NTNU – Trondheim
Norwegian University of
Science and Technology

Investigating the effect of a robotic presence compared to a virtual robot in teaching angles and turn measurements to children

Nicklas Sørlie Utgaard

Jan Tore Stølsvik and Nicklas Utgaard

Investigating the effect of a robotic presence compared to a virtual robot in teaching angles and turn measurements to children

TDT4501 Datateknikk, Master thesis, Spring 2014

Artificial Intelligence Group
Department of Computer and Information Science
Faculty of Information Technology, Mathematics and Electrical Engineering

Abstract

In the later years robotics has seen a huge increase within domestic use, and have now become an affordable tool in the daily life of most people. The goal of this project was to investigate the differences between a physical and virtual robot in terms of increased content knowledge, learning motivation, and interest in science, technology, engineering and mathematics (STEM). To investigate this we conducted an experiment at Trondheim'm International School (THIS), using a quasi-experimental setup with two treatment group, virtual and physical robot. The results showed that there does not exist a statistically significant difference in content knowledge gain, motivation or interest between the robotics group and the simulator group.

Sammendrag

De siste årene har roboter økt i popularitet innenfor vanlige hushold, og har kommet ned på ett prisnivå som gjør robotene tilgjengelig for folk flest. Målet med denne oppgaven var å undersøke forskjellene mellom en fysisk og virtuell robot med tanke på å øke kunnskapsnivået, motivasjonen og interessen for vitenskap, teknologi, ingeniørskap og matematikk (STEM). For å undersøke dette utførte we ett eksperiment hos Trondheim's International School (THIS), hvor vi brukte ett kvasieksperimentelt oppsett med to behandlingsgrupper, virtuell og fysisk robot. Resultatene viste at det ikke fantes noen statistisk signifikant forskjell i økning av kunnskapsnivå, motivasjon eller interesse mellom robot gruppen og simulator gruppen.

Preface

This master thesis was carried out within the Information Management (IF) group under the Department of Computer and Information Sciene (IDI) at the Norwegian University of Science and Technology (NTNU).

Jan Tore Stølsvik and Nicklas Utgaard
Trondheim, July 11, 2014

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

After the technology boom in the 20th century, technology have become extremely important, and thus so have science, engineering and mathematics. It is therefore vital to adapt a continuous effort to improving the education provided within these fields. However from the year 2000 up to 2012, the PISA tests showed a decrease in the average score on the mathematics test (from 500 to 494) overall in Europe [35; 36]. This is also visible in the Norwegian results, indicating a decrease from 499 in 2000 to 489 in 2012 within mathematics and generally a lower score then our neighbouring countries [37].

In 1980 Seymour Papert published the book "Mindstorms: Children, computers, and powerful ideas"[39], where his ideas of an alternative learning environment, which utilizes robotics were presented. It was Papert's belief that educational robotics held a major potential improvement of the current learning environment. Allowing children to interact and construct their their own knowledge in ways previously impossible. This book was the start of educational robotics and many researchers rely on the early work conducted by Seymour Papert within educational robotics. However, they often fail to explain how Papert's theories are applied within their own research.

In the later years robotics has seen a huge increase within domestic use, and have now become an affordable tool in the daily life of most people [21; 19]. A lot of research has been done to show that robotics can be beneficial in an educational setting. The majority of research into educational robotics shows that gains in content knowledge can be achieved using robotics.

1.1   Motivation and goals

Educational theorists believed robotics could be utilized in a learning environment with success, but there has been little incorporation seen throughout the world. Some speculate that the limited adoption is due to lack of empirical evidence on the effect of robotics as a learning tool [48]. Another possibility is that the usage of robotics in education usually has been as a tool to teach students about robotics itself, and thus have formed a narrow field of applicability [42]. The third possibility comes down to the price of robotics equipment, e.g a lego mindstorm kit costs about 650 USD. An interesting field of research is therefore to look at how robotics can be utilized to teach about non-robotics subjects, and perhaps even be used as a motivational or attitude changing tool while still keeping the cost down.

Another common sight while reading about educational robotics is a lack of research regarding the usefulness of a physical robots in itself, as it seems like most people made the jump from traditional teaching to teaching with a robot in one big jump. Several papers also limit themselves by only targeting superficial aspects like fun, engagement, motivation and cooperation. While still basing our approach on Papert's pedagogical theories, we will try to fill some of these holes. We want to test whether relatability and knowledge gains can be improved by using a physical robot compared to using a simulator confined to a local device. We want to take a closer look into the pedagogic aspects of mental model building and investigate the differences in using a physical robot compared to a simulator which is significantly cheaper to develop, buy and maintain.

Since the robotics business is growing all the time and educational robotics has been shown to have great promise we focus this project on an investigation into the potential of robotics in schools and make an effort to expand the ChIRP team's horizon into the world of education. This thesis serves as an introduction into the educational world for the ChIRP team at IDI. We are breaking new ground for the team, establishing collaborations and providing the basis for future work. Because of this we've designed our system implementation and report in such a way that others can learn from and expand upon it.

Project Goals
The goal of this project is to investigate the use of robotics in schools for the ChIRP group. We aim to implement an intuitive way to control the robot with a solid foundation in pedagogic theory. This will require customizations to the ChIRP robot and to test our assumptions with school children. There will not

be enough time to fully test this system but it will serve as a platform, provide guidelines and pitfalls for others to expand on in the future.

In the end we wish to contribute to the discussion of educational robotics both through our software implementations but also the conclusions we draw on robot versus simulator. In this way we intend to further the research in this area.

1.2   Research questions

After the initial literature review (detail in part 2) we found several areas within the research which could benefit of additional research. In an attempt to limit the scope of this project we selected mathematics as our main subject of interest. And in collaboration with people at NTNU we decided to take on one of the research areas which would benefit from more research, physical versus virtual robot. We set out to create a platform which could be used in an educational setting with the ChIRP robot as its core component, with the end goal of providing empirical evidence that could help us answer our research questions.

The formal articulation of the research questions ended up as:

RQ1:   How can using robotics in school help students understand what an angle is and aid in their angle estimation skills?

RQ2:   What are the strength and weaknesses of using a robot compared to a simulated environment?

1.3   Execution of project

This entire project was conducted from February to July 2014. During this time period we've created a pilot-project looking into the use of robotics in an educational setting with a focus on angles and how these can be taught. All we have done leads up to a pilot test at the Trondheim International School (THIS), where we tested our hypothesis and how well suited the ChIRP robot was in an educational setting. The contact established between the ChIRP team and the school will hopefully help foster further research within educational robotics from the ChIRP team.

Literature Reviews
In order to understand educational robotics state of the art we performed a literature study into educational robotics with a focus on mathematics. The setup for

this review can be read in section 2.1. The applicability of robotics in a mathematics setting has proven itself over and over again. The results from the literature study lead us to focus more specifically on angles instead of mathematics in general. We have chosen to focus on angles as it is one of the core concepts within mathematics, students tend to struggle with it, and visualization of angles can be achieved with the help of a robot. After the state of the art literature review we identified the need for an additional literature study. This study was conducted in order to learn more about the current pedagogical theories and test setups, which would help us find a suitable approach to create a solid platform for this project and potential successors.

Design and Development
In order to conduct this project we needed to develop several accessories for the ChIRP and a tablet application for testing in an elementary school. Since this was considered a pilot-project for the ChIRP robot we aimed at designing and developing a prototype with emphasis on expandability, modifiability and portability. We also needed to develop a small circuit to enable bluetooth communication between the robot and tablet application. All in all there was a lot of work to be done before a prototype would be ready for the experiment.

Experiment
To answer our research questions we had to test the robot in a school context. We contacted several schools, had meetings with Ungt Entreprenørskap and other researchers at IDI responsible for attracting students to study at IDI, all of whom were interested in the project. The time frame of the project did however not fit within their existing schedules. Our supervisor (Pauline) had previously been in contact with Trondheim International School (THIS), and they were very enthusiastic about the project, very helpful throughout the whole project, and we were lucky to be able to conduct our experiment with their students. We had 2 meetings with the teacher of the class we were assigned before the experiment itself. In these meetings we discussed our tests and curriculum and got feedback which we used to improve our experiment. It was important for us to include teachers as the end goal is to integrate the robotic activities to normal schools. The experiments were run at the end of the school year because of the long process of developing the system and experiment. Several feedback iterations had to be done with both the software application and the experiment in order to ensure a pedagogical sound foundation. This resulted in a smaller experimental time span than initially intended.

1.4 Report Structure

This report is split into six chapters before including a set of appendices for any information which may be relevant to the reader.

Chapter 1 will provide a short introduction to the project, including motivation behind conduction the project, reseach question and the different phases of the project.

Chapter 2 is dedicated to the literature review and the process used when conducting the literature review.

Chapter 3 presents our design principles and a discussion related to other software solutions, before presenting the system created during this project.

Chapter 4 presents our experimental setup, the procedures used, and a discussion around the reliability and validity of the study.

Chapter 5 provides an overview of all the results and analyses from the experiment.

Chapter 6 is the culmination of this report, and presents a discussion of the whole project before presenting our conclusion and proposals for future work.

LITERATURE AND RELATED WORK

To understand the educational robotics state of the art, focus our goals and define proper research questions for this project, we conducted a literature review. In this chapter we will first present the process, planning and results of the review. Then to aid the reader we've included a short introduction to the theory of constructionism, as this was used in the majority of research within education robotics. Finally, in section 2.4, we'll provide an introduction to different quantitative and qualitative methods of education research available.

2.1 Review setup

The review was conducted in line with the guidelines drafted by Kitchenham's and Khan's guide for writing a systematic review [23; 22]. The initial steps of identifying a need for the review and commissioning a review was however omitted. The phases used during the review was;

Phase 1: Planning

1. Specifying the research question(s)
2. Developing a review protocol
3. Evaluating the review protocol

Phase 2: Conducting the review

1. Identification of research
2. Selection of primary studies

3. Study quality assessment

4. Data extaction and monitoring

5. Data synthesis

Phase 3: Reporting the review

1. Communicating the results through a report.

Initially, we preformed an general search into educational robotics, in order to obtain some fundemental knowledge regarding the subject. In this search we stumbled upon another review written in 2011 by Fabiane Barreto Vavassori Benitti[5], which to a very large extend covers the same topic as initially planned by this review. There was however some minor differences between our research question, but the papers relevant to Benitti proved to a large extend to be relevant for our review as well.

Where Benitti asked general questions like "'*What topics are taught through robotics in schools?*"', "'*is robotics an effective tool for teaching? What do the studies show?*"', and "'*How is student learning evaluated?*"'. We focused on narrowing down the scope of the review, limiting the review to research related to math and how math can be taught in schools using robotics. The research questions used for this review is stated below;

Question 1: How well did the different mathematics concepts get taught with robotics?

Question 2: What advantages or disadvantages except learning gains are there?

Question 3: Were any suggested improvements to the experiments identified?

Question 4: What lacks of research are there in the literature?

The review was conducted in February and March of 2014, with papers retrieved from all the major bibligraphic databases, including, but not limited to, CiteSeer, ACM Digital Library, SpringerLink, ERIC, IEEE XPLORE, Wiley Inter Science, and ScienceDirect. In addition to these databases the search was applied to the google scholars search engine to ensure that we got most of the relevant papers.

The general search query was created using groups of synonyms, concatenated by the `and`/`or` operators before adjustmentint the query to each unique database (e.g making sure the search query was compatible with the search engine at

any given site). The search query used in this review was: `(math or stem or mathematics) and (education or learn or learning or educational or teach or teaching) and (robot or robotics or robots) and (school or k-12)`.

To remove papers that not were of any interest we first applied a set of inclusion criteria, before applying a set of exclusion criteria. In order for a paper to be included in the review it had to pass all inclusion criteria, and not violate any of the exclusion criteria.

IC1 The purpose of the paper is to investigate the usage of robotics in school, where the goal is not to teach about robotics itself.

IC2 The paper should contain some sort of assessment, quantitative or qualitative, of the learning outcome and/or experiences from the study.

IC3 The assessment must address the development of math skills.

IC4 The study should be done in an elementary, middle or highschool context.

IC5 The study should involve the use of physical robots.

EC1 Article does not address the subject of using robotics to teach maths.

EC2 Article does not address, or refer to, any pedagogical foundation.

These criteria diverge from Benitti's review in that qualitative assessments also are included. We justify this by acknowledging the fact that non-immediate returns of educational robotics may be equally important to immediate curricular related returns, and to reflect and investigate this we allow qualitative research to take part of this review.

2.2 Review results

To present the currect state of the art we will go through each of the questions highlighted in chapter 2.1, before presenting our conclusion based on the knowledge we gained throughout the literature study.

How well did the different mathematics concepts get taught with robotics?

The literature review showed that the majority of math concepts present in the elementary and middle school can be taught with the help of robotics. We have

also found several other papers, not included in this study, that teach decimals, geometry, functions, graphs, coordinates, polygons, perimeter and area, symmetry, probability and algebra. This broad applicability of robotics within math have given room for some of the bigger studies presented, as the studies presented by Hussain et al. and the study presented by Lindh and Holgersson. These studies were conducted over the course of a full school year [20; 25]. The results from these long term studies are very important as they may, to a better extent, measure the long term effects of educational robotics. They do however form the minority in terms of study duration, as the majority of the studies in the review had duration comparable to a week. Thus it may be harder to measure anything other than changes in content knowledge alone.

A summery of the different math concepts investigated can be seen in table 2.1.

Article	Math concepts
Barker and Ansorge [2007]	Decimals and geometry
Nugent et al. [2008]	Geospatial and GPS concepts
Nugent et al. [2009]	Geospatial and GPS concepts
Williams et al. [2007]	Physics
Mitnik et al. [2008]	Distance, angles, kinematics, and graph construction
Mitnik et al. [2009]	Graph construction and interpretation skills.
Norton [2004]	Ratio concepts.
Silk [2011]	Proportional reasoning.

Table 2.1: Articles and concepts

Most of the papers presented, and otherwise seen, throughout this literature review provided positive evidence that educational robotics can teach children about math. Out of the twelve selected papers presented in this review we found just two papers that did not provide any significant evidence of positive returns from using robotics (Silk [2011], study 1 and 3). Study 1 was an observational study in a larger study, so it is hard to conclude why the results were not significant, as the researcher did not have full control of the experiment himself. Study 3 observed a robot competition and even before the study it was speculated that the knowledge gain might not be generalizable to regular math concepts. In Silk's forth study he did however find significant evidence of increased math content knowledge.

Even though we focused this literature study on robotics in math education, three of the studies included did not focus on math specifically [4; 33; 34]. They tested broader for SET(science, engineering and technology) or STEM(science, technol-

ogy, engineering and math) learning gains. Therefore math subjects is only mentioned a sub part in their studies.

Barker and Ansorge [2007] reports on a pilot study that examined the use of a science and technology curriculum based on robotics to increase the achievement scores of youth. It was measured by a self created test to measure SET knowledge. They conclude that the activities helped teach math. We identified some concerns related to their approach. The most concerning was related to their analysis and the questions on their test. Analyzing their test we found only two questions directly related to mathematics(”‘what is a ratio?”’, and ”‘what does <mean?”’). The rest of the questions were related to broader topics or closely related to the software used in their study. Because of this we find it hard to support their conclusion that the study helped teach mathematics. A secondary conclusion in their paper stated that the hands-on experiment helped youth transform abstact science, engineering and technology concepts into concrete real-world understanding. It is our opinion that this conclusion is to a much greater extent supported by their evidence, but we reserve the right to be concerned regarding how generalizable these results are since a large part of their test were closely related to their specific software.[4]

Nugent et al. [2008] used instructional activities created by 4-H to increase achievement scores and interest in STEM. 4-H is an organization that aims to develop citizenship, leadership, responsibility and life skills of youth through experiential learning programs. The test they used was based on the test created in [4], which we just discussed. They improved on this test by adding five questions. They claim that there were 9 mathematics questions, which would not be possible based on our previous discussion. They have not discussed the test or included it either, which makes their claim even harder to verify. In addition to this concern they have not included any description of the experimental design or activities. Their conclusion was that youth had significant increases in scores in mathematics(including fractions and ratios), programming concepts, and engineering concepts, while there was no increase in their skills related to the GPS. Though their analysis of the results may be consistent with their conclusions, it is not possible to investigate the underlying factors positive, or otherwise, which may have influenced their results given the information presented in the paper. Because of this it is hard to draw any good conclusions regarding this study.[33]

Nugent et al. [2009] used the 4-H instructional activities, similar to Nugent et al. [2008]. The activities include building and programming robots using lego mindstorms. The test they used was a paper-and-pencil, 37-item, multiple choice assessment, covering topic in computer programming, mathematics(including fractions

and ratios), geospatial concepts, engineering and robotics. A concern regarding this study is that they did not analyze the results of the different concepts separately, and failed to include a good description of the test itself. They reported that the robotics group dramatically increased their scores. But we cannot draw any conclusion regarding mathematics learning gains from these results since we don't know anything about the test specifics.[34]

Overall, the studies which investigated STEM learning gains in general was not very useful for our discussion about leaning gains in mathematics. Although the majority of papers concluded with positive results relating to learning mathmatics, we found a severe lack in explanation of experimental design as well as test design.

Williams et al. [2007] investigated the impact of robotics when teaching physics. Since understanding mathematics and understanding physics is closely related we included this study.

Williams et al. [2007] evaluated the impact of a robotics summer camp on students' physics content knowledge and scientific inquiry skills. To collect data they used mixed methods, because they wanted to not only find out if it was effective, but explore various factors that might have contributed to the impact of the program. That is useful for us and other researchers when designing new experiments. The test consisted of twelve multiple-choice items developed by the team to assess students' understanding of Newton's Laws of Motion. There was a statistically significant difference on the physics content knowledge measure. No statistically significant difference was found when comparing scientific inquiry. The good results in physics might have been influenced by lectures about physics that were held at the summer camp, and not only the robotics activity. However these results are promising in general, the main problem for us is that the context is a summer camp.[48]

At this point we've talked about all the papers where mathematics were a considerable part of the study, but not the sole focus of the study. We now turn our attention to studies directly focused on mathematics.

Mitnik et al. [2009] aimed to develop graph construction and graph interpretations skills. They used a robot for a experimental group and a simulator which acted as the control group. They used the *Test of Understanding Graphs in Kinematics*. This test consists of 21 multiple choice questions. The activities was shown to foster learning in both groups. The results showed that the average increase of the experimental group were almost double that of the control group. This is a very promising study since one of our main concerns when starting this literature

study was if simulators were a just as good, but a cheaper alternative. This study shows that robots are more effective than simulators, at least in their setting.[30]

Norton [2004] investigate students' learning of ratio concepts while building robots with cogs and pulleys. The goal is to explore the learning of ratio when students design and make artifacts in which ratio thinking is embedded. In the experiment students got to work with concepts such as velocity, revolutions, linear measurement, circumference, quantification of gear ratios and quantification of pulley mechanisms. In some lessons, ten to fiftheen minutes was spent in whole group discussions of the underlying theories. They only tested for ratio learning gains. They created their own test with eleven explanatory questions. The students had to write down how they would solve certain tasks. Points were given for completeness. Students improved in their ability to explain science and mathematics concepts on pencil and paper tests. There was not much improvement in their ability to use and explain ratio and proportional reasoning in their construction of artifacts, but this statement was only supported by observation. They concluded that Lego is a powerful manipulative for the learning of mathematical ideas. We believe this study shows great promise for the use of robotics in mathematics teaching.[32]

Lindh and Holgersson [2007] provide interesting data regarding the effectiveness of educational robotics. They reported on the impact of one year of lego training on school performance. The experimental group used LEGO mindstorms activities adjusted to the ordinary school activities, while the control group were students from different schools which took no part in any of the activities. The quantitative test they used was similar to national test in mathematics. They have not mentioned anything else about the test. The question they sought to answer was: *will the robotics activities improve students' performance in mathematics and logical problems?*. With the null hypothesis of, lego do not have a positive or negative effect on students ability to solve mathematical or logical problems, which did not get rejected. So in the end there was no statistical evidence that the average pupil gains from lego training, however when the ANOVA test was performed on subgroups of students, based on previous performance, the null hypothesis was rejected in some cases, namely for the medium good pupils. Indicating that lego training may be useful for some groups of students, but not all. This study is the most relevant study for us as it has the most participants, the longest duration, and was performed in many different schools in Sweden. This means that many of the external factors will be the same as in Norway, seen as the culture if fairly similar. The study did not have good results in learning gains, except for subgroups of students. However, it is not always practical to divide a school class into subgroups and teach in different ways. The fact that it was performed in

schools is very important to us. We are attempting to find out how robotics can be used in schools and none of the previous studies have done so. The other studies have often been set up with small groups of students and many facilitators. This setting is more comparable to a tutor situation, not a school environment. They don't have to face many of the problems that exist in schools, such as conflicts, lack of teachers, many students in one room, chaos and equipment management. The negative results may not be in conflict with the other studies. Almost all the other studies was performed as part of an extracurricular activity. And when they conclude that students learned math compared to the control group this does not imply that they will learn math better through the robotics activities than through regular school activities. [25]

Silk [2011] is a doctoral dissertation. 3 of the 6 experiments presented in the dissertation is included in this discussion. Part 2 was not included because there were issues with the reliability and validity of the test. Part 5 and 6 were not included because they were not relevant for this literature study.

In Part 1 there was a very clear goal to teach students math with robots serving as the context. The activities involved learning how the size of the wheels is related to the distance the robot moves. Four lessons were observed that were part of a larger study. These were chosen because they were the ones highlighting math the most. The test measured the students' ability to solve quantitative problems in robotics and non-robotics contexts that aligned with the instructional goals in the unit, which were proportional reasoning and concepts of measurement. The items were selected from the National Assessment of Educational Progress. There were in total 32 items, 8 items were selected for each concept and 8 items were created for the robotics context by the team for each concept. The test was split in half and one half was given as the pretest and the second as the post test. The test were included and they focus very specifically on math topics, not like the general STEM tasks in the previous studies. There were mixed results for helping students productively engage with and learn about how to control robot movements using math. The students made significant improvement in their problem solving, but the improvement was not in proportional reasoning and was not in the robot context problems either.

Part 3 observed a robot competition. The goal was to measure what students actually learn when they participate in such competitions. Only a few teams used math explicitly in their design solutions, the use of math was found to have a highly variable relationship with design success. The test consisted of 10 multiple-choice and short-answer questions. All questions focused on robot motion. The items were adapted from published sources of problems that assess aspects of

proportional reasoning, and then modified to focus on the robot motion problems. We looked over the questions and they are very similar to standard math tests. The results of the test showed the presence of a significant increase in students' overall problem solving from the pretest to the posttest. Participating in the competition did have some positive impact on students' overall problem solving.

Part 4 implemented an activity where the students should make two robots with different wheels move the same distance. This activity was called RSD(Robot Synchronized Dancing). There were two groups. The experimental group did the RSD activities and the control group prepared for a robot competition. The Problem Solving Assessment was identical to the one in Silk [2011]. Participation in the RSD activities did have a positive effect on students' overall problem solving, but the effect was not reliably different from participating in the competition environment, so it is not clear whether it was the activities or just working with the robots in general that promoted the increase in problem solving. The implementations of the RSD unit helped students improve their understanding of the way the robots work.

Part 1-4 built and improved on the previous studies ending up with part 4 where there were positive effects on students' overall problem solving. One of the conclusions is that just because math is present in an activity, it does not mean that students will learn math. The main reason behind their success in study 4 was the focus on immediately aligning students' activity with the type of thinking that was desired.

We have had some problems in describing why different approaches worked well or not. This is mainly because of the lacking documentations in these studies. Only a few talk about factors that might influence the results. Only a brief overview is given about group size and environment settings and it is hard to draw any conclusions except the ones they chose to put in the report, and even then it is hard to validate these results. Overall the results are quite weak and it is clear that further studies are needed.

What advantages or disadvantages except learning gains are there?

With regards to secondary skills there is a lot greater gap between the results, universally mentioned is however teamwork including social interactions and communication [29; 30; 34; 38]. When working with robots, students tend to get a greater sense of community and start helping each other instead of competing. Students are also eager to help other groups and want to explain how they got

their solution. In Lindh and Holgersson [2007] pupils' learning of cooperating in working groups was of interest. In interviews pupils often said the activities had enhanced the feeling of a community. This is the most relevant study for us and it is promising that collaboration was increased also here, in a school context. Mitnik et al. [2009] had a robotics experimental group and a simulator group acting as the control group. Collaboration was measured by an attitudinal questionnaire and by observation with a focus on four factors shown to foster collaboration is several other studies. The robotics group scored higher on the collaboration questionnaire and a greater amount of collaborative interactions were observed in the experimental group students compared to the simulation group students. In Nugent et al. [2008] results showed positive effects on students' perceptions of the value of teamwork as measured by questions probing their attitudes in working with others and listening to others while problem solving. Nugent et al. [2009] reported non-significant results from a teamwork questionnaire. This study has not described their experimental setup, so it is hard to draw any conclusions on what might have caused these negative results. However they mention that activities happened in robotics summer camps. They speculate that peer relationships within the camp settings can influence the experience of a particular youth.

Students motivation is also affected either positive or negative. In Mitnik et al. [2009] both groups were motivated, but the simulator was based on newness, while motivation in the experimental group was seen to be supported by the students need and possibility to immerse into the activity. Motivation in the experimental group lasted throughout the experiment, while motivation in the control group decreased after two sessions. Nugent et al. [2008] identified an increase in interest and motivation, where pupils working with robots expressed their wish to continue working with robots. Whereas the control group would do the opposite [33].

Many papers write about how interest in STEM can be improved through the use of educational robotics. In addition several robotics competitions like the FIRST robot league are held specifically to improve attitudes toward STEM and encourage students to pursue careers in STEM. Silk [2011] states that competitions effectively blend a focus on excitement and fun with an opportunity to engage in an extended and challenging activity from which to learn valuable STEM-related skills. When compared to a matched comparison group from an existing national data set, alumni participants of the high-school level FIRST robotics competition were more likely to attend college, to major in a STEM-related field, and to expect to pursue a STEM-related career. The results from Nugent et al. [2008] suggest that students do not directly perceive connections between robotics and STEM concepts. When working they become excited about robotics, but not recognize that STEM learning is being integrated into the activities. Middle school students,

in particular, seem to need specific guidance on how they relate. Changing their STEM attitude may be hard if they do not perceive the activity as a STEM activity. Their results show that the students interest in technology increased but it decreased in science.

Control in a real-world context is a factor mentioned in several different ways throughout the literature. Silk [2011] argues that the robots are reliable, manipulable, and inspectable. At the same time they are not simply a made-up or imagined entity, but instead they do real things out in the physical world, and so a person's own intuitive knowledge of the physical world with apply to the robots as well. Papert believed that children could identify with the robots because they are concrete, physical manifestations of the computer and the computer's programs [39]. In effect, the embodied nature of a robot provides a way to depart from the traditional abstract blackboard based teaching scheme to a teaching model in which the student can learn by experimenting on the subjects in the more natural, 3D, real world [30]. Compared with a simulated environment, a physical robot has all the same benefits as the simulator. In addition, an environment based on a physical element may help students to develop stronger affective bonds with it, also developing more situational interest. Thus robotics technologies can include and extend the benefits introduced by simulations. The extended functionality introduced by robots include 3-dimentionality, mobility, and the presence of a real, explorable and measurable setting.

Students get a better sense of the value of mathematics. It is not uncommon for students to feel that what they learn in mathematics class is useless in the real world. In several of the studies they tested specifically for this [44; 34]. The results were positive. The results might just reflect that students now feel math is useful when working with robotics, but at least it got them thinking about one topic where math is necessary in order to succeed.

When testing for other secondary skills the results are to a large extent inconclusive or negative. Hussain et al. and Lindh and Holgersson identifies an insignificant increase in problem-solving, Hussain et al. also identifies an insignificant positive attitude change towards LEGO [20; 25]. For scientific inquiry Williams et al. found no significant difference when comparing the pretest and post test [48]. Though they argue that scientific inquiry may be a process to be learned through long exposure and that their study was to short. Silk [2011] reported a decrease in robotics interest from pretest to posttest. Nugent et al. [2009] had non-significant results for perceived values of GPS/GIS and problem approach. On an attitudinal assessment there was a significant increase in science and robotics task value and self-efficacy in robotics and GPS. Nugent et al. [2008] Positively impacted students' use of a plan to guide their problem solving process.

Were any suggested improvements to the experiments identified?

Almost all papers used a control group that did not do anything. So their scores were, as expected, very stable between tests. This is a concern in our opinion as it does not provide any evidence regarding the effectiveness of educational robotics compared to ordinary tutoring. Generally all the studies' activities were extra curricular activities. The exceptions are Lindh and Holgersson [2007] which tested robotics in schools 2 hours a week instead of traditional activities. In effect this tested robotics against regular school activities. Mitnik et al. [2009] compared the robotics to a simulator. And in Silk [2011] the robotics activity were matched up against a robotics competition, to see if the learning gains came from their test setup and not just generally working with robots.

The learning model used was largely experimental based on constructionism. Barker and Ansorge [2007] used an experimental learning model based on other research where they have 5 steps. Experience, share, process, generalize and apply. They provided the children an opportunity to learn before being told or shown how and then share what they did. Then they considered what was important and tried to generalize the experience. Then they applied this to new situations.

Most of the experiments had students working in small groups. Nugent et al. [2008] does not mention how many students there were per group. The rest state how many students there were per group but there is not much discussion around the choice of group size. In general there was 2-3 students per group[30; 32; 25; 44; 34; 48]. Barker and Ansorge [2007] had as many as 4-5 students per group, but they have not discussed the impact of this decision and there was one teacher per group so they could easily keep control and make sure everyone participated. Mitnik et al. [2009] used groups of three because a previous study found that groups of 4 or more provide collaborative drawbacks. In the robot competition observed by Silk [44] the teams were quite big. The average group size was 7 but the minimum allowed was 1 and maximum allowed was 10. No analysis was done to check if the team size had an impact on learning or collaboration, but that would have been interesting. Lindh and Holgersson [2007], which is the most interesting paper for us, used 3-4 pupils per group. Through the study they identified several factors that could be improved and one was that groups should not be too big, maximum 2-3 pupils per group. If we exclude the irrelevant studies that had a huge amounts of teachers or were in a special context, we find that 2-3 pupils per group is ideal.

Another point that we have mentioned earlier is the fact that many of the studies had one facilitator per group. This is completely unrealistic in a school setting where there might be as many as 30 pupils per teacher. Therefore many of the

activities described in these studies have reduced applicability in a school context. The context in these studies looks more like a tutoring session, and this is not addressed in any of the studies that does this. Lindh and Holgersson [2007] conducted their research in a real school context. They found that having 2 teachers in the room during the robotics activities helped a lot, because then one teacher could help while the other maintained classroom control.

There were several factors to consider when creating tasks. Lindh and Holgersson [2007] found that the tasks must be concrete, relevant and realistic to solve. And that it is important that the pupils can relate the materials to their ordinary school work and their different subjects. Williams et al. [2007] reported that their tasks might have been too easy for the students to require or appreciate the process of scientific inquiry. Almost none of the studies have included their tasks so it is hard to decide how we should design our own.

The purpose of the teacher is mostly one where teachers act as facilitators. They solve doubts about the software and problems when students ask. At least one teachers is therefore needed that can freely be able to go assist one group. It is hard to maintain classroom control at the same time, this is an issue found in a school context and therefore not addressed in most of the studies included here, since most were small experiment with often multiple people involved. Lindh and Holgersson [2007] Did report from school contexts and found that the role of a teacher as a mediator of knowledge and skills was crucial for coping with problems related to this kind of technology. Teachers also acted as resolver of conflicts. Even though groups collaborated well sometimes conflicts occurred.

The time period varies greatly from 1 week to 1 year. With 1 week periods being dominant. However in terms of hours most of the studies were about 40 hours. Even the study that lasted over 1 year only resulted in about 96 hours in total. So the differences in studies is really about how intensive the exposure was.

The sample size is very different from study to study. The by far biggest is Lindh and Holgersson [2007]'s study with 322 students at different schools in Sweden. Nugent et al. [2009]'s had 147 students in total, they were spread on 6 different robotics summer camps. The rest had on average 30 students each.

What lacks of research are there in the literature?

The areas that lack research is partially found from studies' future work chapters, but also from our own analysis. The biggest hole in the literature at this moment is the lack of big sample sizes, as mentioned by Mitnik et al. [2009] and Lindh and

Holgersson [2007]. Lindh and Holgersson [2007] proposes a test run over 2-3 years in school with 3500+ students.

Barker and Ansorge [2007] mentioned the need for research in different populations. Lindh and Holgersson [2007] was performed in Sweden and is the only similar population to Norwegian schools.

Long term studies aimed at discovering whether the activities help foster SET as a career and school subject choices are needed. Research is needed to examine whether the program helps foster positive attitudes towards SET in school and as a career Barker and Ansorge [2007]. More research is needed on whether the program fosters SET careers [4]. It takes commitment and a good followup system to track down students up to 10 years after a study to find out whether they chose to become engineers, or have careers within STEM.

We have found a problem with the control groups' alternative activities. In most cases the control group does nothing, while the experimental group does extracurricular activities. We want to integrate with the school system and look at how a robot can be used in an educational setting. As such we would need information on how a robot compares against regular school activities. Nugent et al. [2009] have addressed the fact that it is not in a school context: "The use of intensive summer camps offer a chance for youth to become more deeply involved in STEM activities than what might be possible in more formal educational setting". When considering if the robot is necessary at all, another question comes to mind: Why should we use robots instead of a simulator? A simulator has many advantages over robotics, such as costs and changeability.

Conclusion

Robotics foster teamwork, motivation, interest in STEM and give students an increased understanding of the usefulness of mathematics. Robots are real, manipulable, inspectable objects, which presents students with an opportunity to develop mental models closely aligned with their preexisting knowledge of the world. In most of the studies seen, the robot only works as a feedback provider to the children. Thus do not utilize all the benefits of the robot and many of the tasks could be fulfilled by the use of a simulator, flash game or mobile application. Which in all cases would be cheaper and easier to acquire for the schools given the robotics landscape seen today. Another benefit is the connection between abstract concepts and physical representation given by robots, this could be especially beneficial within STEM topics as they are usually very abstract and robots may therefore help by giving a concrete understanding of these concepts.

Robotics does not seem to develop students' problem solving approaches. This is tested in several of our selected studies under different names. Problem solving, scientific inquiry, problem approach, students' use of a plan to guide their problem solving process etc. None of the studies did however report a statistical significant increase in these skills.

A big concern for us is the lack of relevant control groups. Even though some studies are placed in schools, the activities are extra curricular, so it does not in any way address if it is better than regular school or other alternative approaches. In other studies there are too many facilitators, making the study irrelevant for school contexts. In addition the studies that did perform it in school contexts found it helpful with more than 1 teacher in the classroom, which might indicate that robotics are too chaotic for regular classrooms.

Most of the research done provided promising results. None have discovered that it worsens learning, but there are examples of it not making any difference for the average pupil compared to traditional methods [25]. It is hard to pinpoint the factors that generates positive results, as the art of teaching and learning is extremely complex and different for each individual. The studies that focused on STEM in general contributed little to our discussion on maths in robotics. Even though they showed good results they did not present their work on mathematics separately from the other STEM concepts. The one study on physics is promising; however it is not math specific. In the mathematics specific studies, robotics was shown to be effective for graph construction and graph interpretations skills, ratios and proportional reasoning.

Even though the literature review showed great promise for educational robotics, few of the papers have collaborated with schools and pedagogs in an attempt to integrate or check how their ideas will integrate into the school curriculum, and this is a concern to us. Schools often have limitations regarding both time and resources and having multiple teachers per classroom, as seen in many studies is not an option. One major denominator in the current research is the lack of details in publications. While general trends and statistics are included in the majority of publications, we still found that detailed discussions related to validity and reliability often were excluded. This made our process of determining the current state of educational robotics extremely difficult.

The literature review indentified several areas where research is needed, though many of them were not feasible to investigate during this project. The main areas being: research utilizing low cost robots, research conducted in a close-to-real school settings, and longtitudal research with large populations. It is our opinion

that research within these areas would greatly benefit educational robotics as a whole.

2.3   Constructionism

As a response to seeing how the majority of research within educational robotics based itself on Papert's theory of constructionism. We've included an introduction to the theory here to aid the reader.

The theory of constructionism is heavily inspired by, but should not be confused with, the constructivist theory of knowledge created by Jean Piaget [7]. Piaget's theory of knowledge is based on the notion that knowledge is constructed in the mind of the person learning(learner) [6]. It argues that knowledge cannot be directly tranferred from a teaching entity to the learner, but that the learner needs to create their own mental models in order to grasp the concepts. To describe this process Piaget articulated two mechanisms describing how knowledge can be internalized by the learner, *accommodation* and *assimilation*. Where the first is refering to the process of reframing one's mental model to fit new experiences, e.g. the learner observers something that does not fit within his or hers mental model, and thus have to change the model. The latter refers to the introduction of new knowledge into the existing mental framework without the need to change the framework itself. Both processes happen through a continous iterative process of redefining and extending the framework of a learner through based on interactions between their environment and their mental models.

In his grant proposal to the National Science Foundation, Seymour Papert describes the difference as such: "From constructivist theories of psychology we take a view of learning as a reconstruction rather than as a transmission of knowledge. Then we extend the idea of manipulative materials to the idea that learning is most effective when part of an activity the learner experiences as constructing is a meaningful product" [40]. The main difference between constructivism and constructionism being their main focus. Constructivism focuses on how children learn, what they are interested in, and what they are able to achieve during different stages of development. Whereas Paperts constructionism focuses on how learning can be accomplished through engaging the learner, and how this facilitates the construction of knowledge [1].

One reason why educational robotics is particulary applicable for the theory of constructionism is because the use of robotics is inherently a construction-based process, e.g. contructing the robot, or using a robot to construct other things. This reason, alongside with Papert's extensive research within education robotics, do,

in our opinion, explain why the theory of constructionism is so popular, and why it is also the basis for this project. The choice of using the theory of constructionism as a basis for the project does however have two major implications in terms of experimental design and intellectual model building.

Intellectual model building

Intellectual models means the structures we have in our mind that we use when we perform tasks such as mathematics. Seymour Papert wrote that through his childhood he used gears as a mental model, even though he wasn't aware of this at that time. Whenever attending school and learning about a new concept like multiplication, he would find a way to relate this concept to the gears.

The mental models described by Papert are usually found individually, and is often related to something the learner finds interesting. One of the goals of constructionism is to help students by providing a culture that is rich in materials and artifacts that the students can build mental models of to make new concepts concrete and simple.

In Papert's book, *Mindstorms: Children, Computers, and powerful ideas*[39], he introduces the notion of a turtle as such an artifact. He introduced and tested the turtle as both a robot and a simulator. In regards to mental models he argued that there would be no difference between the robot and the simulator. The turtle was described as an "object to think with". Although Papert highlight the usage of something similar to a turtle, it need to be mentioned that the theory of constructionism is a generalized theory applicable to any object or artifact. The main example concept used throughout the book is geometry. The turtle can help make geometry simple to learn. Unlike a point which is a very abstract thing, a turtle has (like humans) a position and a direction. This is something the students can relate to and use as a mental model. This way the students may at a later point utilize their mental model of a turtle to solve other problems.

As we experiment with new concepts, our understanding and mental models are changed by our surroundings. Mental models are always adapting through the mechanisms Piaget called assimilation and accommodation. Though its said that one can't teach an old dog new tricks, we argue that our mental models are constantly changing and evolving to fit with new experiences. Mental models do have parallells to how mathematics and physics are taught in school. In introductory courses there are a lot of simplifications in order to make the curricum easy to understand, similary can a mental model contain simplified explanations of, in reality, complex phenomena. When graduating, or getting to a higher stage in your

education, you learn that the thing you learned does not apply anymore, and you have to redefine your mental model to fit the new information. One violation to this parallell is how the knowledge is gained. In traditional blackboard learning the teacher is attempting to transfer their mental models to the student, which makes adapting to new knowledge much harder than if the students were to create their own models from the start.

Experimental design

In the majority of cases where technology, and especially computers, are used in an educational setting, we use it by first showing a concept then asking question about it at a later stage. This is similar to how the regular blackboard and task solving approach of teaching works. In constructionism however, we replace the "'teaching *at*"' with "'assisting them to understand"' while the students themselves are investigating and experimenting using the technology provided. This is done to make students explore and reflect on how they think. They have to think about what instructions to give the robot so that it will behave in the way they want. In addition to this thinking process, when programming the computer the child acquires a sense of mastery over a modern powerful technology. When programming the children learn mathematics as a living language by learning to "talk" to a computer. Papert compares this to learning french while living in France, compared to learning french in french class. In this way we learn how to learn and love math. We continually use math in a way that feels useful, and we avoid math being experienced as a abstract concept that students feel they will never have any use for.

A disadvantage of implementing constructionism in todays educational system is however that it requires a change in the classroom dynamic and curriculum, which may not be welcomed by open arms by everyone. The rather experimental design is often described as "learning without a curriculum". Papert agreed to this by stating; "of course the turtle can help in the teaching of traditional curriculum, but I have thought of it as a vehicle for piagetian learning, which to me is learning without curriculum". It is however clear that Papert sees this a postive thing, rather than a thing that be used to criticize the theory.

The constructionism theory relies on the fact that students are to be active creators of their own knowledge. In order to facilitate this, the teachers need to ask question, help them explore, and than assess their skill level. Teaching without curriculum in this context means supporting children as they build their own intellectual structures with materials and artifacts drawn from the surroundings. This way learning becomes like a guided story of discovery, and not an attempt to transfer mental models between different people. The tasks are designed for students

to hopefully understand how they got the answers by themselves. Teachers avoid most direct instruction, but rather lead the students on a journey through questions, activities, discussions and try to make them verbalize with their discoveries in the correct way, and by doing this create their own mental models.

2.4   Quantitative and Qualitative Methods

Most research methods can be placed in one of two categories: quantitative or qualitative methods. Qualitative research is often used to get an insight into how students perceive something, and their thoughts and feelings about the subject. It is also used to understand their behavior. Quantitative methods have a focus on finding the general correlations, and discover tendencies that give interesting insights for future research or can be supported through other methods. This is done by pointing out special factors that are the cause of a given tendency [46]. If you want a broad understanding of a phenomenon you should use quantitative methods. Testing a hypothesis statistically is an example of quantitative research methods. In this chapter we discuss some methods applicable for our project and at the end we reason about our choice of methods.

Test

Tests are commonly used in schools to find out what the students know and what they have learned. All the papers presented in the pre study applied a pretest and a posttest to support their findings, as this was one of the inclusion criteria.

One advantage of using a test is that students are used to being tested in this manner. It is relatively efficient to get an overview of what the students know, or don't know and what they have learned. This is because a test can be distributed to several students at the same time. One disadvantage is however the difficulty of creating a pedagogical sound test which measures the wanted underlying constructs, and this is especially difficult when venturing into uncharted territories.

Using tests alone has limitations to what can be measured. It is hard to test for deeper kinds of knowledge, as test questions often have only one right answer. A test can therefore only test whether students know the answer or not, even though they might have managed the task with a little assistance. This assistance could both be task specific and general, for example language misinterpretations. With this in mind the question about tests are: does this test really reflect what they understand and what is outside their understanding [3]. A way to work around this would be to include several items in the test that measures the same underlying

knowledge. Interview or an oral test can also be used in this case to get better knowledge of their understanding.

A test should be built on a solid knowledge within the given area to make sure it is accurately reflects the specific concept the researcher is attempting to measure. In other words the test should be valid and reliable.

Observation

Observation is used especially within social anthropology. This type of observation is a set of methods where the researcher participates, openly or hidden, in people's daily lives in a given period of time[46]. In the same way observation can be used in a certain context. The goal of observation is often to see with your own eyes what happens to a person or group in a staged situation or under certain circumstances. According to Thagaard [2003] observation is particularly suited to give information about people's behavior and how they interact with each other, thus it is a good option to measure cooperation.

Observation can be used as the only method in an experiment, but can be, and is often, used as a supplement to other methods. We distinguish between open and hidden observation and spectator, partially participating or completely participating observation. In hidden observation the researcher is present and observing, without the observee being informed of this. The advantage of this type of observation is that the participants don't get influenced and change their behavior as a result of being observed. The downside may be ethical issues and considerations. The difference between non-participating and participating observation is well illustrated by the difference in observing a handball match as a player or spectator.

Interview

Interview is used in situations where you want to study meaning, attitude and experience[46]. According to Thagaard [2003] interview is particularly well used to give information about a person's experience, point of view and self understanding. In an interview situation the researcher often tries to create a relaxed atmosphere, so that the interview resembles a normal conversation revolving around some chosen topics. During the conversation the researcher leads the conversation by asking open or closed questions. The goal is often to get the interview object to talk as freely as possible. In this way it becomes clear what the interview subject means. Their attitudes and experiences comes forward.

An advantage of using interview is the possibility to ask follow up questions and to come close to the interview object. A disadvantage is that it takes a lot of time and the subject might feel less anonymous.

Students might have difficulties with the way questions are written, there can for example be language barriers, and with a short explanation they might be able to answer questions in an oral test that they would have answered incorrectly in a written test. In an oral test there are also opportunities to argue why they approached the way they did and how they solved the task in greater detail. We get to understand their thought process and understand what they have actually misunderstood and understood and not just misinterpreted.

Questionnaire

Questionnaires are in general considered a quantitative method. Ringdal [2001] writes that a questionnaire is a systematic method to collect data from a sample of people. This is done to give a statistical description of the population the sample is collected from. The questionnaire is standardized, which means that everyone get the same questions asked in the same way. A questionnaire is often used to find overall connections and tendencies by pinpointing special factors that are the cause of a given tendency [46].

An advantage of using a questionnaire is that you can perform it on many students at the same time. You also obtain data that you can use for statistical testing and draw conclusions about correlations on.

One disadvantage is that there is no possibility for follow up questions. It might be possible to find connections and correlations between questions, but it is hard to understand why without any other methods as supplements.

Experiment

Experiment is the classical scientific design in science, but it is also used in medicine and in social science under different names. In our case we call it experiment, in other cases it is called trial or intervention [41]. In an experiment the researcher controls the difference between two treatments, X and Y, by deciding when the experimental group should be exposed to X. Control of other factors are secured through randomization. This gives a good foundation for explaining cause and effect. In social science, experiments cannot be performed in such a controlled way. It is impossible to control all the factors that can influence a person or group. But some sorts of experiments can still give valuable knowledge which is difficult to attain through other methods.

The main advantage of an experiment is the possibility to test out something new on real people and see how it works. A disadvantage is the disturbance of the trial population. We intervene with their daily activities. In our case one possible solution could be that the teacher conducts the experiment. An additional disadvantage is the introduction of new people and in some cases exciting tasks, both of which can make the participants alter their behaviour [15; 28].

2.5 Reliability and validity

Reliability and validity are important topics within all types of research. These topics provide a basis for the discussion for whether we are measuring what we think we are measuring and whether these measures can be viewed as consistent and valid. Both within reliability and validity you can find several different concepts addressing different types of reliability and validity. One of example of this can be seen with validity where you often can see papers refer to internal consistency, interal validity or external validity etc.

This chapter will first present a short introduction to what reliability and validity are and some of the most common threats to these concepts. Though some factors relating to validity and reliability have already been mentioned in other chapters, we will include a concluding section about here. This is because these concepts can in various ways been seen as the linchpin of any experiment.

Reliability

When talking about reliability we usually talk about the four key concepts, *equivalence reliability*, *stability reliability*, *internal consistency* and *interrater reliability* [18; 26]. Each estimating different constructs regarding how reliable a study is. It is important to mention that one can not measure reliability, but several strategies exist in order to attempt to estimate how reliable an experiment is.

Equivalence reliability is the extent to which two items measure identical concepts at an identical level of difficulty. This is often measured as the correlation coefficient, measuring the strength of the correlation between the dependent variable and the independent variabel.

Stability reliability is a measure of the instruments stability, e.g how accurate an instrument is. To give an answer to this, one would usually repeat a given test to see if it gives the same results. An analysis with Cronbach's alpha between the first test and the second test can give an indicator to the stability of the test.

Internal consistency is a measure of how well an instructument measures the same underlying concepts. A common way of measuring this is to use Cronbach's alpha between the items on a questionaire.

Interrater reliability is the extent to which raters agree, and is used as a measure of the rating system. Normally this is established by using a Cohen's kappa if there is only two raters, and Fleiss' kappa [17; 43].

Validity

Validity assesses the degree of which an experiment and design measures the concept that the researchers intended to measure. When talking about validity, we usually talk about *internal validity* and *external validity*, each with its own subcategories like *face validity*, *construct validity* and *content validity* [18; 27]. Internal validity refers to how the study was design, organized and conducted, and is important in order to say that an experiment accurately reflects the underlying concepts and constructs. External validity, on the other hand, is important because the results of an experiment doesn't mean much if it is only applicable for the population that participated(e.g. we want the results to be generalised and transferrable). We will not provide you with an exhaustive list of potensial threats to the validity of an experiment, but will here give a short overview over the most common ones.

Internal validity

Some of the threats that may threaten the internal validity of an experiment include *testing effects*, *selection bias*, *experimental mortality* and *diffusion between groups*.

The testing effects is an effect that can be present in any design which include multiple stages. One example of this is the pretest-posttest design, where the results on the posttest can be influenced by the pretest in itself. This can be because that the participants learned from completing the pretest and hence performed better on the posttest.

Selection bias can be another threat to the internal validity of an experiment and refers to differences between groups that may influence the results of the experiment. Though it is most prevalent in quasi-experimental design where the groups are not randomly assigned, it may occur in experimental design as well.

Experimental mortality is a threat that refers to the fact that participants dropping out of the experiment for various reasons. This may become a real potensial

threat in quasi-experimental design if you unintentionally created a group where the people were more likely to drop out of the experiment.

Diffusion between group refers to an effect where one of the group was effected by the other group. In most cases we are interested in the control group, and wether or not they have been affected by the experimental group. The "'contamination"' of the control group may happen in several different ways depending on how the experiment is done. This effect can be seen as a umbrella-term for other threats involving cross-contamination between the groups (e.g. rivalry or demoralization).

External validity

Some threats to external validity have already been covered in the section about internal validity, *selection bias* is one example of this. Some of the more unique threats to external validity include the *"'real-world"' versus "'experimental world"'* and *"'faulty constructs"'* threats, where the first refers to when participants are aware that they are part of a study and may therefore alter their behaviour because of this, or that the environment in which the study is conducted influences the participants. The latter can be a bit more subtle in how it manifests itself, but refers to how well construct have been narrowed down from concepts, and how these constructs are measured.

The "'real world"' versus "'experimental world"' effect is often discussed in terms of the *testing effect*, *experimental effect* and the *experimenter effect* as these effects also influence the internal validity of an experiment. Both the testing effect and the experimental effects have been mentioned previously in the section about internal validity.

The *experimenter effect* refers to the personal biases of the the reseacher influencing the participants and/or the experiment design in such a way that the results would not be valid outside of the experimental setting. This can be as simple as non-verbal cues while observing the participants, and thus giving them a sort of validation that what they were currently doing was correct.

CHAPTER 3

DESIGN AND DEVELOPMENT

This chapter will outline the design principles and guidelines that were indentified in order to create all the necessary tools to complete this project. First we will present the design principles in section 3.1 alongside with why we felt these principles were important to this project. Before moving on to the sections about the software and hardware created during the project, including how the design principles relate to key decision during the creation process (section 3.2 and 3.3).

3.1 Design principles

One of the many goals of this project was to clear the way for future work with the ChIRP robot in an educational setting. In order for us to achieve this goal we were determined to create a system which would not only work for our current setting, but that could be extended and/or modified to fulfill other purposes as well.

In addition to make the system easy to extend and modify we had to enforce a high degree of usability throughout the user interface, this was mainly due to the fact that were going to test the system with children. This got us into thinking about utilizing bluetooth as a means to program the robot as using the arduino development environment was tedious and errorprone in many ways. The notion of usability also prompted a investgation into using preexisting technologies as the Logo programming language and the block programming language (Scratch) defined by MIT's media lab.

As utilizing bluetooth became a viable options for programming the robot we determined that a mobile solution using already existing smartphones and/or tablets

would probably the easiest solution to use for all parties involved. This did however pose another requirement on the system as we did not want to create a vendor specific solution, e.g. forcing the use of android, iphone or windows phone on the users. This can also be seen as a continuation of one of the design goals when the ChIRP robot was created, as it was meants as a cheaper alternative to other existing robots, and forcing a vendor specific solution on schools could result in higher costs.

Based off this knowledge we created five broad design principles for when design choices were to be made. These principles can be seen in table 3.1.

#	Name	Description
R1	Extendability	The system should be easy to extend with new functionality.
R2	Modifiability	The system should be easy to modify. Settings/options should be included to allow for changes at runtime.
R3	Usability	The system should provide an easy to use user interface.
R4	Platform	The system should work on as many as possible platform.
R5	Applicability	The system should be applicable outside the scope of this project.

Table 3.1: Design principles

3.2   Software

This chapter will talk about the software part of the project, how the design principles defined in section 3.1 was applied in order to solve the major obstacles before developing the application, and how the design principles helped us during the development.

The first section will walk the reader through some of our keythoughts during the inital design phase of the application, and why we felt that our solutions fit within the bounderies of the design principles. Following this inital section we've dedicated a own section to the choice of base framework as this was one of the really major decisions during the development process and can in many cases make-or-break an application. The second to last section will go into details about the technical aspect of the application, before we in the last section present the application as it was used during the experiments.

Application of design principles

In order to answer the questions proposed in the thesis while adhering to the design principles outlined we decided to utilize ideas from previous researchers within the field of educational robotics. One of the eariest researchers looking into how robotics could be used in schools was Seymour Papert [39], and as a result of his research the Logo programming language was born. The Logo programming language have since been implemented in numerous programs, including programs as MIT's Scratch, Turtle Art, and StarLogo TNG [16]. Even a modified version of Scratch have created to allow for simple programming of a Arduino(S4A) [9]. Though there exist alternatives to Logo we found that these are often uniquely coupled to a specific robot, f.ex ROBOLAB and Lego Mindstorm robot, and would most likely be close to impossible to adapt for the ChIRP robot. This, coupled with the extensive research conducted using the Logo language and of the language itself [10; 13; 12; 11; 39], made us decide to base our implementation on the Logo language as well.

Some issues did however exist between the current implementations of Logo and our design principles. First of all is the lack of extendability and modifiability. All of them come as prepackaged solution, where modifications to the source is either not possible or extremely complex. This makes adapting the solutions to be usable in multiple settings very difficult. The second issues is their inherent complexity as they all implement the full Logo language specification. While this can be seen as a positive trait, we feel that the complexity of the Logo language in general is too much when using it as a mathematical tool, and would therefore like to have an option to hide some of the complexity. The S4A implementation was created with the arduino in mind, and could have provided a good foundation for our system without the need for major modifications. The issues were however that the inherent complexity in the logo language was always exposed to the user, and that such a solution would require a laptop to create the programs for the robot.

Given the issues identified with the current solutions, we decided to implement everything from the bottom up. As we did not need the full Logo language specification during this project, we envisioned a solution where each command was seen as a separate entity. The intent of this approach was to make the resulting solution extendable, modifiable, and open up for disabling parts of the language. This way we would have the possibility of just using the commands associated with turtle graphics (e.g. `forward`, `backward`, `turn-left`, `turn-right`), and hide away the programming related concepts from the user. As a natural part of basing out solution on the Logo specification we looked at the current solutions.

The Scratch solution[1] provided a block based programming interface for the specifications, something that would fit our needs perfectly in terms of usability and splitting up the language specification. Utilizing block-programming we were able to remove the need for writing correct syntax in order to make a program runnable, and thus increase the usability of the solution drastically.

We proposed to create the system as a two component system, the ChIRP robot and a programming device. As from the get-go we experienced trouble using the arduino integrated developer environment(AIDE), and determined that putting this system into the hands of non-technical personel would most likely deter people for using the system. As previous work with the robot had however shown that bluetooth was a viable option for communication between a mobile device and the robot([8]), we found this to be a viable solution. In addition a solution based on bluetooth communication would allow the system to be vastly more usable as we could remove the wired connections to a laptop, and defined our own interface in a language of our choice. It was therefore decided that the system would use tablets(android, iPad, ect) as its programming device. This did however mean that we had to do alot of work on the application instead of reusing existing code from other projects.

Summary

As the system was created to be used in an educational setting all the way down to lower parts of our educational system, we needed to make it easy to use for everyone, while still maintaining the complexity needed for children of an higher age and/or skillset. We wanted to create a system that could be used to teach fourth graders maths, and still be used to teach tenth graders programming at the same time. This reinforced the need for the system to be extendable, but also have an easy to use user interface. In order to solve this we proposed making several ways of creating the programs in the application, and making our programming modular so that subparts of the language could be disabled when not needed. To make this a feasible solution we envisioned a kernel language consisting of only actuator commands (e.g commands that directly influence the state of the robot), a extended language (which could include branching statements and variables), and a translator able to convert the extended language into the kernel language.

Choice of framework

The first decision to be made were regarding the platform and how to achieve a crossplatform solution. As solutions involving programming a computer connected

---

[1] http://scratch.mit.edu/about/

to the arduino was seen as too complex we focused on frameworks aimed at mobile devices. A native approach was quickly discarded as it would require multiple implementations in multiple languages (E.g. Objective-C for iOS devices, and Java for android devices, etc.). Removing the native approach from our solution pool still left us with several options, including game frameworks and standardized cross platform tools. In a report by vision mobile in January 2013 they present a study involving six hundred developers and their preferences ([47]). In addition to the top five most used frameworks from vision mobile's report and some newer frameworks, we initially included game frameworks into the pool of frameworks. The game frameworks were however discarded relativly fast as they rely heavily on custom built user interfaces and low level programming(openGL/webGL) for performance reasons.

After removing the game frameworks from our solution pool we looked throught the remaining options, PhoneGap(also known as Apache Cordova), Appcelerator, Adobe AIR, Sencha, Qt, DevExpress and Xamarin.

Framework	Prog. lang.	Notes
Apache Cordova	Javascript and HTML	Free of charge under the apache 2 license
Appcelerator	Javascript	High abstraction level. Not free to use in a commercial setting
Adobe AIR	Flash and ActionScript	
Sencha	Javascript	Commercial licensing
Qt	C++	Pay to use
DevExpress	Javascript and HTML	Pay to use. Based on Apache Cordova technology.
Xamarin	C#	Free starter pack, but have usage restrictions

Table 3.2: Overview over potensial cross platform solutions

The frameworks that had commercial licenses or a "'pay to use"' business model were removed from the list of potensial frameworks, which left us with the choice between Apache Cordova, Adobe AIR and Xamarin. We decided to move forward with PhoneGap due to language preferences in the team. Adobe AIR, while stating that one can use javascript, does in fact require the use of Flash and ActionScript for its mobile applications, and the we had very little experience with C#. Apache Cordova on the other hand is build around standard web technologies as javascript, HTML and CSS, which both members of the project have been working with before.

Technical details

With the framework decided we were able to start creating our system. Throughout the process of creating the application we focused on refactoring options, configurations and settings in each component into their own components. These options/settings would at a later stage be gathered together and some were exposed through the user interface in order to make it possible to change the behaviour of the application at runtime. The process of always putting all the configuration and options in the same place made it possible for the system to be extremely extendable and modifiable.

The main part of the application were developed using javascript, HTML and CSS, as these are the primary technologies used when creating application with the Apache Cordova framework. We will therefore keep the focus on these technologies, even though some minor adjustments where done in the native android controller to enable a higher debugging level than the bluetooth plugin had originaly.

Figure 3.1 shows a simple ”‘class diagram”’ of the application. Note that the term class diagram is not completely accurate as javascript is not a object oriented programming language in the same sense as Java, but utilizes prototypical inheritance to reuse common code components. The code for this application however isn't powered by a complicated class hierarchy and thus using the ”‘class diagram”’ is sufficient to show how different components interact.
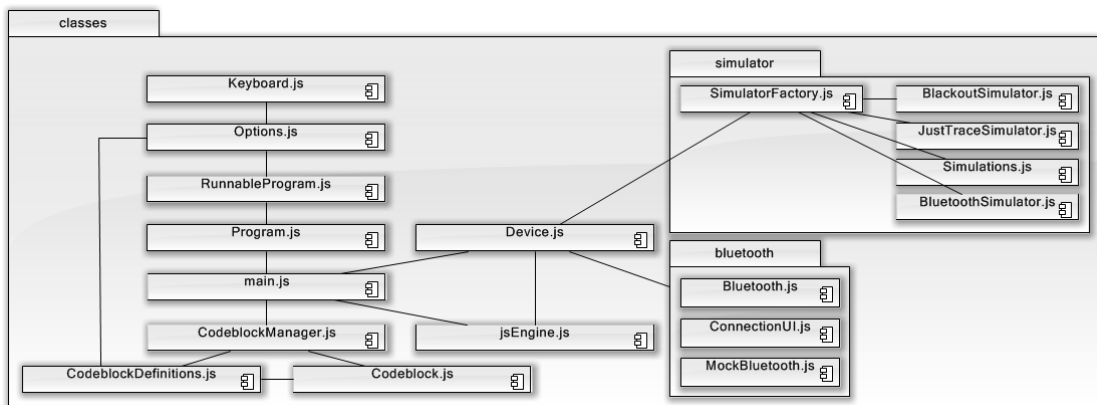
Figure 3.1: ”‘Class diagram”’ of the javascript application.

The class diagram in figure 3.1 shows a total of eighteen different javascript files, most of whom contain a single javascript ”‘class”’, though some do contain several

classes, but these classes are at an abstraction level below what should be exposed by the parent class.

The main responsibilities of each class is shown in table 3.3.

Class name	Descriptions
main.js	Provides the entrypoint for the application. This file does not contain any classes, but have the responsibility of starting the application up, initializing all the other classes and handling the majority of events delegation.
Program.js	Provides the Program class, which is responsibel for the extraction and validation of the program from either the graphical interface or the textual interface.
RunnableProgram.js	Is a wrapper around the Program model responsible for preparsing of the program in the cases where pure javascript have been written.
Options.js	A centralized locations for all application settings. Provides persistent storage of settings, and a simple user interface creator for generating the options menu.
Keyboard.js	A simple class responsible for creating and handling a on-screen-keyboard.
Codeblock.js	Another model class which stores the information of a single command. Marshalling and unmarshalling from and to the graphical and textual interface are handled by this model, but usually called at the *Program* level.
CodeblockManager.js	Responsible for inserting available codeblocks into the header of the application, and helping with the marshalling of textual blocks.
CodeblockDefinitions.js	Provides the definition of all available codeblocks in the system, and how the different codeblocks should be translated into javascript.
Device.js	An abstraction layer for removing any device dependencies and allowing the application to be tested on a computer by mocking different device dependent components.

Continued on next page

Table 3.3 – continued from previous page

jsEngine.js	A generator class responsible for converting the Program or pure javascript into actual commands which can be sent to the robot.
SimulatorFactory.js	The class responsible for keeping track of which of the four simulator types to use. This class listens to changes in bluetooth connectivity in order to allways provide the correct simulator.
***Simulator.js**	Refers to all the differet simulators. These classes takes a RunnableProgram as the argument, and spins up the correct simulator. In cases where a robot is connected they also have the responsibility of sending the commands and keeping in sync with the robot.
***Bluetooth.js**	Refers to both bluetooth classes. Based on a debugging flag in the application one of these classes will be started. MockBluetooth will keep the same to the same interface as Bluetooth, but will use stubbing methods in order to provide a testing surface for the bluetooth functionality.
ConnectionUI.js	The bluetooth connection user interface is dynamically injected if a bluetooth enabled device is detected by the application. The UI is maintained by this class.

Table 3.3: Class descriptions

This architecture and distribution of responsibility was chosen in order to make each component's responsibility as small and concrete as possible, which prevents alot of ripples if one component should change and makes debugging the application easier.

Most of the classes/files in the application are straight forward javascript "'classes"' with a minimal responsibility, which make them straightforward to understand and debug. The exception to this may be the application backbone responsible for translating our internal program representation into a program which can run on the robot and the simulator. Because of this, and all the other interactions between the components, we've included a short sequence diagram and explaination of how the application works when a user tries to start his or her program.

The diagram starts off with the user initiating the `Run` sequence, telling the application to take the currect program and run it. The `Program` class retrieves the program either from the graphical input or from the textual input dependent on which were active at the time when the user started the sequence. The program is at this program replaces by an instance of `RunnableProgram`, see listing 3.1. When the program have been retrieved and converted the event handler requests a simulator from the `simulatorFactory`, which in turn determines the currect `simulator` given the current state of the application(e.g. if a bluetooth connection is actice). With the simulator and runnable program prepared the event handler sends the program to the simulator and lets the simulator take it from there. The simulator will at this point request a program generator from the runnable instance. The program generator will if necessary convert the program into javascript(listing 3.3), sandbox the newly created javascript code and run it to generate the output commands needed for the simulator. The commands(listing 3.2) are at this point presented in a similar fashion to what you would expect from an java iterator (`hasNext` and `next` commands). With the support for native generator coming to javascript with the new ECMAscript 6 standard this is self created generator is most likely to be exchanged in favor of a native approach ([31]). If a bluetooth connection is present at this point and the simulator support bluetooth synchronization it will try to send the commands to the robot.

In order to contain most of the complexity within the application and not on the arduino we translate the commands into simple `move` and `turn` commands for the bluetooth module. A move command would be prefixed with the letter `m`, followed by a either a `+` or `-` indication the direction (positive is forward), and four hexdigits indication the length. I similar structure was used for turn commands with the prefix `t` instead of `m` (positive is a right turn).
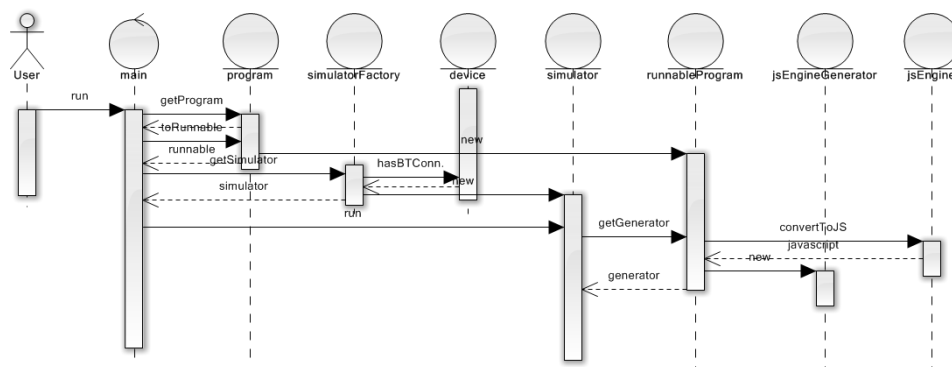
Figure 3.2: Sequence diagram of initialization of a user program

The code snipplets (listings 3.1 and 3.3) were choosen here as an example due to the fact that most of the language currently implemented is showcased in the example. This is however the extended mode of the language, whereas the experiment was conducted without loops, procedures and variables enabled in the language. Another reason for using this rather overly complex example is to give a thorough presentation of the internal representation (e.g. how the halting program is solved by limiting the number of total commands).

It should be noted that the inital solution to translate the user program into robot commands were done by our own simple translater. This did however become infeasble to maintain as the application language evolved and started to touch on more advanced programming structures as loops and procedures.

```
1   SET  NUM 4
2   SET  LENGTH 100
3   SET  ANGLE 90
4   PROC  SEGMENT
5        FWD  LENGTH
6        TL  ANGLE
7   END
8   REP  NUM
9        CALL  SEGMENT
10  END
```

Listing 3.1: Example code for creating a square using variables (Extended language), procedures and loop.

```
1   FWD  100
2   TL  90
3   FWD  100
4   TL  90
5   FWD  100
6   TL  90
7   FWD  100
8   TL  90
```

Listing 3.2: Commands sent to simulator(Kernel language)

```
 1  var icount = 0;
 2
 3  SET("NUM", 4);var NUM = 4;
 4  SET("LENGTH", 100);var LENGTH = 100;
 5  SET("ANGLE", 90);var ANGLE = 90;
 6
 7  function SEGMENT() {
 8     icount++;
 9     if (icount > 30000){return;}
10     FWD(LENGTH);
11
12     icount++;
13     if (icount > 30000){return;}
14     TL(ANGLE);
15     END();
16  }
17
18  REP(NUM);
19  for (var generated_55033 = 0; generated_55033 < NUM;
        generated_55033++) {
20     CALL("SEGMENT");
21     SEGMENT();
22     END();
23  }
```

Listing 3.3: The internal representation of the previous program.

Resulting application

The final version of the application[2] created a easy to use, extendable, crossplatform, and modifiable environment for creating program to the robot regardless of previous experience with programming and robotics. As the application also works as an web application (with the exception of bluetooth) we've deployed the application to web as well[3]. The web application is initialized with a mock instance of the bluetooth module, hence the application behaves exactly as if on a mobile device (appendix E gives an introduction to the usage of the application).

For the non-technical users the application provides a graphical user interface where the user simply can drag the commands around and create the program without ever needing to know anything about programming or internal syntaxes. For more experienced users the application provides a textural interface, where the user can either write the code using the applications own language(listing 3.1), or pure javascript code.

---

[2]v1.1 in https://bitbucket.org/nutgaard/turtle/commits/all
[3]http://folk.ntnu.no/nicklau/turtle/
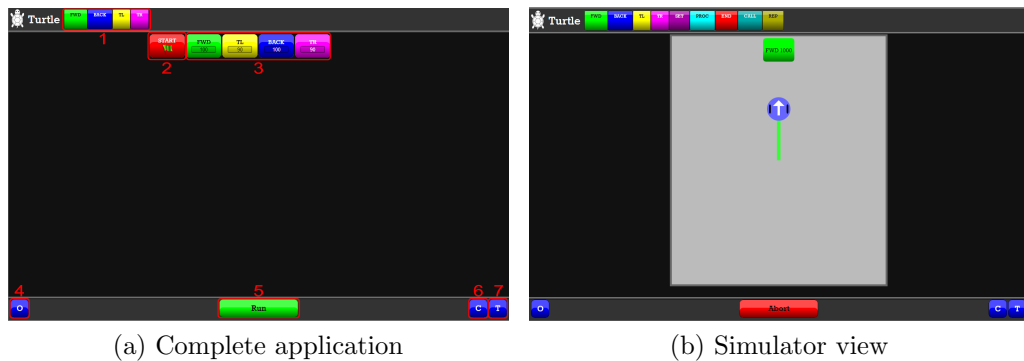
(a) Complete application   (b) Simulator view

In additions to being able to change the complexity of how the program are created
the application provides options for disabling parts of the language. In its simplest
form the application provides only four commands, forward, backward, turn left,
and turn right (figure 3.3a).

For a complete introduction to the usage of the application please refer to ap-
pendix E, where the whole application, including the different simulators, options,
and programming styles, is presented.

3.3 Hardware

The hardware part of the project consisted of initially three separate parts, a
drawing contraption, a battery holder, and a bluetooth shield. In the final version
of the robot we ended up not using the battery holder due to problems with
the manufacturing hardware. In this chapter we will first present the chassis
modification, before moving on to present the circuitry needed for the bluetooth
module.

Chassis modifications

The problem we initially underestimated was how we should make the physical
robot draw on paper. While a front or back mounted pen would be an easy
solution this would not create the angles properly, and thus was not a real viable
options. The clearance below the robot was to small for anything to fit underneath
it, and the battery was placed in such a way that nothing could fit between the
wheels of the robot (see figure 3.4c). In order to fix this problem we designed a
battery holder which could be mounted on top of the robot (see figure 3.3a), but
due to technical problems with the 3D printer we were not able to test this design.
We had to resort to an alternative solution using velcro to position the battery at

the back of the robot and duct tape to add a counterweight at the front of the robot to provide stability (see figures in 3.4).

While this solved the problem of making room beneath the robot we still had multiple issues with getting the robot to draw properly. These issues were mainly due to inaccuracies in the 3D-printing environment meaning the the spring used to push the drawing tip into the paper would get snagged inside its container (figures 3.3b and 3.3c). We also found that the design needs a spring with fine tuned strength. If the spring were too strong it will lift the robot out of balance and only one wheel will touch the ground. If the spring were too weak it will not draw properly. One potensial solution to this problem is to add more weight to the robot. However this may pose new problems as the motors are not strong enough to move a very heavy robot.
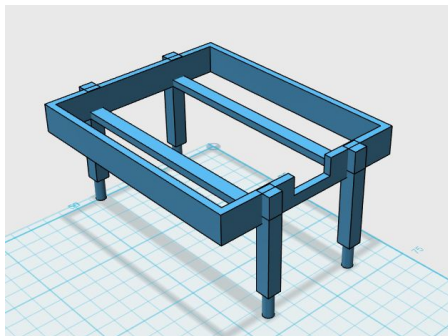
(a) Battery holder
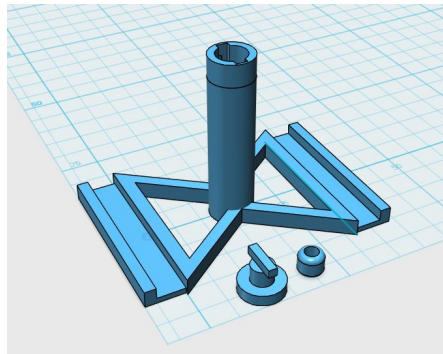
(b) Pencil holder

(c) Printed pencil holder

Figure 3.3: 3D printed utilities
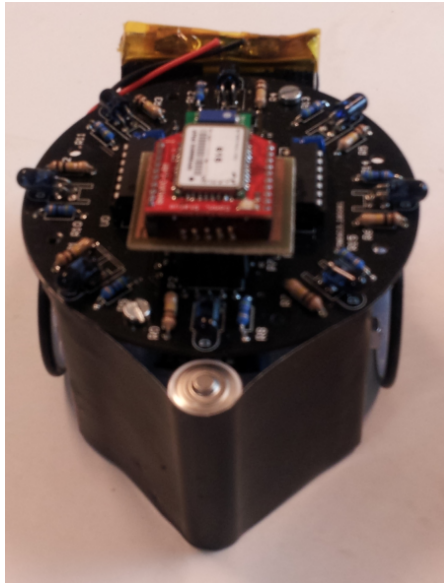
(a) Front of robot
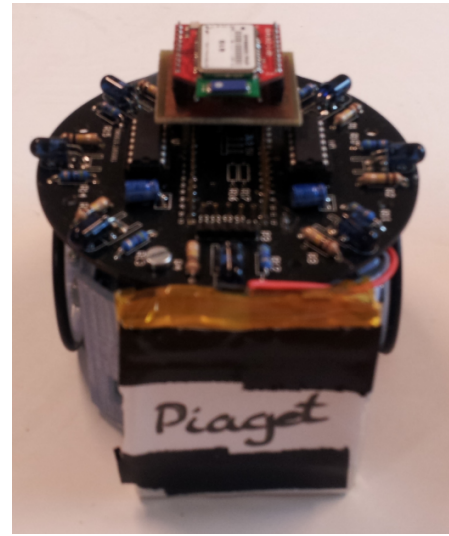
(b) Back of robot
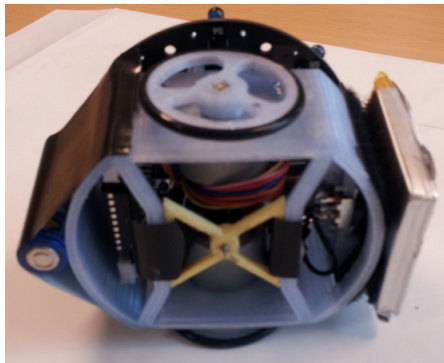
(c) Underside of robot

(d) Robot drawing an angle

Figure 3.4: Pictures of the robot

Bluetooth circuit board

The ChIRP team had previously tested using bluetooth communication between an android device and the robot. We therefore got help in creating the inital circuitry for our bluetooth module(figure 3.6a). This drawing was at this point designed in the Eagle software in order to make it possible to produce at a later stage(3.6b).

The basic functionality of the circuit is to create a mapping between the receiver (**RX**) and transmission (**TX**) signals on the arduino and bluetooth module. In addition to the data signals the circuit board needed to map a 3.3V power supply (**VDD**) and a ground connection (**GND**) to the bluetooth module (see schematic in figure 3.5).
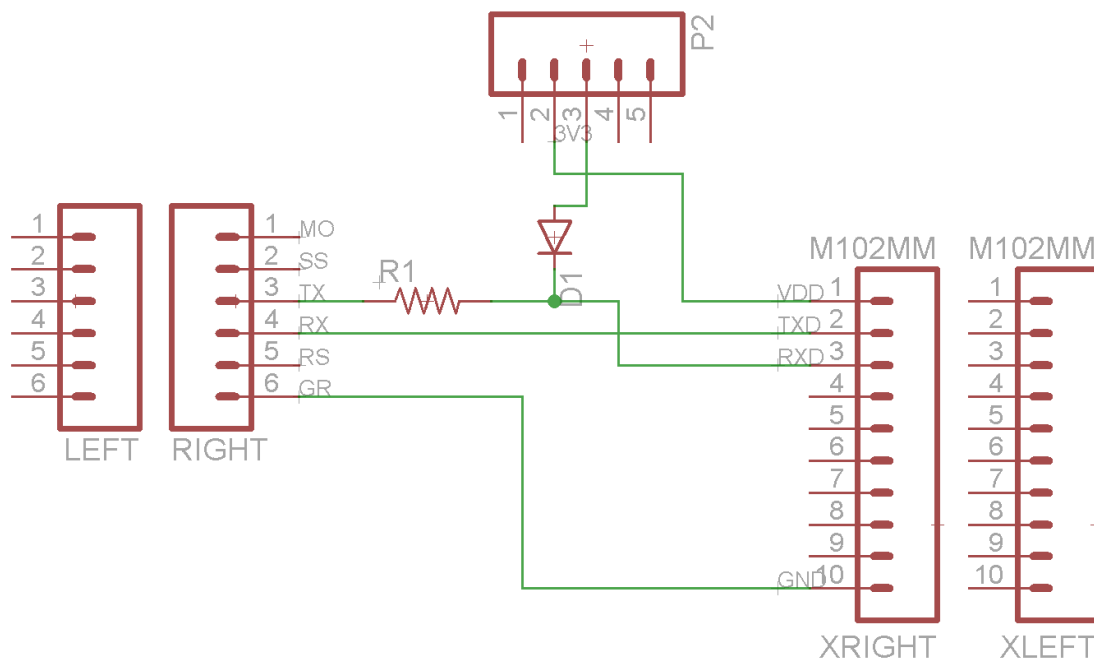
Figure 3.5: Schematic of the bluetooth shield. Components on the right is the bluetooth module.

As seen in figure 3.5 we used a two components between the arduino TX signal and the bluetooth module RX (a 220Ω resistor and 3.3V zener diode). This was because the arduino operates on 5V, while the bluetooth module expected only 3.3V. Mapping these directly would most likely result in the destruction of the

module. It should here be noted that the schematic does in show a normal diode, but the circuit is design for a zener diode and using a normal diode would not work.

It can be seen in figure 3.6d that some soldering around the board was necessary on the prototype. The prototype did however lead us to conclude that the circuit itself was functional and we were able to proceed with creating all the circuit boards that we needed. Figure 3.6e shows the template used in the etching process when creating these boards. While figure 3.6f shows the circuit board alongside with the bluetooth module we used for our robots. The complete assembly of the robot can be seen in figure 3.4d.
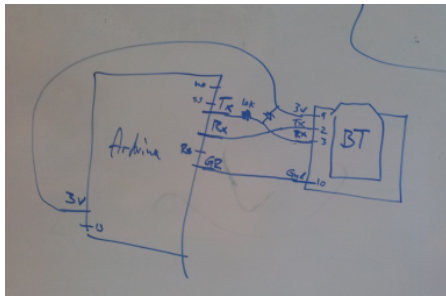
(a) Inital circuit
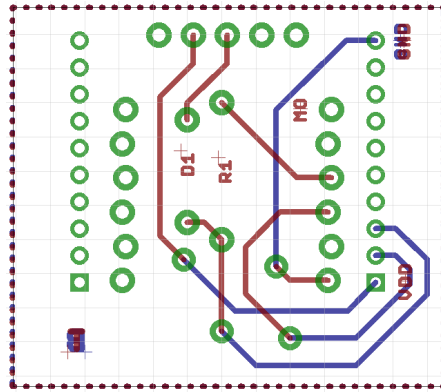
(b) Created in circuit designer
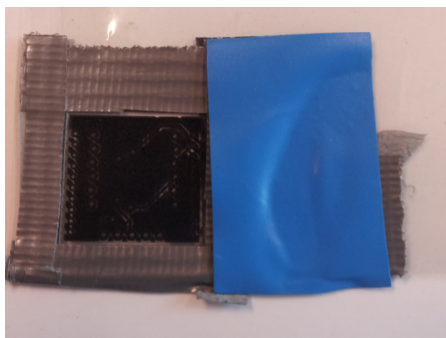
(c) Prototype template

(d) Prototype circuitboard
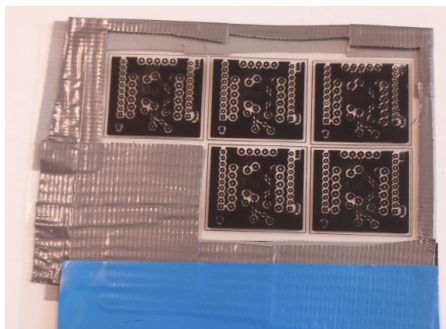
(e) Finished template

(f) Finished circuitboard

Figure 3.6: Circuit boards

RESEARCH METHODOLOGY

The inspiration behind the study was a wish to test the robot as an educational tool in an experimental classroom setup. We wished to increase students' understanding of angles and their angle estimation skills. We also wished to increase the students' cooperation and motivation. As we mentioned in section 2.3, the pedagogic theory we rely on is constructionism, which is quite different from the blackboard and task driven mathematics education, which according to Alseth et al. [2003] happens in most of the classrooms in Norway, including the classrooms at THIS.

4.1 Choice of methods

We chose a mix of qualitative and quantitative methods. This was due to our research questions and pratical considerations like time limitations. The quantitative methods was meant to aid in our attempt to answer the research question: is there a learning difference in students understanding of angles. The qualitative methods should aid in pinpointing strengths and weaknesses in the robot compared to a simulator as well as supplementing the quantitative data findings.

For our first research question "How can using robotics in school help students understand what an angle is and aid in their angle estimation skills?", we wanted to use interview, test and observation. This decision was based on related work discovered through the prestudy, but also our own beliefs that an individual construct their own knowledge within their social environment by building mental models which they can use to think with. We would have liked to use interviews because of its ability to ask followup questions and get a deeper understanding

of how the robotics changed the way students think. However, due to the time limitations we were not able to use this method. The experiment was preformed at the end of a semester as getting in contact with the school, designing the experiment, and designing the prototype system all took a long time. At the time the students schedules were full and would therefore miss something important if they had to be taken out of the class to be interviewed. A test could be administered by the teacher and all the students could do it at the same time. An interview would have taken longer, as we would have needed to interview one student at a time. We tried to compensate for this lack in data by including questions where the students have to elaborate on how they got their answers on the test. We also wanted to observe the class. It is hard to observe without intervening and still understand what the students are thinking, but through watching how they used their protractors we could identify which angle they were measuring, and that shows whether they have understood it or not. We chose testing because this is how students are used to getting tested.

For the second research question "What are the strength and weaknesses of using a robot compared to a simulated environment?", we wanted to use observation, interview and questionnaire. We wanted to use observation to supplement the quantitative results, but also to assess the cooperation and motivation, even though this is hard to observe. We want to test factors found through the prestudy, that might contribute to the robot activity being better than the simulator activity. We will test this through a questionnaire and in-class observations and focus on the important factors of mental models, motivation, cooperation and problem solving approach. We also gather qualitative data about the tasks by asking students to answer how they thought when attempting different tasks in the test. With these results, we will try to amplify the quantitative results and provide another basis for us to draw our conclusions.

4.2   Participants

To analyze our research questions, we tested the robots with a group of fifth-grade students from the international school in Trondheim(THIS). A group of fifteen students (four boys and eleven girls) was selected from a class of twenty to participate in the experiment. Three of the twenty students was on vacation during the experiment period. Two of the twenty did a full score on the pretest and their approach to the problems, as described in the depth questions, were perfect. We therefore concluded that our activities could not teach them anything new. The teacher also suggested that we don't include these two students. She meant that if they were in the group their ideas and help would dominate the

learning gains of the students. Since we want to test the robot activity's ability to teach students the material, these students might interfere with the results. From these fifteen only eleven were able to participate on both sessions. All the students were in the same class and either ten or eleven years old. They had been taught about angles one month prior to our experiment, using a traditional blackboard based teaching scheme. Thus every student had prior knowledge of angles and angle estimation. The next thing they were going to learn was about reflex angles and area of shapes. None of the students had worked in this experimental manner before or worked with this specific kind of technology before.

We originally wanted to include only students who knew nothing about angles. This class was chosen because it was nearing the end of the school year for THIS and this was the only available class. It would have been a perfect choice if the experiment had been performed one month earlier, when they were first introduced to the concepts of angles and degrees. Another alternative would be to do this experiment after the summer vacation, so the knowledge they gained would have had time to sink in and not get different ideas confused or altered by our experiment.

4.3   Experimental design

The selected students were divided into two groups by the teacher in advance. The teacher created the groups with two factors in mind, equality in terms of content knowledge level and by which participants she knew could work well together. We randomly assigned one group to the simulator activity and one to the robot activity. The robotics or experimental group (seven students: three boys and four girls, where one boy and one girl could not attend both sessions) was appointed to work with the physical robot activity. Considering that simulators have shown to be an efficient learning tool [30], and Papert [1980] mention that there is not an obvious difference between mental model building in robots and simulators, the control group (eight students: one boy and seven girls, where two girl could not attend both sessions) worked with an on-screen simulation of the robot, seen from a top down perspective on the tablet(see figure 3.3b). Other than this, the activity remained the same. The choice of control group activity was also influenced by our findings in the prestudy, stating that most studies use a control group that has not done any alternative activities, which we feel is a problem.

We altered which group started first because, as the teacher pointed out, when teaching, you always do best the second time, so we wanted to minimize this effect between the groups. The materials required for this experiment were the robots (one robot per group) for the robot group, the tablet or android phone (one

tablet/phone per group) and protractors. Experimental and control groups had no interaction during the activity sessions, however some of the other simulator group members came over and watched the robotic activity sometimes. The experiment was conducted in the classroom with a teacher and the other group present in the room due to legal issues. The students were free to choose their team mates once the activity started. They then stayed in these groups throughout both sessions. The tasks were the same in both groups.

Groups of two members were preferred. This decision was based on our findings in section 2.2. When we discussed the experimental design with the teachers, they agreed there should be two students per group based on their experience. However to avoid the case of one student working alone and considering our lack of tablets, some groups had to contain three students.

4.4 Procedure

The study was divided into two sessions, one held on a Friday and one on the following Monday. The first session lasted 40 minutes and the second session lasted 45 minutes. Each session consisted of an introduction to the subject, followed by a set of 10 difficulty increasing robot exercises. During the exercise phase the students worked on their own, while we acted as tutors, solving doubts regarding software and the activity itself. If groups were stuck on a single task without making progress we provided subtle hints to help them along. We never gave them the answers or any code, but explained that visualizing the program block by block and think about the effects on the robot. If it was clear that they measured the wrong angle we helped them get on the right track by having them explain how the robot moved. Other than this the students received no external help or instruction. The exercises was designed so that everyone should be able to do something, but no one should be able to do everything. One of the groups had to use a phone instead of a tablet due to unavailability of one tablet. The only difference of our application when used on a phone compared to a tablet is a smaller screen, and it did not seem to make a difference, at least not on the robot group, as they don't use the simulator, which is the part requiring a larger screen.

First lecture
In the first lecture, the robotic activity was the first learning activity that the students in the experimental group did on that day. The control group did other activities with the teacher, not related to our work, before working with the activities.

The content in the first ten minutes was exactly the same. First an introduction of how the application works, then an introduction to supplementary angles and reflex angles. Next followed a 32 minutes session where the students experimented with the robot / simulator and attempted to create paths that were depicted in a task sheet. The tasks are included as appendix C. The exercises of the first session were relatively easy, since their main purpose was for the students to familiarize with the technology. Only open polygons, or paths, where you don't end up where you started, were used in the first lecture exercises. This was done so that students would not get caught up in the precision of the robot or the length of movement. The tasks met our expectations, as everyone managed at least four tasks and the best group finished the last task right before the session ended. When solving these tasks, the students were allowed to use protractors and pens.

Second lecture
When teaching, one always does better the second time. Therefore the control group started first in the second lecture. The session was the first learning activity that the students in the simulator group did on that day. The robotics group had physical education before we swapped groups.

The second lecture started with a small recap on what we did last time. Then followed a short talk introducing regular polygons. We mentioned that, in order to complete a shape and make the robot end up where it started, it needs to turn all the way around during one program run. The students understood that this meant turning $360°$. The recap focused on supplementary and reflex angles. Then followed 45 minutes of robot / simulator task solving activity. The tasks are included as appendix D. In the second session every exercise was a closed shape, and the students had be careful of how far the robot moved in forward or backward block. As in the first lecture the tasks we created for the second lecture also met our expectations, as everyone managed at least four tasks and the best group finished the last task right before the session ended.

4.5   Data collection

In order to assess the learning gains accomplished by the students in angle and turn measurement estimation, a pretest-posttest design was used. The pretest was administered one week before the first session, and the posttest was administered an hour after the second session. The tests were individual and the students were not allowed to cooperate. Both tests were administered by the teacher. The other advantages and disadvantages of robots identified in section 2.2 was measured

based on qualitative observations and quantitative results obtained by means of a post activity questionnaire.

Test

Since we could not find a standardized test that measure angle skills with well defined reliability and internal consistency we had to make our own. The test is also a part of our contribution. Creating a test that uncover what we are looking for is hard, especially when we are looking at a new area of investigation. Pedagogical insight is needed to plan a good pretest and posttest, and we cooperated with the teacher of the experiment class to create our test.

We got inspiration from Clements et al.[11]. They tested similar concepts, and the programming language they used was LOGO. In this research they have created several tests to measure geometry learning gains. These tests have been used in several projects before us [10; 13; 12; 11]. A subset of these tests specifically targeting angles and turn measurements were selected. We then expanded and improved these questions in a session with the teacher, with regards to what the students already knew and what they are supposed to learn next. The finished test consist of twenty questions, divided into six categories (table 4.1). The score range from 0 to 24 and was designed to be completed in thirty minutes. We used the same tasks for both the pre and post test so we could look at differences. The test does not include questions specific to LOGO or the robots, but are general questions to measure angle understanding. Assessments of the tests were conducted after the experiment ended. Each correct answer were given one point, whereas incorrect answers gave zero points. No partial scores were given. Questions added to investigate the though process of the participants were not graded (e.g. question 1, "'how did you solve this?"' and question 11).

Category #	Description
1	General angle understanding
2	Normal(inner) angles
3	Reflex(outer) angles
4	Complementary and angles in shapes
5	Orientation and estimation
6	Calculation with angles

Table 4.1: Categories

Below the test is described step by step and we discuss the decision behind each task and how they were scored. The test can be found in appendix A.

Question 1 was an open ended question meant to assess students understanding of what an angle is and their ability to explain it. There were given no points for this question.

Question 2 and 3 asked the students to draw an angle and then to draw a bigger angle. One point was given for each correct drawing, with a maximum score of 2 points. They were then asked to describe why the second angle is bigger. There were given no points for this answer, it was meant as a qualitative supplement to get a deeper understanding of their though process.

Question 4 was aimed at assessing students ability to differentiate between angles and identify which is bigger in each pair. This question was constructed of 5 tasks, 3 normal, or inner, angles and 2 angles which has highlighted the outer angle. Where each induvidual task were worth one point. These 2 outer angles were added based on the teachers advice, that this is what the students are supposed to learn next. In the first pair, one of the angles has longer sides than the other and seeks to confuse students who don't understand that angle size is measured according to how wide it is, not how much space it occupies. In the second pair, one angle is turned upside down to appear smaller than it actually is and will test students orientation ability. In the third pair the angles are very small and will test the students general estimation skills, in case they cannot answer correctly on the two previous pairs. The 4th and 5th pair are outer angles. The 4th pair has an orientation change to confuse the students while the 5th pair is straight forward to test if they understand outer angles. One point was given for each correct choice, with a maximum score of 5 points.

Question 5 was aimed at angle calculation skills. There are a certain set of geometric properties that the students are supposed to be familiar with already, which we predict will be reinforced by our experiment. One point was given for each correct angle, with a maximum score of 3 points.In addition they are asked to explain how they got their answer, this is not scored.

Question 6 tests the students estimation skills. 2 of the questions have a multiple choice format, while in the 2 last the students need to write an estimate on their own. One point was given for each correct choice or estimate, with a maximum score of 4 points. In the tasks where they have to write their own estimate we have an error margin of $\pm 15°$. The first task is a straight angle which they should be very familiar with. The next has its orientation changed, so does the 3rd. In the 4th angle estimation task the angle is very wide compared to the previous tasks.

Question 7 is the same as question 6, but with outer angles. This will test their ability to measure the outer angle and understanding of what the circle drawn on

the angles mean. One point was given for each correct choice or estimate, with a maximum score of 4 points. There is an error margin of $\pm 15°$. The angles are the same as in question 6, but have had their orientation changed and their order changed.

Question 8 tests the students ability to orient. They must imagine themselves in the position of the arrow and with the direction of the arrow and decide which direction they want to turn. Then they need to estimate the amount of degrees to turn in order to face the black dot. One point is given if both the direction and amount of degrees is correct, with a maximum score of 4 points. There is an error margin of $\pm 15°$.

Question 9 and 10 are also turn estimation tasks. The students must keep track of how much the robot has turned at every step and in which direction it is currently facing. In question 9 one point is given if the answer is 12. In question 10 one point is given if the amount of degrees are 50. The maximum score is 2 points.

Observation

During the experiment we observed the students. We used hidden observation and acted as spectators. We had a focus on teamwork, motivation and problem solving approach during the observation. We got inspiration from Mitnik et al. [2009] and focused on the 4 factors that foster collaboration. Individual responsibility, mutual support, positive interdependence and social face-to-face interactions. To observe problem solving approach we had a focus on how the students used their protractors and if they discussed the problem before starting to program it. As both the simulator group and the robotics group consisted of more teams than there were observers, we confined their working area in such a way that both observers always could see and head every team.

Questionnaire

The questionnaire can be found at the end of the posttest (Appendix B). It was created to test the difference between robots and simulators regarding some of the factors identified on section 2.2. Each of the questionnaire's questions had five possible answers; `--`, `-`, `0`, `+` and `++`; `--` meant "very little" while `++` meant "very much".

The factors focused on in this questionnaire were interest, motivation, entertainment, cooperation and applicability of robotics. All but the last factor tries to

determine the participants perception of the activity as they experienced it. Applicability of robotics was included to get the participants' opinion of how well the activity teached or could teach them about mathematics.

CHAPTER 5

RESULTS

The analyses in this chapter is based on the pretest and posttest results from all the participants that participated in both sessions. To see the questions in the pretest and posttest we refer you to appendix A and appendix B.

The distribution between groups and gender can be seen in table 5.1, the raw scoring data for each participant can be seen in table 5.2, and the means for each group in table 5.3 and with categories in table 5.4.

For the comparison between groups we used a independent-samples t-test in order to determine if there was any significant difference in the gains between the results of the robotics group and the simulator group (section 5.1), while a paired-samples t-test between pretest and posttest scores were used to investigate differences between pretest and posttest scores(section 5.2). In section 5.3, "'Additional analyses"', we have included a set of additonal test, where we further analyse the dataset to see if any significant connections can be found.

All the following section(chapter 5.1, 5.2 and 5.3) have utilized different analyses, but share a common structure in how they are presented. They start with the assumptions related to each analysis, if assumptions are violated, appropriate meaures are used to show that the violation itself does not affect the resulting outcome. At the very end they all end with a small summary of the analysis. The last analysis section (chapter 5.4) in this chapter of the report shows the Cronbach's alpha analysis for measuring the internal consistency reliability and test-retest analysis using Pearson's product-moment correlations between pretest and posttest scores to measure the stability reliability.

	Simulator	Robotics
Participants	6	5
Male	1	2
Female	5	3

Table 5.1: Participants overview

Participant ID.	Raw scores(q1..q24)	Score (Gain)	Group
1 - pretest	111110011111110000111101	17	Simulator
1 - posttest	111110011111110000111110	17(0)	
2 - pretest	111110011111110000110111	17	Robotics
2 - posttest	111110000011110000111000	12(-5)	
3 - pretest	111111111111111111000011	20	Simulator
3 - posttest	111111111111111111001011	21(1)	
4 - pretest	100000011111110000010011	11	Simulator
4 - posttest	111110001111110000111000	14(3)	
5 - pretest	110010001101000010100010	9	Simulator
5 - posttest	111111100011011101100000	14(5)	
6 - pretest	111111111111110011101111	21	Robotics
6 - posttest	111111111111111111111011	23(2)	
7 - pretest	111110011011000000110010	12	Simulator
7 - posttest	111110001011010000100010	11(-1)	
8 - pretest	111110011111100000111000	14	Robotics
8 - posttest	111110011111110000101100	15(1)	
9 - pretest	110001101011101000100000	10	Simulator
9 - posttest	111111100011111000101000	14(4)	
10 - pretest	111110101010111111101111	19	Robotics
10 - posttest	111111111111111111101110	22(3)	
11 - pretest	111110011111111011111111	21	Robotics
11 - posttest	111111111111111101101111	22(1)	

Table 5.2: Results from pre and post test

	Simulator	Robotics	Robotics*	Total
Pretest mean	13.17(4.36)	18.40(2.97)	18.75(3.30)	15.55(4.52)
Posttest mean	15.17(3.43)	18.80(4.97)	20.50(3.69)	18.82(4.40)
Gain	2.00(2.37)	0.40(3.13)	1.75(0.96)	1.27(2.72)
Normalized gain	0.17(0.17)	0.19(0.56)	0.42(0.26)	0.18(0.37)
Sig. Gain (p)	0.09	0.79	0.03	0.152

SD in paranthesis. * outlier removed. Sig.Gain calculated by paired t-test

Table 5.3: Means

Group	Cat. 1(2)	Cat. 2(7)	Cat. 3(6)	Cat. 4(3)	Cat. 5(4)	Cat. 6(2)
Robot-pre	2.00(0.00)	6.60(0.55)	2.40(2.30)	2.60(0.89)	3.20(0.45)	1.60(0.89)
Robot-post	2.00(0.00)	7.00(0.00)	3.40(3.13)	2.40(1.34)	3.00(0.00)	1.00(1.00)
Robot-pre*	2.00(0.00)	6.50(0.58)	3.00(2.16)	2.50(1.00)	3.25(0.50)	1.50(1.00)
Robot-post*	2.00(0.00)	7.00(0.00)	4.25(2.87)	3.00(0.00)	3.00(0.00)	1.25(0.96)
Sim.-pre	1.83(0.41)	4.67(2.07)	1.67(2.42)	2.33(0.82)	1.50(1.38)	1.17(0.75)
Sim.-post	2.00(0.00)	6.67(0.52)	2.33(2.73)	1.50(1.38)	2.00(1.26)	0.67(0.82)

In parenthesis; header: max score, body: standard deviation. * outlier removed

Table 5.4: Means of question categories

5.1   Independent-samples t-test

This test is used to determine if a difference between the means in the two groups exists, and if this difference is statistically significant. An alternative analysis that could be used here is the ANOVA analysis, as this would similar results when the independent variable is dichotomous. In order to use this analysis, six assumption regarding study design and the nature of the data must first be confirmed to hold.

Assumptions regarding study design

The first three assumptions are directed at how the study was designed. The assumptions are; 1) must have one dependent variable measured at the continous level, 2) must have one independent dichotomous variable, and 3) should have independence of observations. In our case all of these three assumptions hold. The dependent variable is `diff`, representing the diffrence from pretest to posttest is a continous variable. The independent variable in this case is wether or not a participant were exposed to the robot or just the simulator, which we have encoded as `robot`. Since `robot` have just two possible values it is dichotomous. The last assumption states that should be no overlap between the different group,

which also holds true for our study. It should however bit noted that participants from the different group could communicate with eachother when they were not actively participating the study(e.g. during the weekend, break, etc) and that this could introduce some errors, but we do not see this as a major infraction of the assumption.

Assumptions regarding the data

The last three assumption are directed at the nature of the data, and will require some extra analysis to verify. The assumptions are; 4) there should be no significant outliers in the two groups, 5) the dependent variable should be approximately normally distributed for each group, and 6) there is homogeneity of variances between the two groups.

Testing for outliers

The forth assumption assumes that there should be no significant outliers in the different group in terms of the dependent variable.
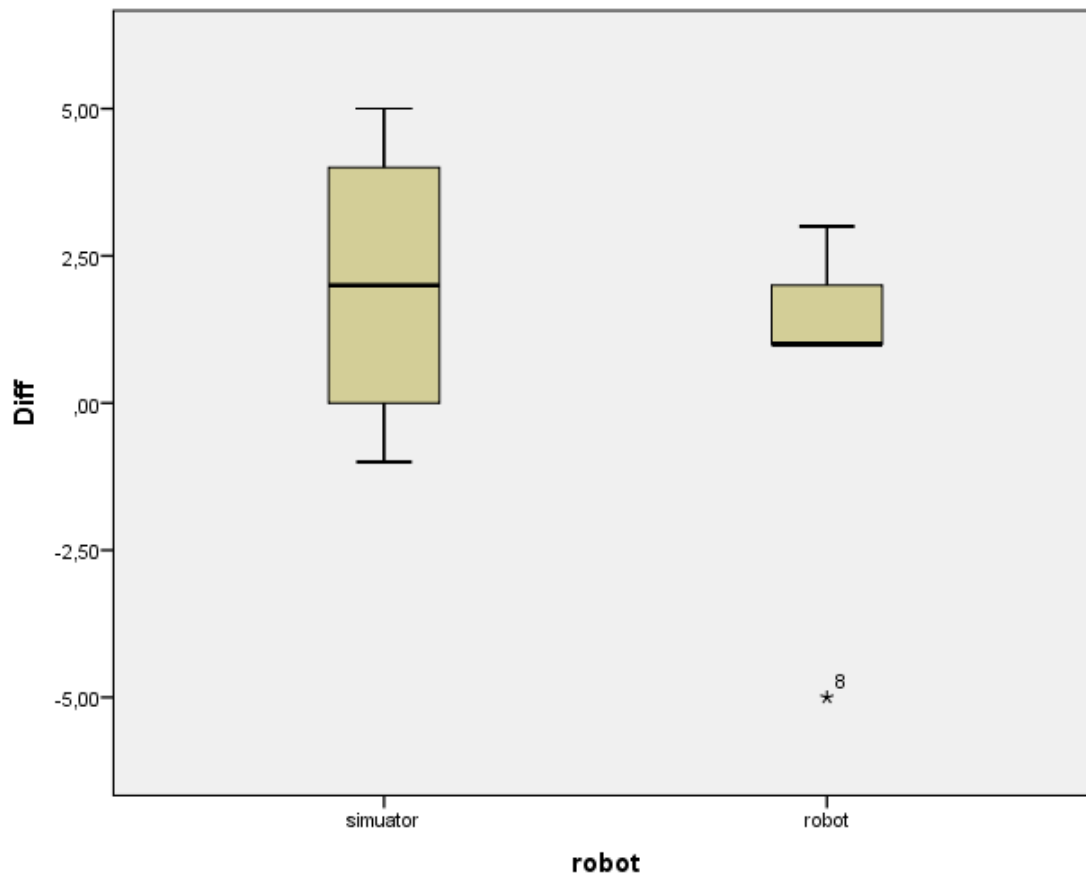
Figure 5.1: Boxplot of gains grouped by treatment

As seen in figure 5.1 an extreme outlier was found in our dataset within the robotics group(The asterix represents an extreme outlier in SPSS). By reassessing the tests, and investigating the circumstances during the pretest and posttest it was determined that the outlier were most likely not caused by errors during assessment, data entry or the tests. As a result of this, a comparison of the results with and without the outlier was conducted (see table 5.5) to see how the outlier effected the analysis. It should be noted that the results without the outliers failed Levene's test for equality of variances ($p = 0.033$) and results in the table are therefore produced by using the Welch t-test. Both do however show that there was no statistically significant difference in the gains between the robotics and simulator group($p > 0.05$ for both results). For good measure we have included two entries in the following table for the robotics group, one where the outlier was included and one where the outlier was not included.

Include outliers	t	df	Sig.	Mean difference	Std. Error difference
No	0.232	7.05	0.823	0.25 ($-2.296..2.796$)	1.078
Yes	0.967	9	0.359	1.6 ($-2.143..5.343$)	1.655

Table 5.5: Comparasion of results with and without outliers, using independent samples t-test. 95% CI

Testing the normality

The fifth assumption assumes that the dependent variable should be approximately normally distributed for each group. This assumption was tested before making the comparisons seen in table 5.5 as this assumption must hold in order to make the comparison valid. In order to check the this assumption hold we calculated the skewness and kurtosis values and their respective z-values, looked at the values produces by using Shapiro-Wilks test for normality and looked at the normal Q-Q plots for each group(see tables 5.6 and 5.7, and figures 5.2,5.3 and 5.4).

Group	Skewness	Kurtosis	Z-skewness	Z-kurtosis
Robot with outlier	-1.838	3.751	-2.013	1.876
Robot without outlier	0.855	-1.289	0.843	-0.492
Simulator	0.000	-1.875	0.000	-1.077

Table 5.6: Z-values for skewness and kurtosis

Group	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Robot with outlier	0.376	5	0.020	0.788	5	0.065
Robot without outlier	0.283	4	-	0.863	4	0.272
Simulator	0.164	6	0.200	0.950	6	0.739

Table 5.7: Test for normality
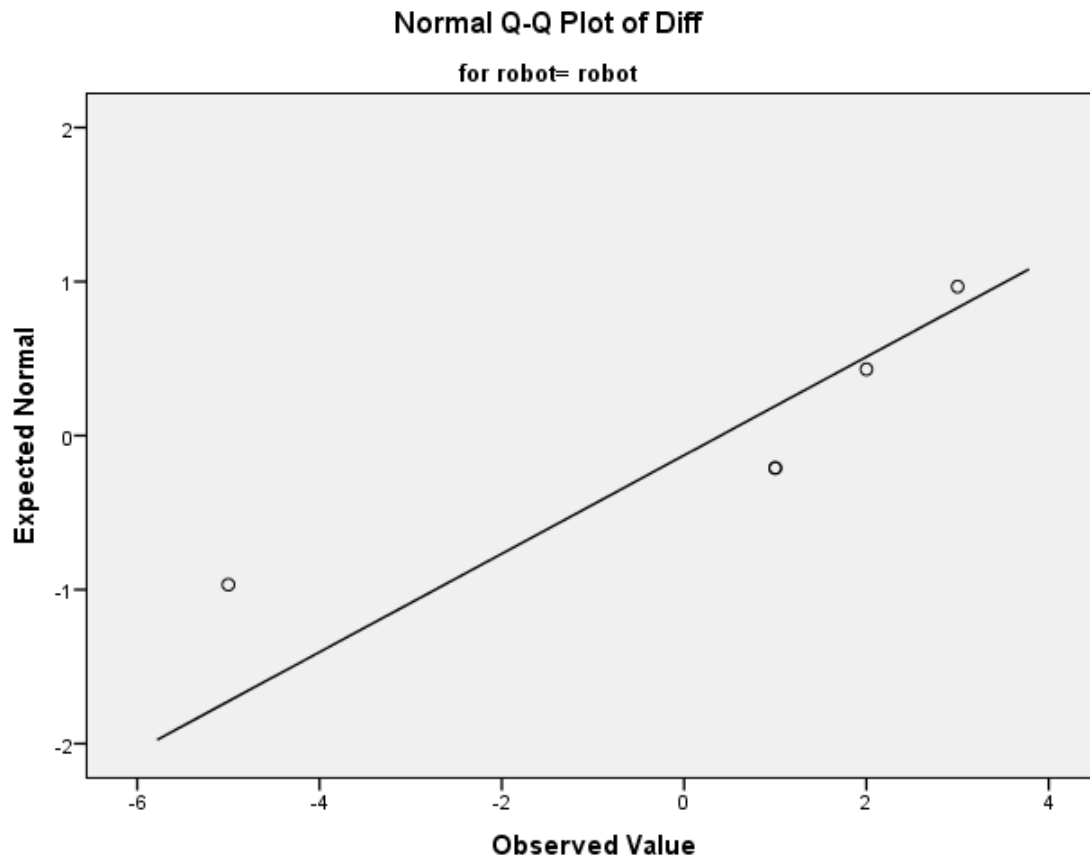
Figure 5.2: Normal Q-Q plot of gains for the robotics group with outliers
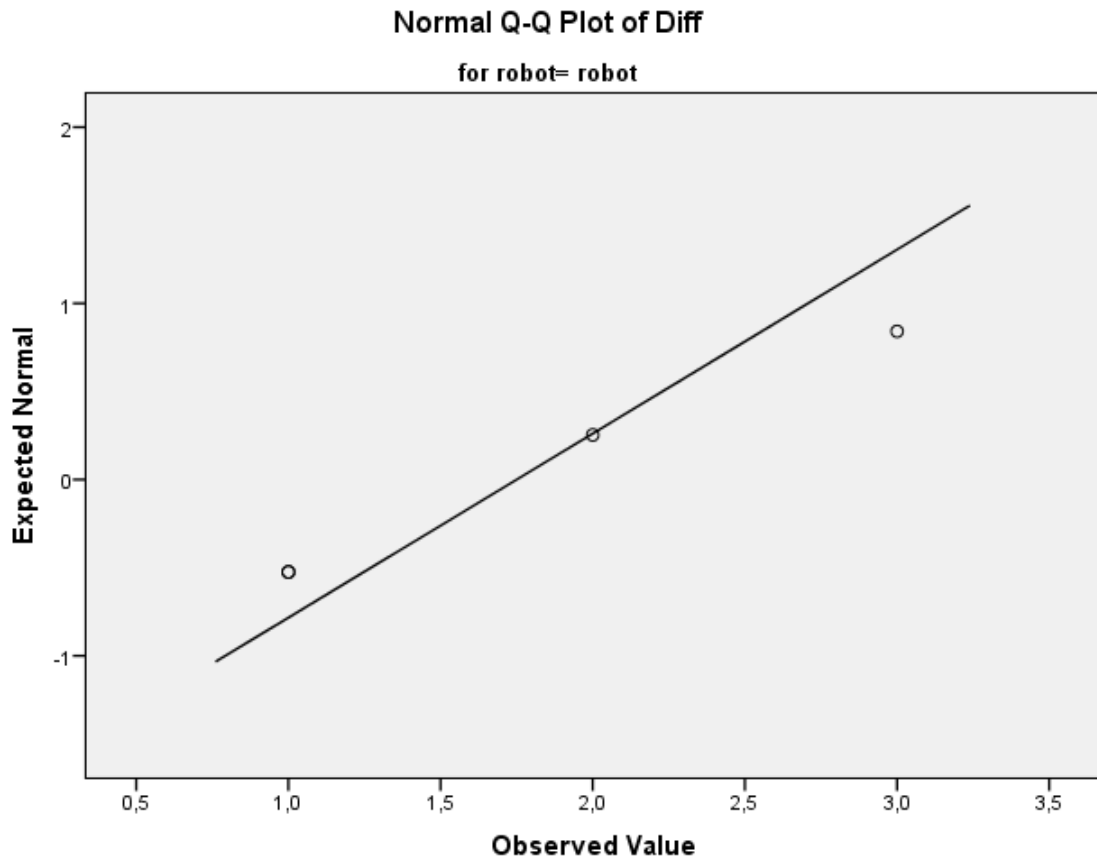
Figure 5.3: Normal Q-Q plot of gains for the robotics group without outliers
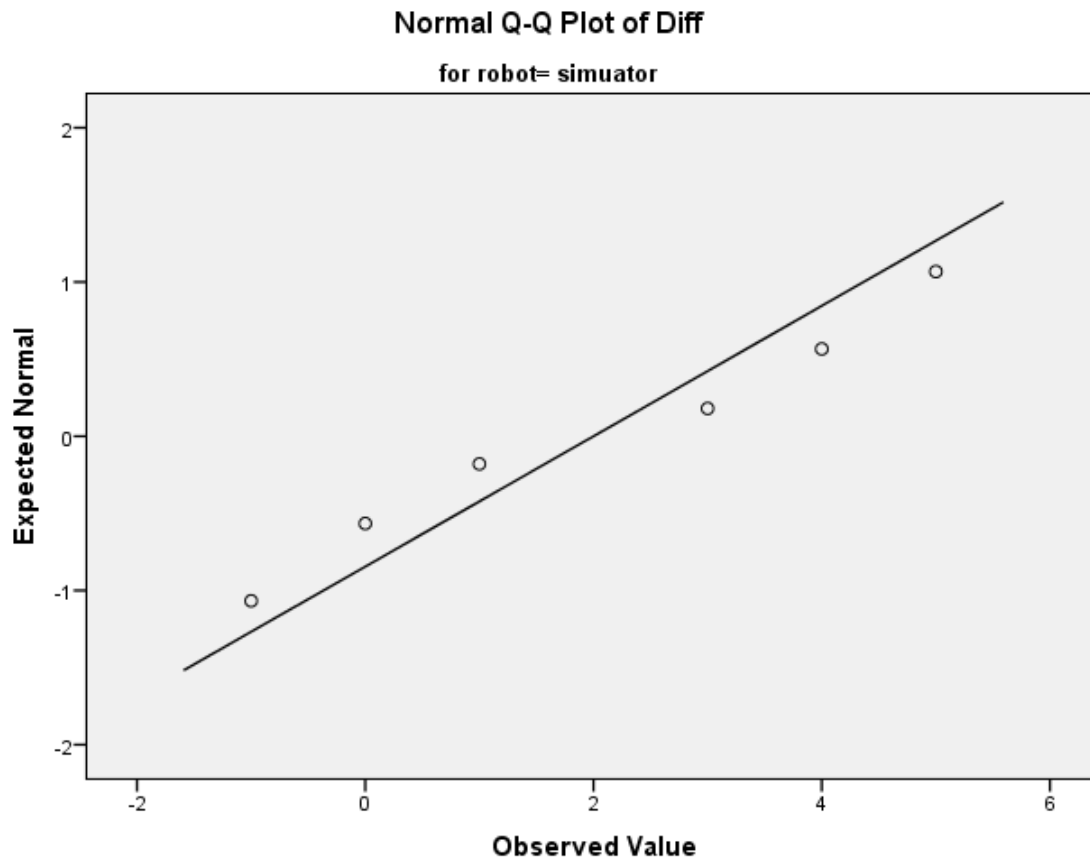
Figure 5.4: Normal Q-Q plot of gains for the simulator group

As seen by this(tables 5.6 and 5.7, and figures 5.2, 5.3, and 5.4) we can be fairly certain that the assumption regarding normality is not violated as all z-values are within the ± 2.58 range (representing a statistical significance level of 0.01) and the Shapiro-Wilk test shows $p > 0.05$ for both groups.

Testing for homogeneity between variances

The last assumption relates to homogeneity of variances within each group. As mentioned when comparing the results with and without outliers, this assumption was violated when the outlier was excluded($p = 0.033$), but held when it was included($p = 0.863$). None of the results do however show a statistically significant difference, and it therefore fair to assume that there is no statistically difference between the two groups.

Analysis summary

The independent-samples t-test was used to determine if there were differences in gains from pretest to posttest between the robotics group and simulator group. One outlier was identified within the dataset (figure 5.1), but did not effect the decision of keeping or rejection the null hypothesis and was thus kept in the dataset during the rest of the analysis. The gains for each group were normally distributed as shown by z-values and Shapiro-Wilk's test ($p > 0.05$), and there was homogeneity of variances when the outlier was included($p = 0.863$). The increase from pretest to posttest was greater for the simulator group ($M = 2.00, SD = 3.13$) than for the robot group ($M = 0.40, SD = 2.37$ with the outlier included, and $M = 1.75, SD = 0.96$ when it was excluded), but not a statistically significant difference, $M = 1.60$, $95\%CI[-2.14, 5.34]$, $t(9) = 0.967, p = 0.359$ with the outlier included, and $M = 0.25$, $95\%CI[-2.30, 2.80]$, $t(7.05) = 0.232, p = 0.823$ when it was excluded.

The analysis presented here have been using the gain from pretest to posttest directly as the dependent variable. A similar analysis was conducted using the normalized gain in order to explain the gain in relation to how much each participant could potentially improve to the posttest. After checking that no assumptions were violated, with the exception of assumption four (no outliers within the groups), we were able to gather the results.

Variable	t	df	Sig.	Mean diff	Std. Error Diff
Pretest	-2.274	9	0.049	-5.23 ($-10.440.. - 0.026$)	2.301
Pretest*	-2.166	8	0.062	-5.58 ($-11.528..0.361$)	2.578
Gain	0.967	9	0.359	1.6 ($-2.143..5.343$)	1.655
Gain*	0.232	7.05	0.823	0.25 ($-2.296..2.796$)	1.078
NormalizedGain	-0.117	9	0.910	-0.03 ($-0.565..0.509$)	0.238
NormalizedGain*	-1.903	8	0.094	-0.26 ($-0.565..0.054$)	0.134

Table 5.8: Results of independent t-test. * outlier removed

5.2   Paired T-test

This analysis was used to determine if a difference between the means in the pretest and the posttest exists for each group, and if these differences are statistically significant. In order to use this analysis, four assumptions regarding study design and the nature of the data must first be confirmed to hold.

Assumptions regarding study design

The first two assumptions when using preparing for the paired t-test are related to the study design. The assumptions are; 1) must have one dependent variable measured at the continous level, and 2) mst have one independent dichotomous variable. These assumptions are the same as the first two assumptions when using the independent-samples t-test, and were shown to hold in section 5.1.

Assumptions regarding the data

The last to assumptions are related to the nature of the data. The assumptions are; 3) there should be no significant outliers in the gain scores, and 4) the distribution of gains in the dependent variable should be approximately normally distributed.

Testing for outliers

This assumption is similar to the fourth assumption in the independent-samples t-test, and as we learned while preforming the that analysis there was an outlier within the robotics group (section 5.1, figure 5.1). In order to see if the outlier affected the results we adopted the same strategy as in the independent-samples t-test and ran the analysis with and without outliers.

Testing for normality

This assumption is similiar to the fifth assumption in the independent-samples t-test. Section 5.1 shows that this assumptions holds with and without the outlier included by presenting z-values (table 5.6), data from Shapiro-Wilk's test (table 5.7), and normal Q-Q plots (figures 5.2, 5.3, and 5.4).

Analysis summary

The paired t-test was used to determine if there were any statistically significant change from pretest to posttest within the two different group. One outlier was identified within the robotics group (figure 5.1), as showed in table 5.9, the exclusion of the outlier did affect the result($p_{robot} = 0.79, p_{robot*} = 0.03$). The test for normality was the same as for the independent-samples t-test and showed that the gain scores within each group were normally distributed. The simulator group did better on its posttest($M = 15.17, SD = 3.43$) then on its pretest($M = 13.17, SD = 4.36$), a statistically non-significant mean increase of 2.0points, $95\%CI[-0.48, 4.48], t(5) = 2.07, p = 0.93, d = 0.84$. The robotics group similarly did better on its posttest($M = 18.80, SD = 4.97$) then on its pretest($M = 18.40, SD = 2.97$), a statistically non-significant mean increase of

0.4points, $95\%CI[-3.49, 4.29], t(4) = 0.286, p = 0.79, d = 0.13$. The results for the robotics group do however drastically change when the outliers were excluded from the analysis, in fact showing a statistically significant mean increase of 1.75points, $95\%CI[0.23, 3.27], t(3) = 3.656, p = 0.035, d = 1.83$.

Group	Paired differences			t	df	Sig.(2-tailed)
	Mean	Std deviation	Std. error mean			
Robot	0.40	3.13	1.40	0.29	4	0.79
Robot*	1.75	0.96	0.48	3.66	3	0.03
Simulator	2.00	2.37	0.97	2.07	5	0.09

Table 5.9: Paired samples T-test of gains. *: outliers removed

Group	Pretest mean	Posttest Mean	Diff. Mean	Diff. SD
Robot	18.40	18.80	0.40	2.37
Robot*	18.75	20.50	1.75	0.96
Simulator	13.17	15.17	2.00	3.13
All	15.54	16.82	1.27	2.72

Table 5.10: Tests means for different groups. *: outliers removed

Group	N	Correlations	Sig.
Robot	5	0.804	0.101
Robot*	4	0.969	0.031
Simulator	6	0.841	0.036
All	11	0.814	0.002

Table 5.11: Paired samples correlations. *: outliers removed

5.3   Additional analyses

The analysis conducted in chapter 5.1 showed that there was no statistically significant difference between the gains of the groups (table 5.5). The analysis did however show that the robotics group had a higher pretest mean score compared to the simulator group(18.40 compared to 13.17), and it is therefore of interest to see if this difference in inital scores affected the study. In order to investigate this we split all the cases into two groups based on their pretest scores, and ran the independent-samples t-test analysis again.

Once again an outlier was identified, the results below include statistics for both the inclusion and exclusion of outliers. The normality assumption was verified to hold through the use of Shapiro-Wilks test and homogeneity of variances was verified with Levene's test. Both assumption held with and without outliers included.

The analysis shows that cases where the pretest score was below 15 have a tendency to achieve a greater increase than those who started with a score above 15(table 5.12 and 5.13). The difference however was not seen as statistically significant (table 5.14).

Group	N	Difference mean	Std. Deviation	Std. Error Mean
$Score >= 15$	6	0.33	2.80	1.15
$Score < 15$	5	2.40	2.41	1.08

Table 5.12: Group statistics for gains when grouping by pretest score, with outliers.

Group	N	Difference mean	Std. Deviation	Std. Error Mean
$Score >= 15$	5	1.40	1.14	0.51
$Score < 15$	5	2.40	2.41	1.08

Table 5.13: Group statistics for gains when grouping by pretest score, without outliers.

Outliers	t	df	Sig	Mean diff.	Std. Error Diff
Yes	-1.295	9	0.228	-2.06	1.596
No	-0.839	8	0.426	-1.00	1.192

Table 5.14: Independent samples test when grouping by pretest score.

5.4 Reliability estimates

Cronbach's alpha

In order to measure the internal consistency reliability of the pretest and posttest we used Cronbach's Alpha as a measurement. The pretest and posttest was divided into six different scales, measuring basic understanding of angles, sizes of inner angles, sizes of reflex angles, sizes of complementary angles and angles in shapes, understanding of the robot turning, and maths involving angles. A list of tasks and the categoriy they affect is listed in table 5.15.

The task indicators, `q1..q24`, refer to the tasks in the pretest and posttest where the participants could score point. The first task, `q1`, would be the second task on the pretest; *draw an angle*. The fourth task on the pretest include five subtasks, these are represented as the task indicators `q3` to `q7`, and so on. The same categories were used for the posttest as well with task indicators from `qq1` to `qq24` to avoid any confusion.

Category	Description	Tasks
Category 1	General angle understanding	q1, q2
Category 2	Normal (inner) angles	q3, q4, q5, q11, q12, q13, q14
Category 3	Reflex (outer) angles	q6, q7, q15, q16, q17, q18
Category 4	Complementary angles	q8, q9, q10
Category 5	Orientation and estimation	q19, q20, q21, q22
Category 6	Calculations with angles	q23, q24

Table 5.15: Tasks and categories

Table 5.16 and 5.17 shows the Cronbach's alpha for eachs of these categories for the pretest and posttest respectively. We've added a column for describing other noteworthy observations during this analysis as well. The cronbach's alpha for the whole pretest was $\alpha = 0.812$, for the posttest $\alpha = 0.860$, and for all tests results $\alpha = 0.908$. All of which were above the 0.7 threshold as recommended by [14] and [24], indicating a high degree of internal consistency. The alpha value when comparing the pretest scores to the posttest scores was $a = 0.897$, indicating a high degree of stability. The comparison of pretest scores to posttest scores are further analysied in section 5.4 where Pearson's product-moment correlations are calculated.

Category	Alpha	Notes
Category 1	*	q1 was excluded because it had zero variance. Could not calculate with just one variable.
Category 2	0.739	q12 showed negative correlation with the rest of the variables. Deleting it would give an alpha value of 0.815
Category 3	0.873	deleting q6 would give an alpha value of 0.877
Category 4	0.703	q9 was excluded because had zero variance
Category 5	0.613	q20 had low correlations with the other variables. Deleting it would give an alpha value of 0.750
Category 6	0.556	
Whole test	0.812	Low *Corrected Item-Total Correlations* for the following variables: q2, q6, q7, q10, q12, q15, q19, q20 and q23

Table 5.16: Cronbach's alpha for pretest

Category	Alpha	Notes
Category 1	*	q1 and q2 was excluded because they had zero variance. Could not calculate with zero variables.
Category 2	*	q3, q4, q5, q11, q12 and q14 was excluded becayse they had zero variance. Could not calulate with just one variable.
Category 3	0.962	Low correlation for q17. Deleting it would give an alpha value of 0.978
Category 4	0.909	
Category 5	0.384	Low *Corrected Item-Total Correlation* for every variable except q21
Category 6	0.714	
Whole test	0.860	Low *Corrected Item-Total Correlation* for q19, q20 and 22

Table 5.17: Cronbach's alpha for posttest

Test-retest correlations

In order to measure the stability reliability we used the Pearson's product-moment correlations between the pretest scores and posttest scores. The initial analysis showed indicated that the relationship to be linear with both variables normally distributed(Shapiro-Wilk, $p > 0.05$). There was a strong positive correlation between the pretest scores and posttest scores ($r(9) = 0.814, p = 0.002$).

Validity and Reliability in the experiment

In the literature section (section 2.5) we provided an introduction to some of the common threats to reliability and validity. This section will try to tie these potensial threats to the experiment conducted during this project.

In the introduction to reliability we mention four key concepts that often are used to describe how reliable an experiment is. The *interrater reliability* is in our case not relevant as it addresses the differences between two researchers rating the participants using qualitative measures. Since this project relies on quantitive data and the only introduction for subjective opinions to contaminate the results was during the scoring process for the pretest and posttest. In order to mitigate this we conducted the scoring process side by side, setting up strict rules for what we would consider a valid answer.

The remaining three concepts, *internal consistency*, *equivalence reliability* and *stability reliability*, are however of interest to this project. For a more detailed description of the estimation process we refer you to chapter 5.4, where the results from the pretest and posttest are analysed using Cronbach's alpha and Pearson's product-moment correlations. The result from these analyses showed a alpha value above 0.8 for the pretest and posttest, above 0.9 for all the results combined, and a strong positive correlation between the pretest scores and posttest scores. These analyses indicate a high degree of internal consistency and stability.

The discussion related to validity is in many ways more complicated then that of reliability as many of the threats to validity can only be verified through being critical towards yourself, and not a scientific analysis.

One of the benefits to validity during this project is the relatively short timespan between the pretest and posttest, which help mitigate several of the threats to validity. This is however a double-edged sword as the short timespan may make some threats more likely to reduce the validity. In our opinion the threats that were most likely to have been interfering with our study are the *testing effect*, *selection bias* and *diffusion between groups*[1]. Some threats that haven't been mentioned during the introduction as *history effects*, *maturation*, etc. have intentionally been overseen as these effects are more likely to happen during longitudinal studies and thus it is very unlikely that they have played a role during this project.

The testing effect, more specifically the learning effect, may however have played a role during experiment as we used the same test for both the pretest and the posttest. The learning effect describes the effect where participants gain knowledge

---

[1]We use this as an umbrella term for when participants interact or are affected by eachother

from completing the pretest, and therefore score higher on a posttest. In order to mitigate this effect there was not given any feedback on the pretest results to the participants and the pretest was given a week in advance of the experiment itself. We grew more confident that this effect had not played a role in our experiment when comparing the pretest results with the posttest results, as it was seen that the participants in some cases would get a question wrong on the posttest where they had previously answered correctly on the pretest. It should however be noted that these results could also be because the experiment confused the students.

Selection bias is also a potensial threat to this study as the groups were put together by the teacher and then randomly assigned to either the robotics group or the simulator group. The analysis in chapter 5.1 showed that the pretest mean score of the robotics group were five points higher on average (table 5.3). The groups were created by the teachers as at the time when the groups were created we were left with very few students, and any random assignment could possibly skew the distribution of highly skilled participants into one group. It was therefore suggested that the teacher, which have an in-depth knowledge about each participant and who they may work well with. The analyses in part 5 (Results) shows that the average gain between the groups were very equal if outliers were pruned from the data set. This may indicate that the gains were equals between the group, but it may also indicate that the selection bias promoted more learning in one group over the other. There is no way of determining the effect of this, we therefore urge the reader to keep this fact in the back of the head.

The last big potensial threat indentified to this experiment was the umbrella term *diffusion between group*. This effect may have manifested itself throughout several different ways, it could be simple interactions between the participants in the groups during the break time or during the weekend between the two sessions. Another possibility is that the participants of one group became more motivated or less excited because of the other group, e.g. the participants of the simulator group may have given the little extra in order to show that they should be allowed to use the robot, or (and perhaps more likely) they may become demoralized because they were in the simulator group and felt that the robotics group was the better alternative. During the sessions we had some indications that some of the participants in the simulator group felt unjustly placed in the simulator group, and thus may not have utilized their full potensial. This was addressed by the teacher with the children after the first sessions, but we are uncertain about how this may have affected the results.

5.5   Observation

The observation during the experiment was performed in order to understand the students' thought process, and how they worked and cooperated during the experiment. Here we report interesting findings.

The students appeared to have more fun with the robot activity than the simulator activity. Some students even asked if they could stay and work with the robots during the break after the first robot session. The teacher mentioned this as well. She said that when they were done they were all excited and told her how fun it had been. Although it is nice that the students are having fun this might pose a problem if it shifts their focus away from the actual learning. One of the groups started playing and doing other activities than the exercises. The same group started filming the robots in the second lecture. After a while the teacher intervened and told them to put the phone away and focus on the tasks. This problem was not observed in the simulator group. The teacher had told the students about the experiment and that it would involve robots before we did the experiment. Some students in the control group could also see the other activities in the first lecture, since we were sharing the classroom with them, separated only by a whiteboard. This could have caused a negative effect in the control group students, because they were expecting robots, but got to work only with the simulator. However in the second lecture when the teacher told them that the simulator group goes first, several students shouted YES!

In the robot group on the first day everyone discussed the problems with each other. No students seemed uninterested or sat by idle. In one group there was one student who understood the concepts very well. This team did not cooperate very well. The good student tried to include the others but even though they controlled the tablet, they did not get to experiment much as he knew the answer to most of the problems. In the second lecture one of the group members seemed quite uninterested. This student was on the group with the student who knew most of the concepts. Most of the students asked for help among their fellow students. Some even asked questions to other groups. They did not specifically ask how they solved the task, but they asked if they had gotten similar errors to what they were facing at that time.

The different groups used different approaches. The choice of approach did not differ much or seem do be dependent on whether they were in the experimental group or the control group. A trial and error approach was present to some degree in most groups. Some groups would create a plan at first by measuring all the angles and programming everything step by step, but when the robot did not create the path they wanted, they tried again or changed the program slightly and

tried again, without reflecting on why it did not work. Many groups created the shape in exercise 3 when attempting to create the shape in exercise 2 in the first lecture. We hoped this would cause some reflection, but many of the students simply checked exercise 3 as done, and then could not figure out how to create the shape from exercise 2. The students who talked it over and measured the angles before programming managed to do more tasks than the teams that did not. After a while students started to used their hands to keep track of which direction the robot would face at a certain point in the program. This is also reflected in the depth questions from the test, where several students wrote that they used their hands to aid estimating the angles.

The students were very engaged and worked the whole duration of the activity. Sometimes the students would get frustrated when they could not find the solution, but then they would return to the task when they had thought a little bit about it, or let someone else on the group use the tablet for a while. During the experiment we overheard many students come to realizations about which way and how much to turn. They would say "Oh, now I see" or "It has to turn 135! not 45!".

Even though some students tried to explain how they got the answer they were lacking the necessary vocabulary to express themselves. They did not describe the robot's movements or their actions with correct mathematical expressions. This is illustrated by quotes like: "We forgot to take it up!", overheard when a student didn't move forward after a turn.

One benefit of the robots physical element became clear when the students had to create alternative programs for the robot, to create the same shape. A physical robot can be placed at an initial starting position and with an initial starting direction at the users choice. With the simulator this was not an option and students in the simulator group therefore thought of 2 identical programs, with the exception of turning right 90 degrees at the start, as different solutions. For example they created `TR FWD TL FWD` as an alternative to `FWD TL FWD`, because the simulated robot moved in a different direction.

The students were allowed to use protractors, but they did not use it the way we intended. The correct way to use the protractor would be to measure the angle between the robot's movement path and the next line in the path. Instead they measured the internal angle of the shape and then had to do calculations to figure out how far to turn the robot.

Regarding the difficulty of the exercises, the closed shape exercises in lecture 2 turned out to be a lot harder for students to manage compared to the open polygons or paths as we call them from lecture one. We thought the first 3 shapes would

be relatively easy since they were regular polygons in which all sides and angles are equal. The program you have to make is merely the same step 3,4 or 5 times. Even though it was hard, the students seemed to grasp the ideas pretty fast. We could have introduced the idea of a repeat block, but that would have taken away from the main activity, and again we did not have that much time.

conclusion   The robots were more fun for the students than the simulator. This might have positive and negative effects, as we observed students working with the robot loosing focus on the tasks. Both the robot and simulator foster collaboration, but if there is a significant skill difference between the students it might be a problem. This student will know the answers before the others, and they wont be able to experiment, fail and reason about these failures. One solution to this problem is tasks requiring multiple group roles. In this way everyone has to do one job and the task can't be done by just one student. A trial and error approach was present in most groups. We wanted the students to reason about their failures and why the robot behaved the way it did. One solution to this problem can be to make them explain their solution as a part of the task. This would however require many teachers per classroom, which is often not realistic. One alternative solution is having students prepare a presentation of the solution and have one random group present their solution and explain it at the end of the session. We unfortunately did not have time to do this. The robot's physical element was shown to be more intuitive in understanding what makes programs different from each other. The students did not use their protractors correctly. They measured the inner angle instead of the supplementary angle which exist between the robots direction vector and the next path in the program. A protractor could have been used in the introduction, but we wanted them to figure it out by themselves. However this appeared to be harder for them than anticipated.

5.6   Depth questions

In addition to the scored items on the test there were a few questions asking how the students thought or how they found the answer. The findings from these questions are presented here. The answers did not vary between experimental and control group students, so all results are presented at the same time here.

In question 1 nobody managed to accurately describe what an angle is on the pretest. The definition of an angle is "A shape, formed by two lines or rays diverging from a common point". Most students answered this question by drawing an arrow pointing to a drawing in question 2 and wrote "this is an example of an angle". However one student was close to the definition. This student wrote:

"an angle is two or more lines that meet into a point". The error being two or more lines when it should be just two lines. In the posttest there was a major improvement in all almost all the replies. 4 students got the question classified as wrong. Everyone else wrote that it is 2 lines that meet or comes together. 3 of the students also mentioned that the two lines must share a common point.

In question 2 everyone drew an angle that were classified as correct in the pretest. In question 3 everyone except one student drew a larger angle in the pretest. However their explanations of why it was bigger were not accurate. Only 3 of the students were classified as correct. They wrote: "this is bigger because its more open", "it is bigger because it's more open" and "it has a bigger opening". In the posttest everyone got 2 points for drawing an angle and drawing a bigger angle. The explanations had improved significantly. 4 students did not provide accurate descriptions of why it was bigger, but the remaining students were classified as correct. Their explanations are grouped by the mentioned properties. 1 student said the lines are longer away, 1 said longer out, 2 said that it had a bigger opening, 1 student said that the angle was wider and the last 2 students explained it by saying the angle in task 2 was $\frac{1}{4}$ of a circle while the angle in task 3 were $\frac{1}{2}$ of a circle.

In tasks 5-10 there were a lot of different answers. We primarily wanted to change the way they think, but most of the explanations remain the same from pretest to posttest. We will focus on the explanations that changed from the pretest to posttest. The biggest changes were found in question 5 and 6. The findings are presented below by quoting what they answered in the pretest and posttest and discussing these results.

In question 5 most students used the same approach to the problem in both the pretest and the posttest. From the students that did change their approach we can see that 2 students started using their hands to aid in the estimation. One student went from a calculation approach to an estimation approach and one student went from an estimation approach to a calculation approach. It is hard to conclude with a positive or negative overall effect when there is no trend across students to what effect the experiment had on their approach. The changes are discussed in detail below.

Pretest: "by estimating and looking at the other angles for example"
Posttest: "45+45 = 90 +90 = 180"
Discussion: This student has gone from an estimation approach to a calculation approach to the problem. The student has realized that there is 180 degrees in a triangle even though we never explicitly mentioned this.

Pretest: "example in the question number 2. I just knew that if I plussed 60+60 degrees it would be 120"
Posttest: "I used my fingers to measure the angles"
Discussion: This student had correct answers to questions 5.2 and 5.3 in the pretest, but they were wrong in the posttest. We can see that the though process was wrong in the pretest. The student got 120 degrees on question 5.2 by adding 60 to 60. Even though the student now got the tasks wrong, the student now has a more correct approach, even though the approach we wanted to teach them was to subtract 60 from 180.

Pretest: "45+45=90"
Posttest: "The two other angles are smaller so I just guessed"
Discussion: This student has gone from a calculation approach to an estimation approach. This resulted in a wrong answer on task 5.1 in the posttest. The student had a correct answer to 5.1 in the pretest.

Pretest: "I looked at the small circles in the corners to see which one is biggest"
Posttest: "I measured with my fingers to see how big"
Discussion: This student used estimation in both pre and post test; however in the posttest the student started using his/her hands.

In question 6 most students, like in question 5, used the same approach. From the students that did change their approach we can see that they predominantly changed into a calculation approach. One student improved his/her estimation by using his/her hands. Overall the experiment have improved the students approach to estimating the angles in question 6. The changes are discussed in detail below.

Pretest: "I looked at which angle fits which number"
Posttest: "I measured with my fingers"
Discussion: This student answered correctly on question 6.4 in the post test. The student answered wrong on this question in the pretest. The rest of the results are the same. Using his/her fingers have improved his/her ability to estimate angles.

Pretest: "I plussed 180 and 180 and got the answer"
Posttest: "I knew that a straight line is 180 so I took about 35 away and found the answer"
Discussion: This student went from a wrong answer on 5.4 in the pretest to full score in the posttest. The approach taken in the posttest is correct, although he/she should subtract 45 degrees, not 35, but it is an estimation task after all. The approach taken in the pretest is wrong and his/her approach has improved during the experiment.

Pretest: "With my eyes"
Posttest: "I plussed 45+45+45 = 135"
Discussion: This student got a full score on both the pretest and the posttest; however the approach taken in the posttest is a calculation approach and considered better than the estimation approach used in the pretest.

Pretest: "I used the other numbers on number 6 to help me. And measuring with stripes with my fingers"
Posttest: "I took and divided it into 90 degrees and saw it was 30 degrees + it together"
Discussion: This student had marked a 90 degree angle on the 135 degree angle and estimated the degree of the remaining angle to be 30 degree, although it was 45. However he/she used a calculation approach on the posttest instead of an estimation approach used in the pretest.

In questions 7-10 there was no change in the students approach. Students still struggle with these tasks, as they did in the pretest. They measure the inner angle and not the outer angle which is marked. We covered what the outside angle is and how it is marked in our introduction, but it is a technical issue that is hard to learn by experimentation. If they did not understand it when we described it, it would be hard to discover this on their own. The only noteworthy change was in question 7, where a student explained his approach with "360-90". In the pretest this student simply pointed an arrow to question 7.2 and wrote that it these angles are the same. In question 8 there was general many wrong statements and explanations. The students did not manage to accurately describe how they thought. We would suggest having conversations or interviews with the students to understand how they think. Only one student had a full score on the post test, but this student had a full score on the pretest as well. The other students struggled with both the amount of rotation and the direction they should rotate. In question 9 most students used a mathematical approach, either writing that 12X30 = 360 or 360/30 = 12.

Question 11 was not scored but included to assess the students' attitude towards educational robotics. We saw an increase in the student attitude on this question. Two students answered blank on both the pretest and the posttest. 5 students answered "yes" on both pretest and posttest. One student answered "sometimes" on both pretest and posttest. The last three students changed their attitude positively. They went from "no", "maybe" and "I don't know" to "yes", "yes, we have" and "I think so".

5.7   Questionaire

Overall the questionaire did not yield very significant results. There was about the same spread in both robot and simulation group. Surprisingly there was a difference in question 17: Do you think imagining a robot turning can help you when working with angles? For the robot group 2 answered 0 and 2 answered + giving a total score of 2. In the simulator group 3 answered ++ and 1 answered + giving a total score of 7. This is the main factor we thought would be prevalent in the robot group. As seen in figure 5.5 there was a slightly higher score for the robot group in the questions regarding interest, motivation and entertainment value when summed together. Question 15 and 16 scored the same for both groups.
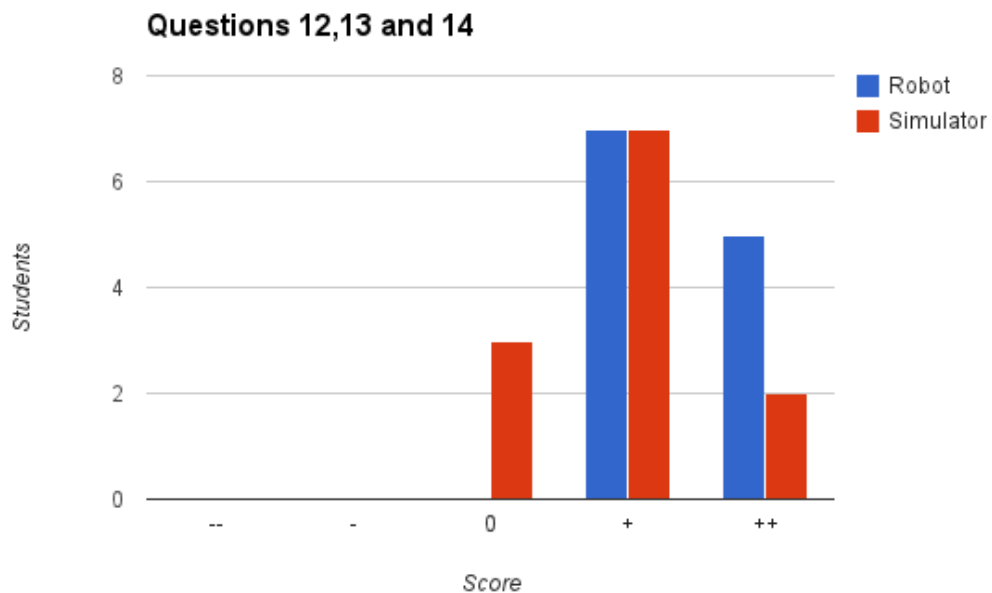
Figure 5.5: Histogram of answers in questions 12,13 and 14

DISCUSSION, CONCLUSION AND FUTURE WORK

6.1 Discussion

In this chapter we try to answer our research questions based on the insights gained through the experiment.

How can using robotics in school help students understand what an angle is and aid in their angle estimation skills?
Angles and mathematics in general can be taught with robotics by using the robot as an object to think with, and create mental models around. It is also an excellent object to use in an experimental classroom design, as the answer is not straight forward and experimentation is needed to create most programs. Through our experiments we found that both groups using robots and simulators performed better on the posttest than on the pretest. However these results were statistically insignificant. When an outlier from the robotics group were excluded it revealed a statistically significant mean increase.

Both the robot and simulator fostered collaboration. We encountered a problem during the experimentation phase. If students in one group had different levels of understanding, the experimental classroom design failed, because one student dominated the group and had all the answers. The rest of the team members then did not get to experiment, fail and reason about these failures. More challenging exercises might be needed for these students, but it is hard to make customized tasks to different students in the same class, or in the same group. A different solution might be to force group roles as in Mitnik et al. [2009] [30], to make the experiments interesting for everyone in the group, even though some students

understand the concepts better.

Regarding problem solving, a trial and error approach was present to some degree in all groups and the students used their protractors wrong. The activities we presented did not alter the students' approach. We found that overall students need clear instructions and should be presented with opportunities to explain how they solved their tasks, or they might just apply a trial and error approach. However, asking every group to explain their solution is a time consuming process, and hard for teachers while at the same time maintaining classroom control. One possible solution is to ask the students to write down an explanation of their program and choose a random group to present this explanation at the end of the lesson.

Both the robots and simulators were fun to work with for the students. Teaching is easier when students are having fun and are engaged.

Through the depth questions it became clear that the activities helped students understand and describe what an angle is, as everyone managed to describe an angle accurately in the posttest.

Our results are consistent with previous research presented in section 2.2. Robotics foster teamwork, motivation, and give students an increased sense of how robotics can utilize mathematics. We could not change the students' problem solving approach. This is a hard process and a clear step by step approach should possibly be explained to the students before letting them experiment on the robots.

What are the strength and weaknesses of using a robot compared to a simulated environment?

Overall there was a big difference between using robots and using a simulator. The increase from pretest to posttest was greater for the simulator group than for the robot group, but not a statistically significant difference.

Through the questionnaire we found that the robotics group though the activities were more fun and interesting. This was also supported by our observations during the experiment. We also found that the students in the simulator scored higher on the question: "Do you think imagining a robot turning can help you when working with angles?". This result was surprising, as we thought the robotics group would score higher on this question.

Robots appear to be more fun, but this is not necessarily a good thing. Students were more preoccupied and did not listen to everything we said. In addition some students did not focus on the tasks given to them. Thus maybe hindering their learning gains from the experience.

When using a physical robot students did not differentiate between two identical programs with the exception of a turn left or right at the start, as they could have done that themselves by putting the robot in that direction before starting the program. In the simulator group this was seen as two different approaches, even though they could have rotated the tablet in the same manner, this was not obvious to them as it was in the robotics group. This attests to the positive benefits of using a physical robot instead of a simulated environment.

6.2 Conclusion

This project serves as an introduction into the educational world for the ChIRP team at IDI. We performed a literature review to understand the educational robotics state of the art with a focus on mathematics. This also serves as an introduction to the pedagogical aspects, and the research methodologies needed when conducting educational research. We have created a test to measure knowledge in the area of angle estimation, calculation and understanding. The test was found to be reliable and valid. To perform these experiment we created chassis extensions for the robot and a bluetooth module platform which can be used for several other components as well. We have created a tablet application to control the robot with, but also serves as a simulator. This application was created with a focus on expandability and cross platform functionality for use in future research.

We investigated how a robot can be used to teach mathematics to school students. We have also found strengths and weaknesses of using a robot compared to a simulator. The robotic activities did help students content knowledge about angles, and did aid in their understanding and ability to explain what an angle is. However the difference from a simulator was not statistically significant.

Through this project we have also established a relationship between IDI and the international school in Trondheim which can help other researchers when expanding on our project. In addition our project serves as an introduction in itself for researchers wanting to look into educational robotics.

6.3 Future work

The main lacking area in research at this moment is a large scale study over a long time period. We also propose a project testing either the robotics activities or simulator activities versus a regular school curriculum over a longer time period. The student tasks in another experiment should take into consideration the problems we have encountered. The student tasks could be created with more complex

goals, maybe something requiring a whole lesson to solve. Students should be put in a position where they have to explain their approach and reasoning about their solutions, and specific group roles should be created so that one student cannot dominate the experimentation experience of all the other students. For future projects more of the ChIRP functionality should be used, including its distance sensor functionality and possibly create new extensions for use in schools. For such projects the chassis and application might need further improvements, and we would suggest altering the entire chassis for such purposes, as the process of adapting extensions for the existing chassis proved difficult.

BIBLIOGRAPHY

[1] Ackermann, E. (2001). Piaget's constructivism, papert's constructionism: What's the difference. *Future of learning group publication*, 5(3):438.

[2] Alseth, B., Breiteig, T., and Brekke, G. (2003). Endringer og utvikling ved r97 som bakgrunn for videre planlegging og justering-matematikkfaget som kasus.

[3] Andersen, F. Ø. and Frost, J. (2009). *Evaluering-i et dialogisk perspektiv*. Cappelen Akademisk Forlag.

[4] Barker, B. S. and Ansorge, J. (2007). Robotics as means to increase achievement scores in an informal learning environment. *Journal of Research on Technology in Education*, 39(3).

[5] Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3):978 – 988.

[6] Bodner, G. M. (1986). Constructivism: A theory of knowledge. *Journal of chemical education*, 63(10):873.

[7] Cakir, M. (2008). Constructivist approaches to learning in science and their implications for science pedagogy: A literature review. *International journal of environmental and science education*, 3(4):193–206.

[8] ChIRP (2013). Chirp demo ntnu researchers night.

[9] Citilab (2013). S4a - scratch for arduino.

[10] Clements, D. H. and Battista, M. T. (1990). The effects of logo on children's conceptualizations of angle and polygons. *Journal for Research in Mathematics Education*, pages 356–371.

[11] Clements, D. H., Battista, M. T., and Sarama, J. (2001). Logo and geometry. *Journal for Research in Mathematics Education. Monograph*, pages i–177.

[12] Clements, D. H., Battista, M. T., Sarama, J., and Swaminathan, S. (1996). Development of turn and turn measurement concepts in a computer-based instructional unit. *Educational Studies in Mathematics*, 30(4):313–337.

[13] Clements, D. H. and Meredith, J. S. (1993). Research on logo: Effects and efficacy. *Journal of Computing in Childhood Education*, 4:263–90.

[14] Devellis, R. F. (2003). Scale development: Theory and applications second edition (applied social research methods).

[15] Fernandes, E., Fermé, E., and Oliveira, R. (2006). Using robots to learn functions in math class. In *Proceedings of the ICMI 17 Study Conference: background papers for the ICMI*, volume 17.

[16] Foundation, L. (2011). Logo web page.

[17] Gwet, K. (2001). Handbook of inter-rater reliability. *Gaithersburg, MD: STATAXIS Publishing Company*, pages 223–246.

[18] Howell, J., Miller, P., Park, H. H., Sattler, D., Schack, T., Spery, E., Widhalm, S., and Palmquist, M. (2012). Reliability and validity.

[19] Hsiu, T., Richards, S., Bhave, A., Perez-Bergquist, A., and Nourbakhsh, I. (2003). Designing a low-cost, expressive educational robot. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 3, pages 2404–2409. IEEE.

[20] Hussain, S., Lindh, J., and Shukur, G. (2006). The effect of lego training on pupils' school performance in mathematics, problem solving ability and attitude: Swedish data. *Journal of Educational Technology & Society*, 9(3).

[21] Kara, D. (2003). Sizing and seizing the robotics opportunity. *Presentation in RT Los Angeles by Robotics Trends, USA*.

[22] Khan, K. S., Ter Riet, G., Glanville, J., Sowden, A. J., Kleijnen, J., et al. (2001). *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews*. Number 4 (2nd Edition). NHS Centre for Reviews and Dissemination.

[23] Kitchenham, B. A. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.

[24] Kline, R. B. (2005). Principles and practice of structural equation modeling.

[25] Lindh, J. and Holgersson, T. (2007). Does lego training stimulate pupils ability to solve logical problems? *Computers & education*, 49(4):1097–1111.

[26] Ltd, L. R. (2012a). Reliability in research.

[27] Ltd, L. R. (2012b). Research quality.

[28] Miller, D. P., Nourbakhsh, I. R., and Siegwart, R. (2008). Robots for education. *Springer handbook of robotics*, pages 1283–1301.

[29] Mitnik, R., Nussbaum, M., and Soto, A. (2008). An autonomous educational mobile robot mediator. *Autonomous Robots*, 25(4):367–382.

[30] Mitnik, R., Recabarren, M., Nussbaum, M., and Soto, A. (2009). Collaborative robotic instruction: A graph teaching experience. *Computers & Education*, 53(2):330–342.

[31] Mozilla Developer Network (2014). ES6 Generators.

[32] Norton, S. J. (2004). Using lego construction to develop ratio understanding. *Mathematics education for the third millennium: Towards 2010*, pages 414–421.

[33] Nugent, G., Barker, B., and Grandgenett, N. (2008). The effect of 4-h robotics and geospatial technologies on science, technology, engineering, and mathematics learning and attitudes. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, volume 2008, pages 447–452.

[34] Nugent, G., Barker, B., Grandgenett, N., and Adamchuk, V. (2009). The use of digital manipulatives in k-12: robotics, gps/gis and programming. In *Frontiers in Education Conference, 2009. FIE'09. 39th IEEE*, pages 1–6. IEEE.

[35] OECD (2000). Database - pisa 2000.

[36] OECD (2012a). Database - pisa 2012.

[37] OECD (2012b). Results from pisa 2012, key finding for norway.

[38] Owens, G., Granader, Y., Humphrey, A., and Baron-Cohen, S. (2008). Lego® therapy and the social use of language programme: an evaluation of two social skills interventions for children with high functioning autism and asperger syndrome. *Journal of autism and developmental disorders*, 38(10):1944–1957.

[39] Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas.* Basic Books, Inc.

[40] Papert, S. (1989). Constructionism: A new opportunity for elementary science education.

[41] Ringdal, K. (2001). Enhet og mangfold. *Bergen: Fagbokforlaget.*

[42] Rusk, N., Resnick, M., Berg, R., and Pezalla-Granlund, M. (2008). New pathways into robotics: Strategies for broadening participation. *Journal of Science Education and Technology*, 17(1):59–69.

[43] Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86(2):420.

[44] Silk, E. M. (2011). Resources for learning robots: Environments and framings connecting math in robotics.

[45] Thagaard, T. (2003). Systematikk og innlevelse: en innføring i kvalitativ metode.

[46] Tjora, A. (2012). *Kvalitative forskningsmetoder i praksis.* Gyldendal Akademisk.

[47] VisionMobile (2013). Developer economics 2013: The tools report.

[48] Williams, D. C., Ma, Y., Prejean, L., Ford, M. J., and Lai, G. (2007). Acquisition of physics content knowledge and scientific inquiry skills in a robotics summer camp. *Journal of Research on Technology in Education*, 40(2).

Appendices

APPENDIX A

PRETEST

NAME:_____

1. What is an angle? Try to explain with words.

2. Draw an angle

3. Draw a bigger angle. Explain why it is bigger.

4. Which angle is bigger in each pair? Circle the one you think is biggest.

5. How big is the missing angle measurement? Write the answer on the line.

60°

90°

?=_____

?=_____

60°

45°   45°

?=_____

How did you find the answer?

_____

_____

6. Estimate the angles

☐   30°
☐   60°
☐   90°
☐   120°
☐   180°
☐   270°
☐   330°

☐   30°
☐   60°
☐   90°
☐   120°
☐   180°
☐   270°
☐   330°

Write an estimate

——————

Write an estimate

——————

How did you estimate this angle?

_____

_____

7. Estimate the angles

- ☐ 30°
- ☐ 60°
- ☐ 90°
- ☐ 120°
- ☐ 180°
- ☐ 270°
- ☐ 330°

- ☐ 30°
- ☐ 60°
- ☐ 90°
- ☐ 120°
- ☐ 180°
- ☐ 270°
- ☐ 330°

Write an estimate

————

Write an estimate

————

How did you estimate this angle?

_____

_____

8. How many degrees and in which direction (right or left) should you turn the spinner arrow to aim it directly at the center of the black dot? Circle the direction you choose and write how many degrees on the line.

LEFT / RIGHT

How many degrees?

_____

LEFT / RIGHT

How many degrees?

_____

LEFT / RIGHT

How many degrees?

_____

LEFT / RIGHT

How many degrees?

_____

How did you think when you solved these tasks?

_____

_____

9. A robot turns 30 degrees every time it turns. How many turns must the robot make before it is facing in the same direction as it was when it started? How did you get your answer?

10. A robot turns 40 degrees left, then it turns 20 degrees right and then 30 degrees left. How far and in what direction has the robot turned in total from where it started? How did you get your answer?

11. Do you think you can learn about angles and degrees using robotics?

APPENDIX B ———————————————————————

POSTTEST

NAME:_____

1. What is an angle? Try to explain with words.

2. Draw an angle

3. Draw a bigger angle. Explain why it is bigger.

4. Which angle is bigger in each pair? Circle the one you think is biggest.

5. How big is the missing angle measurement? Write the answer on the line.

60°

90°

?=_____

?=_____

60°

?=_____

45°   45°

How did you find the answer?

_____

_____

6. Estimate the angles

□   30°
□   60°
□   90°
□   120°
□   180°
□   270°
□   330°

□   30°
□   60°
□   90°
□   120°
□   180°
□   270°
□   330°

Write an estimate

_____

Write an estimate

_____

How did you estimate this angle?

_____

_____

7. Estimate the angles

☐ 30°
☐ 60°
☐ 90°
☐ 120°
☐ 180°
☐ 270°
☐ 330°

☐ 30°
☐ 60°
☐ 90°
☐ 120°
☐ 180°
☐ 270°
☐ 330°

Write an estimate

————

Write an estimate

————

How did you estimate this angle?

8. How many degrees and in which direction (right or left) should you turn the spinner arrow to aim it directly at the center of the black dot? Circle the direction you choose and write how many degrees on the line.

LEFT / RIGHT

How many degrees?

_____

LEFT / RIGHT

How many degrees?

_____

LEFT / RIGHT

How many degrees?

_____

LEFT / RIGHT

How many degrees?

_____

How did you think when you solved these tasks?

_____

_____

9. A robot turns 30 degrees every time it turns. How many turns must the robot make before it is facing in the same direction as it was when it started? How did you get your answer?

10. A robot turns 40 degrees left, then it turns 20 degrees right and then 30 degrees left. How far and in what direction has the robot turned in total from where it started? How did you get your answer?

11. Do you think you can learn about angles and degrees using robotics?

In the questions below mark the box that matches how you feel.
- - means very little.
+ + means very much.

12. How interesting was the activity?
    ☐ − −
    ☐ −
    ☐ 0
    ☐ +
    ☐ + +

13. How motivated did you feel during the activity sessions?
    ☐ − −
    ☐ −
    ☐ 0
    ☐ +
    ☐ + +

14. How entertaining was the activity?
    ☐ − −
    ☐ −
    ☐ 0
    ☐ +
    ☐ + +

15. Do you think robots can make mathematics more interesting?
    ☐ − −
    ☐ −
    ☐ 0
    ☐ +
    ☐ + +

16. Throughout the activity, how much did you cooperate with your teammates?
    ☐ − −
    ☐ −
    ☐ 0
    ☐ +
    ☐ + +

17. Do you think imagining a robot turning can help you when working with angles?
    ☐ − −
    ☐ −
    ☐ 0
    ☐ +
    ☐ + +

APPENDIX C

EXPERIMENT 1 TASKS

Use the robots to draw these paths

1.

2. $135°$

3.
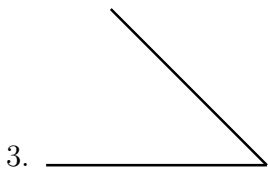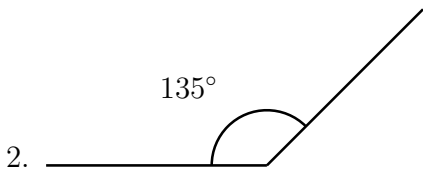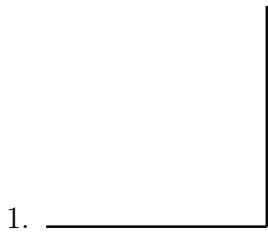
4. Create the paths from task 1-3 again, but use a different program than you did last time. (Hint: you can use the reflex angle)

5.

6.

7.

8.

9.

10.
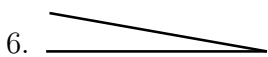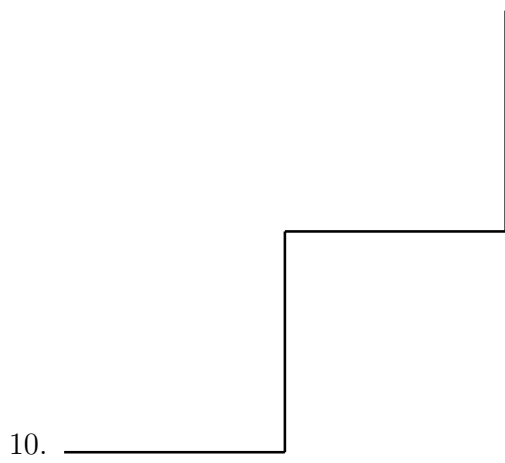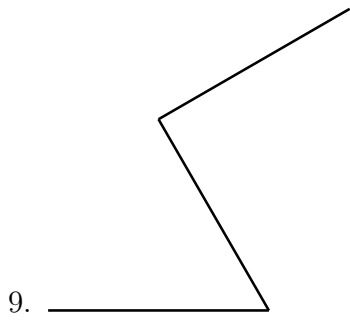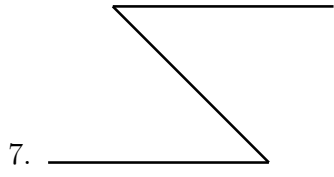
11.
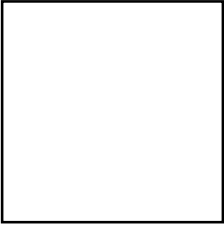
APPENDIX D
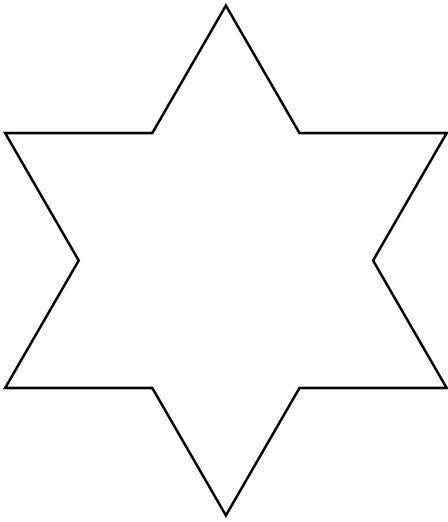
EXPERIMENT 2 TASKS

Use the robots to draw these shapes

1.

2.

3.

4.

5.

6.

7.

8.
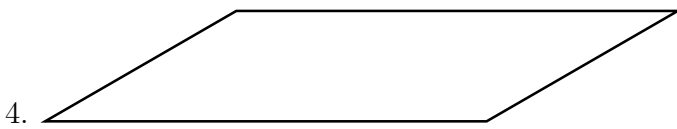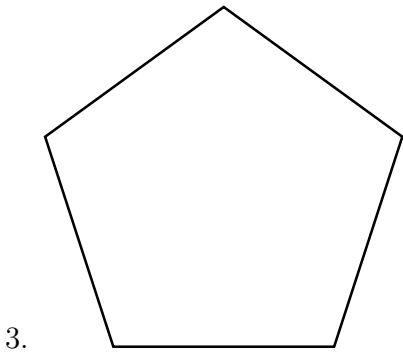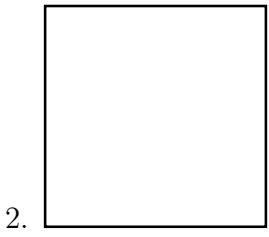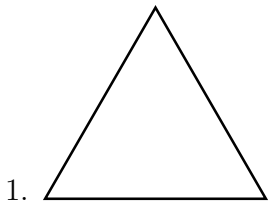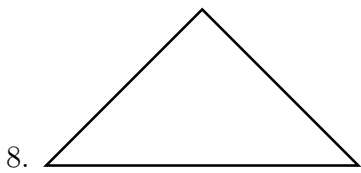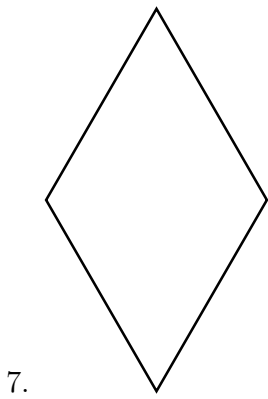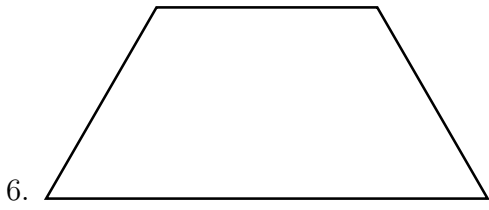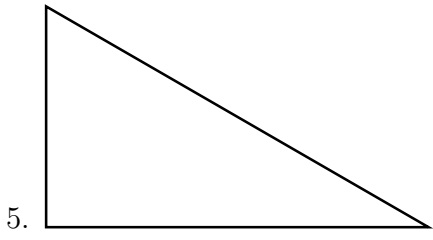
9.
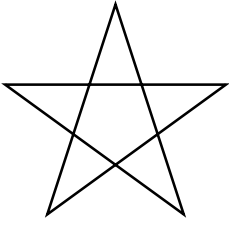
10.

APPENDIX E

APPLICATION INTRODUCTION

TurtleBot

1   The TurtleBot application

At its very core is the TurtleBot application a combined integrated development environment and simulator, enabling children to explore the world of robotics and programming. The application was created with extendability in mind, having two distinct methods for creating programs, the possibility of extending the existing block language, several programming languages, and different visual setups depending on your preference and needs.

1.1   Basic screen layout

The application is divided into three major components; *header*, *body* and *footer*(figure 1).

In order to make the application easy to use by people unfamiliar to programming and robotics it utilizes a block programming interface as its default input method. With this input method the user can drag *codeblocks* from the header, down to the body/programming area. After the blocks has been placed in the programming area they may be moved around to change the program or deleted by dragging them back to the header.

As seen in figure 1 the application starts with four codeblocks the first time. These blocks are the *move-blocks* in the application. Everytime one of these blocks are executed by the simulator it will move the robot around.

In order for you to get more familiar with the layout we have highlighted the different components, and provided a simple description of them below figure 1.

Figure 1: Graphics view.
NB: button 6 has been moved to the left side

1 This is the available *codeblocks* for your *program*.

2 The *startblock*. This is a special kind of *codeblock*. It may not be moved or removed, and must always be the first *codeblock* in the *program*.

3 The *program*. This part consists of one of more *codeblocks*, and are the part which tells the robot what to do. An explaination of the different codeblocks can be seen in section 1.3.1 and section 1.3.2.

4 The *Options* menu.

5 The *Run* button. Pressing this button run the *program*.

6 The *Clear* button. This button will remove all blocks except the *startblock*.

7 The *Input mode* button. This button will change the input view. If you are working in the *graphics view* it will change to the *textual view*, and vise versa.

1.2 Views

The header/programming area in the application has three different views, which all serve a different purpose. The view shown in figure 1 is the *graphics view*, here programming is done by drag'n'drop as explained. The second view is the *textual view*, as seen in figure 2. The textual view

provides the textual counterpart to the graphics view, and when used with block programming you are able to switch between the two without losing your program. The view also supports an alternative programming language for more experienced users. The textual view is considered harder to understand, and it is therefore recommended that most people start with the graphics view. The final view is the *simulator view*. This view takes on many different visual appearances, depending on the state of the application and different options. In its most basic form it will look like figure 5, but depending on what has been done previously in the application it may have changed. A list of all the different simulator appearances can be seen in section 1.4.
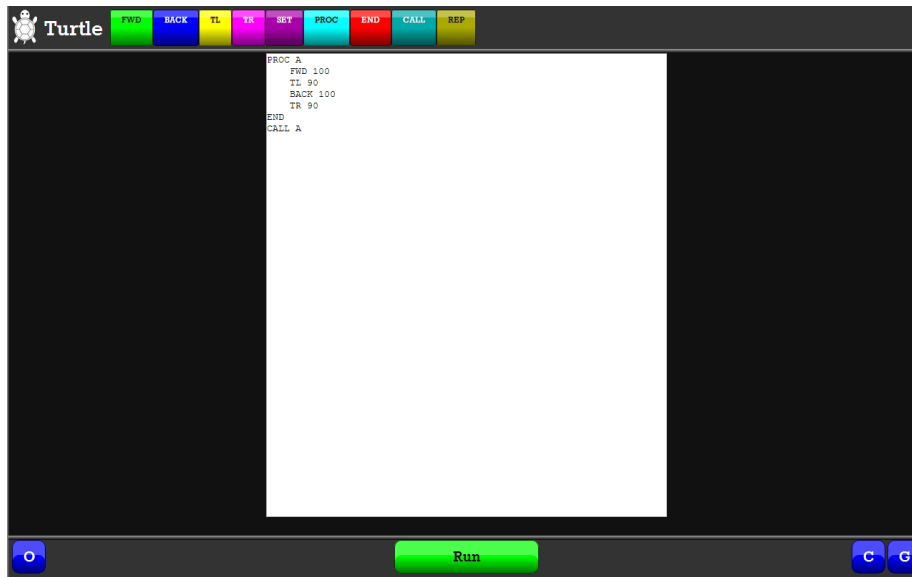

Figure 2: The textual view when extended mode is on.

1.3   Programming

1.3.1   Normal mode

When starting the application for the first time, it will start in what is called *normal mode* showing the *graphics view*. In *normal mode* you have four *codeblocks* available to you:

<div align="center">

FWD - Forward   **BACK - Backward**
TL - Turn Left   **TR - Turn Right**

</div>

FWD and **BACK** makes the robot move forward and backward respectively. The input value determines how far the robot will go. **TL** and **TR** makes the robot turn either left of right. The input value determines how many degrees it will turn.

In figure 1 you can see an example where all these four *codeblocks* are used. The resulting picture from this *program* can be seen in figure 3. For examples of textual program please see section 1.6.
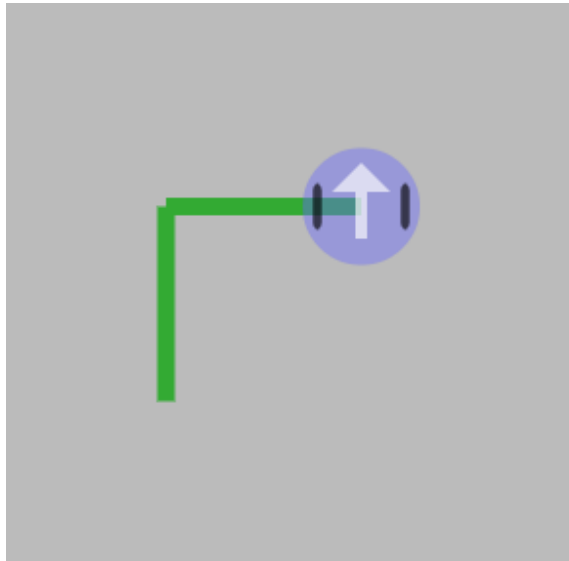
Figure 3: The result from running the program in figure 1

1.3.2 Extended mode

In *extended mode* you will see five new *codeblocks* appear as available (figure 4). These blocks are related to programs that need *variables*, *procedures* and *loops*. The usage of these blocks and how they work is beyond the scope of this introduction, but it should be easy to figure out for anyone with previous experience with programming.

Figure 4: The main view when extended mode is on.

1.3.3 Javascript mode

The javascript mode is a purely textual mode, and only recommended for those who have mastered using the extended mode in textual form, or have previous knowledge about programming. This mode utilized an architectural artifact within the application to allow for writing pure and real

javascript code directly into the application. To use the javascript mode; change your view to the textual view and add `//#javascript` as the first line and you are good to go.

1.4   The simulator

As mentioned previously the simulator has many different appearances which depends on the current state of the application. Here is a list of the current simulator modes:

1.4.1   Always simulator

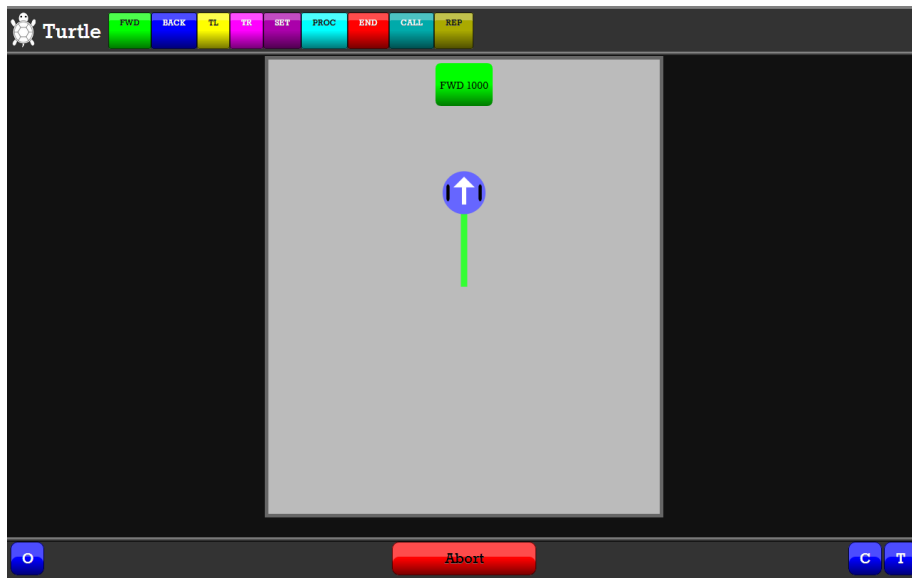With the *always simulator* mode the simulator is shown regardless of the state of your application.

Figure 5: The simulator view.

1.4.2   Bluetooth adaptive headsup display

The *bluetooth adaptive headsup display* changes its appearances depending on wether the application is connected to a robot via bluetooth or not. In the case where it is not connected to a robot it will show the default simulator view, as seen in figure 5. When the application is connected it will just show the current task of the robot, as seen in figure 6

Figure 6: The bluetooth headsup display.

1.4.3 Bluetooth adaptive blackout

The *bluetooth adaptive blackout display* has similar characteristics to the bluetooth adaptive heasdup display, but even removes the heasdup display. While the program is runnin all focus will be at the robot as the screen will at this point not show any useful information.
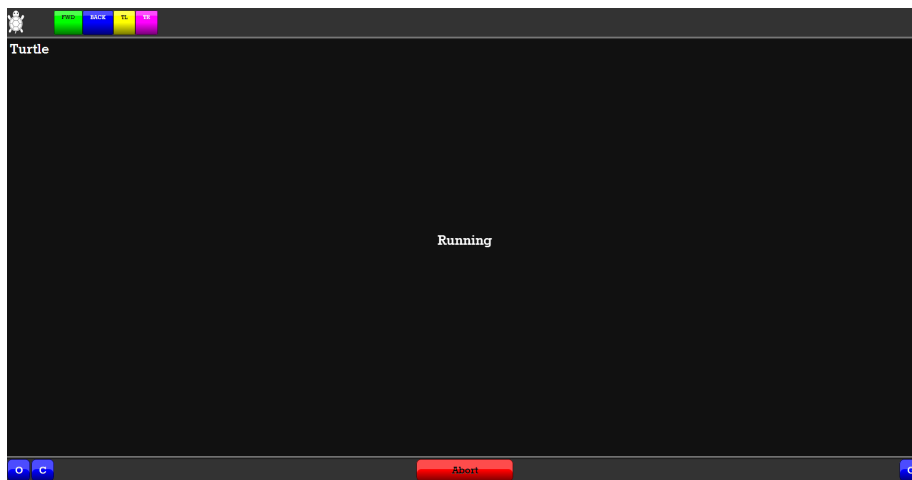

Figure 7: The bluetooth blackout display.

1.4.4 Just trace simulator

This mode is does not send any information to the robot even if it is connected. Instead this mode can be used for rapid debugging of your program. This mode skips all fancy animations that other modes may have, and draws the a trace of the robots movement as fast as it can. The end image will be very similar to those of the normal simulator.

1.5 Options

The *options panel* allows you to change the behaviour of the application at runtime, change the inital setup, and connect to an external robot. If the yellow *Connect* button does not appear then

check the bluetooth settings on your device, as this button only appears if the application detects a device with bluetooth.
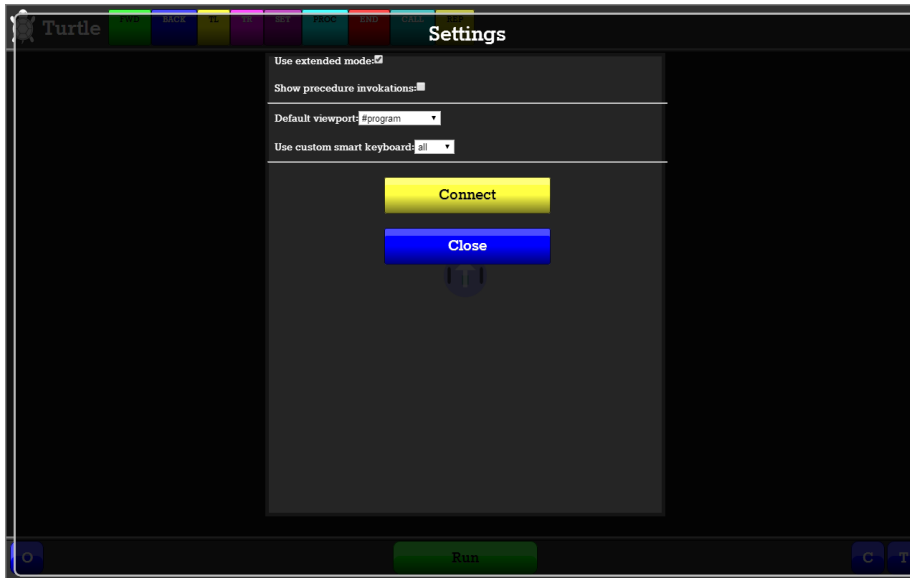

Figure 8: The options view.

The following attributes are exposed through the options menu:

Attribute	Description	Default
Use extended mode	Allows the user to toggle between *normal* and *extended* mode.	unchecked(normal mode)
Show procedure invokations	Determines wether or not to show the blocks from the extended mode during simulation.	unchecked(dont show)
Default view	Allows the user to set the *textual view* as default.	#program(graphics view)
Smart keyboard	Turns on different modes for the on screen keyboard.	none(dont show)
Simulator selection	Allows the user to change which simulator mode that are used.	bluetooth adaptive headsup

1.5.1 Connecting to a robot

To connect to an external robot; click on the *Connect* button and wait for a list of robots to show up on the screen(figure 9), click on the desired robot and wait for the connection to happen. If everything works fine the *Connect* button should turn green, this is an indication that the application is in contact with the robot(figure 10).
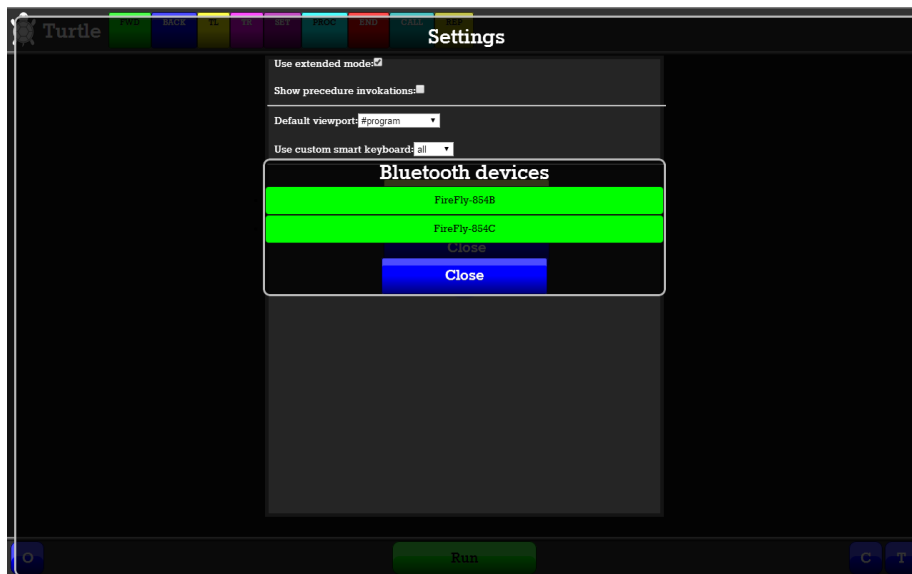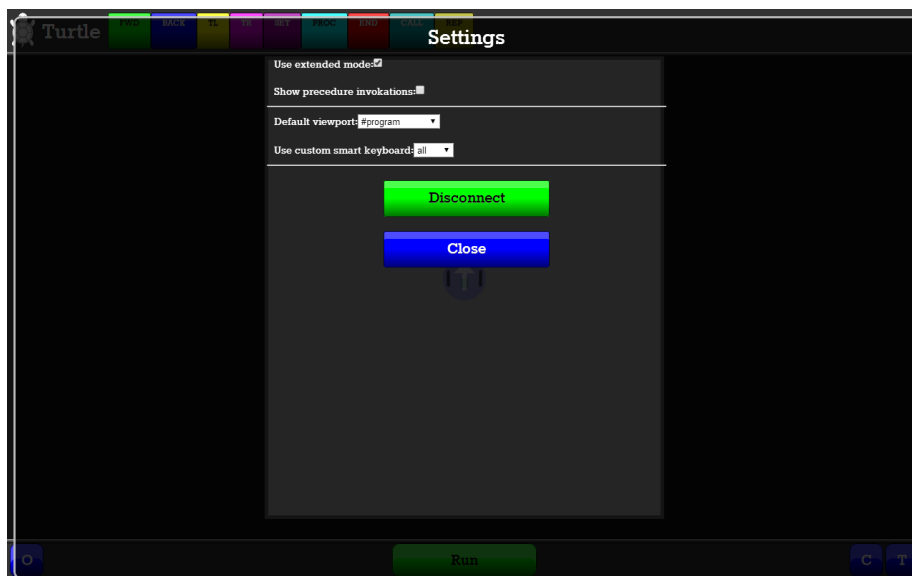
Figure 9: The list of available robots.

Figure 10: The options view when connected to a robot.

1.5.2 On screen keyboard

This is fallback solution, if the native keyboard doesn't work correctly.

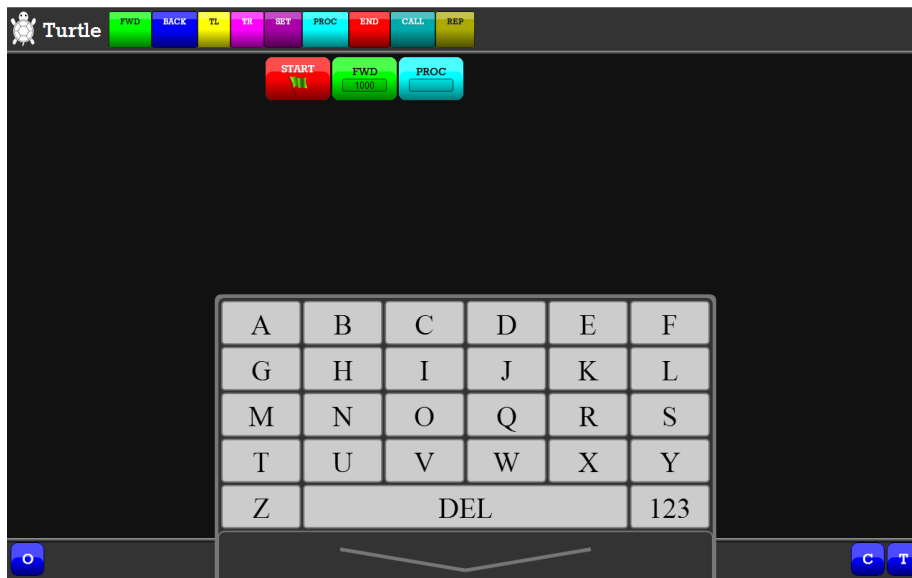Figure 11: The on screen keyboard for numeric input.

Figure 12: The on screen keyboard for alphanumeric input.

1.6   Textual code examples

All the example below will draw a square with side length of 100, and return to the original starting position and facing the same way as when it started.

1.6.1 Normal mode

```
1  FWD 100
2  TL 90
3  FWD 100
4  TL 90
5  FWD 100
6  TL 90
7  FWD 100
8  TL 90
```

Listing 1: Normal mode.

1.6.2 Extended mode

```
1  REP 4
2    FWD 100
3    TL 90
4  END
```

Listing 2: Extended mode using loops.

```
1  PROC A
2    FWD 100
3    TL 90
4  END
5  REP 4
6    CALL A
7  END
```

Listing 3: Extended mode using procedures and loops.

```
1   SET NUM 4
2   SET LENGTH 100
3   SET ANGLE 90
4   PROC SEGMENT
5       FWD LENGTH
6       TL ANGLE
7   END
8   REP NUM
9       CALL SEGMENT
10  END
```

Listing 4: Extended mode using procedures loops and variables.

```
1  var icount = 0;
2
3  SET("NUM", 4);var NUM = 4;
4  SET("LENGTH", 100);var LENGTH = 100;
5  SET("ANGLE", 90);var ANGLE = 90;
6
7  function SEGMENT() {
8     icount++;
9     if (icount > 30000){return;}
10    FWD(LENGTH);
11
12    icount++;
13    if (icount > 30000){return;}
14    TL(ANGLE);
15    END();
16 }
17
18 REP(NUM);
19 for (var generated_55033 = 0; generated_55033 < NUM;generated_55033++) {
20    CALL("SEGMENT");
21    SEGMENT();
22    END();
23 }
```

Listing 5: The internal representation of the previous program.

1.6.3 Javascript mode

```
1  //#javascript
2  var num = 4;
3  var length = 100;
4  var angle = 90;
5  function segment(){
6     FWD(length);
7     TL(angle);
8  }
9  for (var i = 0; i < num; i++){
10    segment();
11 }
```

Listing 6: Javascript mode