Comparison of Errorless Learning and Response Cost to simple Trial and Error learning in an automated learning environment

Brynjar Ólafsson

Thesis of 60 ECTS credits
Master of Science (M.Sc.) in Computer Science

June 2017
Comparison of Errorless Learning and Response Cost to simple Trial and Error learning in an automated learning environment

by

Brynjar Ólafsson

Thesis of 60 ECTS credits submitted to the School of Computer Science at Reykjavík University in partial fulfilment of the requirements for the degree of
Master of Science (M.Sc.) in Computer Science

June 2017

Supervisors:
Hannes Högni Vilhjálmsson, Supervisor
Associate Professor, Reykjavík University, Iceland

Berglind Sveinbjörnsdóttir, Supervisor
Adjunct, Reykjavík University, Iceland
Copyright
Brynjar Ólafsson
June 2017
Comparison of Errorless Learning and Response Cost to simple Trial and Error learning in an automated learning environment

Brynjar Ólafsson

June 2017

Abstract

Feedback is an important part of the learning cycle. Computers make immediate (corrective) feedback possible which increases the number of learning opportunities. Relatively little research has been done on feedback in automated learning environments. We believe that it is important to know what type of feedback is best to use, the timing of the feedback and how much feedback should be given. In this project we will be looking into the different types of feedback and if there is any difference in performance and enjoyment when learning with them. This project compared three different methods of feedback: Errorless Learning, Response Cost and Trial and Error, in an automated learning environment based on Discrete Trial Training. 24 nine year old children from two classes were randomly distributed to the three different feedback conditions. A comparison of how fast the participants learned to recognise and discriminate between three similar bacteria showed that the Response Cost condition was significantly slower that the other two conditions. A comparison of how enjoyable each feedback condition was showed little to no difference. This can be an important finding since many educational computer games use Response Cost as a feedback method for teaching.
Samanburður á villulausri þjálfun og fórnarkostnaði við einfalt happa og glappa nám í sjálfvirku kennslu umhverfi.

Brynjar Ólafsson

Júni 2017

Útdráttur

Endurgjöf er mikilvægur hluti lærdóms. Tölvur gera okkur kleift að fá tafarlausa endurgjöf sem eykur fjölda lærdómstækifæra. Tiltölulega lítið er um rannsóknir tengdum endurgjöf í sjálfvirkum kennslumhverfum. Við teljum mikilvægt að vita hvernig endurgjöf sé best að nota, tímasetningu endurgjafar og magn endurgjafar. Í þessu verkefni eru þráð mismunandi aðferðir við endurgjöf bornar saman: Villulaus þjálfun (e. Errorless Learning), fórnarkostnaður (e. Response Cost) og einfalt happa og glappa nám (e. Trial and Error) í sjálfvirku kennslu umhverfi byggt á aðgreindum kennsluæfingum (e. Discrete Trial Training).

24 niður grunnskólanemendur úr tveimur bekkjum var skipt upp af handahófi í þráð hópa miðað við endurgjöf. Samanburður á hversu hratt þátttakendur náðu að læra að þekkja og greina á milli þriggja svipaðra bakteria sýndi að fórnarkostnaðsaðferðin var hægari en hinar tvær. Samanburður á hversu skemmtilegt var að læra eftir aðferð síndi litinn mun. Þetta er hugsanlega mikilvæg uppgötvun þar sem margir kennsluleikir nýta sér fórnarkostnaðsaðferðina.
Comparison of Errorless Learning and Response Cost to Simple Trial and Error Learning in an Automated Learning Environment

Brynjar Ólafsson

Thesis of 60 ECTS credits submitted to the School of Computer Science at Reykjavik University in partial fulfilment of the requirements for the degree of Master of Science (M.Sc.) in Computer Science

June 2017

Student: 
Brynjar Ólafsson

Supervisors: 
Hannes Hógni Vilhjálmsson
Berglind Sveinbjörnsdóttir

Examiners: 
David James Thue
Marta Kristín Lárusdóttir
The undersigned hereby grants permission to the Reykjavik University Library to reproduce single copies of this Thesis entitled *Comparison of Errorless Learning and Response Cost to simple Trial and Error learning in an automated learning environment* and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the Thesis, and except as herein before provided, neither the Thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author’s prior written permission.

09.05.17

Date

Brynjar Ólafsson

Master of Science
I dedicate this to my family.
Acknowledgements

A prototype that was used for this work was funded by 2016 RANNÍS student innovation fund “Sjálfvirkni í aðgreindum kennsluæfingum”.

Specially thanks to my supervisors Hannes Högni Vilhjálmsson and Berglind Sveinbjörnsdóttir and the examiners David James Thue and Marta Kristín Lárusdóttir.
This dissertation is original work by the author, Brynjar Ólafsson.
## Contents

Acknowledgements xvii  
Preface xix  
Contents xxi  
List of Figures xxiii  
List of Tables xxv  

<table>
<thead>
<tr>
<th>1</th>
<th>Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Motivation</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Problem Statement</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>Structure of Thesis</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>Background</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Applied Behaviour Analysis</td>
<td>3</td>
</tr>
<tr>
<td>2.2</td>
<td>Response Cost</td>
<td>4</td>
</tr>
<tr>
<td>2.3</td>
<td>Errorless Learning</td>
<td>5</td>
</tr>
<tr>
<td>2.4</td>
<td>Trial and Error</td>
<td>6</td>
</tr>
<tr>
<td>2.5</td>
<td>Learning Styles</td>
<td>6</td>
</tr>
<tr>
<td>2.6</td>
<td>Feedback</td>
<td>7</td>
</tr>
<tr>
<td>2.7</td>
<td>Discrete Trial Training</td>
<td>8</td>
</tr>
<tr>
<td>2.8</td>
<td>In Summary</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
<th>Related Work</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Feedback in automated teaching applications</td>
<td>10</td>
</tr>
<tr>
<td>3.2</td>
<td>Intelligent Tutoring Systems</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4</th>
<th>Apparatus</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Design Considerations</td>
<td>12</td>
</tr>
<tr>
<td>4.2</td>
<td>Quick Overview</td>
<td>12</td>
</tr>
<tr>
<td>4.3</td>
<td>Design Process</td>
<td>14</td>
</tr>
<tr>
<td>4.4</td>
<td>Detailed System Description</td>
<td>16</td>
</tr>
<tr>
<td>4.4.1</td>
<td>General</td>
<td>16</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Errorless Learning</td>
<td>20</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Response Cost</td>
<td>22</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Trial and Error</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>Method</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Conditions</td>
<td>24</td>
</tr>
<tr>
<td>5.2</td>
<td>Participants</td>
<td>26</td>
</tr>
<tr>
<td>5.3</td>
<td>Measurement</td>
<td>28</td>
</tr>
</tbody>
</table>
6 Results
6.1 Preface .......................................................................................................................29
6.2 Learning Effectiveness of System ............................................................................31
6.3 Hypothesis A: Difference in Time to Learn ............................................................32
6.4 Hypothesis B: Difference in Time Between Clicks ..................................................39
6.5 Hypothesis C: Difference in Enjoyment ....................................................................41

7 Discussion ..................................................................................................................43

8 Conclusion & Future Work ..........................................................................................45

References .....................................................................................................................47

Appendix A .......................................................................................................................49

Appendix B .......................................................................................................................52
List of Figures

4.2.1 The Trial Screen .................................................................13
4.2.2 The Star Reinforcement ....................................................13
4.2.3 The Skill Menu ..................................................................14
4.4.1.1 The Opening Screen, Choose a User Screen and Make New User Screen ..................17
4.4.1.2 The Choose a User and the Options menu for that User ..............................................18
4.4.1.3 The Options Menu and the Skill Menu .......................................................................18
4.4.1.4 The Skills Menu and the Options menu .....................................................................19
4.4.1.5 The Skills Menu and the Trial Screen .......................................................................19
4.4.1.6 A User Getting a Gold Star ......................................................................................20
4.4.2.1 The Full Prompt .................................................................................................20
4.4.2.2 The Medium Prompt ..........................................................................................21
4.4.2.3 The Small Prompt .................................................................................................21
4.4.2.4 No Prompt .............................................................................................................22
4.4.3.1 Response Cost Example 1 ...................................................................................22
4.4.3.2 Response Cost Example 2 ...................................................................................22
6.1.1 Total Time It Took Participants To Learn Flowers .........................................................30
6.1.2 Session Count When Learning Flowers .......................................................................31
6.2.1 Typical Learning Trend For Participants .....................................................................32
6.3.1 Total Time It Took Participants To Learn Bacteria .......................................................33
6.3.2 Total Time It Took Participants In Both Classe To Learn Bacteria ............................36
6.3.3. Session Count When Learning Bacteria ....................................................................39
6.4.1 Average time between clicks when learning bacteria ....................................................40
List of Tables

List of Figures

5.1.1 The difference between conditions..................................................................................26
5.3.1 The variables that were recorded and their description.................................................28
6.3.1 Summary of the time to learn bacteria by condition.........................................................33
6.3.2 Levene's test of homogeneity for the first class..............................................................34
6.3.3 The ANOVA for the total time to learn bacteria..............................................................34
6.3.4 Tukey HSD multiple comparison of the time to learn bacteria ........................................35
6.3.5 Summary of the time to learn bacteria by condition for both classes.............................36
6.3.6 Levene's test of homogeneity for both classes ...............................................................37
6.3.7 The ANOVA for the total time to learn bacteria for both classes......................................37
6.3.8 Tukey HSD multiple comparison of the time to learn bacteria for both of the classes ...38
6.4.1 Summary of the average time between clicks when learning bacteria by condition .......40
6.4.2 The ANOVA for the time between clicks when learning bacteria for both classes .........41
6.5.1 Average user rating for users by condition......................................................................41
6.5.2 A single factor ANOVA for the user rating ..................................................................42
Chapter 1

Introduction

1.1 Motivation

Computers and now tablets are more common in schools all over the world (e.g. Balanskat & Garioia, 2010). This increase in technology in the school system has sparked the demand for more and better software applications for students and classrooms. There has been an increase in software applications available for most fields of education such as general class management applications like Google Classroom, Edmodo and Class Dojo, a personalised learning resource such as Khan Academy and pure educational applications like Duolingo and Scratch Jr. and many more. Even though many of these educational applications are very promising and are undoubtedly useful, most current research does not offer any robust evidence of positive or negative impacts on pupils’ learning or what methods/procedures are best for students to learn by. It is widely recognised that feedback is an important part of the learning cycle (e.g. Karpicke & Roediger, 2007; Roediger & Karpicke, 2006a, 2006b). Feedback in this sense being information about how we are doing in our efforts to reach a goal such as learning a new concept. Computers drastically change classroom dynamics when it comes to feedback since feedback can be given immediately. Because feedback can be given immediately the number of learning opportunities increases. This makes it important to do it right because the effect of the feedback happen more often and quicker. If for example the feedback method is deterring the damage could already be done before someone can intervene.
1.2 Problem Statement

We believe that it is important to know what type of feedback is best to use, the timing of the feedback and how much feedback should be given. In this project we will be looking into the different types of feedback and if there is any difference in performance and enjoyment when learning with them.

1.3 Structure of this Thesis

In Chapter 2 we will start by introducing a few ideas and core concepts, which creates a context and helps you understand what this work is about. Then in Chapter 3 we will give you an overview of some related work. In Chapter 4 we describe the apparatus we made for this experiment before describing the method of the experiment in Chapter 5. We then proceed to Chapter 6 where we list the results we achieved. In Chapter 7 we discuss our results and our thoughts on the experiment before going into Chapter 8 for our final conclusion and future work.
Chapter 2

Background

2.1 Applied Behavioural Analysis

A great deal of research has been done in the field of Applied Behaviour Analysis (ABA) on children with special needs (e.g. Howard, Sparkman, Cohen, Green, & Stanislaw, 2005; Lovaas, 1987). This line of research has led to a better understanding of what methods are effective in teaching skill acquisition for all children. ABA is the application of behavioural principles that over time, increase or decrease targeted behaviours. Behaviour is analysed in these three steps: the antecedent, the behaviour and the consequence, also known as the ABC’s of behaviour. In an educational setting the consequence could be looked at as feedback and the whole cycle (the ABC’s) as a feedback loop. The frequency of behaviour can be either increased with reinforcement or decreased with punishment. Reinforcement/punishment can either be positive or negative. The term positive in this sense means that something was introduced/added while negative means that something was removed. An example of a positive reinforcement, which is a rather common occurrence in the school system, is when a student is asked a question (the antecedent), he or she gives the right answer (the behaviour) and gets a praise (the consequence). Given that the praise is a pleasant experience the likelihood of that behaviour happening again in the future increases. In some cases tokens are used instead of or with the praise as a positive reinforcement (as the feedback) (e.g. Myles, Moran, Ormsbee & Downing, 1992). Tokens can be used as a way of paying children for completing a task. The children can then use these tokens later on to
engage in desired behaviours which would not otherwise be accessible. Tokens also help motivate students since they provide a visible record of improvement.

2.2 Response Cost

Although reinforcement is most commonly used in skill acquisition, the principle of punishment can also be observed in one common method called Response Cost. Response Cost is a method that entails negative punishment to reduce undesired behaviour. In Response Cost a desired item is taken away as a consequence of undesired behaviour in order to decrease undesired behaviour. Feedback in the form of reinforcement and punishment are common in computer games. This feedback usually comes in form of points and loss of points. Tokens can take on various forms, sometimes as hearts for example, where the player tries to avoid certain situations or items that would make her loose a heart (the Response Cost) while other situations or items would make her gain a heart.

There is no denying that a lot of learning can take place when playing computer games but that does not mean that we should blindly apply game conventions when developing educational games. We believe it is useful to link core mechanisms such as feedback to well established theoretical models and approaches. Although it does not guarantee desired result it could help in figuring out when a certain type of feedback/method of learning suits a learner. It could be different depending on the learners knowledge. If the learner knows nothing about the subject and cannot infer anything from her/his existing knowledge it could be better to make the questions easier. When the learner has gotten some idea of the subject a more challenging type of teaching could take place. This is hard to balance because it is important to maintain a positive affective engagement. If the game is too
hard/challenging it could make the learner avoid learning with the game altogether. The same applies to a game that is not challenging enough and the learner may lose interest because of that. The most effective teaching method will not do any good if no one wants to play it.

### 2.3 Errorless Learning

Perception can play a role in our performance. In 1975 Hiroto & Seligman showed that failed learning opportunities can generalise to other areas in a phenomena called learned helplessness. In their experiments, subjects were presented with puzzles. One group got regular solvable puzzles while the other group was presented with puzzles that had no solutions. Participants were then presented with similar situations but with new types of problems to solve. The problems in this phase were identical, so each group had an equal chance of solving the problems but the group that previously got unsolvable puzzles did significantly worse. Another ABA method for skill acquisition called Errorless Learning tries to tackle this problem. Errorless learning can be effective for learners who frequently make mistakes, who lack confidence (or are too anxious), and/or who do not remember their learning experiences and the feedback that they receive as in the case of Alzheimer's patients (Clare, Wilson, Carter Roth & Hodges, 2002) and people with cognitive impairments (Sohlberg, Ehlhardt & Kennedy, 2005; Clare & Jones, 2008). This method is used a lot with children with autism since they do not learn as successfully from their mistakes (that can sometimes provoke problem behaviour such as tantrums, aggression, and self-injury) as typically developing children may, but instead continue to repeat them. Unlike other teaching procedures where opportunities for initial mistakes are allowed and then corrected through prompting, the teaching procedures are designed in such a way that reduces the chances of
the learner making any errors or mistakes. As each skill is taught, children are provided with a prompt immediately following an instruction. The immediate prompt reduces the chance of incorrect responses. To promote independence the prompts are then systematically faded to allow the learner an opportunity to provide correct responses on their own. This is a teaching approach that is rarely applied directly in educational games.

2.4 Trial and Error

Trial and error is sometimes contrasted with Errorless Learning. In it the learner attempts a task and then benefits from feedback, whether the attempt was correct or incorrect. Edward Thorndike (1898) devised a classic experiment in which he primarily used cats in a puzzle box to empirically test the laws of learning.

2.5 Learning styles

There are a number of theories on learning styles, some emphasise different ways students approach problems and others on the preference of auditory, visual or kinaesthetic teaching material (see Felder and Silverman 1988, Felder 1993, Felder and Spurlin 2005, Kolb 1984, Dunn, Dunn, & Price 1979). Even so there is little evidence that there is any efficacy of most learning style models (Curry 1990; Pashler, McDaniel, Rohrer & Bjork, 2009) and according to Stahl (2002) there has been an “utter failure to find that assessing children's learning styles and matching to instructional methods has any effect on their learning.” An experiment by Krätzig and Arbuthnott (2006) showed no correlation between learning style and performance and concluded that people’s judgments represented preferences rather than superior skills. According to the National Research Council (1999) there is no universal best teaching
practice. Instead you can look at the methods of teaching as tools where each tool depends on the task at hand and the materials one is working with. It would be interesting to know if the same applies to feedback or if there is some inherent difference in learning with different kinds of feedback.

2.6 Feedback

Feedback plays a critical role in student’s retention of learned information. There seems to be a general consensus that testing enhances memory for the successfully retrieved items (see Karpicke & Roediger, 2007; Roediger & Karpicke, 2006a, 2006b). Testing not only gives teachers a sense of how well their pupils are doing but also provides feedback for the student herself making it a learning event in itself. When it comes to the difficulty level of tests researchers seem to disagree. Some researchers argue that the greater the effort required for retrieval, the greater the benefit to subsequent memory retention and therefore the most benefit is to be gained from taking a difficult test (e.g., Carpenter & DeLosh, 2006; Pyc & Rawson, 2009). Proponents of Errorless Learning (e.g., Skinner, 1958; Terrace, 1963) say that a difficult test brings with it the risk that the learner may make many errors which may increase the chance of answering incorrectly in the future. A recent study suggest that when students make an error on a multiple-choice test, that error tends to persist on later tests (Marsh, Roediger, Bjork, & Bjork, 2007). This increased chance can be reduced significantly by providing corrective feedback after the test (e.g. Metcalfe & Kornell, 2007; Pashler, Cepeda, Wixted, & Rohrer, 2005). It has even been shown that corrective feedback is more beneficial than spending the same amount of time in a read-only condition (e.g., Kornell, Hays & Bjork, 2009). This makes an automated version of discrete trial training (see 2.4) a
suitable platform for learning. Since it is automated, the number of learning opportunities can be maximised and the discretisation reduces the difficulty of each learning opportunity.

2.7 Discrete Trial Training

Discrete Trial Training (DTT) is an ABA procedure and is considered one of the leading methods used to teach children with autism (Smith, 2001). The main benefits of this system according to Smith (2001) is that the trials are short allowing for numerous learning opportunities. In DTT skills/tasks are broken down into smaller, structured and more teachable components where each component is taught separately. The skills/tasks are then built-up using discrete trials that teach each step one at a time. In DTT all trials have a very specific set of steps that are well defined and every response is registered. Because it is registered the teacher knows what the student knows and what should be taught next. An example of a trial is as follows:

1. Discriminative stimulus: The student is shown an number of pictures. A simple example could be a trial with three pictures of the colours red, yellow and blue. The teacher presents a brief, clear instruction or question such as point to red.

2. Prompt: A prompt is given at the same time as the discriminative stimulus to help the student perform the response. As the child progresses, the teacher gradually fades out and ultimately eliminates the prompt.

3. Response: The student responds to the discriminative stimulus.

4. Consequence: Given a correct response the student receives a positive reinforcement which may be in the form of a verbal praise and/or through a token economy. Given an incorrect response the teacher gives a simple emotionless no and the student is corrected by prompting the right response.
5. Inter trial interval: A brief pause (1-5 seconds) before presenting the cue for the next trial. The scripted format makes the procedure an excellent candidate for an educational game. In a computerised form the trials can be even shorter since all the registration is automated and the next trial can be represented immediately.

2.8 In summary

Optimising educational programs can have a huge advantage for students. Students spend 10 mandatory years (more in other countries) in elementary schools. Usually we have around 30 students in a classroom learning the same material using the same method. Although this system works for the majority of students we still have students that do not fit into this system and eventually drop out. The goal of this project was to build an automated learning environment, based on a common teaching method: Discrete Trial Training (DTT), and use it to compare three different teaching approaches: Errorless Learning, Response Cost learning and simple Trial and Error learning. First we want to see whether students were able to learn in this environment. We then want to see if there is a difference in performance or enjoyment between any of these methods. If we can find a difference it could have a direct impact on how to design educational software.
Chapter 3

Related Work

3.1 Feedback in automated teaching applications

A variety of feedback methods can be seen in recent popular educational applications. Games like Duolingo¹, iKnow² and Mindsnacks³ all use Response Cost in the way that the user has to press the correct answer or else he/she will loose a life. A version of Errorless Learning can be seen in Dr. Browns Autism Apps⁴. There the user can press a button which makes the right answer bigger. Another educational application directed at children with autism is called See.Touch.Learn⁵ which implements the simple Trial and Error learning method. All of these are award winning educational applications with different learning methods and feedback. These software applications differ of course in more than just the feedback method but our aim is to try and make all other variables the same to find out if there is a performance difference to these different feedbacks.

¹ https://www.duolingo.com/
² http://iknow.jp/
³ https://www.mindsnacks.com/
⁴ http://www.drbrownsapps.com/
⁵ http://www.brainparade.com/
3.2 Intelligent Tutoring Systems

Instead of a one size fits all strategy of delivering content to a passive learner, intelligent tutoring systems (ITSs) are computer-based educational systems that aim to provide immediate and customised instruction or feedback to learners (Murray, 2003; Manske & Conati, 2005). The ITS model contains four components: the domain model, the student model, the teaching model, and a learning environment (Freedman, 2000). One of the fundamental components of ITSs (VanLehn, 1988) is a model of the student’s knowledge. This model is used to individualise and to adapt the instruction to the student’s needs. This is similar to Discrete Trial Training as described above. The student model also provides error flagging and corrective feedback. Corbett and Anderson (2001) showed that immediate feedback/correction faired better than when the student had more control over the timing of feedback and error correction. There is some worry that immediate feedback prevents students from learning error-detection skills and may discourage metacognitive self-monitoring processes more generally (Chi, Bassok, Lewis, Reimann and Glaser, 1989). We believe that this is a skill that can be taught independently and we will therefore not elaborate on this issue here. It is possible that feedback is just a preference as it is with learning preference but it could also be the case that different feedback works differently depending on the learner’s knowledge of the subject. One of the challenges for ITSs is to try to figure out when the user is having trouble with a subject and try to provide timely feedback in those cases, sometimes trying to steer them back to the correct solution track. Since we are using a discretised teaching method this will be less of a challenge for us since answers are either correct or incorrect, and can be tallied to give an indication of learning.
Chapter 4

Apparatus

4.1 Design Considerations

For the purpose of studying feedback methods in an automated teaching environment we needed to develop our own software where everything could be kept the same while changing only the feedback. We used Discrete Trial Training for the model of our design. We designed this game for an iPad because many schools in Iceland have either a class iPad which the class members can take turn in using or have adopted an iPad per pupil policy. The intended users were all elementary school children because if you have an application that is understood by a seven year old student it is most likely going to be understood by an older child. The development platform we decided to use was Unity 3D since Unity is a multi-platform engine giving us the option to deploy to Android as well in the future.

4.2 Quick Overview

The system is an automated learning environment where each user completes a number of trials as can in Figure 4.2.1. In each trial the user gets a verbal and a written instruction to point at a certain stimulus. The user gets a cheerful jingle followed by stars as reinforcement if they pick the right answer as can be seen in Figure 4.2.2. The user receive a neutral buzz sound and the screen shakes (as a basic feedback) if they pick the wrong answer. All users have to get eight or more independent (without a prompt) correct answers to get a golden star.
Two golden stars in a row mean that the user has reached the learned criteria for that stimulus group as can be seen in Figure 4.2.3.

**Figure 4.2.1** The first trial for a user learning the flower stimuli group. The screen has nine empty slots for coins that will either be golden (correct answer) or silver (wrong answer or correct with prompt)

**Figure 4.2.2.** The stars reinforcement. The stars get bigger and rise up and out of the screen
4.3 Design process

The design of this application was an iterative design and development process. For the development of this product we received a grant from the Icelandic Student Innovation Fund (í. Nýsköpunarsjóður námsmana). As part of this grant we developed a prototype that was ready for testing with users. This prototype was used to teach the recognition and discrimination of the Icelandic names of three similar birds (Common Snipe, European Golden Plover (Summer Plumage) and Meadow Pipit), three similar flowers (Bellflower, Heath Violet and Heart's Ease) and three similar bacteria (Salmonella, Listeria and Klebsiella). These were the same stimuli as were then used in the main study. The reason why we tried to have the stimuli in each group as similar as possible was to have the experiment challenging and to show the effectiveness of the system. The reason for these stimuli groups
was that the experimenter wanted to limit the chance of the participants knowing the stimuli beforehand.

The prototype was tested on six, five to seven year old kids and one 13 year old. In this experiment the sessions were counted to see how well they were doing. The younger kids did as follows. When using Errorless Learning the participants learned birds on average in 8 sessions, flowers in 9 session and bacteria in 9 session. When using response cost learning the participants learned birds on average in 10 sessions, flowers in 10.5 session and bacteria in 11 session. The control group learned birds on average in 12 sessions, flowers in 11 session and bacteria in 13 session. Even the five year olds were finishing the experiment in around 20 minutes so it seemed like the application was suitable for at least ages five and up.

The testing of the prototype exposed some fundamental usability flaws in the software. We decided to add timing of participants to the session count. The session count is too discrete and does not give enough information about how the participant did. We did not know if the participant spent 10 minutes on each session or 10 seconds. In the prototype the prompts were a red arrow and a red box around the correct answer. The idea behind that colour was that it would attract attention to it. This turned out to have an aversive effect on some of the participants since they would associate red with danger. The red square was therefore transformed to a yellow glowing box. The red arrow was changed into a picture of a pointing finger. In the prototype there was no way for the participants to repeat the verbal instruction which posed a problem for the non-reading participants or if the participant missed the verbal instruction. A voice bubble button, to repeat the instruction, was added into the final version of the project. The systematic fading of the prompts in the Errorless Learning condition were different. The boxes were removed before removing the red arrow. This was changed since the red arrow seemed to be the most noticeable prompt.
were faded after two consecutive correct answers which seemed too much and was changed to one. The reason why it seemed to much is that the number of stimuli in each group is three. That means you need at least 18 trials before you can start answering independently. If you are a fast learner and have already acquired the skill of discriminating and recognising the stimuli in fewer trials you still need to finish the 18 trials with the prompt as was the case with the 13 year old.

After fixing everything that was revealed during the initial user test, a formal pilot study with eight participants born in 2010, which was the youngest class of students in Icelandic Elementary schools, was performed. This pilot showed that that age group did not really suit the experiment. There can be a huge gap in cognitive development and maturity between children born early in the year versus those born late in the year. The time to complete for these participants was between nine and 30 minutes. The variance was not thought to be due to the different learning conditions but due to age. Most of the eight participants took over 25 minutes to complete which can be a long time to sit for such a young person. Some of the participants kept stopping in the middle of a session to tell the experimenter about all sorts of things. That not only made the time between answering skewed but also distracted the participant. One participant just liked the stars and sounds and kept pushing random buttons.

4.4 Detailed System Description

4.4.1. General

When the game is started the opening screen appears. In it the experimenter can make a new user by pressing the “Nýskrá” button, choose a user by pressing “Velja Notanda” button, and
send all the acquired data as a csv file by email by pressing the “Email button” as can be seen in Figure 4.4.1.1.

**Figure 4.4.1.1.** The opening screen with the choose a user, make a new user and Email button, the pick a user screen and the make a new user screen.

In the pick a user menu the experimenter can choose a user. That opens up that users menu where the experimenter can log that user in, change the feedback condition or delete the user as can be seen in Figure 4.4.1.2
Figure 4.4.1.2. The choose a user screen featuring only one student, “Brynjar” and the options menu for the user Brynjar. The user menu for each user has a log in (i. Innskrá) button, a dropdown menu for different feedbacks, and delete user (i. Eyða notanda) button.

If the experimenter choses to log in the user the skills menu opens up. Here the experimenter explains the experiment to the user and hands him the iPad.

Figure 4.4.1.3. The options menu for a user “Brynjar”. If the experimenter chooses to log in (Innskrá) the skills menu opens up. The skills menu is what the participant sees for the first time when playing the game.
Figure 4.4.1.4. The skills menu and the options menu for the skill menu. By pressing the cog wheel an options menu opens up where it is possible to change the condition with a drop down menu and to log out the user by pressing “Útskráning”. Logging out takes the user to the main menu.

Should the user press any of the stimuli group icons a new session starts with that stimulus.

Each session has nine trials. In the end of the session the user goes back to the skills menu.

Figure 4.4.1.5. The skills menu and the Trial screen.

If the user gets eight or more independent (without a prompt) correct answers (do not have to be in a row) the user receives a golden star as can be seen in Figure 4.4.1.6.
Figure 4.4.1.6. The user “Brynjar” after getting eight independent correct answers resulting in a golden star in the skills menu.

4.4.2 Errorless Learning

Errorless Learning differed from the other conditions in the way that in this condition the user would get a prompt. Figures 4.4.2.1 to 4.4.2.3 show a typical user progression in the first session when learning the bacteria stimuli group with the Errorless Learning condition. To simplify we only showed the progression of the Klebsiella bacteria and where all of the users answers were correct. The prompts are systematically faded with every correct answer.

Figure 4.4.2.1. A screenshot of the full prompt. Here the participant is asked to point to Klebsiella and gets a finger pointing to the right answer as well as a glowing yellow box.
**Figure 4.4.2.2.** A screenshot of the medium prompt. Here the participant is asked to point to Klebsiella. Now the finger has disappeared but the glowing box remains.

**Figure 4.4.2.3.** A screenshot of the small prompt. Here the participant is asked to point to Klebsiella. Now the finger has disappeared and the glowing box has become smaller.

Since each session consists of nine trials the user should have a no prompt status, as seen in Figure 4.4.2.4, in the end of the session (given that the user got all answers right with a prompt). This would give the user a silver star as explained above. The next session would then give the user a chance to get a golden star since the user can get an independent correct trial with all three stimuli. An independent correct trial gives the user a golden coin and eight or more golden coins give the user a golden star. No matter the prompt status, if the user picks the wrong answer in the errorless learning the user will be fully prompted to the right answer. In the case of the user picking the wrong answer in two trials in a row the user goes
Figure 4.4.2.4. A screenshot of the no prompt. Here the participant is asked to point to Klebsiella. Now the finger and the box have disappeared. This is what the each trial looks for both the Response Cost condition and the Trial and Error condition.

up a prompt status as follows:

- No prompt goes to small prompt.
- Small prompt goes to medium prompt.
- Medium prompt goes to full prompt.

4.4.3 Response Cost

In the Response Cost condition the participant would receive a gold coin if he chose the right answer and lose a gold coin if he or she chose the wrong answer (given he/she had any) as can be seen in Figures 4.4.3.1 and 4.4.3.2.

Figure 4.4.3.1. A user with two correct answers after two correct trials learning with the Response Cost condition.

Figure 4.4.3.2. The user chose the wrong answer and lost a gold coin.
4.4.4 Trial and Error

The Trial and Error condition serves as a baseline. In the Trial and Error condition the participant does not receive prompts, gets a gold coin for a right answer and does not lose a gold coin if a wrong answer was chosen, but gets a silver coin instead.
Chapter 5

Method

As mentioned above the goal of this project was to see if an automated version of Discrete Trial Training of learning environment was a suitable learning platform. What we mean by that is that all participants could properly recognise and discriminate similar stimuli. We further wanted to see if there was any performance difference between the three feedback conditions described above. By performance we mean the total time and the number of sessions it took the participants to reach the learned status, which is two golden stars in a row (see Chapter 4). We also think there will be a difference in the time between each chosen answer between feedback. The main reason for that is because the punishment in the Response Cost condition. We think that the participants might think more about each answer to avoid the punishment. It is also possible there is a difference in how enjoyable the participants depending on the feedback condition. Therefore the hypotheses being tested were three.

- The first hypothesis $H_A$ states that there is a mean difference between different feedback conditions in how fast the participants get a learned status (explained below). The null hypothesis $H_{A0}$ states that there is no mean difference between different feedback conditions in how fast the participants get a learned status
- The second hypothesis $H_B$ states that there is a mean difference between different feedback conditions in the average time between clicks. The null hypothesis $H_{B0}$ states that there is no mean difference between different feedback conditions in the average time between clicks.
• The third hypothesis $H_C$ states that there is a mean difference between different feedback conditions in how enjoyable the participants thought the game was. The null hypothesis $H_{C_0}$ states that there is no mean difference between different feedback conditions in how enjoyable the participants thought the game was.

5.1 Conditions

The performance of students learning by three different conditions was compared. All of the conditions are the same except for one variable as can be seen in Table 1. In all three conditions the participant got a cheerful jingle followed by stars (as reinforcement) if they picked the right answer. In all three conditions the participant received a neutral buzz sound and the screen would shake (as a feedback) if they picked the wrong answer.

1. **Errorless learning.** In the Errorless Learning condition the participant was prompted with a finger pointing at the correct answer as well as a large golden glow around it. The prompt was faded with every right answer. First the finger disappeared. Then the golden glow got smaller. After three correct answers in a row the participant would get an independent trial. If the participant chose the wrong answer the participant was fully prompted to the right answer. If the participant selected the wrong answer twice in a row the prompts were reinstated and increased (for every 2 wrong answers in a row).

2. **Response cost:** In the Response Cost condition the participant would lose a gold coin if he or she chose the wrong answer (given he/she had any).

3. **Trial and error:** The Trial and Error condition served as a baseline variable. In the Trial and Error condition the participant did not receive prompts and did not lose a gold coin if a wrong answer was chosen.
All three conditions and their different characteristics can be seen in Table 5.1.1.

Table 5.1.1.
The difference between conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Reinforcement</th>
<th>Prompts</th>
<th>Feedback</th>
<th>Punishment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errorless Learning</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Response Cost</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Trial and Error</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

Note. All conditions differ by one element to single out what factor might make a difference. Feedback in this case means that the participant was informed that he chose the wrong answer.

5.2 Participants & Procedure

For the main study, 17 students born in 2007 were recruited at an elementary school. The experiment was reported to the data protection authority and parental consent was given.

Eight of them were male, nine were female. One of the participants declared to the experimenter before hand that he had a hard time learning because of his attention deficit disorder. All participants were allowed to discontinue a session or leave the experiment at any time. One female participant asked to leave after nine sessions of learning bacteria with the Errorless Learning condition. One female participant had to leave because her parents were there to pick her up. Three of the participants were polish and spoke Icelandic at different levels (one almost spoke no Icelandic, one spoke moderate Icelandic and one spoke without a discernible accent). A between participant design was used to examine which of these three learning conditions was the most effective in teaching novel stimuli. Participants were randomly distributed to three groups based on what learning condition they would learn by. To eliminate learner effects all participants in each groups started with a different stimuli group. For example, the first participant in group one would start with birds, the second
participant in group one would start with flowers and so on. To reduce time of day effects the first participant was from group one, the second one from group two and so on. Before starting the game all students were shown a picture of the items to be learned. If a participant could name any of the organisms the participant would not learn that group of organisms but instead proceed learning the next one since knowing one third of the objects would skew the data immensely. The participant were then read a scripted explanation of the game mechanics (see Appendix A). The application provided a speech bubble that could be pressed to repeat the question. The experimenter sat in another corner of the room and told the participants that he would answer any questions they had about the mechanics of the game but had to do work. This was done to be less imposing on the participants. After testing the 17 participants it seemed like the Errorless feedback was having a considerable effect on the time it took learners to learn the bacteria stimuli. We therefore decided to add another class of students that would only learn the bacteria. From this class we got 9 students, four were male and five were female.
5.3 Measurement

Measurements were performed by the application. The variables that were recorded are summarised in Table 5.3.1.

Table 5.3.1.
The variables that were recorded and their description

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session Count</td>
<td>How many sessions it took to gain two golden stars in a row (the learned criteria). Each session contains 9 trials.</td>
</tr>
<tr>
<td>Average time between clicks</td>
<td>The average time between each button press (each answer) in a session, including the time it took to press the first answer.</td>
</tr>
<tr>
<td>Total time</td>
<td>The total time it took to reach the learned criteria. This does not include time spent in the menu.</td>
</tr>
<tr>
<td>Time stamp of each answer</td>
<td>The date and time of each answer.</td>
</tr>
<tr>
<td>The response of the participant</td>
<td>What answer the participant gave.</td>
</tr>
<tr>
<td>The correct answer</td>
<td>What the correct answer was when the participant answered.</td>
</tr>
<tr>
<td>Total number of correct answers</td>
<td>The current total number of correct answers.</td>
</tr>
<tr>
<td>The total number of wrong answers</td>
<td>The current total number of wrong answers.</td>
</tr>
<tr>
<td>The total number of prompts</td>
<td>The current total number of prompted answers (assisted). Only applicable in Errorless Learning.</td>
</tr>
<tr>
<td>The user rating</td>
<td>The participants evaluation of how enjoyable the game was on a scale of 1 to 10. Participants answered on google forms (see Appendix B).</td>
</tr>
</tbody>
</table>
Chapter 6

Results

In this chapter, we will present the results….

6.1 Preface

When looking at how long it takes students to achieve a learned status, we see that the session count highly correlates with the time to learn in minutes. Therefore the analysis that follows will focus on the time to learn. Most of the participants (ca 70%) could recognise one of the birds, the European Golden Plover, and therefore data from that stimuli group was excluded. A few students (ca 33%), pointing at different pictures of flowers, said that they knew that one of the them was called Fjóla. Two of the flowers end in the word Fjóla but the students could not say the full name of the flower. This could have happened because the children kept talking about the game in class (according to the teacher). Another reason could be because Fjóla sounds like the colour of the flowers (Fjólublár, e. violet). A comparison of all the data acquired for the flowers stimuli group showed little to no difference in the time it took participants to reach the learned criteria (barring one outlier) as can be seen in Figures 6.1.1 and 6.1.2.
Figure 6.1.1. The total time it took each participant to reach the learned criteria in the flower stimuli group by learning method. There are only four participants in the Response Cost group because one of the participants could name Bellflower (Campanula).

We therefore decided to focus the analysis on the bacteria stimuli group. Furthermore, initial analysis indicated that the Errorless feedback was having a considerable effect on the time it took learners to learn the bacteria stimuli. This seemed worth looking into, but due to the limited number of subjects, it was hard to draw strong conclusions. Therefore we decided to add one more class of subjects, in the bacteria stimuli group only. It is possible that there is an inherent difference between these stimuli groups since the participants are more likely to have had exposure to flowers than bacteria. We also expected the names of flowers to be easier to grasp since they contain the names of Icelandic colours while it is unlikely that
Figure 6.1.2. The number of sessions it took each participant to reach the learned criteria. There are only four participants in the Response Cost group because one of the participants could name Bellflower (Campanula).

the participants would have heard the names of bacteria. This makes participants completely novice to bacteria stimuli and they have a harder time inferring from existing knowledge. We will revisit the question of stimuli creation in our final discussion.

6.2 Learning Effectiveness of System

As mentioned above one of the things we wanted to see is whether this automated version of automated Discrete Trial Training was an effective platform for teacher-less learning. All of the participants, in all conditions, reached the learned criteria for all of the stimuli groups without teacher intervention and therefore we are quite confident that this approach to automating learning, for certain kinds of knowledge (i.e. recognition and discrimination) is an
effective one. Figure 6.2.1 shows a typical learning curve for a participant learning the bacteria stimuli group with the Errorless Learning condition.

![Graph showing percentage of right, wrong, and prompted answers](image)

**Figure 6.1.2.** The typical trend for participants when learning with the Errorless Learning condition. The graph shows the percentage of right, wrong and prompted answers after each answer. A right answer is an independent correct trial (without a prompt).

### 6.3 Hypothesis A: Difference in Time to Learn

As mentioned above we wanted to see if there is a difference between any of the conditions in how fast the participants reached the learned criteria. When learning the bacteria stimuli group it looked like the participants did much better learning with Errorless Learning as can be seen in Figure 6.3.1.
Figure 6.3.1. The total time it took all participants in the first class to reach the learned criteria when learning the bacteria stimuli group, grouped by learning condition.

As can be seen in Table 6.3.1 the average time to learn with the Errorless Learning condition was more than twice as fast as the Trial and Error condition and more than three times faster than the Response Cost condition.

Table 6.3.1.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errorless</td>
<td>5</td>
<td>7.67</td>
<td>1.53</td>
<td>0.2</td>
</tr>
<tr>
<td>Response Cost</td>
<td>5</td>
<td>27.18</td>
<td>5.43</td>
<td>9.06</td>
</tr>
<tr>
<td>Trial and Error</td>
<td>5</td>
<td>17.37</td>
<td>3.47</td>
<td>7.07</td>
</tr>
</tbody>
</table>

A Levenes test of homogenity shows that the distributions have homogenous variance as can be seen in Table 6.3.2.
A single factor ANOVA showed that the difference between the groups was not significant at the 5% level although it was really close (0.064) as can be seen in Table 6.3.3.

Table 6.3.2.
Levenes test of homogenity for the first class

<table>
<thead>
<tr>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.844</td>
<td>2</td>
<td>12</td>
<td>.20</td>
</tr>
</tbody>
</table>

We therefore proceeded to do a Tukey’s honest significance difference (HSD) test. As we can see in Table 6.3.4 the difference is mainly between Errorless Learning and Response Cost.

Table 6.3.3.
A single factor ANOVA for the total time to learn bacteria

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>38.04</td>
<td>2</td>
<td>19.02</td>
<td>3.495</td>
<td>0.064</td>
<td>3.885</td>
</tr>
<tr>
<td>Within Groups</td>
<td>65.31</td>
<td>12</td>
<td>5.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>103.35</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6.3.4.

Tukey HSD multiple comparison of the time to learn bacteria

<table>
<thead>
<tr>
<th>(I)Feedback</th>
<th>(J)Feedback</th>
<th>Mean Diff (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95 % Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errorless</td>
<td>Response</td>
<td>-3.9</td>
<td>1.48</td>
<td>.05</td>
<td>-7.84, .36</td>
</tr>
<tr>
<td></td>
<td>Trial</td>
<td>-1.94</td>
<td>1.48</td>
<td>.42</td>
<td>-5.88, 2.00</td>
</tr>
<tr>
<td>Response</td>
<td>Errorless</td>
<td>3.9</td>
<td>1.48</td>
<td>.05</td>
<td>-0.36, 7.84</td>
</tr>
<tr>
<td></td>
<td>Trial</td>
<td>1.96</td>
<td>1.48</td>
<td>.41</td>
<td>-1.97, 5.90</td>
</tr>
<tr>
<td>Trial</td>
<td>Errorless</td>
<td>1.94</td>
<td>1.48</td>
<td>.42</td>
<td>-2.00, 5.88</td>
</tr>
<tr>
<td></td>
<td>Response</td>
<td>-1.96</td>
<td>1.48</td>
<td>.41</td>
<td>-5.90, 1.97</td>
</tr>
</tbody>
</table>

Even though the difference between the Errorless Learning condition and the Trial and Error condition was almost double we cannot say with any confidence that there is a difference in the time it took them to learn, perhaps because the sample size was so low. We can on the other hand say with 95% confidence that there is a difference between the Errorless Learning condition and the Response Cost condition in how fast they reach the learned criteria.
Figure 6.3.2. The total time it took participants in both classes to reach the learned criteria when learning the bacteria stimuli group, grouped by learning condition. There are eight participants in each group, the last three in every group are from the second class.

When we added the second class of students (three new students for each condition) it looks like the striking results from the first class was more likely due to random distribution than some inherent difference as can be seen in Figure 6.3.2. Figure 6.3.2 shows that all three participants from the second class, learning with Errorless Learning, did considerably worse than the first class. As can be seen in Table 6.3.5 the Response Cost condition continues to do considerably worse.

Table 6.3.5.
Summary of the time to learn bacteria by condition for both classes

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errorless</td>
<td>8</td>
<td>25.58</td>
<td>3.20</td>
<td>5.48</td>
</tr>
<tr>
<td>Response Cost</td>
<td>8</td>
<td>48.72</td>
<td>6.19</td>
<td>11.73</td>
</tr>
<tr>
<td>Trial and Error</td>
<td>8</td>
<td>24.28</td>
<td>3.03</td>
<td>4.64</td>
</tr>
</tbody>
</table>
A Levene's test of homogeneity show that the distribution is homogenous as can be seen in table 6.3.6.

Table 6.3.6.

Levene's test of homogeneity for both classes

<table>
<thead>
<tr>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.01</td>
<td>2</td>
<td>12</td>
<td>.159</td>
</tr>
</tbody>
</table>

A single factor ANOVA, however, shows that there was in fact a significant difference between the conditions at the 5% significance level as can be seen in Table 6.3.7.

Table 6.3.7.

A single factor ANOVA for the total time to learn bacteria for both classes

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>50.51</td>
<td>2</td>
<td>25.255</td>
<td>3.467</td>
<td>0.05</td>
<td>3.47</td>
</tr>
<tr>
<td>Within Groups</td>
<td>152.99</td>
<td>21</td>
<td>7.285</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>203.51</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We therefore proceeded to do a Tukey’s Honest Significance Difference (HSD) test. The test shows that we can with higher than 90% confidence say that the Response Cost differs from both Errorless Learning (p = .09) and Trial and Error (p= .07).
Table 6.3.8.
Tukey HSD multiple comparison of the time to learn bacteria for both of the classes

<table>
<thead>
<tr>
<th>(I)Feedback</th>
<th>(J)Feedback</th>
<th>Mean Diff (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errorless</td>
<td>Response</td>
<td>-2.99</td>
<td>1.35</td>
<td>.09</td>
<td>-6.39  - .41</td>
</tr>
<tr>
<td></td>
<td>Trial</td>
<td>.16</td>
<td>1.35</td>
<td>.99</td>
<td>-3.24  3.57</td>
</tr>
<tr>
<td>Response</td>
<td>Errorless</td>
<td>2.99</td>
<td>1.35</td>
<td>.09</td>
<td>-.41  6.39</td>
</tr>
<tr>
<td></td>
<td>Trial</td>
<td>3.16</td>
<td>1.35</td>
<td>.07</td>
<td>-.25  6.56</td>
</tr>
<tr>
<td>Trial</td>
<td>Errorless</td>
<td>-.16</td>
<td>1.35</td>
<td>.99</td>
<td>-3.57 3.24</td>
</tr>
<tr>
<td></td>
<td>Response</td>
<td>-3.16</td>
<td>1.35</td>
<td>.07</td>
<td>-6.56  .25</td>
</tr>
</tbody>
</table>

Anecdotally, the same distribution can be seen with the session count for the participants when learning bacteria (Figure 6.3.3).
Figure 6.3.3. The session count it took participants in both classes to reach the learning criteria when learning the bacteria stimuli group. There are eight participants in each group, the last three in every group are from the second class.

Since the one way ANOVA shows significance at the 5% level we reject the null hypothesis $H_{A0}$ for hypothesis $H_A$ and accept the alternative hypothesis that there is a difference between the conditions.

6.4 Hypothesis B: Difference in Time Between Clicks

When it comes to hypothesis B, that there is a difference between the average time between clicks, there seems to be no difference between conditions as can be seen in Figure 6.4.1.
Figure 6.4.1. The average time between clicks for participants in both classes when learning the bacteria stimuli group. There are eight participants in each group; the last three in every group are from the second class.

As can be seen in Table 6.4.1, the Errorless Learning condition was the fastest while the Response Cost and Trial and Error condition are almost the same.

Table 6.4.1.
Summary of the average time between clicks when learning bacteria by condition

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errorless</td>
<td>8</td>
<td>19.95</td>
<td>2.49</td>
<td>0.38</td>
</tr>
<tr>
<td>Response Cost</td>
<td>8</td>
<td>22.65</td>
<td>2.83</td>
<td>.29</td>
</tr>
<tr>
<td>Trial and Error</td>
<td>8</td>
<td>22.38</td>
<td>2.8</td>
<td>0.49</td>
</tr>
</tbody>
</table>

A single factor ANOVA showed that the difference between the groups was not significant at the 5% level, as can be seen in Table 6.4.2.
Table 6.4.2.

A single factor ANOVA for the average time between clicks when learning bacteria for both classes

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>553</td>
<td>2</td>
<td>.276</td>
<td>713</td>
<td>0.502</td>
<td>3.47</td>
</tr>
<tr>
<td>Within Groups</td>
<td>8.139</td>
<td>21</td>
<td>388</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8.691</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since there is no mean difference to be found in the average time between clicks we cannot reject the null hypothesis $H_{B0}$ for the second hypothesis $H_{B}$.

### 6.5 Hypothesis C: Difference in Enjoyment

Hypothesis C states that there is a mean difference between different feedback conditions in how enjoyable the participants thought the game was. Participants in general were very enthusiastic about playing the game. The average user rating for all users was 8.4 (Female 8.25, Male 8.5). The difference between conditions can be seen in Table 6.5.1.

Table 6.5.1.

Average user rating for users by condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Average user rating</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errorless Learning</td>
<td>8</td>
<td>9.13</td>
<td>0.98</td>
</tr>
<tr>
<td>Response Cost</td>
<td>8</td>
<td>8.13</td>
<td>5.27</td>
</tr>
<tr>
<td>Trial and Error</td>
<td>8</td>
<td>7.88</td>
<td>6.70</td>
</tr>
</tbody>
</table>
A single factor ANOVA showed that there was no significant difference between the groups as shown in Table 6.5.2. We therefore cannot reject the null hypothesis $H_{c0}$ for hypothesis $H_c$.

Table 6.5.2.

A single factor ANOVA for the user rating

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>$F$</th>
<th>P-value</th>
<th>$F$ crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>7</td>
<td>2</td>
<td>3.5</td>
<td>0.811</td>
<td>0.458</td>
<td>3.467</td>
</tr>
<tr>
<td>Within Groups</td>
<td>90.625</td>
<td>21</td>
<td>4.315</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>97.625</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 7

Discussion

As explained above there was difference between the groups in how fast they reached the learned criteria. The Response Cost condition did significantly worse than the other conditions. When looking at the difference in the average time between clicks and enjoyment the results show no difference. This can be an important finding since many educational computer games use Response Cost as a feedback method for teaching. This would be acceptable if the thrill of the challenge of maybe loosing a coin makes learning more enjoyable but our findings do not support that. This could possibly be because of design differences. In the games we mentioned above that use Response Cost the player starts with hearts and the session ends when you loose all your hearts. In our design you collect coins and loose them by choosing a wrong answer. It is possible that their design is more enjoyable. It is also possible that the performance of feedback varies between subjects and the participants self-efficacy in that subject. This could for example make Errorless Learning more suitable when the user has no idea about what is being taught and needs more help while Response Cost could be more suitable when the user has some idea.

Both the flower and bird stimuli groups suffered from the lack of control of the participants exposure to the objects to be learned. Even though both classes were from the same school the second class did the experiment in the schools library where there were people occasionally walking by while the first class had access to a private room. This could possibly explain why the average of the first class did slightly better. Since the sample size is small the variance in data is big and therefore it is possible that a clearer difference could be
seen with a larger data set and a longer test period. The experimenter told the participants that he had made this educational game which can make the participants eager to please and should be changed for future studies. A novelty effect also has to be taken into consideration since there can be a difference in how open the participants are to new situations. There is also the case of the Hawthorne effect, a slight tendency of individuals to modify their behaviour if they know they are being observed (see McCarney, Warner, van Haselen, Griffin, & Fisher, 2007) but this effect should even out for all participants. Since we are not comparing our results to standard teaching methods it is hard to tell if all of the groups did good or bad. The same can be said about the user rating. Did the participants rate the game so high because it was a break from a monotonic school schedule or was it in fact so enjoyable? The participants in the errorless group seem to have a marginally shorter average time between each click. That is likely because some of the participants seemed to press the right answer immediately when presented with a prompt instead of using the prompt as a help to make the connection between the instruction and the answer. This could also be eliminated with a longer test period (given that the participants would figure out how to use the prompt properly).
Chapter 8

Conclusion & Future Work

This thesis introduced an automated version of Discrete Trial Training that implements three different feedback procedures: Errorless Learning, Response Cost and Trial and Error. Our results show that an automated version of Discrete Trial Training is a suitable platform for learning, and the Response Cost condition performed worse than the Errorless Learning and the Trial and Error conditions.

There is a lot of variance in the data so future research would need to include more students and test them for a longer period to get rid of any novelty effect and any difference that comes forth because the participants are unsure about the controls or uneasy about doing an experiment. This could be done by having a training session for the participants where they learn something else. To be sure all participants have equal exposure to the stimuli we suggest making new stimuli. These stimuli could be odd shapes that we give names.

To see if students do have a preference to feedback a pre-screening would need to be devised. This test would split students into groups based on their preference. Each student would then be randomly assigned to a feedback method. We could then see if students do better when learning with their desired feedback method versus learning with a non desired one. If this is the case it is possible (with enough data) to make profiles for the user and change the feedback method depending on the user, the users knowledge base and the subject at hand.
A new method could be produced, which is a mixture between Response Cost and Errorless Learning, where participants would lose a token for doing wrong but could get less of a penalty if they would ask for help.

This thesis connected three different fields of study: Computer Science, Psychology and Education. As technology becomes an ever bigger part of the educational system and the learning experience we believe it is important that the design of educational software be linked to well versed pedagogical and psychological theories.
References


Chi ….Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems


Appendix A

The scripted explanation for the participants in Icelandic.

“Sæll/ sæl ég heiti Brynjar Ólafsson og ég var að búa til námstölvuleik. Mig langar að fá þig til að prófa hann. Mig langar að byrja á því að spyrja þig hvort þú kannist við eitthvað af þessum atriðum á meðfylgjandi myndum.

Ef ekki þá er þetta atriðin sem þú munt læra núna. Þú þarf að klára hvert atriði áður en þú byrjar á því næsta. Til að klára atriði þarf þú að fá tvær gullstjörnur í röð. Til að fá gullstjörnu þarf þú að ná átta eða fleiri gullpeningum af niú.”

1. Texti fyrir Errorless hóp:

   “Til að byrja með færð þú hjálp. Hjálpin er í formi gyllts kassa og fingurs sem bendir á rétt svar um leið og það er leisið upp. Í hvert sinn sem þú velur rétt svar með hjálp færð þú silfur stjörnu og hjálpin minnkar. Ef þú velur rétt án hjálpar færðu gullpening. Hvert rangt svar gefur silfurpening.“

2. Texti fyrir Response cost hóp:

   “Hvert rétt svar gefur gull pening. Ef þú velur rangt svar þá missir þú gull pening sem þú ert búinn að vinna þér inn.”

3. Texti fyrir Control hóp:

   “Hvert rétt svar gefur gull pening. Hvert rangt svar gefur silfurpening.“

Þegar nemandi er búinn hvort sem það er að hann valdi að hætta, kláraði öll atriðin er tíminn var búinn.

   “Mig langar að þakka þér fyrir að prófa námstölvuleikinn minn en áður en þú færð að fara er eitt sem á vil biðja þig um að gera. Hér fyrir framan þig er skali á bilinu 0 og 10.
Skalinn virkar alveg eins og hitamælir en nú mælir hann hversu skemmtilegt þér fannst að spila leikinn. 0 táknar ekkert skemmtilegur á meðan að 10 táknar skemmtilegast í heimi.”

The scripted explanation for the participants in English.

“Hi, my name is Brynjar Ólafsson. I have created an educational computer game and would like you to try it for me. I want to start by asking you whether you recognise any of these items. You will be learning these items now. You have to complete each category before starting the next. To finish a category you have to get two gold stars in a row. To get a gold star, you have to reach eight or more gold coins out of nine.”

1. Text for errorless group:

"To begin with, you will receive a prompt. The prompts are in the form of a glowing box around the right answer and finger pointing to it as soon as it is read. Every time you choose the correct answer with the help you get a silver star and the prompt will be faded. If you choose the right answer without a prompt you get a gold coin. Each wrong answer gives you a silver coin."

2. Text for the cost Response group:

"Each correct answer gives a gold medal. If you choose the wrong answer, you will lose a gold medal you have earned.”

3. The text of the Control group:

"Each correct answer gives a gold medal. Each wrong answer gives a silver coin."

When a student is finished whether he chose to stop, finished all the items or the time was finished.
"I want to thank you for trying out my educational computer game. Before you leave I would like you to do one thing for me. Here in front of you is a scale between 0 and 10. The scale works just like a thermometer but now the scale represents how entertaining you thought the game was. Zero represents not entertaining at all while 10 represents as entertaining it can get. “
Figure b1. A question about how entertaining the game was. Zero represents not entertaining at all while 10 represents as entertaining it can get.