals (i.e., eyeball shape) can be relevant to our eye-tracking system.

We used two methods to minimize the loss of eye-gaze data. First we used a secondary 17” monitor to display eye-gaze traces in real-time. The secondary monitor allowed the investigator to monitor the user’s eye gaze traces while the user was playing which allowed the investigator to detect when the eye-tracker was failing to pick up eye data due to excessive head movements. When we detected loss of eye-gaze data, we asked the subject to shift their position until the eye-tracker started collecting eye-gaze data again.

Secondly, we introduced an eye-tracker familiarization period before the calibration of the eye-tracker. We developed this based on an observation we made when one user who spent more time looking at the pre-calibration screen also had reduced data loss. This period consisted of time for the subject to get more familiar with the eye-tracker and better understand how their actions affect the ability of the eye-tracker to collect data. The eye-tracker has a pre-calibration display (as shown in Figure 5.1) that allows the subject to see their eyes being captured (the two white dots in Figure 5.1). At the bottom of the display, the bar is either colored green (when the eye-tracker is able to capture data), yellow (when there is diffi-
culties with detecting eye gaze but it is collecting some data) or red (when it is unable to capture the eyes). On the right side, it shows how far the subject’s eyes are away from the monitor. When the eyes get too close or too far, the eye-tracker encounters difficulty in identifying the eyes and therefore the bottom display will display yellow and then red.

![Calibration](image)

**Figure 5.1:** Pre-calibration screen which allows subjects to become more aware of the capabilities of the eye-tracker.

We instructed the subjects to move around in their seat to see how far they could move before the bottom bar turned red. We also gave the example of resting their head on the hand as another reason why the eye-tracker would have problems and had the subjects try this out to see the effect. We then gave the subjects some time to “play” with this display. By having the subject move around while observing this screen, they are able to become more aware of how much movement they can make and what actions (i.e. resting their head on the hands) can cause difficulties in capturing eye gaze data. It also made it easier for them to understand when we requested that they shift positions (when we detected the eye-tracker stopped collecting data). We found that not only did subjects find this fun; it also was effec-
tive in improving the quality of the data being collected. We found that this combined with active monitoring of the eye gaze data using the secondary display helped improved the quality of data collected.

5.2 Assessment Tools

5.2.1 Pre-test and Post-test on Factors and Common Factors

The pre-test and post-test the students completed were identical. This test is shown in Appendix C. The tests were originally developed for a previous study [22]. The test consists of two sections. The first section consists of 16 multiple choice questions on the factorization of a number. Students need to select the factors of the number from a list. The second section consists of 5 common factor questions which allow us to examine the student’s knowledge of the concept of common factors. Like the factorization questions, students are given a list where they can select the common factor of the two numbers.

10 of the numbers found in the factorization section also correspond with numbers in the common factor questions which allow us to assess student’s knowledge of the common factor concept by looking at their pattern of response to the factorization and common factor questions. The other 6 questions correspond with numbers that appear in the first four mountains of the game.

The test was marked by giving one mark for each correctly circled response and taking one mark away for each incorrectly circled response. As a result, it is possible for a student to receive a negative score on a test when they selected more incorrect answers than correct ones. The highest possible grade on this test was 31.
5.2.2 Student’s Subjective Assessment of the Game

After the post-test, each student was given a written questionnaire (as seen in Appendix D). The questionnaire was based on one used in a previous Prime Climb Study [22], but was modified to reflect changes made to the agent (i.e. the name of the agent changed) and included 7 questions which were originally part of the post-test in the previous study. First, we asked questions related to the agent, and the student’s opinion on how the agent performed and how the agent could be more useful to them. Next we asked some general questions towards their attitudes towards help and playing the game. Next we asked when they did (or did not) have fun while playing and lastly we asked when they learned math while playing. Most questions allowed students to respond on a Likert scale where 1 indicated “Strongly Disagree” while 5 indicated “Strongly Agree”, but there were some open-ended questions which allowed us to gather feedback on how the subjects thought that the game / agent could be improved. In addition to using the information to guide future development of the game (i.e., how can we apply their suggestions), we chose to use one question “I wanted help when I was stuck” as a measure of the student’s attitude towards help.

5.3 Eye-Tracker Data

Eye-gaze information is provided in the form of fixations (i.e., eye-gaze remains at one point on the screen) and saccades (i.e., eye-gaze moves quickly from one fixation point to another), which we can use to derive attention patterns of the user. Both fixations and saccades are made up of multiple samples from the eye-tracker (i.e. fixations are made up of all consecutive samples that relate to the eye focusing on a single place on the screen while saccades are made up of all consecutive samples between two fixations). In this section we will first describe some details of the raw data we collected from the eye-tracker, then how we
processed the raw data into measures which we can analyze. We will also give an overview of measures we can derive as well as the subset we decided to focus on. Lastly we will discuss data validity and how we determined if we had sufficient information to analyze each hint.

5.3.1 Data Collection

The Tobii eye-tracker stores data in the form of three files that are useful to us. First, the All data file provides the data in the form of individual samples from the eye-tracker. Each sample gives the coordinates of the raw eye data, the pupil size, and the quality of that sample (whether that sample had data for one or both eyes). In addition, it includes the length and coordinates of the fixation which that sample is associated with. In addition to the samples, it records any event (i.e. mouse clicks, keyboard presses, user created custom flags, etc.) which occurred. The next two files are subsets of the All data file which allows for easier data processing. The Fixation data file provides a summary of the fixation information, i.e., the fixation timestamp, duration and coordinates. Last, the Event data provides all key presses, mouse clicks and allows us to flag events such as when the mountains change, when the hint opens and closes, and when the player uses the Magnifying Glass.

5.3.2 Defining Relevant Eye-Data

As the game interaction can be long (up to 65 minutes) and dynamic, we need to divide the game play into portions that are easier to analyze. As we are interested in investigating if and how students are using the hints provided to them in game play, we chose to focus on the portions that include the time that the hint is available to the player. Each hint was tagged with both its sequential order as well as the hint type it was associated with. These short time periods provide a fairly static view of the game as players are unable to make any
moves during this time frame. This simplifies analysis as we don’t have to account for objects moving during the interaction. We originally used the Prime Climb game logs to generate the times for the hint start / end but we found that these times were not adequately synchronized with the Tobii eye-tracker log files providing the eye-tracking data. This is due primarily because the two log files were stored on two different computers (the computer hosting the web game and the computer the subject was playing on). This resulted in a varying mismatch between time stamps which made the results less accurate than desired. As a result, we manually tagged each hint opening and closing using the Tobii Studio software.

5.3.3 Areas of Interest

We defined four Areas of Interest (AOI from now on), that represent the main four components of the game. We will use these AOIs to analyze the attention behavior of these subjects with respect to the received adaptive hints and identify relevant attention patterns from the Tobii eye-gaze data.

These AOIs are shown in Figure 5.2. The Tool AOI coincides with the magnifying glass, whereas two different AOIs are defined over the dialog box used to display hints: Hint and Hint Close. Hint contains the text of the hint message while Hint Close contains two buttons called “More Hints” and “Resume Playing” which allow subjects to request more hints or close the hint and resume playing the game. We separated the hint dialogue box into two AOIs as we wanted to distinguish the actions of observing the hint from the decision to either get another hint or close the hint display. The last AOI is the Mountain which encloses the Prime Climb Mountain currently being displayed.
Figure 5.2: Prime Climb screen showing the four areas of interest: Hint, Hint Close, Mountain and Tool.

Now that we have narrowed the focus we have to determine first what eye-tracking measure to use and then whether we have enough data for each hint to provide a reliable analysis of the data. This is because the arrangements described in section 5.1.1 helped but did not eliminate actions that can cause errors in the eye-gaze data.

5.3.4 Eye-Tracking Measures

Goldberg et al. [15, 16] describes a set of basic eye-tracking features which can be used for comprehensive eye-data processing. Using the software developed for processing eye-tracking data, we are able to extract the following measures: number of fixations, total fixation time, fixations/word, and scan path lengths for the overall interaction (i.e., the time a hint is open). In addition to the above measures, we can also extract the measures: time to first fixation, time to last fixation, and number of transitions between individual AOI.

Two standard measures used by eye-tracking research to measure overall attention [15,16] are number of fixations and total fixation time (i.e., the total number and time a sub-
ject’s gaze rests on the Hint AOI for each displayed hint). In fact, we found that they were highly correlated (Pearson r of 0.93 for S1 and 0.95 for S2). We chose to use total fixation time rather than number of fixations as it also allowed us to compare the time spent with the time one would expect a person to take to read the hint [19]. One weakness to both these measures is that while they give a measure of overall attention to the hints, they do not provide detailed information on how a hint is actually processed (e.g., it cannot differentiate between a player who stares blankly at a hint vs. one who carefully reads each word). Furthermore, it is not ideal to compare attention to the different types of hints in Prime Climb because they have different lengths on average. Thus we also use the ratio of fixations per word (fixations/word from now on) which is a measure that is independent of hint length and gives a sense of how carefully a subject scans a hint’s text. This measure is 1 when the subject fixates once on each word of the text and decreases when a reader starts skipping words. Between these two complementary measures, we are able to look at both the general attention to the hint as measured by fixation time as well as a measure of the amount of care the students are taking with the text with the fixations/word measure.

In addition to the two measures described above, there are several other measures we could have chosen to use. The scan path length measures the Euclidean distance between fixations in pixels. Scan path length can be used to compare scanning behavior where longer scan paths can indicate less efficient scanning behavior of the users. In the work of Goldberg and Kotval [15], they found that when they compared the scan path length of a good and poor interface, the poor design lead to a longer scan paths. However in our case, scan path lengths comparison between hints would not be ideal as it would also be affected like fixation length by the varying length of our hints. Time to first fixation on the Hint AOI in particular might
be useful in detecting how well our hints are at drawing the subject’s attention to the hint but is not able to measure how much attention they gave to the hint. *Time to last fixation* has a similar issue that it does not give any indication of the amount of attention to the hint. Lastly the *number of transitions* between AOIs can be useful in looking at the specific attention patterns involving hints. One use for this measure is when looking at the effect of the *Tool* hints in encouraging the use of the Magnifying glass but in the case of the two subjects, there were no transitions between the *Hint* AOI and *Tool* AOI. It also would be interesting to see how much they go back and forth between the mountain and other AOIs. We also could combine some of these measures however we chose to start looking at just the first ones.

### 5.3.5 Data Validation

As mentioned in previous sections, some eye-gaze data was lost when the students were playing, generally due to excessive head/body movement, which had to be expected as we were using children who have high energy and do not sit in the same position for the entire experiment. This data loss is in the form of invalid samples which do not contain any useful data.

While we knew we had loss over the entire period, what we needed to establish the amount of eye-data lost during the *segments* of interest (i.e., when hints are open). We also need to determine where to set the threshold to exclude segments from analysis due to lack of eye-gaze data.

Two methods developed in our lab to determine validity are time based (i.e., a segment is determined to be valid if the proportion of time we have valid samples is greater than a preset threshold) or gap based (i.e., a segment is valid as long as there is no gap in the sequence of samples greater than a certain length).
We determined validity using the time based method as overall amount of data collected is more important than gaps which might cause hint exclusion. To use this method, we need to determine the threshold by plotting the percentage of segments that get rejected for different threshold values.

![Figure 5.3](image-url)  
**Figure 5.3** Percentage of segments discarded for different threshold values.

We will look at the results for all subjects we collected eye-tracking data, though this chapter will focus on the first two subjects. Figure 5.3 shows the percentage of segments that would be excluded from analysis for each subject given different possible threshold values (i.e. the higher the threshold is set; the more segments will be rejected). We see that one subject has a severe loss of data in which over 60% of the segments had no data collected at all, we therefore chose to exclude this subject from future analysis as the low amount of eye-tracking data would make analysis of this subject unreliable. Based on the remaining samples, we choose to set 75% as our threshold for excluding a segment from analysis (i.e. if a segment contains less than 75% of the possible eye gaze samples, it would be consider invalid). At this threshold, we find that the majority of the subjects have a low percentage of rejected segments (less than 20%), while the worst 4, we still have over half of the segments
for analysis. Therefore, this threshold is a good compromise between providing a good proportion of valid segments while excluding those segments which the data is missing for a large proportion of time. In the case of the two subjects which we will discuss in the next section, a total of 20 segments (13%) were excluded and in most cases, these consisted of segments for which the eye-tracker collected little to no data.

5.3.6 Processing of Eye-Tracker Data

We used two programs to process the raw eye-tracker data into the measures we discussed earlier. The first was a C# program created by the author which used the log files as well as eye-tracker log files to extract the relevant eye-tracking measures (fixation time and count) into a format which could be exported into an excel spreadsheet. This program reported for each hint, the total fixation time and count for each AOI and the total fixation time and count for the entire hint. Fixations per word were then calculated in the excel spreadsheet. It also reported the percent of eye-tracker data present for each hint (total fixation time divided by the time the hint was open).

Later, PETL (Python Eye-tracking Library) was created in our lab for the processing of eye-tracker data on another project. We decided to switch to this package as it extracted more eye-tracker features than the C# program and because we wanted to improve the consistency of the eye-tracking processing done between projects. We used the C# program to generate the results in Chapter 5 while PETL was used for the results in Chapter 6.

As PETL was developed for another project in our research group, we needed to make some adaptations to this program to account for the differences in the projects. One important change was to include fixations which were partially contained within the time frame of a hint. The original version only included fixations that were entirely contained
within the time period of interest (i.e. when the hint is visible on the screen) while we extended it to include those that more than 50% of its duration is contained within the time period of interest. 50% was chosen as it would prevent one fixation from being in more than one segment if two segments were consecutive. This was important as many of our hints were of very short duration so excluding these fixations greatly reduced the amount of data for these very short hints. Next we adapted PETL to use the event data log to generate the time segments which correspond to when the hints are open. After providing the program with the eye-tracking log files as well as a file defining the AOIs, the program outputs the information into a format that can be exported in an excel spreadsheet. Once we had the information generated by PETL, we calculate the fixations per word and added some additional information, including whether the hint was caused by a correct or wrong move, whether the move made after the hint was correct or wrong, whether the hint provided was specific to the numbers in the previous move, whether the hint was user requested, the length of the hint message as well as the expected time an average speed reader would take to read the hint. We also labeled each hint on whether it occurred in the first half or second half of the game (based on time).

5.4 Knowledge Test Results

As mentioned earlier, each subject took a pre-test on number factorization as well as a post-test after playing the Prime Climb game. The first subject (S1) scored 30 out of 31 total points in the pre-test, and got full marks in the post-test. The second subject (S2) scored 22/31 in the pre-test, and 25/31 in the post-test.
5.5 Initial Comparison between Fixation Time and Hint Length

We start our investigation by comparing each subject’s attention to each hint type with the expected reading time based on the hint’s length. The agent’s adaptive hints can be divided into three types: Tool hint, which are on average 15 words long with a range between 14 and 16 words; Bottom-Out hints, which are on average 17 words with a range between 11 and 25 words; and the longer Definition hints, which are on average 36 words long with a range between 25 and 49 words. The amount of time it would take an average-speed reader to read the text, assuming a reading speed of 3.4 words per second, would be 4.4, 5 and 10.5 seconds respectively (based on the work by Just and Carpenter [19]).

![Figure 5.4: Comparison between the mean fixation time on the three types of hints and the expected reading time.](image)

Figure 5.4: Comparison between the mean fixation time on the three types of hints and the expected reading time.

We start our analysis by comparing each subject’s attention to each hint type with the time expected to read the hint based on the hint’s length. We use total fixation time as a measure of overall attention. As we can see in Figure 5.4, S1 spent more time looking at the longer Definition hints but there is not statistically significant difference between the reading time for the three hint types (as tested by ANOVA, F(2,54)=0.8668, p=0.42). In addition, S1
shows fixation time that is much shorter than the time an average-speed reader would need to read the hints (expected reading time from now on) but the high standard deviation in each of the categories indicates a trend of selective attention. In the case of S2, we did find a statistically significant difference between the fixation times for the hints which reflected their length difference (as tested by ANOVA, F(2,68) = 3.681, p = 0.03). Pairwise comparisons with Games-Howell adjustment for samples with unequal variance show a statistically significant difference between the Definition and Tool hints (p = 0.022) and between Definition and Bottom-Out hints (p = 0.035). In both cases, S2 spent more time on Definition hint than Bottom-Out and Tool hints. As with S1, we also see that the standard deviation is very high suggesting a similar trend of selective attention. This gave us a direction to look at the pattern of attention for each subject.

5.6 Patterns of Attention to Hints

5.6.1 Subject 1

As mentioned previously, we saw a large standard deviation in the time spent on hints. We can visualize this by showing the total fixation time on each individual hint (separated by hint type). Figure 5.5 shows the pattern for S1. The x-axes represent the sequential hint number in each category. Also shown on the graph is a dotted line corresponding to expected reading time, as well as a line showing the median split on total fixation time for this dataset. We calculate the expected reading time by taking the number of words in the hint and dividing it by the reading speed of 3.4 words per second [19]. We use the median split as a rough indication of subject-specific low vs. high attention. We also marked with a triangle on the x-axis to indicate when the user received a hint due to a wrong move.
Figure 5.5: Total fixation time spent by subject 1 on a) Definition, b) Bottom-Out, and c) Tool hints. Dotted line shows the expected time an average-speed reader would take to read that hint. Dot/dash line shows the median reading time for S1.

For about the first half of the displayed Definition hints, there is a pattern of attention being high for one hint, and low for the Definition hint given as the next step in the hinting cycle. Given that two subsequent definition hints provide two different ways to explain what number factorization is, this pattern suggests that receiving two alternate definitions may be perceived by this subject as being redundant. Subject attention then decreases substantially
for most of the second half of the Definition hints provided. As seen earlier, the total fixation time per hint for this subject is generally below the expected reading time, for all hint types. This could be either due to the fact that the subject is not reading as carefully as she should when she is attending to the hints, or that the subject is a faster reader. To discriminate between these two cases, we look in more detail at the fixations in the high-attention hints (i.e., hints with total fixation time above the median split), to ascertain if there is a difference in patterns for the fixation time higher than the expected time (of which there is one each in Definition and Bottom-Out hints). We use the measure fixations/word as a measure of how much attention the subject puts on each word.

We next looked at all the high attention hints given and divided them into two categories, those above the expected reading time and those below the expected reading time. We found a statistically significant difference between the fixations/word for the high attention hints above the expected reading time ($M = 0.593; SD = 0.067$) and below ($M = 0.300; SD = 0.155$) (two sample t-test, $t = 2.615, p = 0.014$). This result may suggest that the subject is not reading as carefully as she can for those high-attention hints below the expected reading time. However if we look at the scan path in Figure 5.6 (i.e., the complete sequence of fixations generated by the subject while the hint was open) for Definition hint #15 (the only Definition hint with total fixation above expected) we see a sequence of rather scattered fixations between words in different lines, with some fixations returning to previously visited regions. In contrast, the scan path for most of the other high-attention Definition hints (which were below the expected reading time) show a more regular reading pattern from left to right (e.g. Definition hint #9 in Figure 5.7), suggesting that the subject is reading these hints carefully while the higher fixations/word of Definition hint #15 is most likely due to confusion. This is
supported by the more regular pattern seen in Bottom-Out hint #7 (Figure 5.8) which is the only other hint where the subject spent more than the expected reading time on.

Figure 5.6: S1’s scan path for Definition hint #15 showing a more scattered pattern of attention (left). Enlargement of hint region of this hint (right).

Figure 5.7: S1’s scan path for Definition hint #9 showing a more regular reading pattern (left). Enlargement of hint region of this hint (right).
Figure 5.8: S1’s scan path for Bottom-Out hint #7 (left). Enlargement of hint region of this hint (right).

To summarize, gaze data analysis for individual definition hints for subject 1 indicates that this subject is a fast reader (reading times below the expected), who does pay attention to a selection of the definition hints, at least at the beginning of the interaction (Definition hints before #15 with reading time above median)

While attention to definition hints decreases with time, attention to tool and bottom-out hints reaches its low in the middle of the interaction, but picks up again towards the end (see Figure 5.5b and 5.5c). A possible explanation for these trends is that definition hints become less useful overtime, as mountains get more difficult (i.e. include larger numbers), because the subject is already familiar with the factorization definitions by now and the generic examples in the hint don’t help directly with understanding the outcome of the current moves. However, apparently the subject still needs help dealing with the larger numbers, so she does read short hints when they appear and specifically attends to bottom-out hints because they provide the information needed to understand the outcome of the previous move.
As far as Tool hints are concerned, we checked whether tool hints that were actually followed (i.e. that resulted in the subject accessing the magnifying glass after receiving the hint) had been attended more carefully. S1 followed 56% of tool hints received. Of these, 4 had total fixation times below the median split (hints 3, 5, 8, 10 in Figure 5.5c), suggesting that tool hints successfully act as reminders for magnifying glass usage even when the subject simply glances at them.

To conclude our analysis of S1’s attention to hints, we checked if move correctness influences attention to hints. We compare the subject’s attention to hints received after a wrong or a correct move, using the fixations/word measure to take into account length variations in hints of the same type (e.g., Definition hints given in response to correct move are longer than Definition hints for incorrect moves, while Bottom-out hints for correct moves are shorter than Bottom-out hints for wrong moves).

Table 5.1: Statistics on the mean fixations/word for each hint type for the two subjects.

<table>
<thead>
<tr>
<th></th>
<th>Tool Hint</th>
<th>Definition Hint</th>
<th>Bottom-Out Hint</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixations/word - correct moves (SD)</td>
<td>0.340 (0.185)</td>
<td>0.191 (0.141)</td>
<td>0.424 (0.094)</td>
</tr>
<tr>
<td>Fixations/word – wrong moves (SD)</td>
<td>0.274 (0.054)</td>
<td>0.158 (0.146)</td>
<td>0.243 (0.170)</td>
</tr>
<tr>
<td><strong>S2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixations/word – correct move (SD)</td>
<td>0.547 (0.293)</td>
<td>0.383 (0.286)</td>
<td>0.595 (0.311)</td>
</tr>
<tr>
<td>Fixations/word – wrong move (SD)</td>
<td>0.487 (0.336)</td>
<td>0.201 (0.211)</td>
<td>0.271 (0.269)</td>
</tr>
</tbody>
</table>

The first two rows in Table 5.1 show the results for S1. We found no statistically significant differences between correct and wrong moves for Definition or Tool hints but there were statistically significant differences for Bottom-out hints (M = 0.424 SD = 0.094 for cor-
rect moves; $M = 0.243$ SD = 0.170 for wrong moves; two sample t-test, $t = 2.143$ p = 0.039). These results show that, perhaps surprisingly, S1 takes more time to read hints for correct moves than those for wrong moves in the case of Bottom-out hints. A possible explanation for this finding is that receiving hints after a correct move is somewhat unusual, thus these hints stimulated the subject’s curiosity and subsequent attention, perhaps with the hope that the hints could be tips to further improve S1’s performance with the game.

5.6.2 Subject 2

In the case of S2, we again see a very high standard deviation in total fixation time. In fact the plots of fixation time for individual hints show that, while there is essentially no fixation time on some of the hints, fixation time for others is substantially higher than the expected reading time (see Figure 5.9). This is especially true for Definition hints, where for 8 of the hints, fixation time is well above the expected reading time (see Figure 5.9a). This data suggests that this less knowledgeable subject is willing to devote substantial attention to some of the Definition hints in order to make sense of them. S2, however, also shows the patterns of alternating attention to successive Definition hints shown by S1 (up to about hint #26), confirming that our strategy of presenting both types of Definition hints for every hinting cycle may be over kill. Like S1, S2 also shows the pattern of consistent diminished attention to Definition hints later in the interaction (with the exception of one spike), confirming that Definition hints appear to lose subject attention when mountains get more difficult.

Contrary to S1, who only received unsolicited hints, S2 also requested hints while playing, however of the 2 hints requested (Definition hints 17 and 37), she only spent substantial time on the first requested hint. This result is interesting because it shows the added value of eye-tracker information, even if the presence of interaction events that may be con-
sidered highly informative. That is, requesting a hint should be a fairly strong indication that a subject will pay attention to that hint, but this was not the case for S2, and eye-tracking data was able to detect this discrepancy.

Figure 5.9: Total fixation time spent by subject 2 on a) Definition, b) Bottom-Out, and c) Tool hints. Dotted line shows the expected time an average-speed reader would take to read that hint. Dot/dash line shows the median reading time for S2.
In the case of *tool* hints, S2 shows a pattern of consistent attention between hint 7 and hint 10 (see Figure 5.9c), which corresponds to a period of frequent usage of the magnifying glass with 4 accesses out of the total 18 performed by S2. Overall S2 followed *tool* hints less frequently than S1 with 6 out of 21 (29%) hints followed. Two of these hints were below the median split, showing, as was the case for S1, long attention is not necessary for the tool hints to be effective reminders.

For *bottom-out* hints (Figure 5.9b), S2 shows a rather consistent pattern of selectively paying attention to about half of them throughout the course of the interaction (fixation time at or above the median split), while just glancing at some of the others and spending a much higher time (about 12 second) on the fourth *bottom-out* hint provided, which could be an indication of confusion.

As with S1, we conclude our analysis of S2’s gaze data by checking if move correctness influences attention to hints. We compare the subject’s attention to hint received after a wrong or a correct move, using the fixations / word measure as we did for S1. The numbers for S2 are reported in the last two rows of Table 5.1. Although there were no statistically significant differences, the trend is similar to the one we found for S1 in that both gave more attention to hints caused by correct moves than wrong moves. For *Bottom-out* hints, there is a marginally significant difference in fixations per word, where hints caused by correct moves have higher fixations per word than hints caused by incorrect moves (*M* = 0.595 *SD* = 0.311 for correct moves; *M* = 0.271 *SD* = 0.269 for wrong moves; two sample t-test, *t* = 1.842 *p* = 0.088).
5.7 Discussion

In summary, we found that both subjects showed a pattern of alternating attention to successive Definition hints, indicating that at presenting both types of Definition hints for each hinting cycle may be redundant. We also found that attention to these hints decrease over time, indicating that subjects find them less and less useful as the interaction proceeds. Another similarity between the two subjects is that Tool hints can trigger usage of the magnifying glass even when they are simply glanced at. Finally, both subjects spent more time on hints caused by correct moves. We also saw that the level of attention was different between the two subjects, with the less knowledgeable subject showing a trend to be spending more time on definition hints than the more knowledgeable one.

While we need more data to draw firm conclusion, the trends mentioned above already allow us to refine indications from previous studies that attention to the Prime Climb agent hints may be scarce. While attention is scarce for some of the hints, it is certainly not for all, showing that some subjects are willing to look at didactic hints even in the context of game play.

One question we need to investigate to improve hint effectiveness is which factors influence a subject’s decision to attend a hint or not. One obvious factor is whether the hint is generated at the appropriate time, i.e., whether the model of student learning that drives hint generation is accurate. This is generally hard to assess since it requires knowing whether the numbers involved are known or unknown by the subject. This information is only known for a subset of all moves, namely the moves that involve numbers that appear in the pre and post-test. In addition, we are also interested in looking at other factors that may affect whether the subject attends the hint or not. Some of these factors can include move correctness (was the
hint caused by a correct move or incorrect move), time of hint (did the hint occur in the first or second half of the investigation) and their attitude towards receiving help and hint type.

We are also interested in learning whether attending to the hints affected their gameplay either by making fewer mistakes after receiving a hint or using the magnifying glass tool after receiving a Tool hint. We will look at these issues in more detail with the study described in the next chapter.
Chapter 6: In Depth Analysis of Eye-Tracking Data

In this chapter, we expand on the work we started in Chapter 5. As we saw in Chapter 5, students showed selective attention to hints. Here we are going to look at what factors influence a subject’s decision to attend a hint or not. Next we will look at how attention to the hints affected game play. We will then conclude with a discussion of results.

6.1 Experimental Design & Data Collection

We collected eye-tracking data from twelve participants (six female) from grades 5 and 6 (six players in each grade). The mechanism was the same as discussed in section 5.1 and our results include the two initial subjects discussed in Chapter 5.

The collection of eye tracking data and extraction of hints from the data remains the same as discussed in section 5.3. We used the same measures for analysis and a discussion of data validity for all subjects is found in section 5.3.5.

6.2 Descriptive Statistics on Game Play

In this section we will look at some of the general statistics collected from the game play. In particular, this will include participant learning from pre-test to post-test, the amount of time spent on the game, move statistics as well as some data on Magnifying Glass usage.

6.2.1 Pre- and Post-test Scores

We did not observe any improvement from pre to post-test performance, with participants scoring an average of 74% (SD = 31%) in the pre-test, an average of 72% (SD = 31%). From Figure 6.1 we can see that some subjects did very well but there were several that performed very poorly on both tests. Except for one student, the test scores between pre-test and post-test were very similar.

---

2 There was one additional subject that was excluded from the study due to a lack of eye-gaze data for that subject as discussed in section 5.3.5. While some subjects may have known each other, students participated at different times and only played with an investigator.
post-test were very similar. In general when a student went down in score, it was usually due to missing one of the factors of a number that they got correct on the pre-test rather than giving an incorrect answer (i.e. circling a number that isn’t a factor). This could suggest that player fatigue might have contributed to the poorer results or that the subjects did not take as much care on the post-test as they could have.

![Figure 6.1 Pre and post-test scores for the 12 subjects given as percentages.](image)

6.2.2 Time Spent on the Game, Number and Type of Hints

The study’s game sessions lasted 33 minutes on average ($SD = 15$ minutes). Consistent with previous Prime Climb studies, subjects rarely asked for help. One subject asked for four hints, two subjects asked for hints twice and two other subjects requested one hint. Prime Climb, however, generated unsolicited hints frequently: an average of 51 hints per player with an average frequency of 37 seconds ($SD = 44$ seconds). Thus, lack of system interventions can be ruled out as a reason for the lack of learning seen. If anything, it is possible that the hints happened too frequently which can cause them to interfere with game play and lead subjects to ignore them. In Figure 6.2, we can see the average time between the hints given varies between the subjects. It should be noted that for two of the subjects who received the most frequent hints were the subjects with the lowest pre- and post-test score which indicate that lack of knowledge was a factor in these subjects receiving frequent hints.
6.2.3 Moves

In addition to the looking at frequency of hints and test performance, we looked at the number of wrong moves made by the subjects and the proportion of wrong moves that occurred after the subject received a hint. From Figure 6.3 we see the percentage of wrong moves they made as well as the proportion of wrong moves that were made after receiving a hint. On average, subjects made incorrect moves 17.5% of the time ($SD = 4.5$). In addition the percentage of those wrong moves which were made after receiving a hint was 37.7% ($SD = 11.4$). So here we see that there was more variation in their actions after receiving a hint than in the percent of wrong moves they made.

Figure 6.3: Wrong moves made by the subjects including the proportion of wrong moves that were made after receiving a hint (solid blue portion) which average 37.7% ($SD = 11.4$).
6.2.4 Magnifying Glass Usage

We were also interested in how the subjects took advantage of the support tools that we provide in the game, namely the magnifying glass (as seen in Figure 6.4). We found that on average, subjects used the magnifying glass 18 times ($SD = 20$). This large standard deviation indicates that while some players used the tool frequently, others used it very infrequently. In fact, 2 subjects did not use the magnifying glass tool at all, while 3 more only used it 5 times. We also looked at the behavior of the players after they received the Tool hint. We found that on average, subjects used the magnifying glass after receiving a Tool hint 20% of the time ($SD = 18$). This suggests for at least some subjects, that Tool hints can be effective in triggering the use of the magnifying glass.

![Percent of Tool Hints Followed](image)

Figure 6.4: Percent of Tool hints that were followed by the use of the magnifying glass.

6.3 Factors that Affect Attention

In this section we look at which factors influence a subject’s decision to attend a hint or not. One obvious factor is whether the hints generated were justified, i.e. whether the probabilistic student model that drives hint generation accurate in assessing a student’s num-
ber factorization knowledge. Next we look at additional factors that may influence student attention to hints in our dataset:

- **Move Correctness** indicates whether the hint was generated in response to a correct or wrong move.
- **Time of Hint** divides the hints into two categories depending on whether they occurred in the first or second half of the student’s interaction with the game based on the median split over playing time.
- **Hint Type** reflects the three categories of Prime Climb hints: Definition, Tool, and Bottom-out.
- **Attitude** reflects student’s general attitude towards receiving help when unable to proceed on a task, based on student answers to a related post-questionnaire item, rated using a Likert-scale from 1 to 5. We divided these responses into three categories: Want help, Neutral, and Wanted no help, based on whether the given rating was greater than, equal to, or less than 3 respectively.
- **Pre-test score** represents the student percentage score in the pre-test as an indication of student pre-existing factorization knowledge.

We look at these factors in relation to two complementary measures of attention: **Total fixation time** (i.e., the total time a student’s gaze rests on the Hint AOI of each displayed hint) and **Fixations per word** (i.e., the number of fixations on the Hint AOI of each displayed hint divided by the number of words in that hint). As mentioned earlier, total fixation time gives a measure of overall attention to hints, but it doesn’t differentiate between a player who stare blankly at a hint versus carefully reads each word. Furthermore, it is not ideal to compare attention to the different types of hints in Prime Climb because they have different
lengths on average (15 words for Tool hints; 17 words for Bottom-out hints; 36 words for Definition hints). Our second chosen metric Fixations/word is a measure that is independent of hint length and can give us a sense of how carefully a student scans a hint’s text.

6.3.1 Impact of Model Accuracy

One factor that can affect a subject’s attention to hints is whether the hints were justified, i.e. whether the probabilistic student model that drives the hint generation is accurate in assessing a student’s number factorization knowledge. Unfortunately, we can only answer this question for the numbers tested in the post-test, which are about 10% of all the numbers involved in the Prime Climb game. The model sensitivity on post-test numbers (i.e., the proportion of actual positives which are correctly identified as such) is 89%, indicating that the model generally did not underestimate when a subject knew a post-test number and thus it likely triggered justified hints on them.³

It should be noted, however, that for post-test numbers the student model is initialized with prior probabilities derived from test data from previous studies. For all the other numbers in Prime Climb, the model starts with generic prior probabilities of 0.5. Thus the model’s assessment of how student factorization knowledge on these numbers evolves during game play is likely to be less accurate than for post-test hints, and may generate unjustified hints.

In summary, it appears that the hints were justified at least in the case of the numbers involved in the pre/post-test.

³ While we found specificity to be lower than desired (46%), the effect on the student is that she would receive fewer hints which would reduce student interruption and should not be relevant on whether the subject attended the hints she did receive.
6.3.2 Additional Factors which Affected Game Play as Measured by Hint Fixation Time

In addition to model accuracy, we looked at the following additional factors: Move Correctness, Time of Hint, Hint Type, Attitude, and Pre-test score that may influence student attention to hints in our dataset. We start our analysis looking at total fixation time on a displayed hint as a measure of attention. We ran 2(Time of hint) by 3(Hint Type) by 2(Move correctness) by 3(Attitude) general linear model with pre-test score as a co-variant, and total fixation time as the dependent measure. In addition to statistical significance, we report partial eta-squared ($\eta^2$), a measure of effect size. To interpret this value, 0.01 is a small effect size, 0.06 is medium and 0.14 is large.

6.3.2.1 Results

We found main effects for both Time of Hint and Attitude but we don’t discuss them in detail because they are further qualified by the detected interactions. The interaction effects we found are:

- **Attitude** and **Time of Hint**. $F(2,447)=5.566$, $p=0.004$, $\eta^2 =0.024$. (see Figure 6.5, Left).

  Fixation time for those with a neutral help attitude dropped from being the highest among the three groups in the first half of the game to being very low in the second half. For students who do not want help, fixation time is the lowest of the three groups in the first half of the game, and drops even lower during the second half. In contrast the fixation time for those who wanted help did not change.
Figure 6.5: Interaction effects between (Left) *Time of Hint and Attitude*; (Middle) *Time of Hint and Hint Type*; (Right) *Move Correctness and Hint Type*.

- **Time of Hint and Hint Type**, $F(2,447) = 5.963$, $p=0.003$, $\eta^2=0.026$. (see Figure 6.5, Middle). Fixation time drops for all hint types between the first and second half of the game. The drop, however, is statistically significant only for *definition* hints, suggesting that these hints become repetitive and were perceived as redundant despite the inclusion of examples illustrating the definitions.

- **Hint Type and Move Correctness**, $F(2,447)=3.435$, $p=0.033$, $\eta^2=0.015$. (see Figure 6.5, Right). Players had significantly higher fixation time on *definition* hints caused by correct moves than of those caused by incorrect moves. There is also a significant difference between fixation time on *definition* hints after correct moves and the other two types of hints after correct moves, but this difference is likely an effect of *definition* hints being longer, as we discussed earlier. There were no statistically significant differences between fixation times on correct vs. incorrect moves for the other two hint types. We find the result on *definition* hints somewhat surprising, because we would have expected hints following correct moves to be perceived as redundant and thus attended less than hints following incorrect moves. It is possible, however, that
the very fact that hints after correct moves were unexpected attracted the students’ attention.

6.3.3 Additional Factors which Affected Game Play as Measured by Fixations per Word

To gain a better sense of how students looked at hints when they were displayed, we ran a general linear model with the same independent measures described above (*Time of Hint, Hint Type, Move Correctness, Attitude, and pre-test scores*) with fixations/word as the dependent measure.

6.3.3.1 Results

We found the following significant main effects:

- **Hint Type.** $F(2,447)=31.683, p<0.001, \eta^2=0.124$. (see Figure 6.6, Left). *Definition* hints (Avg. = 0.17, $SD = 0.22$) had a statistically significant lower fixations/word than either *Tool* (Avg. = 0.35, $SD= 0.38$) or *Bottom-out* hints (Avg. = 0.34, $SD = 0.32$), possibly due to the fact that students tend to skip the actual definition part of the hints, which does not change, in order to get to the factorization examples at the bottom.

- **Attitude.** $F(2,447)=6.722, p=0.001, \eta^2=0.029$. (see Figure 6.6, Right). Students who wanted no help had the lowest fixations/word (Avg.= 0.25, $SD = 0.30$), significantly lower than the other two groups. The difference between the help (Avg. = 0.36, $SD = 0.38$) and neutral group (Avg. = 0.31, $SD = 0.28$) is not significant, but the trend is in the direction of the help group having higher fixations/word than the neutral group.
Figure 6.6 Main effects of: (Left) Hint Type; (Right) Attitude.

- **Pre-test score.** $F(1,447)=6.614$, $p=0.01$, $\eta^2=0.015$. (see Figure 6.7). Subjects with the lowest (below 65%) and highest (above 94%) pre-test scores had fewer fixations/word than subjects with intermediate scores. For high knowledge subjects, this effect is likely due to the fact that the hints were not justified. We can only speculate that, for the low knowledge subjects, the effect may be due to a general lack of interest in learning from the game.

Figure 6.7: Main effect of Pre-test Score.
In addition to the main effects, we also found two interaction effects, both involving *Move Correctness* (see Figure 6.8).

- **Move Correctness and Hint Type.** $F(2,447)=11.141, p<0.001, \eta^2=0.013$. (see Figure 6.8, Left). Fixations/word on Bottom-out hints drops significantly between those given after a correct move (Avg. = 0.48, $SD = 0.27$) and those given after an incorrect move (Avg. = 0.19, $SD = 0.22$). This result confirms the positive effect that *Move Correctness* seems to have on attention to hints found in the previous section for definition hints. Here, the effect possibly indicates that students are scanning Bottom-out hints for correct moves carefully in order to understand why they are receiving this detailed level of hint when they are moving well.

- **Move Correctness and Time of Hint.** $F(1,447)=3.922, p=0.048, \eta^2=0.009$. (see Figure 6.8, Right). This shows that fixations/word drop significantly between hints for correct moves given in the first and second half of the game, suggesting that the aforementioned surprise effect of hints for correct moves fades as the game progresses.

![Figure 6.8: Interaction effects between (Left) Move Correctness and Hint Type; (Right) Move Correctness and Time of Hint.](image)

68
6.3.4 Discussion

All the factors that we explored (Time of Hint, Hint Type, Attitude, Move Correctness and Pre-test Scores) affect to some extent attention to the Prime Climb hint, and the results of our analysis can be leveraged to improve attention to these hints. We found, for instance, that attention to hints decreases as the game proceeds, and the drop is highest for definition hints, suggesting that these hints are too repetitive and should be either varied or removed. If a student has an existing attitude toward help, this attitude generates consistent patterns of attention to hints throughout the game (low attention for those who do not want help, higher attention for those who do). This result suggests that general student attitude toward receiving help should be taken into account when generating adaptive hints, and strategies should be investigated to make hints appealing for those students who do not like receiving help. Similarly, strategies should be devised to make students with low knowledge (as assessed by the student model) look at the hints, since our results indicate that these students tend not to pay attention, although they are the ones who likely need hints the most. We also found that students with a neutral attitude toward help had much less consistent attention behavior than the students who wanted help and the students who did not. The neutral students showed quite high attention to hints in the first half of the interaction, but dropped almost to the lowest in the second half, confirming that the Prime Climb hints should be improved to remain informative and engaging as the game proceeds. In the next section, we show initial evidence that improving attention to hints as discussed here is a worthwhile endeavor because it can improve student interaction with the game.
6.4 How Attention Affected Game Play

In the previous section we were looking at the factors that affected attention to the hints. In this section, we look at whether attention to hints impact subjects’ performance with Prime Climb. In particular, we focus on the effect of attention to hints on the correctness of the subsequent player’s move. We also looked into factors that affect whether they decide to use the magnifying glass after receiving a *tool* hint.

6.4.1 Move Correctness after Hints

As our dependent variable, *Move Correctness After Hint*, is categorical (e.g., the move is either correct or incorrect), we use logistic regression to determine if *Fixation Time*, *Fixations per word* and *Hint Type* are significant predictors of *Move Correctness After Hints*. One weakness in using this analysis is that the data points in our dataset are not independent, since they consist of sets of moves generated by the same subject. Lack of independence can increase the risk of making a type 1 error due to overdispersion (i.e., ratio of the chi-square statistic to its degrees of freedom is greater than 1), but this is not an issue in our data set. ($\chi^2 = 6.41 \text{ df } = 8$)

Table 6.1 shows the results of running logistic regression on these data, indicating that *Fixations per word* is the only significant predictor of *Move Correctness after Hints*. The odds ratio greater than 1 indicates that as fixations/word increases, the odds of correct moves also increase. This suggests that when players read the hints more carefully, their next move is more likely to be correct.

The results of the logistic regression also indicate that the type of hint subjects pay attention to does not impact move correctness. This finding is consistent with the fact that, in Prime Climb, *Bottom-out* hints do not provide direct information on what to do next; they
only explain how to evaluate the player’s previous move in terms of number factorization, and this information cannot be directly transferred to the next move. Still, it appears that some form of transfer does happen when students pay attention to the hints, and helps them make fewer errors, suggesting that if we increase student attention to hints, leveraging the results discussed in previous sections, we may increase the overall level of learning triggered by Prime Climb.

Table 6.1: Logistic regression results for Move Correctness after Hint.

<table>
<thead>
<tr>
<th></th>
<th>95% CI for Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B (SE)</td>
</tr>
<tr>
<td>Fixations/word</td>
<td>0.98 (0.44)</td>
</tr>
</tbody>
</table>

Note: $R^2 = .01$ (Cox & Snell), .02 (Nagelkerke). Model $\chi^2(1) = 5.42$, $p = 0.02$

6.4.2 Magnifying Glass Usage

Next we looked at the Tool hints to see whether attention to the hints can affect the subject’s decision to follow the hint’s advice and use the magnifying glass. We again use logistic regression as our dependent variable, Use Magnifying Glass, is also categorical (either they use the magnifying glass or not). We are interested to see if Fixation Time and Fixations per word are significant predictors of Use Magnifying Glass. In this case, we did not find that either of these measures were significant predictors of the use of the magnifying glass. This seems to support that for some subjects, the Tool hints can be effective in triggering magnifying glass usage even if the subject just glances quickly at the hint. A possible reason for this is that the Tool hints are nearly all the same so reading one Tool hint may be enough to get the message and the subsequent hints merely act as reminders.
Chapter 7: Conclusion and Future Work

In this thesis, we set out to achieve one overarching goal. We will now review whether we have met this goal.

Goal: Explore how user attention patterns to user adaptive hints are impacted by factors relating to user’s existing knowledge, hint type, hint timing, and attitude towards getting help in general. We are also interested in exploring how attention to hints affects subsequent game play.

In Chapter 4, we discuss the design and re-implementation of the new Prime Climb which will provide us with a more modular and flexible game which will allow future researchers the ability to make alternate versions to test research hypotheses. This facilitated us in conducting a user study which allowed us to fulfill our goal of exploring the user attention patterns which we investigated in Chapter 5 and 6. We started by investigating how two subjects were processing the hint received from the agent. We found that subjects were showing selective attention to the hints. We also found that their attention to hints (especially Definition hints) decrease over time which indicate that subjects find them less and less useful as the game proceeds. We also observed that subjects spent more time on hints caused by correct moves.

Next we then expanded the investigation to 12 subjects where we looked in more detail at the factors which affect attention. We found that all our factors (Time of Hint, Hint Type, Attitude, and Move Correctness) affected to some extent attention to the Prime Climb hints. We also found that if a subject had an attitude towards help, this attitude generates consistent patterns of attention to hints throughout the game. In the case of the subjects with an attitude of not wanting help, we see low attention to hints while those who wanted help
showed higher attention. We found that those with a neutral attitude towards help had a much less consistent behavior in relation to the hints; they pay quite high attention to the hints in the first half of the game but dropped in the second half. We also observed that subjects with low domain knowledge (as assessed by the pre-test) tended to not pay attention to the hints although they are the ones who likely need the hints the most. We also saw evidence that paying attention to the hints had a positive effect on move correctness after receiving a hint. This is a promising sign that if we can improve attention to the hints, we may increase the overall level of learning from playing Prime Climb.

This work contributes to existing research on student use and misuse of adaptive hints in intelligent tutoring systems by looking at how subjects react to hints when they are provided unsolicited by the system as opposed to explicitly requested by the student or obtained by gaming strategies. There are two additional aspects that are innovative in this work. The first is that we focus on adaptive hints provided by an edu-game which is in a context that is especially challenging to provide didactic support because it can interfere with game playing. The second is that we use eye-tracking data to analyze student attention. We found that attention to hints is affected by a variety of factors related to users’ existing knowledge, hint timing, hint context and attitude towards getting help in general.

The next step in this research will be to leverage these findings to improve the design and delivery of Prime Climb hints. First we need to look to see if we can improve the accuracy of the student model, specifically to improve the specificity as the model is weak at detecting what players do not know. This is important as if we could better identify when the
player does not know a number, we may be able to help the player learn more as we provide help when they need it. On the other hand, if we are better at detecting what players do not know, we will also trigger more hints. But we already see diminishing attention to hints, especially the *Definition* hints, which means that providing even more hints may be undesirable. Either way we need to investigate whether providing fewer (though hopefully more justified) hints would increase player attention to the hints.

In addition, there are still areas of improvement we can make with the re-implemented Prime Climb game. Currently Prime Climb provides players with unsolicited hints and then allows the player to receive more hints at that time. The original Prime Climb game also contained a help menu system which allows players to request help as they are playing. If we re-implement the help menu, we could provide support on demand as well as unsolicited. We can also re-implement an affective model which was developed in the past. This would allow us to provide support based on user affect. Lastly we can provide an administrative application which would make it easier for researchers to generate reports (i.e. log file extracts, summary statistics of game play) and modify variables without explicit knowledge of the underlying code.

Another place where we could improve hints is to use the hints to teach players tools that can help them determine if two numbers share a common factor. For example, all even numbers share a common factor (of 2) and reminding players of this fact may improve their move choices. There are also similar rules for other numbers (e.g. 3, 4, 5, 8, and 9). By providing hints which remind players of these concepts, we may help them play better, especially when the numbers involved get larger.
We also plan to extend the Prime Climb model to use eye-tracking data in real-time for assessing if and how a student is attending to hints and react accordingly. If we can detect when players are not attending to hints, we can provide interventions to help them refocus.
Bibliography


Appendices

Appendix A  Permission Forms

Research Participant Consent Form

Dear Parent/Guardian,

My name is Cristina Conati and I am an Associate Professor in the Department of Computer Science at the University of British Columbia. I am working together with other researchers, Mary Muir, Minela Gutica, and others (UBC, computer science), on a research study exploring the potential of computer tools for helping students learn mathematical and scientific concepts and skills.

In order to carry out our research, it is necessary to work with a large number of students in a variety of formal and informal settings, including classrooms and summer camps in British Columbia and our Research Lab located in the Department of Computer Science at UBC. Over the next year our research team will be conducting a number of inter-related but distinct studies focusing on the relationships between learning, attention and individual differences. For each study, students will be using computer-based educational tools and activities in particular settings. A variety of methods will be used to record data, including field observations by researchers, surveys and questionnaires, and computer log files. We would also like to make a video recording of each session, but will not do so without written consent (see attached Research Participant Video Release Form). Students might be interviewed and the interview videotaped (see attached Research Participant Interview and Audio Recording Release Form).

Your child has been selected to participate in a study which will investigate issues related to the influence of a pedagogical agent and adaptive interventions on children attention and the way they learn when they interact with an educational game, Prime Climb. This game aims at teaching mathematical concepts such as factorization and prime numbers, material that is part of your child’s current curriculum. We have a demonstration available that students can play with, with which your child will interact for about 30 to 45 minutes. One or two researchers from one group will be present in the laboratory to interact with students while they use the software, and make sure the sensing and recording devices are used in a proper way. The total time of participation (including instruction, questionnaires, and interviews) will be up to 90 minutes.

Please be advised that the identities of all students who participate will be kept confidential, since pseudonyms will be used in all reports released to any persons other than the researchers.

This research has been granted ethics approval by the UBC Office of Research Services (project # H04-80596). If you have any inquiries concerning the procedures or your child’s role in this study, please feel free to e-mail me at conati@cs.ubc.ca, for further clarification or information. In addition, if you have any concerns about your child’s treatment or rights as a research subject, please call the Research Subject Information Line in the UBC Office of Research Services, at 604-822-8598. Sincerely,

Cristina Conati
Associate Professor, Department of Computer Science
University of British Columbia

________________________________________________________

Please detach here and return bottom portion

________________________________________________________

Student’s name: ________________________________

I have received a copy of the participant consent form for the research study involving the educational game Prime Climb (part of the research project # H04-80596).

I __________________________ have read the above form and consent / do not consent (circle one) to have my child participate in the study as described...

Student signature: __________________________ Date: __________

(only a parent/guardian needs to sign this form if they do not wish to give consent)

Parent/Guardian signature: __________________________ Date: __________
Research Participant Video Release Form

Dear Parent/Guardian,

In the course of carrying out our study, we would like to record each session using a video-camera in order to help us with our data analysis. When releasing the results of our research, we would also like to use either still photographs or short clips (no more than 10 seconds) taken from the video recording made during your child’s participation in our study.

However, we recognize that both video recording and subsequent use of images can be a delicate issue in some cases and we will not record the study session or use any images of the participant without consent from both the student and parent/guardian. Any parent/guardian or student has the right to refuse consent, and such withdrawal will not in any way jeopardize the student’s involvement in normal classroom activities. If you indicated on the research participant consent form that you do not consent to your child’s participation in this study then you do not need to return this video release form.

This research has been granted ethics approval by the UBC Office of Research Services (project # H04-80596). If you have any inquiries concerning the procedures or your child’s role in this study, please feel free to e-mail me at const@cs.ubc.ca, for further clarification or information. In addition, if you have any concerns about your child’s treatment or rights as a research subject, please call the Research Subject Information Line in the UBC Office of Research Services, at 604-822-5998.

Sincerely,

Cristina Conati
Associate Professor, Department of Computer Science
University of British Columbia

---------- Please detach here ---------------

Student’s name: ___________________________

I have received a copy of the participant video release form for the research study involving the educational game Prime Climb (part of the research project # H04-80596).

I __________________ have read the above form and consent / do not consent (name of parent/guardian) (please circle one) to my child’s session of the study being recorded by video-camera.

Student signature: __________________ Date: ______________
(only a parent/guardian needs to sign this form if they do not wish to give consent)

Parent/Guardian signature: __________________ Date: ______________

Student’s name: ___________________________

I have received a copy of the participant video release form for the research study involving the educational game Prime Climb (part of the research project # H04-80596).

I __________________ have read the above form and consent / do not consent (name of parent/guardian) (please circle one) to images or short video-clips to be used in published results.

Student signature: __________________ Date: ______________
(only a parent/guardian needs to sign this form if they do not wish to give consent)

Parent/Guardian signature: __________________ Date: ______________

Version 1.1 October 2010. Author: Mary Mair maryannie@cs.ubc.ca

Page 1 of 1
Research Participant Interview and Audio Recording Release Form

Dear Parent/Guardian,

In the course of carrying out our study, we would like to interview and audio record your child in order to obtain more data for our study.

However, we recognize that audio recording can be a delicate issue in some cases and we will not interview and record the participant without consent from both the student and parent/guardian. Any parent/guardian or student has the right to refuse consent. If you indicated on the research participant consent form that you do not consent to your child's participation in this study then you do not need to return this interview and audio recording release form.

This research has been granted ethics approval by the UBC Office of Research Services (project #H04-80596). If you have any inquiries concerning the procedures or your child's role in this study, please feel free to e-mail me at conati@cs.ubc.ca, for further clarification or information. In addition, if you have any concerns about your child's treatment or rights as a research subject, please call the Research Subject Information Line in the UBC Office of Research Services, at 604-822-8398.

Sincerely,

Cristina Conati
Associate Professor, Department of Computer Science
University of British Columbia

--------------------------------------------------------------------- Please detach here ←---------------------------------------------------------------------

Student's name: ____________________

I have received a copy of the participant video release form for the research study involving the educational game Prime Climb (part of the research project #H04-80596).

I __________________ have read the above form and consent / do not consent (please circle one) to my child's interview.

Student signature: ____________________ Date: ____________

(only a parent/guardian needs to sign this form if they do not wish to give consent)

Parent/Guardian signature: ____________________ Date: ____________

---------------------------------------------------------------------

Student's name: ____________________

I have received a copy of the participant video release form for the research study involving the educational game Prime Climb (part of the research project #H04-80596).

I __________________ have read the above form and consent / do not consent (please circle one) to the child's interview to be audio recorded.

Student signature: ____________________ Date: ____________

(only a parent/guardian needs to sign this form if they do not wish to give consent)

Parent/Guardian signature: ____________________ Date: ____________

Version 1.1 October 2010. Author: Mary Mourayyan@cs.ubc.ca
Research study on Educational Computer Games

Looking for kids in grade 5 or 6

Why? To participate in a research study involving a math computer game (Prime Climb) designed to help kids learn number factorization.

Where/When? Participants would need to come to the Computer Science department for about 75 to 90 minutes after school, day to be arranged with you.

What? Participants will:

- Write a short test involving number factorization
- Play the game for a max of 40 minutes (or as long as they are comfortable)
- Write a post-test, and fill a questionnaire.

Each participant will receive a $10 gift certificate from Kids Books.

Parents interested in having their kids participate should contact Mary Muir.

This project is conducted under the supervision of Prof. Cristina Conati, Department of Computer Science, University of British Columbia (conati@cs.ubc.ca)

Ethics Approval by UBC Office of Research Services (project # H04-80596)
Appendix C Pre-test

PRE-TEST

Factors
- circle all the factors of the number
- you may circle more than one answer
- if none of the factors are listed, circle “none”

1) The factors of 15 are
   2 3 5 7 11 none

2) The factors of 30 are
   2 3 5 7 11 none

3) The factors of 25 are
   2 3 5 7 11 none

4) The factors of 42 are
   2 3 5 7 11 none

5) The factors of 14 are
   2 3 5 7 11 none

6) The factors of 49 are
   2 3 5 7 11 none

7) The factors of 9 are
   2 3 5 7 11 none

8) The factors of 27 are
   2 3 5 7 11 none

9) The factors of 11 are
   2 3 5 7 11 none

10) The factors of 33 are
    2 3 5 7 11 none

11) The factors of 31 are
    2 3 5 7 11 none

12) The factors of 36 are
    2 3 5 7 11 none

Example:
The factors of 50 are
   2 3 5 7 11 none

Because
   $2 \times 25 = 50$
   $5 \times 10 = 50$

Example:
The factors of 7 are
   2 3 5 7 11 none

Because
   $1 \times 7 = 7$
PRE-TEST

13) The factors of 81 are
   2 3 5 7 11 none

14) The factors of 97 are
   2 3 5 7 11 none

15) The factors of 89 are
   2 3 5 7 11 none

16) The factors of 88 are
   2 3 5 7 11 none

Common Factors
- circle all the common factors of the two numbers
- you may circle more than one answer
- if these numbers have no common factors, circle “none”

1) 15 and 30 share ____ as common factors.
   2 3 5 7 11 none

2) 25 and 42 share ____ as common factors.
   2 3 5 7 11 none

3) 14 and 49 share ____ as common factors.
   2 3 5 7 11 none

4) 9 and 27 share ____ as common factors.
   2 3 5 7 11 none

5) 11 and 33 share ____ as common factors.
   2 3 5 7 11 none

Example:
12 and 18 share __ as common factors
   2 3 5 7 11 none
Because
2 is a factor of 12 and 18
3 is a factor of 12 and 18
# Appendix D Questionnaire

**Prime Climb Questionnaire**

**Student Number:**

**Gender:** Boy/Girl

**Questions:**
- Please circle the answer that best suits you

### Part I

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Occasionally</th>
<th>Sometimes</th>
<th>Usually</th>
<th>Always</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>
### Prime Climb Questionnaire

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>The agent intervened too often.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>13</td>
<td>The agent did not intervene enough.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>14</td>
<td>I liked the agent.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>

How could the agent be more useful for you?

How could the agent be more fun?

How could the agent be more interesting?

What other help could the agent provide to help you play this game?
Prime Climb Questionnaire

Student Number: 

Part III
Do you agree with the following statements about your game experience?

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I wanted to beat my partner when I was playing the game.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>I felt upset when I fell down the mountain.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td>I became curious about math (number factorization) by playing the game.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>I wanted help when I was stuck.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>5</td>
<td>I always wanted to be ahead of my partner when I was playing the game.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td>I wanted to work out difficult problems on my own.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>7</td>
<td>I wanted to learn math by playing the game.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>8</td>
<td>I didn’t mind falling down the mountain.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>I did not want anyone’s advice when deciding what to do.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>I didn’t want to think about math (number factorization) when I was playing the game.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>11</td>
<td>I didn’t mind the agent giving me advice.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Prime Climb Questionnaire

Student Number: __________

Part IV
Do you agree with the following statements about your game experience?

<table>
<thead>
<tr>
<th></th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Didn't Happen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I wanted to have fun when I was playing the game.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>I had fun when I finished a level.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>I had fun when I was climbing successfully (no falling).</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>I had fun when I was climbing quickly without falling.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>I had fun when one climber was swinging from the rope.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>I had fun when I fell.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>I had fun when my partner fell.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>I had fun when there were many falls in a row.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>I had fun when I used the magnifying glass.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>I didn’t care whether I had fun or not.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Please tell us any other time you had fun while playing the game.

Please tell us any other time you did not have fun while playing the game.
Prime Climb Questionnaire

Part V
Do you agree with the following statements about your game experience?

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Didn’t Happen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I learned math when I made a successful move.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>I learned math when the agent told me to think about why I failed.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td>I learned math when the agent told me to think about why I succeeded.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>I learned math when the agent reminded me about the common factor rule.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>5</td>
<td>I learned math when the agent suggested using the magnifying glass.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td>I learned math when I followed the agent’s advice.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>7</td>
<td>I learned math when I got to the bigger numbers.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>8</td>
<td>I learned math when the agent said something I already knew.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>I learned math when the agent said something I didn’t understand.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>I learned math when I made an unsuccessful move.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
<td>N/A</td>
</tr>
</tbody>
</table>

I was already good at factorization before I played the game. 1 2 3 4 5

I also learned math when I:

I was stopped from learning math by:
Appendix E  Server States

The game server has several states which it can be in during the game play:

- **START**: In this state, two players have joined a game instance and both send ready signals to the server. The server then goes into the **READY** state.

- **READY**: In this state, the server builds the game state for a mountain and sends the information to both players. Once completed, the server goes into the **IDLE** state.

- **IDLE**: In this state, the server is waiting for a client to make a move. Either player can make a move, whichever’s move is received first, gets to make the next move. There are four possible states, the game can move to which are **MOVING**, **FALLING**, **RESET**, **GOAL**

- **MOVING**: In this state, a player has made a correct move. The server stops accepting any new requests until the move is done. Once the move is made, both clients get an update due to the move and server state returns to IDLE.

- **FALLING**: In this state, a player has made an incorrect move but there is at least 1 valid move upon making a fall (i.e., at least one number the player is swinging over does not share a common factor with the number the partner is on). The server sends the falling message for the client to animate and waits for the player to pick a hexagon to continue. In this state, the only action available is to choose a hexagon. Once a hexagon is chosen, the state returns to IDLE.

- **RESET**: In this state, a player has made an incorrect move but there is no valid move upon falling. This could be due to being too close to the bottom of the mountain (which would cause the player to fall off the mountain) or due to the case where all three hexagons the person is swinging across contain a common factor with the hexagon the partner is on. The server resets both the user’s positions to the bottom of the mountain in one of the possible valid starting points of that mountain. The server then returns to the **IDLE** state.

- **GOAL**: In this state a player reaches the top of the mountain. The server then sends an update to each client to show that one player reached the top of the mountain, and then goes into the **READY** state for the next mountain. If there are not any more mountains left, it exits the game and goes to the **EXIT** state.
• *EXIT*: In this state, the game is finished. The signal for the game animation for completion is given to both players. The server sends a message back to both clients to go to the lobby screen.
Appendix F  Study Raw Data

F.1  Play Time, Hint Number and Time between Hints

Table A.1 Playtime, total hints and average time (SD) between hints for each player.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Playtime in minutes</th>
<th>Total Hints</th>
<th>Average time between hints (SD) in seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>64</td>
<td>66</td>
<td>55.8 (47.3)</td>
</tr>
<tr>
<td>S2</td>
<td>65</td>
<td>86</td>
<td>44.9 (50.1)</td>
</tr>
<tr>
<td>S3</td>
<td>25</td>
<td>41</td>
<td>36.5 (36.2)</td>
</tr>
<tr>
<td>S4</td>
<td>55</td>
<td>72</td>
<td>41.0 (47.0)</td>
</tr>
<tr>
<td>S5</td>
<td>29</td>
<td>30</td>
<td>54.8 (81.1)</td>
</tr>
<tr>
<td>S7</td>
<td>20</td>
<td>36</td>
<td>32.7 (35.2)</td>
</tr>
<tr>
<td>S8</td>
<td>22</td>
<td>30</td>
<td>41.8 (35.6)</td>
</tr>
<tr>
<td>S9</td>
<td>15</td>
<td>24</td>
<td>33.1 (45.0)</td>
</tr>
<tr>
<td>S10</td>
<td>19</td>
<td>29</td>
<td>36.1 (39.4)</td>
</tr>
<tr>
<td>S11</td>
<td>25</td>
<td>82</td>
<td>17.4 (15.1)</td>
</tr>
<tr>
<td>S12</td>
<td>35</td>
<td>34</td>
<td>59.1 (49.2)</td>
</tr>
<tr>
<td>S13</td>
<td>17</td>
<td>77</td>
<td>12.0 (9.2)</td>
</tr>
</tbody>
</table>
F.2 Pre and Post-test Score and Proportional Learning Gain

Table A.2 Pre- and post-test scores as well as proportional learning for each player.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Pretest (%)</th>
<th>Post-test (%)</th>
<th>Proportional Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>97</td>
<td>97</td>
<td>0</td>
</tr>
<tr>
<td>S2</td>
<td>74</td>
<td>81</td>
<td>0.25</td>
</tr>
<tr>
<td>S3</td>
<td>100</td>
<td>97</td>
<td>-0.03</td>
</tr>
<tr>
<td>S4</td>
<td>94</td>
<td>90</td>
<td>-0.5</td>
</tr>
<tr>
<td>S5</td>
<td>97</td>
<td>94</td>
<td>-1</td>
</tr>
<tr>
<td>S7</td>
<td>52</td>
<td>55</td>
<td>0.07</td>
</tr>
<tr>
<td>S8</td>
<td>90</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>S9</td>
<td>65</td>
<td>68</td>
<td>0.09</td>
</tr>
<tr>
<td>S10</td>
<td>100</td>
<td>94</td>
<td>-0.06</td>
</tr>
<tr>
<td>S11</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>S12</td>
<td>94</td>
<td>84</td>
<td>-1.5</td>
</tr>
<tr>
<td>S13</td>
<td>39</td>
<td>19</td>
<td>-0.25</td>
</tr>
</tbody>
</table>
F.3 Moves

Table A.3 Moves made by subject including percent of wrong moves and number of wrong/correct moves made after a hint.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Number of moves</th>
<th>Percent of wrong moves</th>
<th>Number of wrong moves made after a hint</th>
<th>Number of correct moves made after a hint</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>184</td>
<td>18.5%</td>
<td>12</td>
<td>54</td>
</tr>
<tr>
<td>S2</td>
<td>197</td>
<td>20.8%</td>
<td>21</td>
<td>64</td>
</tr>
<tr>
<td>S3</td>
<td>131</td>
<td>14.5%</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>S4</td>
<td>186</td>
<td>23.1%</td>
<td>21</td>
<td>50</td>
</tr>
<tr>
<td>S5</td>
<td>119</td>
<td>13.4%</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>S7</td>
<td>125</td>
<td>18.4%</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>S8</td>
<td>114</td>
<td>17.5%</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>S9</td>
<td>134</td>
<td>10.4%</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>S10</td>
<td>129</td>
<td>13.2%</td>
<td>4</td>
<td>25</td>
</tr>
<tr>
<td>S11</td>
<td>186</td>
<td>25.8%</td>
<td>24</td>
<td>58</td>
</tr>
<tr>
<td>S12</td>
<td>148</td>
<td>12.8%</td>
<td>4</td>
<td>34</td>
</tr>
<tr>
<td>S13</td>
<td>174</td>
<td>21.3%</td>
<td>19</td>
<td>58</td>
</tr>
</tbody>
</table>
### F.4 Magnifying Glass Usage

Table A.4 Magnifying glass usage for each subject

<table>
<thead>
<tr>
<th>Subject</th>
<th>Number of Times player used Magnifying Glass</th>
<th>Number of Tool hints</th>
<th>Percent of Tool hints followed by Magnifying Glass usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>57</td>
<td>16</td>
<td>56%</td>
</tr>
<tr>
<td>S2</td>
<td>18</td>
<td>21</td>
<td>29%</td>
</tr>
<tr>
<td>S3</td>
<td>0</td>
<td>10</td>
<td>0%</td>
</tr>
<tr>
<td>S4</td>
<td>48</td>
<td>17</td>
<td>18%</td>
</tr>
<tr>
<td>S5</td>
<td>43</td>
<td>7</td>
<td>43%</td>
</tr>
<tr>
<td>S7</td>
<td>20</td>
<td>9</td>
<td>11%</td>
</tr>
<tr>
<td>S8</td>
<td>5</td>
<td>7</td>
<td>0%</td>
</tr>
<tr>
<td>S9</td>
<td>0</td>
<td>6</td>
<td>0%</td>
</tr>
<tr>
<td>S10</td>
<td>5</td>
<td>7</td>
<td>29%</td>
</tr>
<tr>
<td>S11</td>
<td>5</td>
<td>19</td>
<td>16%</td>
</tr>
<tr>
<td>S12</td>
<td>7</td>
<td>8</td>
<td>25%</td>
</tr>
<tr>
<td>S13</td>
<td>9</td>
<td>19</td>
<td>11%</td>
</tr>
</tbody>
</table>
Appendix G  More Details on Specificity and Sensitivity

In contrast, the model specificity (i.e., the proportion of actual negatives which are correctly identified as such) is 46%, indicating that the model will overestimate when a subject did not know a post-test number. In this case, the effect would be that the model would miss out on a chance to provide a hint when a hint was justified. While the specificity is lower than desired, the effect on the student would receiving fewer hint which would reduce student interruption and should not be relevant on whether the subject attended the hints they did receive. We can also look at the specificity and sensitivity for each student (see Figure A.1). Here we see that the specificity is much more variable than the sensitivity. In two cases (S1, S3), the subjects did not make any mistakes on the factorization questions (and as a result specificity could not be calculated) while in two other cases (S9, S10) the model was unable to identify any of the numbers that the player got wrong on their post-test as unknown. One weakness with the measure of specificity is that in a few cases, the subject made few (or no) mistakes in the factorization questions which lead to very few data points for the calculations.

Figure A.1: Sensitivity and specificity of the model for each student.
One reason this might occur is that the post-test may not measure the student’s knowledge as well as it could. In many cases when the post-test score was lower than the pre-test score, it was a result of a subject failing to include one of the factors of a number which they correctly identified in the pre-test. While it could be that they made a guess on the pre-test, it seems possible that player fatigue may have led them to take less care on the post-test than they had on the pre-test.