UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering, Mathematics & Computer Science

AI as the assistant of the teacher: an adaptive math application for primary schools.

Lotte J. van Dijk

M.Sc. Thesis

October 2021

Supervisors:
dr. ir. M. van Keulen
prof. dr. K. Schildkamp
dr. S. Wang

Computer Science &
Interaction Technology
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Preface

I started my academic career in the education field by studying Educational Science combined with primary school teaching. I soon realized that there is still much work to do when it comes to technical innovation in the educational field. I therefore made it my personal goal to bring technical innovation to the educational field by studying the interaction between Technology and Computer Science. Coming from social sciences it was quite a challenge to pick up two technical masters, however with hard work and a passion for technology I pulled through. From the beginning of my master's, I knew I wanted to develop an application for the educational field as my thesis subject. Luckily I found my supervisor Maurice, who shares my enthusiasm for this topic. From my experiences as a teacher I noticed that there was a need for adaptive materials. Not materials that work with completing levels or performing on easy normal or hard, but an application that truly adapts towards the level of the user. With that in mind I started this adventure. I learned a lot about prototyping, designing, programming, and writing. For the designing part I sat together with Chris Vermaas, who has challenged me in many ways. Thank you for all your enthusiasm and critical notes.

I'm truly grateful that I have been given the opportunity to work on this topic for so long with the support of my supervisors, Maurice van Keulen, Kim Schildkamp and Shenghui Wang.

I hope that I can continue working on this project long after I graduate, so it can be used by the public and help students at their own level.

Summary

In this thesis the question, ”how to design and develop an AI assistant that adapts towards the zone of proximal development of the students, while reducing the workload of the teachers“, is answered. This is done by looking at the design characteristics of such a system. To answer this research question, a problem analysis has been done. It focuses on various complex fields such as the classroom, teachers, children, and learning, supported with literature research. Three tasks of the teacher are presented herein: administration, analysis and content creation. The possibilities of an AI helping with these tasks will be presented leading to a global design for an adaptive application in primary schools and a prototype for an adaptive math application. The primary focus of these designs lies on the possibility to stay within the zone of proximal development. This is done by developing an adaptive algorithm that increases and decreases the difficulty level without interference of the teacher. Therefore, the student can always practice math exercises at the level they are currently at. Next to the adaptive behavior, the design also shows the advantages for the teacher as the teacher gets more insight in the level and the progress of the children. The results are analysed by the application and the teacher gets an interface with notifications and an overview of the progress of the individual, the class or per subject. The adaptive algorithm is tested by simulations and shows the envisioned behaviour. It is able to find the level of the student and adapts based on the interaction. There are different parameters that can be changed in order to change the speed of the adaptation. This can be useful as children are learning at different speeds. The algorithm itself is generic and can be implemented with different subjects, as long as the subjects have exercises that can be structured in a hierarchical way. Overall, the global design and the prototype shows a possible answer to the research question. The next step would be to conduct a user study in a classroom with the prototype.

Contents

Chapter 1

Introduction

Are computers taking over the world? To what extent are we dependent on computers? Almost everyone carries their small personal computer 'smartphone' with them anywhere they go. And when you step into your car, there is a computer that makes sure you have a safe journey. It is not always visible, think about a grocery store and all the computer systems, like payment and logistics, just to make sure that you can fulfil one of your basic needs: food. We are surrounded by computers in our daily lives and are increasingly more dependent on them.

The increasing dependency on computers is also visible in primary schools. The administration is done on the computer with the help of a student tracking system, sometimes in combination with a partially paper administration. In addition, most children work quite often on a computer to practice skills or create content and the digiboard brings the online world into the classroom.

The general idea is that computers are not taking over the classroom, but they are there for us to make our lives easier. But how is it possible that teachers in the Netherlands are feeling an increase in workload [1]? Is the AI not implemented correctly to lower their workload? What if the AI would truly be the assistant of the teacher? What would that look like?

First of all, a teacher needs to keep a good administration of the progress of children. The administration has become more complex with the introduction of student tracking systems. Sometimes it looks like that if we can track something, we have to track it. Most of the input for the tracking systems originates from paper assignments and observations, resulting in administration tasks where the teacher has to analyse and type the results into the computer by hand. Except for some standardized tests, there is no automation involved. Even when children are practising online, the data that is generated is often not stored or shared with the tracking system - a missed opportunity for automatization. An AI might be able to ease administration tasks, as computers are well equipped at logging data.

Secondly, a teacher has the task to analyze work to keep track of progress. As the memory of a teacher is limited and a class can consist of 30 children, the AI could help with this task. An AI can easily signal stagnation in the learning process based on multiple data sources. Or give useful visualizations that help with keeping track of the progress of the whole class and the individuals in that class.

Thirdly, a teacher needs to make decisions on the content of his or her education. At the moment, methods in the form of books help the teacher to give guidance. The drawback from these books is that it is static and will not adapt towards the children that are working with the method. Children are often divided into a finite amount of groups (mostly 3 or 5) based on their level. For all lessons concerning one subject, they get practice materials based on the predetermined level. For example, a student that is in the lowest group of mathematics and is practising a sub-subject that went really well, will still get easy practising materials for that sub-subject, even though the student could solve it at a higher level.

It's quite a lot of work for a teacher to divide the children into these groups per lesson, therefore the children are fixed for a longer amount of time and only evaluated once in a while. This is where the AI could assist the teacher, as the AI can divide the groups per education activity. When the activity is online, the content can then be adapted in real-time by increasing or decreasing the level based on the interaction and the previous results. This behaviour can be compared with a teacher creating practice materials based on the previous responses and looking over the shoulder. Individual guidance is not feasible in a normal classroom setting, but an AI can do this rather easily. This would mean that the AI helps the teacher to become more adaptive towards the individual student.

There are three different tasks where an AI could assist the teacher. The children would benefit from that collaboration as they would receive education that is better adapted to the level of the child. This adaptation is desirable to keep the children in the zone of proximal development. The zone of proximal development is developed by Vygotsky [2] and exercises in this zone are assumed to lead to an optimal learning process as they are not too hard or too easy but still challenging. As this zone is not static or dynamic, the teacher or program needs to make predictions about the location of this zone based on previous results and observations.

The general idea of this thesis is to discover the possibilities of using AI as an assistant of the teacher to help with making the education more adaptive towards the student, hence staying in the zone of proximal development. As this is a broad topic, it is made more concrete by describing one application that could help with the three tasks that are named above. One specific subject, mathematics, is used to develop a prototype of this application. The application is one of the possible answers on how to use AI as an assistant of the teacher. The focus would be to create an application that is adaptive toward the level of the user. Therefore information about the user needs to be gathered and analyzed.

This leads to the following main research question for this thesis:

“How to design and develop an AI assistant that adapts towards the zone of proximal development of the students, while reducing the workload of the teacher?”

The work is validated with the realisation of an adaptive application for teaching math. The main question can be divided into sub-questions, the three sub-questions are focusing on the general idea of the adaptive application, the last sub-questions is focusing on the development of a specific prototype:

1. What are the requirements of the adaptive application for this purpose?
2. How can the application be designed to assist the teacher in the administration, analysis and content creation?
3. How could the adaptive application be used in the classroom?
4. How to develop an adaptive application for mathematics for primary school children where the content is dynamically created based on the individual and current performance?
    - How can the application dynamically create arithmetic questions in real time, based on the individual and current development?
    - How can the application adapt towards the level of the student?
    - What are the influences of different parameters on the behaviour of the adaptive algorithm?

The research question is answered by creating and describing an global design of an adaptive application and building a prototype that creates arithmetic exercises and adapts based on the user interaction. Chapter 2 focuses on the problem analysis by conducting a literature study on the following topics: the classroom, adaptation, AI and AI as assistant of the teacher. The last topic focuses on the three tasks of the teacher that were mentioned in the introduction, namely, administration, analysis and content creation. The chapter concludes with a summary and conclusion.

Chapter 3 focuses on the first three sub-research questions. This chapter starts with the goals and the requirements of a global design for the adaptive application. Conclusions from chapter 2 are used to draw these requirements. From the requirements, a design is presented. This design is validated by the use of two different quality measurements. The first one focuses on different instruction principles to enhance learning and the second one focuses on the quality of an online application. The chapter concludes with a summary and conclusion.

Chapter 4 focuses on the last sub-research question. A concrete prototype is proposed and built. Details on the design choices are given. The prototype is validated with the use of a simulation and a group discussion. Also a field study was designed with the user of a neural network but due to the Covid-19 pandemic this was not executed. The design of the field study can be found in appendix A.

Chapter 5 discusses the most important topics that together answer the main research question. First starting off with the three tasks of a teacher: administration, analysis and content creation. There are many possibilities to support a teacher with these tasks with AI. The possible usages in the classroom and the prototype are discussed thereafter. As AI brings many ethical and privacy issues, a section in the discussion is devoted to this topic. As part of recommendation for future improvements of the application, the possibilities for emotion recognition and the usages of other data sources is discussed.

The chapter concludes with future work where the main recommendation is to do an effectiveness study with the prototype in the classroom. Chapter 6 concludes the thesis with the main findings. The thesis presents different possibilities to use AI in the classroom and gives a framework of an application that can help the teacher with administration, analysis and content creation for different subjects. The main advantage is that the application can be fully adaptive towards the level of the student, hence in the zone of proximal development. The prototype is a concrete example of the proposed system and focuses on math exercises.

Chapter 2

Problem analysis

This chapter gives background information and an overview of the literature research that is needed to build an adaptive application. The first section focuses on the concept of the classroom, with the environment, actors and processes. Thereafter a brief history of AI and a definition is given that holds for this thesis. Also a few machine learning algorithms are reviewed for the adaptive behavior of the application. The section thereafter focuses on adaptive learning in general, gives an overview of the didactic of mathematics that is implemented in the prototype, and discusses different existing applications with an adaptive component for education. Subsequently, the usages of AI in the classroom as an assistant of the teacher is further discussed. The AI can help with the administration, analysis and content creation. Different examples of these tasks are given. A few of these examples are used in the adaptive application. This chapter concludes with a summary and conclusion that is the basis for the global design and the prototype. This problem analysis is not an extensive literature research, as is covers many different and complex topics, a choice is made which information to provide. For example, much more can be said about learning, however the information that is used in the global design is presented here.

2.1 The Classroom

The classroom as a concept is viewed from three different perspectives. First, an overview of the classroom as an environment is given. This is related to the previous research that has been conducted. Thereafter, the actors in the classroom are discussed. The focus lies on the teachers, students, developers and content creators. Lastly, the primary process in the classroom, learning, is described. From a more pedagogic point of view, learning in the classroom is reviewed.

2.1.1 Environment: The classroom

The classroom is a complex environment with many aspects. In 1970 Lacey [3] called the classroom 'The Black Box of Schooling', as little was known, about the processes inside the schools. For many years educational science did not focus on the classrooms [4]. Experiments in a controlled laboratory-like environment were more popular [4].The classroom was studied with input-output models, forgetting about causal mechanisms [4]. The focus lays on social class and social structures to explain education achievements and social inequality [4]. By the end of the 1960s, the research field started to show more interest in the classroom processes in order to understand the relationships between social class and school achievements [5]. Eventually, more research focused on explaining the Black Box of Schooling, to discover social mechanisms that can't be discovered with quantitative statistical techniques [4]. The focus shifted towards the more theory-driven instead of method-driven, making the approach more realistic [6]. This meant that the context and explained mechanism from other studies are prioritized to clarify before gathering the empirical data.

The classroom used to be a physical location where students and teachers come together with the primary goal to educate the students. The word classroom has become a synonym for education [4]. With the current development, the classroom has been extended into the online environment. Because of the COVID-19 pandemic, the physical classroom was not available anymore in many countries, and the education forcefully continued online. As schools are opening up again, it will be a great opportunity to research which part to keep online and which part to continue in the physical classroom.

The current situation in the Netherlands

This part is written to give a brief overview of the current education system in the Netherlands. To give the reader the opportunity to compare the education system in the Netherlands with his or her own system and to make a possible bias visible. Important to note is that most literature about education is also focused on Western society.

It is important to understand the current situation of the education system, as any implementation of AI or any other educational innovation will bring change to the education system. Therefore, the initial situation in the Netherlands will be described. The Dutch education system is built upon 3 levels: primary school, high school and college. In this thesis the focus will mainly be on the primary school, however, most of the proposals made in this thesis can be generalized to high school and college with small adjustments. Primary school starts at the age of 4, and parents are obligated to send their child to school from the age of 5 up to and including 12 [7]. For example during the primary school period, the first 8 years of the educational career, a child needs to have at least 7.520 hours of education by law [8]. A teacher in the primary school is teaching all subjects to one class. In contrast to teachers in higher education, as they are teaching one or a few subjects and teach this to multiple classes. During primary school time, differentiation is done within the classroom by the teacher, for example there are special lessons and activities for gifted students or special needs students. All children of the same age group will be in one class except for children with complex special needs. Although, if possible, these children will join the regular primary school with some extra support or facilities. After primary school, the children will be divided into different levels. Some schools combine multiple levels but eventually, children will finish one level. This level gives access to the next institution. For example, VWO is needed to get access to the university. The Dutch education system gives the opportunity to go to a higher level horizontally if students are performing well. This makes it possible to even get to university when starting at a lower level.

All levels are conducted with an exam. The first two levels are exams made by external organisations such as the Cito eindtoets for primary schools and Centrale Examens for high school. How these levels are reached is in the hands of the schools. They have the freedom to choose how to educate as long as the academic achievement of the students is average. Most schools choose a method for a subject that comes with books, exercises, tests, a teacher manual and computer software from a few different publishers. In primary school, there are a few different education concepts about teaching, most prominent are Montessori, Jenaplan, Dalton and Vrije school [9]. These education concepts bring different ideas about how to teach and how to use teaching materials. Without going into details into all education concepts, they also bring different ideas about the role of the teacher and students. For example a teacher could be a coach for the children or focus on giving the children more responsibility for their own learning process. When AI is brought into the classroom, it is important to take these different education concepts into account or focus on one or a few.

2.1.2 Actors: Teacher, students, developers and content creators

When evaluating the effectiveness of an AI in the classroom, it is important to understand who is using and building the AI [6]. In this case, there are multiple actors such as teachers, students, school directors, parents, developers, user interface designers and many more. This creates a sort of power play and some disagreements as not all actors want the same thing [6]. Even though there are multiple actors, our focus lies on the actors that are directly involved, namely the teachers, the students, developers and content creators. The role of a teacher in the classroom is rather broad. The teacher is teaching one or more subjects to the students depending on the level. In primary school the teacher teaches all subjects and in higher education a teacher focuses on one or a few subjects. It is assumed that the teacher has a solid knowledge base in those subjects [10]. Transferring this knowledge is the real challenge. The teacher has to come up with a strategy to transfer this knowledge by doing activities with the students. These activities differ per subject and teaching style, but normally consist of (interactive) lessons and exercises. During the activities, the teacher has to make sure that the classroom is an environment that makes learning possible by actively engaging the students [11]. The teacher is giving the students guidance in their thinking process and tries to shape and expand their thinking [12] [10]. Students should be feeling respected and secure.

The student role is to participate in activities that make the student grow in social and cognitive skills and in knowledge. Students are unique and therefore have different needs in the classroom. Most students are still growing up and going through different stages in life while receiving education at different points in time [13]. There is much more to say about the (psychological) development of students. For further reading Slater and Bremner wrote an extensive book about developmental psychology [14]. The developers and content creators are not the intended users but really important when developing an application. For the developer it is important that the application is built in such a modular way that it is easy to add extra features and that code can be reused. This modularity makes it possible to work on one part developer while another developer is working on another part in parallel. Thus a developer doesn't have to understand the whole application to be able to add a new feature to the application. For the content developer it is important that it is easy to add new content and structure this content accordingly as the amount of content is ever growing. They have a huge influence on the adaptive behavior of the application as they have to set standards for different levels of the content. These different levels will be used by the application to base the adaptation on.

Issues with the workload of teachers

There are some problems emerging in the education system in the Netherlands. Teachers are reporting increasing work pressure and there is an increase in the number of burn-outs each year. In the Netherlands, 17% of the working class is reporting burn-out issues [15]. Teachers are far above this average with 27,4%. They report that they have to do too much work. The burn-out complaints are related to the work pressure that rests on the teachers. In the last couple of decades, the workload of teachers has increased. Since 2017, teachers have been protesting against the high work pressure and the low salaries. They argue that the amount of administration and non-teaching tasks is too much and that they don't have enough time to focus on the students anymore. Every small detail needs to be logged and for every child that is different than average, an (individual) plan needs to be written mostly by the teacher. Teachers claim that they don't have enough time to focus on the most important part of teaching, the children.

One of the solutions that the teachers are demanding is more people in the classroom to reduce the load. Another problem emerging over the last few years is the lack of teachers in general. Schools have trouble finding qualified teachers as retention rates are plummeting. It would be innovative to look at the possibilities to reduce the workload of teachers with the use of AI and this might lead to not needing more people in the classroom after all. In essence, the question would be 'how can AI contribute to the quality of teaching?'.

2.1.3 Processes: learning

One of the focuses of the education system is to enhance learning. Learning is something quite primitive, from birth on you start learning by discovery. This automatic process can be influenced by the environment [14]. For example, the language that a child learns to speak depends on the environment. As we are aware of this environment and its influence on the child, schools are trying to shape a good environment that gives the children the possibility to develop even more. As a society, we developed norms and ideas on what a child should learn. Therefore, different curricula were developed that give guidance on what a student should learn. For example, in the Netherlands, the curriculum for primary schools is created by Stichting Leerplan Ontwikkeling (SLO), called Tussendoelen en Leerlijnen (TULE). This document includes goals per primary school grade. However, it doesn't say much about how this learning should take place. As learning is a complex process, there are different theories about learning, such as behaviorism, cognitivism, constructivism [16].

Without going into details of each theory, there are a couple of elements that might be important to enhance the learning process. First of all, repetition and practice. In learning psychology, there is consensus that around 7 repetitions are necessary to consolidate knowledge [14]. Secondly, emotion could enhance the process [17]. This is why people remember their wedding day but can't remember an insignificant rainy Tuesday in September. Thirdly, engagement or motivation could also be an enhancing feature for learning. Although the mechanism behind this still remains unclear as research has shown mixed results [18].

All elements above can be influenced by the teacher. For example, even a dull subject can be made into a fun exercise that children enjoy. But that depends heavily on the teaching skills and the imagination of the teacher. As the teacher mostly chooses the activities that the children will undertake, the teacher can build in repetition moments and activities. For example, when teaching math, the teacher can start with a small repetition activity tendered towards the previous lesson(s). A short recap to activate the previous memories [19]. Television production also uses this principle by giving a small recap of the previous episode(s). They even take the next episode into consideration by giving a recap that focuses on elements that are important in the next episode.

When translating this information to AI as an assistant of the teacher, there are a few things to point out. It would be beneficial to create a system that can help with practicing and keep track of the number of repetitions, as well as facilitating the creation of engaging and motivating content for students. This can be achieved by using information about the interests of the student that also reflects into the practice materials that the AI is offering to the student.

To summarize, this gives the following guidelines for the application:

- The application should give exercises that differentiate per student in level and in proceeding speed. This means that exercises will be practiced multiple times with different time intervals until the student has consolidated the knowledge. The AI component is figuring out when the student has consolidated the knowledge or how much practice is needed from which exercises.

- The application should be able to add personalized features to make the students more enhanced. This can be done by using different themes in the application or using certain topics in generated exercises.

- The application should give the teacher insight in the progress and the level of the individual students, and which and how many exercises are practiced with corresponding results.

2.2 Adaptation

As discussed in the introduction, exercises that are given to students should coincide with their zone of proximal development [2]. As this zone is constantly changing while learning and not static, preferably, education is also adapting in a flexible way. Current solutions focus on providing a static way of adaptation by offering different levels of content. Children are stuck at this level for a while thus, this does not comply with the individual needs of the student. When being assigned to a certain level of math, it might happen that certain math topics are too hard or too easy but the child still has to practice on the assigned level. A teacher might signal this and switch the level for one lesson however this is quite a time consuming task for a teacher to monitor 30 students at the same time. To help the teacher with this task, it is important to understand how adaptive learning might take place. Therefore, the next paragraph focuses on adaptive learning in general, not necessarily with the help of a computer. Thereafter the didactic of teaching math is reviewed as these ideas are implemented in the prototype. This section concludes with an overview of other educational applications with an adaptive component.

2.2.1 Adaptive learning

In this report the following definition will be used: adaptive learning is customizing the learning process towards the individual learner and responding to the behaviour of the individual learner to increase the learning efficiency.

Adaptive learning has been a topic of research far before applications were developed for studying, as adaptive learning does not necessarily need a technical aspect. In a one to one situation, a teacher is also adapting towards the level of the child. However, with technological developments, it is easier to exploit the principles on a wide scale. Atkinson and colleagues have been pioneers in developing an adaptive learning scheme [20]. Over time, different adaptive learning schemes have been developed [21] [22]. Different aspects are taken into account and used as a parameter for the estimation model. The parameters differ per domain and learner. Examples of parameters are response accuracy, the history of the learning sessions and response time.

Mettler, Massey and Kellman [23] have developed an adaptive learning system that targets the response time and accuracy for solving multiplications. They proposed a sequencing algorithm as there was a finite amount of sums and gave each sum a priority score to establish a sequence. The priority score was based on the accuracy, response time, number of trials and some constants. The priority score made sure that there were other sums before a sum came back. This space was variable. They found that with their improvements, current state of the art technology-based adaptive learning systems improved the learning outcomes even more. [23]

Another way of adaptive learning is making the subject personalized. In the research of Walkington [24], they used the interest of the student to improve the learning outcome. As the questions that were asked by the intelligent tutoring system were related to a topic that they showed interest in before. Especially for the group that struggled, learning in the online environment showed quite some improvement compared to the control group.

To make adaptive learning possible with technology, there are two types to identify. The first type is adaptive learning that is based on a rule set created by an expert and/or based on prior research and experimenting from the psychology field. The second type is adaptive learning that is based on data collection. For example, Pearson and Knewton are collecting big data from educational applications [25]. Even though collecting education data goes back over a century, [25] with the use of technology it becomes easier to capture this kind of data. However, before this data becomes useful for adaptive learning, the data needs to be analyzed. This can be done in several ways. First,data visualisation will give insights into different groups of students and the level of the students. Prediction models are built to capture the best learning strategy per student or student group. These prediction models can range from simple regression analysis to complex deep neural networks. More pedagogical expertise of educators is being replaced by data scientists and based on big data [25]. Tien optimised a data-driven tutoring system and found that giving personalized feedback in the form of hints to students had a positive influence on the process of learning. It even came close to the level that human tutors can achieve with the students.

A combination of both types would be interesting; on one hand, an application that is based on learning theories and on the other hand a data-driven application, where theory would be the basis to create the application and give it a head start. The data that will be gathered and be used to improve the application's adaptive behaviour by searching for patterns. It might even shed new light on existing learning theories. In this thesis the above approach is used, where the application is developed based on theory but will be improved with a data-driven approach in the future. Meaning that a fully functioning application is created that forms a framework to gather data and eventually as part of future work, will be improved by using this gathered data. This can be done by data analysis or even by building a neural network for predicting the level of students.

To summarize, this gives the following guidelines for the application:

- The application should be able to adapt based on the interaction with the application. Therefore, the application should be able to predict the level of the student while the student is progressing. The application should be aiming at in the zone of proximal development.

- The application should be able to give feedback/support to the student based on the interaction with the application. With the goal to increase the level of the student.

- The application should be able to adapt to different students. Meaning that the application can support students that learn faster or slower than an average student.

2.2.2 Didactic background of arithmetic

Didactics can differ per country. This part is written with the Dutch didactic in mind as the application is designed for the Dutch primary school system. However, this information can still be generalized to school systems of different countries. When children learn arithmetic there are four different stages to identify; informal, representation – concrete, representation – abstract and formal [26] [27] [28]. An overview of the stages and some examples can be found in figure 2.1. Important to note is that the informal stage is the stage where the smarties are in the physical world where the children can feel and play with the smarties and even eat the smarties, it is not an image.

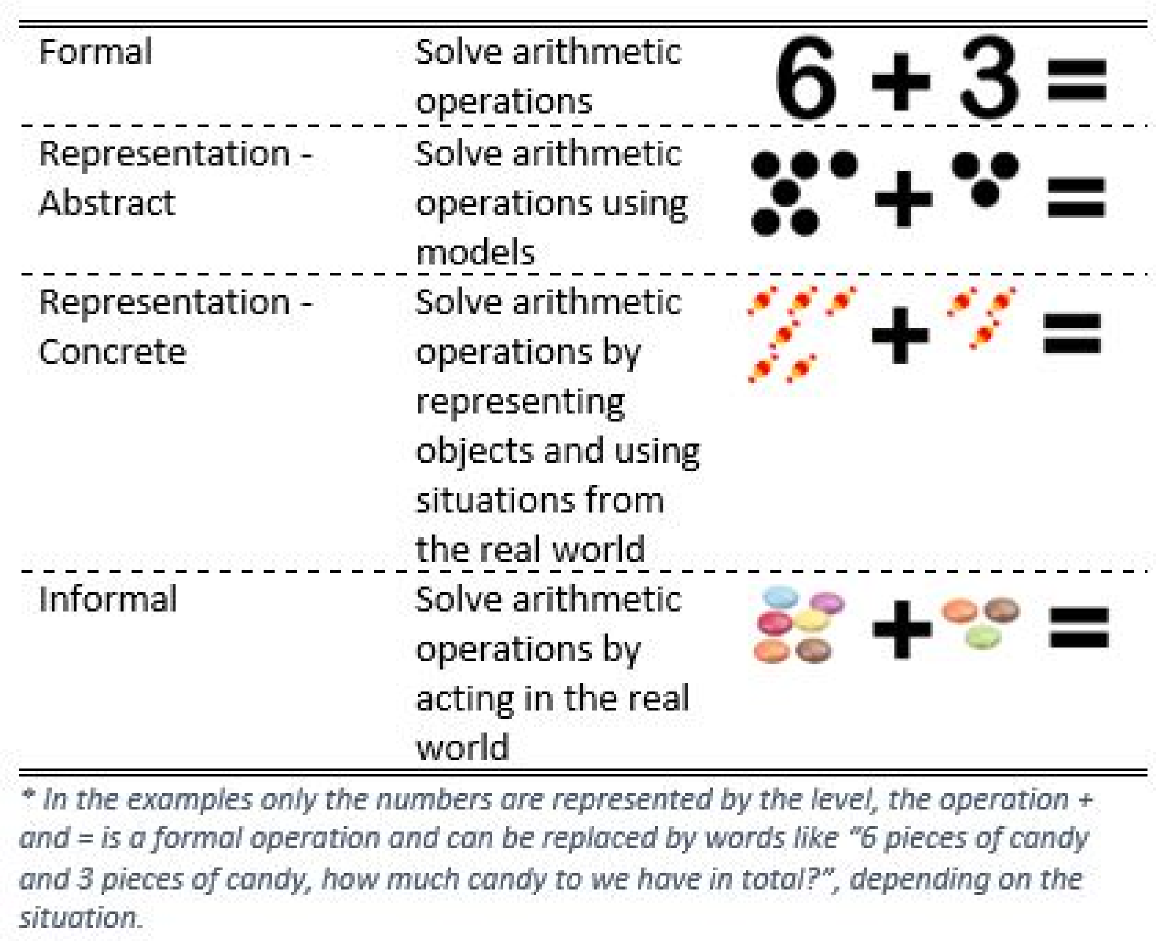


Figure 2.1: Stages of solving arithmetic

The first stage,referred to as the informal stage, is when children are counting in the real world, and amounts are concrete. Normally, this is combined with gestures like pointing when counting. When entering the representation or concrete stage, an abstraction level is added to the arithmetic. For example, children can count things on paper where the representation varies from looking like a real-life object to a rather abstract object also called the representation – abstract stage. Eventually, children will come to the formal stage where they can solve exercises in a rather abstract way, with digits as abstract signs which can also be combined to numbers representing larger quantities/values. Going through the stages is not a linear process, and not all arithmetic subjects are in the same stage at the same time. This means that children can be in the formal stage for addition and subtraction but the representation – abstract stage for multiplication and division. It can also happen that children need to take a step back and return to a lower stage to understand the exercise at hand. Even adults do this sometimes when facing mathematical problems, for example when calculating how much paint you need for a room and you draw the walls flat on the paper to make a representation of the formal sum at hand.

The application must support the different stages. As the informal stage is merely happening in the real world, a computer can't facilitate this stage as it is always some kind of representation, therefore this stage is not addressed in the application. The other stages will be considered and it should be possible to switch between a stage within a practice session.

When practising addition and subtraction, children must get insights into what this means [29]. In the first place, what are digits and what can you use digits for. What does it mean when you have to group items and combine these groups. And what kind of abstract representations does a child need to be able to apply strategies for addition and subtraction.

To get insights into arithmetic, children start with counting and connecting these amounts in the real world [29]. At some point, children need to connect the symbol of the numbers with counts and amounts. After this children will start to count with jumps (like 2-4-6 or 5-10-15). This can be done by giving only 2 euro coins and counting the value. Counting will be connected with addition and subtraction. Before the symbols are introduced, the context around the exercise will challenge the child to start solving the sum. An example is, you have 5 smarties, if you give your brother 2 smarties, how many smarties do you have left? Eventually, the symbols for additions and subtraction are used and a more formal way of asking questions. When the sums are getting more complex, a strategy to solve the addition or subtraction is needed together with an abstract representation.

The strategy central in this application is the smart dividing of numbers and the representation of the numeric line. When adding or subtracting values it is easier to divide the second value for example into hundreds, tens and single digits [28]. After splitting, adding or subtracting first the hundreds, thereafter tens and lastly, the single digits, simplifies the exercise at hand. This is to minimize the cognitive load. When the addition or subtraction is passing tens or hundreds (for example 9+3), the unit is split again into two single digits (into 1 and 2 in this example). To make this splitting more insightful, the numeric line is added from the representation – abstraction stage. The numeric line looks like a ruler and gives a representation of consecutive numbers. On this numeric line, the starting value can be viewed along with the added or subtracted value together representing the sum at hand.

To summarize, this gives the following guideline for the prototype that focuses on the generation of arithmetic exercises:

- The application should support the different stages of the learning process when practising addition and subtraction. While doing this, the used strategy should give insight into how addition and subtraction works. This is done by creating elements that tend to connect different levels of abstraction, furthermore, the numbers should be divided to provide a strategy to help the child solve the exercise.

2.2.3 Existing applications

In this section, two applications are discussed that are focused on mathematics, Squla and the Rekentuin. Thereafter Duolingo and WRTS are discussed, they focus on language education. One of the applications already on the market is Squla. Squla offers online games to practice all primary school subjects. Different levels are offered and the teacher can adjust the level of the students. The games are developed by game developers and education experts.

Furthermore, the ”Rekentuin” (Math garden) provides up to 26 mini-games to practise their mathematics skills [30]. The games are adapting towards the level of the individual child. Furthermore, instructional videos are also adaptive. This makes sure that the students will always practice at their current level. The content is connected to the goals of the Dutch government for primary school children. This is displayed in a sort of route that the children have to take. The student can choose between 3 levels where the levels are connected to the predicted score. For example in the lowest level, it is expected that the child will get 90% of the exercises correctly. The child is motivated to pick different games that belong to different arithmetic domains, as the plants in their garden are connected to the amount of practice that they have done. Plants are happy when the student is practising a lot and unhappy when not practising. The analysed data that the application gathers is available for teachers.

Duolingo is an adaptive application for learning languages. Within boundaries of a certain predetermined set of exercises, given exercises are based on the interaction with the program. When making the same mistake often, that exercise will come back during a practice session until the user doesn't make that mistake anymore. Certain practice sessions can only be accessed when other practice sessions are completed.

WRTS is a program that lets you create your own content (words and translation) or use content from other users and study book publishers. It has the same adaptive behaviour, namely repeating exercises that the user finds difficult. Therefore it looks at the practice sessions and based on the incorrect responses the practice session will adapt.

Overall there is a lack of literature regarding the efficiency of these and other applications. The ”Rekentuin” has some papers published and works together with researchers that focus on the educational aspects of the application. There seems to be a gap between applications for commercial use and applications for research purposes. Applications described in papers that focus on adaptive learning, do not make it to the public often. It would be interesting to create an application that has the quality to be commercialised but also has a strong research basis. This is to give the children the best of both worlds.

Based on the above the application should have to following design guidelines:

- The application should be able to be incorporated into the classroom by supporting the teacher instead of being a standalone application where students can practice.
- The application should be built with the use of literature research and should be able to support (independent) research when the application is live.

2.3 AI

In this section a general description of AI is given. As AI is a broad concept, it is narrowed down to the aspects of AI that can be used in this thesis. Thereafter, algorithms for prediction are discussed that can establish the adaptive behavior of the application. The use of AI is tangled with many privacy issues, therefore a dedicated paragraph is written about this topic.

2.3.1 What is AI?

At the Dartmouth conference [31] in 1956 the term “Artificial Intelligence” was first used. As a handful of scientists came together to discuss their work on how to make machines behave intelligently. These scientists came from different backgrounds: mathematics, psychologists, electrical engineers, working in both industry and university. Five of the attendees: Allen Newell (CMU), Herbert Simon (CMU), John

McCarthy (MIT), Marvin Minsky (MIT) and Arthur Samuel (IBM) eventually became the founders and the leaders in AI research [32]. The term “Artificial Intelligence” was coined by John McCarthy [31]. On the question of what AI is, he answered: “It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.” [33]. Of course this raises the question, what is intelligence? “Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines.” [33]

In the past 60 years, different subfields have emerged such as neural networks, computer vision, robotics, speech and natural language processing, machine learning and many more. When people talk about AI, it is not always clear which sub-field they are talking about. Just like AI has been influenced from different backgrounds, the sub-fields are not stand-alone fields as they are dependent on each other. For example, natural language processing can use machine learning or neural networks and robots use computer vision

In this thesis AI is used for three different aspects of the application that assists the teacher. First the application adapts to the level of the user when presenting exercises. For this task an new algorithm is developed that is inspired by Q-learning, an algorithm used for self-learning agents. Secondly, the application can generate exercises by using techniques from Natural Language Processing, for example to make ‘fill the gap’ exercises. Thirdly, the application presents data in a smart way to the teacher. It analyses the data and based on the results it reports back to the teacher by displaying useful graphs or by giving notifications. The goal of using AI in this thesis is to adapt the level of exercises based on the analyzed data from the interaction with the adaptive application to stay in the zone of proximal development of a student.

2.3.2 Algorithms for prediction and adaptive behaviour

When looking at algorithms that can be used for adaptive education, there are two main requirements. FirsT, it should be able to handle the input data from the user that is interacting with exercises. Second, the algorithm should give output that can be translated into an exercise.

One of the least complex algorithms that can be used is the (multi) linear regression model, to model a student or a student group and give a level of the next exercise as output. The drawback of linear regression is that not all expected relationships are linear and can therefore not be captured by this type of regression.

When relationships are not always clear or linear, a neural network can be used as a nonlinear regression model. When feeding enough training data, the neural network will discover the hidden patterns in the data. Based on new data the neural network can predict the best exercise. A drawback of a basic neural network is that most of the time it is designed for a specific task [34]. Furthermore, a great deal of labeled data is necessary for training. A specifically labeled dataset may be hard to find and require human labor to label by hand. Statistical algorithms could be a more time efficient option as they also perform well on prediction tasks.

An interesting form of a neural network is the recurrent neural network (RNN). This network works well with time series data and could therefore be implemented in real-time in the application. Multiple studies have used the RNN for predicting educational results [35], [36], [37], [38]. A traditional neural network does not consider the history and the relationship between past data and current data [35]. When predicting what level a student will achieve, information about how a student has been achieving in the past and how fast the student is progressing is important information to predict the future. RNN has been proven to be an effective model to predict final grades with an accuracy of 90% [36]. Even when the data given to the RNN is not from the same course, the RNN does a better job in predicting results in another course than when using regression analysis [35]. Therefore, an RNN is a good candidate for real-time applications that try to predict the level of the student in real-time.

There is not always enough data available to train a neural network properly. One of the strategies that can be used is ”one-shot” learning [39], this makes it possible to predict a class based on a single labelled example, where normally hundreds or thousands of examples are necessary. This is done by exploiting previously learned knowledge on similar tasks. By using this pre-knowledge, the system can learn much faster, i.e., with one or a few labelled examples. However this pre-knowledge also needs to be established before using one-shot learning. One-shot learning has a positive side in that it is less prone to over-fitting, is much faster and can even outperform some state of the art deep neural networks.

When data is scarce, reinforcement learning is a good option as it does not require any labelled input and output pairs. Reinforcement learning [40] can handle sequential decision-making problems when there is limited feedback. It tries to optimise a sequence of decisions based on the previous interactions and random exploration. The environment is modelled by states and actions as a Markov Decision Process. Reinforcement learning can be model-based and model-free.

However, the environment described in this thesis is not fixed as the student performance will differ in time. Also, a student can answer the same question differently, making it hard to model the student. It could be possible to create a model-based algorithm with a reward function, like reinforcement learning. The algorithm rewards harder questions with a higher reward than easier questions. However, when the student is at the limit of their level, the questions will not be answered correctly thus leading to a penalty. When the algorithm tries to optimise the score and is still exploring towards higher levels, an increase of the level of the student would not be a problem for the algorithm.

A specific algorithm in the domain of reinforcement learning is Q-learning [41]. This algorithm is model-free and it can find the optimal policy for any finite Markov decision process. It is trying to maximize the total reward by exploring the environment and updating results in a Q-table. This algorithm comes close to the requirements for the application, being able to formulate a model based on no prior knowledge. However, the environment (the user) is not stable as the optimal policy changes as the user is learning. Exploration in Q-learning decreases over time; this is useful as you start with no knowledge and want to quickly create a model. However, the amount of exploration should not be decreased infinitely as the environment is changing over time, exploration is needed to keep up with this change. Exploration should be limited in some way as it would lead to unwanted behaviour. For example, a child that barely understands 2+3, should not be asked to solve equations such as 594 + 143 as part of the exploration.The exploration should be aimed at equations like 6+8 that are just above the level of the child. This would lead to the formulation of a model that works locally with the Q-learning algorithm. This Q-learning algorithm is the inspiration for the adaptive algorithm that is developed in this thesis.

To summarize, this leads to the following design guidelines:

- The application should use a prediction algorithm inspired by Q-learning that decides which exercise must follow to maximize the learning efficiency. This exercise should be in the zone of proximal development.

- The application should give a variety of exercises, most in the zone of proximal development but also some easier and harder exercises. The latter are to check if a student has forgotten or respectively has already mastered an exercise. If so, the application should act upon that information and increase or decrease the level of the exercises.

- The application should work with a clustering of similar exercises to ensure that little repetition of the same exercise occurs. This guideline is specific for the prototype with the math exercises, as the focus lays on acquiring a skill, not remembering an answer. On the contrary, when drilling words it might be useful to add this repetition.

2.3.3 Privacy issues

Most AI solutions require and/or acquire data. The EU has made regulations about the use of AI and the use of data. For data protection of EU citizens, the EU formulated the General Data Protection Regulation in 2016 [42]. This regulation regulates the processing of personal data and the free movement of this data to protect EU citizens. Rules about which data to collect, how and how long to store it and consent. The general idea is that personal data should be processed lawfully, fairly and in a transparent manner. The EU is also working on laws on the usages of AI [43]. They made a proposal on how to regulate the usages of AI. In this proposal they defined AI as follows:

"'Artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

Annex I:

(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;

(c) Statistical approaches, Bayesian estimation, search and optimization methods."

Only focusing on the techniques is insufficient, therefore the proposal also focuses on the risk that AI might bring. For example, face recognition to open your phone might be preferred, but the same technology used to track the movement of individuals might not be. High risk applications need to comply with more regulations than low risk applications. This shows how important it is to formulate a certain goal to use the AI technique, this goal also needs to be ethical. As developers it would be wise to analyse the possible risks of the software that they are building, how easy can the software be used for unethical applications and what can be done to prevent that usage?

The application proposed in this thesis needs to comply with the General Data Protection Regulation, meaning that the data that will be gathered needs to have an explicit goal that justifies the collection of certain data. This needs to be communicated with the public to make it transparent. Consent is also really important, especially because the target audience are children.

Even though there are regulations about the usages of data and a proposal for regulation on AI, there is also a more ethical side of using data and AI. For example, when building the adaptive application certain risks will rise. If the system is malfunctioning, what are the possible consequences for the children working with the system? If the system is not in the zone of proximal development and therefore only offers very easy or difficult questions, this could disturb the development of the child. Therefore, what are the safety nets in the application and what is the role of the teacher? These questions are answered in chapter 5: the discussion, after the design of the application is explained in chapter 3.

To conclude, from the above the following design guideline is drawn:

- The application should have an explicit goal and justification for all data collection. The application is collecting data to be used for the adaptive behavior application and to display the progress of the students to the teacher. Data can also be collected for research to improve the application and to discover its efficiency. The latter is future work. In both cases consent needs to be acquired.

2.4 AI as assistant of the teacher

An AI has the potential to assist the teacher in different ways. First, an AI can store more information than a human does, this is especially useful for administration tasks. Section 2.4.1 discusses this topic in more detail. Second, an AI can process more information in less time than a human, making it possible to support the analyzing tasks of the teacher. Section 2.4.2 elaborates more on this topic. Third, an AI can create content on the fly based on the individual level of the students faster than a human. Section 2.4.3 explains how this content can be created. All these tasks can be made more efficient with the use of AI. It doesn't mean that the AI will completely overtake the tasks of the teacher, but when used at the right moment and at the right amount it should support the teacher in the classroom to make the education more adaptive and lighten the workload. These are not the only possibilities to use AI in the classroom, for example, an AI embodied as a robot could also support social interactions or become an assistant in the classroom. For future reading on this topic, Belpaeme et al [44] has written an extensive review about the role of social robots in education.

2.4.1 Administration

When teaching, keeping track of the progress of each student is a challenge. The learning process is happening in the heads of the students and only the ”output” can be analysed to discover if a student understands something. Additionally, a teacher is not able to keep a record of each individual ”output” of a student in their head. Therefore, good administration is important. It gives the opportunity to look back and to get a good overview of progress. It also makes it easier to analyze the ”output”.

To build a good administration, it is necessary to gather data. A primary source is the results of practice sessions and tests. But also more complex sources like behavior and state of the mind could be gathered. As there is a lot happening in the classroom at the same time and the teacher is not writing observations down the entire time, maintaining good administration of 30 children is a challenge.

AI could help ease this task. There are different possibilities, depending on what data needs to be gathered and where this data is coming from. For example, data from an online application is easier to capture than data from an unstructured personal paper notebook. The latter source has more steps in between before it can be saved in the administration, such as gathering the sources, scanning the notebooks, and retrieving the data from written text with, for example, Natural Languages Processing.

Therefore, the application described in this thesis would be an online application to ease the administration task for the teacher as it is completely automatized. The focus lies on logging the progress of the children and at what level they are at which point of time. This can be done by using data about the interaction with the system, such as correctness and response time. Additional mouse movement and keyboard strokes can also be logged by the application. This data has low complexity but can enclose a higher level of information. For example, the combination of mouse movement and keyboard strokes can be used for stress detection [45] or the speed of typing can be used for emotion detection [46]. Complex data like emotion, how to detect this and most important how to use it will be part of the future work. Some ideas around this topic are discussed in chapter 5.

2.4.2 Analyzing

One of the tasks of a teacher is keeping track of the progress of the students. This is mostly done by analyzing the student in real-time in the classroom or afterwards based on tests and exercises. The goal of the analysis in this thesis is to be able to give the students education that is in their zone of proximal development. Unfortunately, the attention of a teacher can only be focused on one student at a time. On the contrary, an AI can monitor multiple students simultaneously, as processes in an AI can work in parallel. AI is therefore strong in analyzing students in real-time. The analysis can happen at different levels, such as group, student, subject, session and

exercise level.
When working on the computer, the AI can correct the answers in real-time and give appropriate feedback. The strength of the AI can be combined with the strength of the teacher. The teacher can monitor the progress in real-time through the AI, for example: in an interface on the computer. When needed, the teacher can step in and give feedback or explanation that is tailored to the student. The AI can give information that it has gathered about the student and the progress, helping the teacher to give the correct feedback. Therefore, the teacher can focus more on teaching and individual guidance instead of monitoring who needs help. Feedback could also be given by the AI, however, research showed that oral feedback with the opportunity to discuss is more effective than feedback that is only given on paper [47].
Furthermore, an AI is good at finding outliers. This is useful when looking at the progress of students. Students that are inconsistent with their progress, need attention from the teacher to figure out what is causing this. For example, the AI can easily give notification to the teacher when a student is struggling and the teacher can act upon that.
One step further would be that the AI can analyze and recognize types of errors. The AI attempts to cluster the same types of errors, and the teacher can create lessons that are focused on the specific error or misconception of the students. Making the lessons more effective and useful. The research of An and Wu [48] showed also the importance of understanding misconceptions in mathematics. In this research, teachers are asked to analyze homework by identifying errors, analyzing reasons for the errors, designing approaches for correction and taking action for corrections. By doing so they improved their education in general as they better understood the student's thought patterns and enhanced their pedagogical content knowledge.
To take advantage of the analyzing capabilities of the AI, the communication between teacher and AI should run smoothly. Meaning that the results from the analyzed data should be communicated quickly and efficiently. This can be done with the use of an interface that can give real time notifications to the teacher through the use of a device. For example, the teacher will get a notification when a child is stuck for a while and probably needs help. Or when a teacher is preparing a math lesson, it will have an overview with predictions on which misconceptions the children might have. The AI tries to assist the teacher with the analyzing tasks without the teacher losing his or her autonomy. The teacher still remains in control but is assisted by AI.
The analysis described above focuses on the teacher, however it is also possible to direct feedback to the student that is based on the analysis of the system. To make the adaptive behavior possible, the system has to analyse the results as well.

Based on the above, the application should have to following design guidelines:

- The application should give insights in the real-time progress in the classroom on class level, subject level and individual level to the teacher. This monitoring should help the teacher to maintain a concise overview of the progress of the children and would make it easier to conduct good interventions that help the students.

- The application should give notifications and nudges to the teacher when students are struggling in the teacher interface. These notifications are based on the progress of the students, for example when the level of the student significantly drops, the teacher will be notified.
- The application should correct the exercises in real-time and give new exercises right away that are adapted towards the student. Next to correction the application should also give a certain level of support to the student if necessary. For the prototype the numeric line will be used to give the students support for solving the exercises. If this is not enough, the application will drop the level or the teacher is expected to interfere.

2.4.3 Content creation

Teachers sometimes create or modify the content that they use in the lessons. Content creations make it possible to adjust lessons specific to the level of the students, hence the zone of proximal development, making the lessons more efficient. In contrast, the content of books is static, except for different levels and paths through a book, there is not much differentiation. However, differentiation is a time-consuming process. Especially with the increasing workload of teachers, it is not always feasible to create their content. This is where AI can make a difference. In fact, AI can create content in real-time and adjust to the individual student or the whole class. There are different types of content that an AI can create and different techniques to do this. When looking at primary school, two types emerge: conceptual knowledge and problem-solving skills. For content creation in general, it is important to create a type of reusable framework, [49] especially because development costs are high. This framework could be used by multiple subjects without the need to create a new framework for each individual subject. Most subjects can be turned into concept graphs, semantic nets or ontologies and give organising structure to the subject [50]. This can be used for content creation as well. Additionally, a model can be created of the student about which parts of this concept graph are mastered and which part still needs to be taught. By using concept graphs, content can even be created and in real-time.

Content that is created in real-time with the goal to personalize and that is tailored to the needs of the student is also called personalized E-learning or Adaptive E-learning. There is a need for this type of learning because students are unique and take different routes to learn. They work at a different pace and with a different approach. Personalized E-learning could help to make learning more productive and be optimized for each student [49].

From the above and the two paragraphs below, the following design guidelines are drawn:

- The application should be able to generate content dynamically by creating templates for exercises. In these templates words or numbers can be changed to make many different exercises based on one template.

- The application should be able to support different types of exercises and subjects. However, for the prototype, the focus lies on the automatic generation of arithmetic exercises. This is done by randomly generating exercises that possess certain properties, such as crossing the ten (for example 9+7).

Conceptual knowledge

Not all subjects are suitable for the creation of content but for every subject that has a set of rules,it is to some extent possible. This is referred to as conceptual knowledge. For example, when learning grammar and filling in the correct verb, it is possible to change the sentences to get different exercises. “I ... you” can be filled in with many verbs (love, hate etc.) or a slight change of sentence: “I ... my dog”. These small changes make it possible to dynamically create content. This is also possible with math exercises, where the content of the sum or the numbers used can be changed. These rule-based questions have one thing in common, there is a finite amount of answers correct. This makes it easy to check the answer with an AI.

But what if the exercises become more complex? For example, exercises that are drawn from running text. There are two options: content creation based on Natural Language Processing or with the help of a human. The first depends on the understanding of the sentences. For example the sentence: ”The Second World War lasted from 1939 to 1945” could be turned into “What lasted from 1939 to 1945?” or “How long did the Second World War last?”.

The way knowledge is communicated is mostly through the use of running text. This is easiest to understand for humans, in contrast, for AI this is much harder. It would be a possibility to save knowledge as structured data. This would make it easier for an AI to create content dynamically. This is where the human can step in, to create the structured data. Even though it is possible to gather this data for example from text or the internet, it might be useful to have a human to check the quality and filter out possible mistakes.

Problem-solving

The primary goal of education is to acquire problem-solving skills according to Holohan, Melia, McMullen and Pahl [50]. Even though conceptual knowledge might be the basis of these problem-solving skills, the content created to practice these skills is different. It challenges a student to make connections and to apply (conceptual) knowledge. For example, what are the similarities between World War I and World War II? An AI should have an understanding of both World Wars to generate or check an answer. This understanding has to be programmed in an AI. To illustrate the complexity of this understanding, imagine that the AI wants to compare the amount of death for both wars with each other. The AI needs to have an understanding, when there are 10 million deaths and compare it with 10 million + 1000 deaths, the AI should say this is similar. However, when this same AI would compare 1000 deaths to no deaths, it would probably also say it is similar. This is because the AI has to follow some kind of rule. In the example, it would base its judgement on the absolute difference to form a rule. In fact, the absolute difference is the same in both examples, however, how a human would perceive it is different. That is a kind of intelligence that is hard to implement in AI and what makes it hard to generate content that practices problem-solving skills.

A way around this problem is using a search engine like google. There are tons of insight questions on the internet with answers created by humans. If an AI can find these questions and present them to the user, the user will get the feeling that the AI is indeed intelligent, but what it really does is just finding information, it does not understand the information itself. The danger of this approach is the lack of control. On the internet, there is an increasing amount of disinformation and conspiracy theories, especially around topics that are sensitive to political views. For an AI it is difficult to distinguish between this information. It is important to consider what needs to be controlled or checked by humans and what does not. Limiting the number of sources that an AI can use to only trusted sources and a system with feedback loops could help to tackle these inaccuracies.

2.5 Summary

Three tasks of a teacher are formulated that can be supported by AI, namely, administration, analyzing and content creation. Educators spend a lot of their time on administration tasks. The most significant advantage of an AI assistant is that it could reduce the amount of time spent on administration. Possibly, an AI could simply overtake the administration tasks. However this has one big disadvantage: the lack of control. AI systems, especially new systems are prone to error. These errors could possibly bring harm to the student. For example, if incorrect data suggests that a student should repeat the school year. The primary focus should be on assisting the teacher. Meaning that the teacher is still in control and can still check the system once in a while. Important side note, administration might not be a purely AI powered activity as not all logging tasks need to involve AI. However this administration is still the basis for analyzing and content creation tasks.

When there is a solid base of information, the AI can help the teacher with analyzing. This can range from making graphs to making complicated predictions about future performance. Another advantage would be time reduction. A further advantage would be that an AI could correlate data independently of the educator. This could give new insights in theories about learning and could improve the quality of their education. For example, the effect of time spent on instruction or practice compared to the results. Some classes might thrive from more instruction time and others need to spend more time on practice to increase learning results.

A final advantage would be that the AI system can analyze results in real-time. This allows the possibility to correct misconceptions from students within the practice session. Teachers can be notified by the system if a student is struggling and the teacher can act upon that information. A possible disadvantage could be that the teacher is relying too much on the system, not everything has a causal link. Therefore, the teacher should always be critical towards results presented by the AI. When it is possible to analyze results and draw useful information, the AI could help the teacher with content creation. The advantage would also be the reduction of time needed to create content that is adapted towards the students. It takes a lot of time to create new content. The development of an AI that can dynamically create content is also time consuming, however when the system is in place, new content can be created in real-time that is adapting towards the level of the user. The same system can work in multiple classes as well. If new content is added, all educators working with the system can get access to this content. A challenge would be to get the right content and structure it in such a way that the AI can generate new content. This content needs to be structured in a hierarchical way, in order to be able to adapt towards the level of the student.

Overall, existing technologies are available that make AI assistants a possibility for educators. This gives multiple advantages. Using an AI as an assistant can provide feedback to the teacher on individual learning progress, with this feedback they can adapt their instruction, thereby improving the quality of teaching and learning in the classroom. Another advantage is that the students will receive higher quality education that is better adapted towards the level of the student, as the content could be generated in real time based on their current level and progress. Finally, using the assistant could potentially reduce the workload of a teacher and allow the teacher more time to focus on teaching. The challenge would be to build a system that still gives the teacher control and that a teacher can check, to prevent creating a black box system. In chapter 3, a possible design of such an assistant is presented and in chapter 4 a concrete prototype is presented. Overall these advantages may lead to an AI system that makes it possible to provide targeted education in the zone of proximal development, thereby providing the student education that is tailored to their individual needs, level and speed.

The design guidelines mentioned in the above sections will be used to design the application and the prototype. Not all guidelines are included in the design of the prototype as the focus lies on the application's adaptive behavior and the interface of the student. This means that the interface of the teacher with the overview of the progress of the students and the notification systems will be more abstract compared to the other elements. From the guidelines, the goals for the application and the prototype are formed. This leads to concrete requirements for both.

Chapter 3

Global design: an Adaptive Application

From the problem analysis the goals and requirements for the global design are formulated, these can be found in section 3.1 and 3.2. This leads to the textual description of the global design in section 3.3 with the following topics: interface, content creation, adaptation and feedback, the usages of the AI assistant and the challenges for implementation.

3.1 Goals

The application has three main stakeholders: teachers, students and developers with different goals. The goals are described below per stakeholder.

Teachers

- Being able to make the education activities more adaptive towards the individual students, hence staying in the zone of proximal development.
- Reduce the workload of the teacher to spend more time on teaching instead of administration, analysis and content creation.

Students

- Receiving practice materials that are adapted to the level of the individual student, hence staying in the zone of proximal development.

Developers

- Building a modular application that can be updated with the newest techniques.

Content creators

- Building a modular application that can be updated with new content that can also be connected with older content.
- Adding content with a hierarchical structure that can be used to dynamically create exercises that can adapt to the individual needs of the student.

3.2 Requirements

The goals of the stakeholders are translated into requirements of the system. The requirements are grouped per topic.

Adaptive

- The application should adapt based on the level of the students without the need of interference from the teacher.
- The application should give the teacher the possibility to influence adaptive behavior in order to make the application usable in different educational situations. For example when the teacher wants to practice one specific exercise type instead of letting the system decide the content.

Administration

- The application should keep an administration of the results. This administration should be easily accessible.
- The application should give insights into the progress of the children.

Analysis

- The application should analyze the results and give conclusive results to the teacher.
- The application should give the teacher nudges and notifications that help the teacher monitor the students.

Content

- The application should support different content
- The application should be able to generate content based on the level of the student.

Development

- The application should be modular, meaning that extra functionalities can be added easily.
- The application should consist of interfaces for students and an interface for the teacher.

Usages

- The application should be simple to use and intuitive.
- The application should be able to support different educational activities and settings such as practicing, testing or activating initial knowledge.

3.3 Global design of the AI assistant

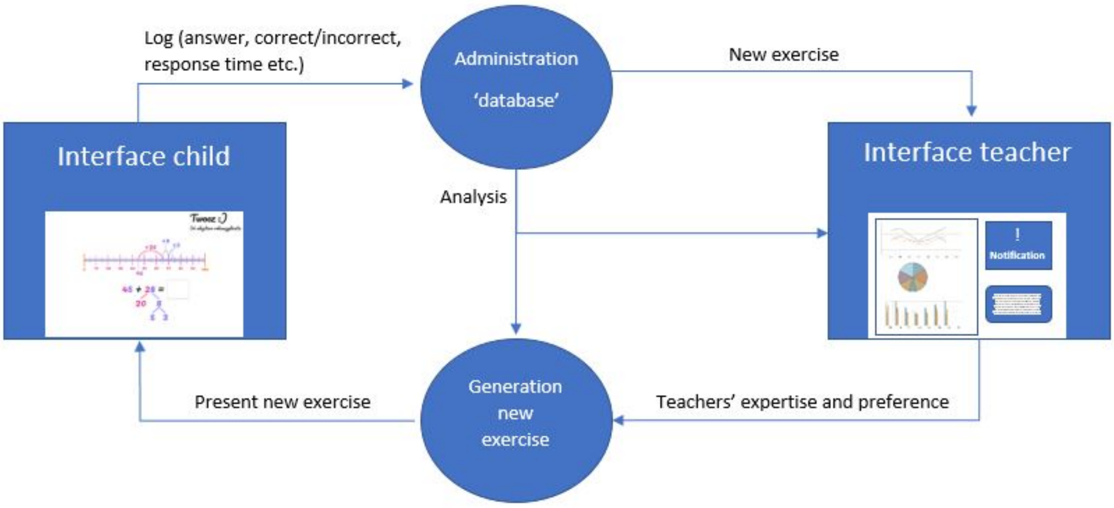


Figure 3.1: Global design

In figure 3.1, the global design of the AI assistant is visible. The application consists of two interfaces, one for the teacher and one for the student/child. More information about the interfaces can be found in section 3.3.1. From the student interface information about the interaction is sent to the database. This information is analysed by the AI assistant and is sent to the teacher's interface to provide insights on the student's progress and provide notifications. These notifications assist the teacher by monitoring student activity. An example of a notification could be a message that a certain student is falling behind or has incorrectly answered a certain amount of exercises. The analysed data is sent to the application that chooses the questions for the students. This is where the adaptation of the exercises takes place.

From the interface of the teacher there is also a possibility to influence the system that chooses the next exercise and therefore influences the adaptive behavior. An example would be that the teacher would change the learning rate of the system for a student that learns faster or slower than average.

The AI assistant is to be embodied in an online application mainly because it allows for flexibility as computers/tablets are already integrated into schools, making it easy to implement. This online application has a few advantages. For one, every interaction with the system can be logged for administration purposes. Because the application is online, a strong server can run in the background, making many analyzing tasks possible even in parallel. This server is also responsible for the content creation that can be done in real-time and therefore can adapt to the level of the student. The application can be filled with different subjects by content creators making it usable in every situation, even if the subject given was not part of the application before. As the content is shared with a general system, others can also use the new content for their educational level. Section 3.3.2 will go into depth about how this content can be created. When the student is interacting with the system, the application is adapting and gives feedback to the student. This is explained in section 3.3.3 in more detail. After it has been made clear what the application is capable of, the usage of such an application in the classroom is discussed in section 3.3.4. Even though there are a lot of advantages of using an AI as an assistant, there are still some challenges to face. These are described in section 3.3.5.

3.3.1 Interface

The application will consist of separate interfaces for students and the teacher. An example can be seen in figure 3.1. The interface for the students focuses on presenting content in the form of exercises to the user and is adapting towards the level of these users. The user can answer these exercises by typing or using the mouse or touchscreen. The interface should be attractive and intuitive as well as provide prompts so users can ask for help.

The teacher's interface focuses on the progress of the users. An sketch of what this might look like is given is figure 3.2. This interface shows data that has been analyzed for conclusions and recommendations. A teacher can select the results from the whole class, one subject, one student and even the results from a certain practice session. The AI will give the teacher some suggestions about which children are doing well and which children need further assistance. The AI tracks the level of the student, for example, a student that is weak in math will not be flagged if he or she performs at the same level, only when the level is decreasing or increasing significantly.

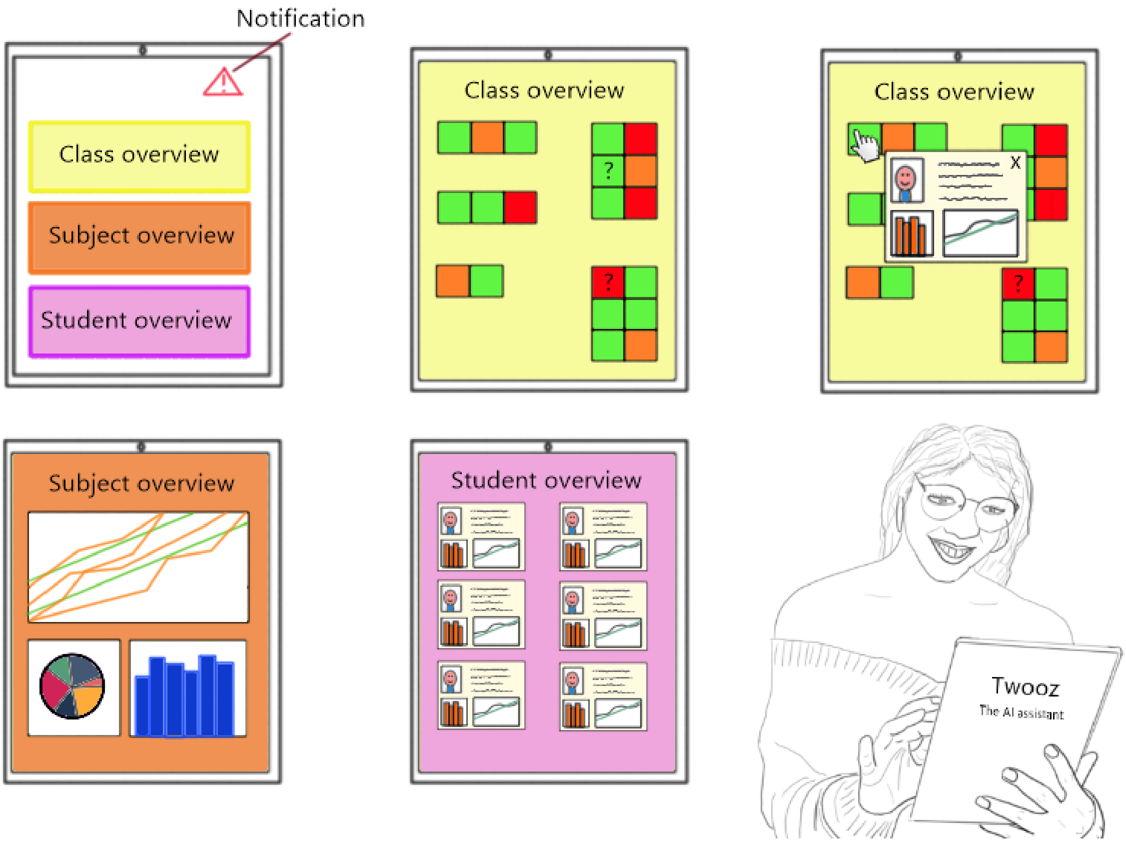


Figure 3.2: Sketch of the interface of the teacher

Furthermore, the teacher's interface can be used for customization of the program. The teacher can modify the settings of the program; settings such as the speed at which the program should adapt or to specify which exercises should be practised in succession can be modified. In case the adaptation is too rigid or too slow, the teacher can change this rate in this interface. Thus, the teacher is always in control and the program is really just an assistant. Influencing which exercises come next may seem to contradict the adaptive component, however it gives the teacher the possibility to let the student practice certain exercises that belong to a certain instruction or lesson, thereby making it possible to let the program focus on certain exercises. This opens a wide range of possibilities to use the application in multiple ways in the classroom. More on this in section 3.3.4.

3.3.2 Content creation

Some applications are designed for only one purpose, making it difficult to use different content. This program attempts to make a structure that can be used even when the content is different. This is done by creating sort of templates per exercise that can be filled with different content. The content is created as dynamically as

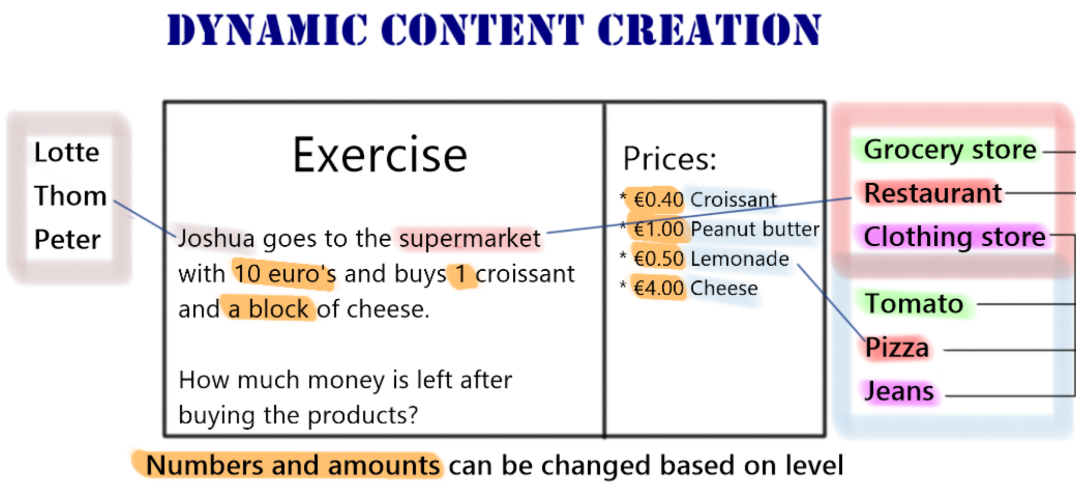


Figure 3.3: Dynamic Content Creation example

possible. Making sure that the questions a student receives will differ per training session. An example of the possibilities of dynamically created content is given in figure 3.3. This is a context-rich problem with multiple steps. There are different possibilities to generate this question. One of the easiest ways is to change the name in the story. More complex changes are the store and the products as these words are linked together. A restaurant has different products than a clothing store. These words need to be linked together. The price of the products can also differ per exercise, making a completely new sum to solve. Also, the price of the products need a price range per product, as a croissant for 100 euros is not realistic. These are constraints that the program can take into account when creating the exercise. Next to the different options in price, the program will take the level of the student into account and find a fitting exercise for that student. For example, exercises that use the numbers 1 and 3 are easier to solve than when using the numbers 0.99 and 2.99.

The principle of making a “web of words” that belong together is widely used in Natural Language Processing. Together with information about the function of the word, this principle can also be used when creating language exercises such as grammar.

The prototype described in the next chapter consists of a specific example of content: Arithmetic exercises without any context. But the general idea is that different content can be implemented in the application. Some might not be using all the features such as dynamic content creation and will therefore consist of static questions, but they can still be implemented in the program.

The most challenging exercises are the ones without a fixed amount of answers, as multiple sentences are hard to check for an AI. The program can still ask these types of questions however, they will need to be checked by the teacher or parent. When there is a mix of questions this should not be an issue, as the program can still adapt in real-time and adapt on the hand-checked answer in a later stage. It still decreases the amount of checking for the teacher.

As there are different types of questions that can be asked, different subjects can be implemented in the program. For every subject a different level per student is possible, making the program adaptive on multiple subjects. Some of the levels may even be influencing each other, for example, the level of reading influences the level of grammar but also the level of context-rich math questions. As the question has different dimensions, this is something that the program can also use to give the exercises in the zone of proximal development. More information on the zone of proximal development is given in section 3.3.3.

Whatever the content is, it needs to be structured internally. First of all the content needs to belong to one or more subjects (Math, English, Biology etc.), within these subjects, branching subjects will emerge. Thereafter, the content will need to be connected to other content. It needs this structure/hierarchy so that the application can use it to decide upon which content to give to the user to make the application adapt towards the level of the user. The content is filled in in such a way that the AI can create multiple exercises. Techniques used are inspired by the NLP techniques to create sentences and by using rules from mathematics. The adaptive behaviour is done by looking at the progress of the user and taking previous results into account. This is not only done by looking at how well the exercises proceeded but also things like response time can be taken into account. How these features relate to each other, differ per subject and even per branch subject. The program is set up in such a way that these parameters can be changed per content unit at any time by the developers and in a bit more limited way by the teachers.

There are different types of content, the most important distinction being exercises vs support information. The exercises provide a question to be answered while support information is connected towards the exercises. If the user knows and understands that information, the exercises related to that support topic can be correctly made. For example, the numeric line that will be used in the prototype is support information and the sum is the exercise. The numeric line is explained in detail in section 4.4.4.

3.3.3 Adaptation and feedback

Adaptation is one of the key elements of the application and is heavily influenced by the zone of proximal development by Vygotsky [2]. He defines this zone as the following: ‘the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined under adult guidance or in collaboration with more capable peers" (p. 86) [2]. While giving exercises in this zone, it is assumed that the learning progress is optimal. The application attempts to move to this zone of proximal development by adapting based on data retrieved from earlier interactions. This data could consist of a combination of accuracy score, response time or even mouse movement, and emotional state of the student. When the application reaches the limits of independent problem solving, it gives the child a bit of guidance in the form of feedback. This feedback is dependent on the subject at hand. For example for solving simple arithmetic, displaying a numeric line would be considered guidance. If this guidance is also failing, there is still guidance from the teacher as they can then keep a close eye on the progress of the student. If the exercises are too hard, thus outside the zone of proximal development, the application will lower the level of the exercises.

Guidance is important while being in the zone of proximal development. This can be achieved by giving feedback while solving the exercise at hand. Most feedback given in schools is delayed feedback by correction of the end product. Ball [12] argued that it is wrong to only give feedback at the end product. It is important for the learning process to also look at the steps in between even when the answers are correct. This can also help with making misconceptions visible and encourage correct strategies.

For good feedback or guidance, you need to have an understanding of how to solve a certain exercise, which steps to take and what kind of errors can be made thus making feedback generation more complex as there are different correct ways to solve an exercise. The program can predict common mistakes, however, this prediction needs to be designed by hand and is highly influenced by the content. Common mistakes like switching one number or letter are quite easy to analyze just like switching the minus for a plus sign. When these types of errors are spotted, the program will give a hint like: 'did you switch the sign?'. These hints will be given in speech and text, making it accessible for strong and weak readers. Furthermore, the program will indicate if possible where the user should pay more attention by using an animation towards the point of interest. For more complex feedback, general rules will be less applicable. Therefore, this will need to be designed per category of type of exercise. As mentioned before, the predictions of common mistakes and misconceptions will need to be designed by hand. This is partially true. There are machine learning techniques available that are adept at grouping data. They might also be able to group mistakes caused by misconceptions. However, this is future work as it includes data collection, data labelling and an algorithm that will need to be developed.

3.3.4 Usage of the AI assistant

The primary focus lies in using this application in the classroom. Extensive research suggests that fully guided instruction is an effective instructional strategy for novice and intermediate learners [51]. Therefore, we look at ten research-based principles for guided instruction [52] and discuss how this application can assist.

Begin a lesson with a short review of previous learning: Daily review can strengthen previous learning and can lead to fluent recall As the application adapts towards the level of the user, it automatically practices recently learned exercises. It aims to present exercises that are correct a majority of the time but still have incorrect responses. These would be exercises that were previously explained and practised. This is the zone of proximal development. Furthermore, as mentioned before, the application can be customized. The teacher can select specific exercises that the children first need to solve before moving on.

Present new material in small steps with student practice after each step: Only present new material incrementally, and then assist students as they practice this material. The adaptive behaviour only increases the level of the exercises if the previous exercises are almost mastered. Within the exercises, there are different levels of support. The support level decreases as the students are progressing. The teacher may also choose to fix the type of exercises that they have to practice and only activate the adaptive support within the exercise type. The teacher will get an overview in real-time on how the students are progressing and assist the students in person if necessary.

Ask a large number of questions and check the responses of all students: Questions help students to practise new information and connect new material to their prior learning. The idea is that after instruction and some practice, the application can be used to practise at their own speed. The responses can easily be checked in real-time by the teacher in their own interface. The application makes it easy to give the students a large number of exercises and all responses are checked immediately. As the exercises are generated, there is no limit in how many exercises a student could practice.

Provide models: Providing students with models and worked examples can help them learn to solve problems faster. The teacher can give instruction based on these models and worked examples. The application will use the same models when giving support. An example of the usage of such a model would be the numeric line when practising simple arithmetic. This principle is explained in detail in section 4.4.4.

Guide student practise: Successful teachers spend more time guiding students' practice of new material. The goal of the application is to reduce the workload of the teachers, to give them more time to spend on guidance while the students are practising with new materials. For example, teachers do not have to manually check homework and analyze the results. The application checks the answers in real-time and gives an overview of the analyzed results to the teacher. It even gives notifications and recommendations about which students might need guidance.

Check for student understanding: Checking for student comprehension at each point can help students learn the material with fewer errors. It may be assumed when a student has a high accuracy score on a category, the student fully understands how to solve the exercises. The application itself does not have an understanding if the students are taking the correct steps for the correct reasons. For example, a student can solve an exercise on their calculator and therefore always gives the right answer. This is an important task for the teacher and the application can only give an indication but no real conclusions.

Obtain a high success rate: Students need to achieve a high success rate during classroom instruction. The adaptive algorithm is programmed in such a way that the students will be correct most of the time, hence it is searching for the zone of proximal development. Therefore obtaining a high success rate, but still, they don't practice at a too easy level.

Provide scaffolds for difficult tasks: The teacher provides students with temporary support and scaffolds to assist them when they learn difficult tasks. This behaviour is included in the application by using different levels of support per category. Based on the progress this support is decreased or even increased when the student is decaying in level. Furthermore, the teacher receives feedback from the application concerning the progress. This feedback can be used by the teacher to provide directed support and scaffolds.

Require and monitor independent practise: Students need extensive, successful, independent practise for skills and knowledge to become automatic. The application is strong in independent practice, as the students are working at their own level as the exercises are created based on their current level. The teacher can easily monitor the progress of all students even though they may work on different levels.

Engage students in weekly and monthly review: Students need to be involved in extensive practice to develop well-connected and automatic knowledge. The application is suitable for review as it automatically gives some exercises that they have already mastered. When they make mistakes in these exercises, it will automatically increase the chances of facing such an exercise again. Giving the student more possibilities to practice this exercise. The further the student is progressing in the program, the lower the chance that easily solved exercises will reoccur as they are assumed to be well established. As mentioned before, the teacher may also influence which content needs to be practised. If it is noticed that there is a knowledge gap, the teacher can direct the program to practice a certain type of exercise more.

From this overview, it is clear that the application has the opportunity to support the teacher with these instruction principles. To make these principles practical, one example will be given of a realistic lesson where the application is integrated. This lesson will follow the Explicit Direct Instruction (EDI) strategy [53]. This strategy is used in many schools and has a strong research-based foundation. For example, the 10 instructional principles [52] are integrated into EDI.

Every lesson starts with the objective of the lesson. For example: “At the end of the lesson, I know how to solve additions that cross 10, such as 8+5=13”. The teacher sees in the overview that 2 children can already solve exercises that cross the 10 and challenge them to practice with the application during the instruction as they already have mastered the concept, or gives them a completely different task to work on.

The teacher’s goal is to activate knowledge that is related to this objective. The teacher could use the application for practice with addition problems that do not cross the 10. The teacher could restrict the application’s adaptive behaviour, and it should not give exercises that cross the 10 or even harder exercises. It can still adapt the support level of the exercises, making sure the success rate stays high. The instructor would keep track of the results by looking at the their interface.

If the teacher notices that the exercises are too hard for the children, they could still change the lesson plan and spend more time on these exercises as they are necessary to master before moving on. When this activity is successful, the instructor will begin explaining the objective and conveys the skills necessary. They give a few guided examples of exercises. These exercises are solved together with the teacher, for example on the digiboard. When progressing, the teacher leaves out a few steps as the children need to take these steps on their own.

The instructor can choose a certain exercise in the application with a certain support level and ask the children to solve this exercise. The instructor will get an overview of the results and can explain more if necessary even to only a few children. Thereafter the students will practice a designated amount of exercises related to the objective and if they are performing well, they can practice on their own with the adaptive behaviour of the application. Children that are struggling will receive more instruction and guidance from the teacher. The application can analyze their results and provide recommendations for their instruction, for example, what type of mistakes were made by the majority of children. Eventually, all children will work with the fully adaptive application. This can have the consequence that some children will practice the exercises that do not cross the 10 more than the exercises that cross the 10 as they still need more practice in those exercises before proceeding. However they are still practicing within their own zone of proximal development.

As can be shown in the example, the application can be useful for real-time signalling. A teacher keeping track of 30 students at once will likely miss out on things. The application can assist with potential oversights. The application is not replacing the rounds of the teacher through the class, as the social component is also important. That said, the application would give the teacher a head start. Based on the analysis of the application, it is clear what the student is struggling with and what the student has mastered already. The application could even suggest what type of misconception the student could have. Based on this information the application could even suggest activities that would be useful to tackle this misconception.

Lastly, the application could also be used as a homework assignment. As the program is giving questions that are in the zone of approximate development, the student is given questions that are challenging but not too difficult. The application can be customized towards a certain homework assignment just like with the normal practice sessions. This gives the teacher the flexibility to let the students practice what is important at the moment. This might be useful if they have just learned something new or as preparation for the next lesson.

3.3.5 Challenges for implementation

One of the challenges that teachers could face with this program are students intentionally making errors to lower the difficulty of exercises. This could be countered in different ways. First, give students the responsibility and make sure that they understand the consequences. This could be done by educating them about the program, that it will adapt towards the students level and therefore give questions that are a bit challenging for the student. When they are too challenging, discuss this with the teacher instead of intentionally giving incorrect answers. Another solution would be to limit how much the level can decrease. For example, there are certain categories that a student has mastered and would never have to go back to that level. Instead of fixing the lowest level, the application could also notify the teacher when the level is dropping dramatically compared to the earlier interactions.

When working with the application, it will require investment from the teacher.he teacher should invest some time in the program to understand all the settings and customization. The program should also make the lessons more efficient and ultimately save time. The application is mostly online and uses a computer,however not all teachers are comfortable with using technology, especially new technology. Therefore, it is important to give the teacher the possibility to follow training and to get help afterwards. When new features are introduced, the training could be resumed as well. This also means that the interface of the application should be intuitive and easy to use. During the first use, the application should explain itself by giving a small tour through the basics. In addition, the design of the application should be simple and intuitive in order to make the application easy to work with even without (much) training. As the workload of the teachers is high, the amount of training time should be as minimal as possible. Short explanatory videos could also help.

3.4 Summary

In this chapter a global design is proposed to create a framework for applications that can be used in an educational setting. Three tasks of a teacher are identified: administration, analysis and content creation. Each task is discussed in detail on how AI can help with this task. These tasks are the basis for an application that can help the teacher with monitoring the students and help to adapt the education towards the level of the student. The global design shows that multiple subjects can be supported, and different types of exercises can be generated. The students are provided with adaptive content that adapts to the zone of proximal development. The teacher can use the application as a second set of eyes in the classroom, as it assists with the administration of the subject that the application provides. It also assists with monitoring the students by providing different results of the analysis. The teacher can get real time notifications about the progress of the student and can be nudged by the system when a student likely needs help. The application gives a good overview of the progress of the class. This global design can be used as an inspiration to develop an application. In chapter 4 such an application is developed with the global design as starting point. The focus lies on an application that can help with arithmetic in primary school.

Chapter 4

The prototype: adaptive math application

With a top-down approach, the idea of an adaptive application to assist the teacher is detailed, resulting in a practical working prototype. The prototype is an embodiment of a more general idea that can be used in a variety of situations. This global design is described in chapter 3. First the goals of stakeholders are given in section 4.1, this is an addition to the goals described in the previous chapter, section 3.1. These goals are worked out in requirements in section 4.2, these requirements are also a refinement from the requirements that are discussed in the previous chapter, section 3.2. These requirements and goals are also described in more detail in a running text in section 4.3. Thereafter, the application that is developed as a prototype is described in detail in section 4.4. The design of the prototype is discussed in detail on a usability level and a technical level. The focus lies on the algorithm that makes the application adaptive. The prototype is validated in section 4.5 with a simulation and a group discussion. The simulation is held to understand how the adaptive algorithm of the prototype is responding in different situations. A group discussion is held with a multidisciplinary team about AI in the classroom and the proposed application, resulting in a summary with the prominent findings of this group discussion.

4.1 Goals

The next goals are additional to the goals described in section 3.1 and are formulated specific for the prototype. These are drawn from the design guidelines in the problem analysis but only focuses on the functionalities of the prototype.

Teachers

- Reducing the workload of the teacher by providing a tool that they can use to give adaptive math exercises.

- Being able to make the educational activity: practicing arithmetic, more adaptive towards the individual students.

Students

- Receiving practice mathematics materials that are adapting towards the zone of proximal development of the student.
- To practice addition and subtraction at speed.
- Giving the students guidance when they have trouble solving the exercise.

Developers

- Building an example of a modular system that can easily be expanded to different subjects and where the back-end is ready for implementing the teachers interface.

Content creators

- Creating a structure for arithmetic exercises that can be used by the adaptive algorithm.
- Dynamically creating arithmetic questions based on a certain level.

4.2 Requirements

The goals of stakeholders are translated into requirements of the prototype. The requirements are grouped per topic. These are quite similar to the requirements in section 3.2 but made more specific for the prototype that focuses only on the interface of the student and the adaptive behavior of the application. Therefore, the interface of the teacher is neglected in this design.

Adaptive The application should adapt the math exercises based on the level of the students without the need of interference from the teacher.

Administration The application should keep an administration of the results. This administration should be easily accessible.

Analysis The application should analyze the results and adapt based on these results the level of the exercises.

Content The application should be able to generate arithmetic exercises based on the math level of the student.

Development The application should be modular, meaning that extra functionalities can be added easily.

Development The application should consist of an interface for the students.

Usages The application should be simple to use and self explanatory.

For the prototype, the focus lies on mathematics, more specifically simple arithmetic: addition and subtraction. This focus provides a few advantages: first these sums are easy to generate and to check the correctness of the answers. Second, for this subject most students need plenty of repetition, as it is the basis for further arithmetics and mathematics in general. Third, arithmetic exercises are representative for a real-world exercise type. Therefore it would be useful to develop an application that makes it easy to practice addition and subtraction while the exercises are fully adapting to the level and behaviour of the student.

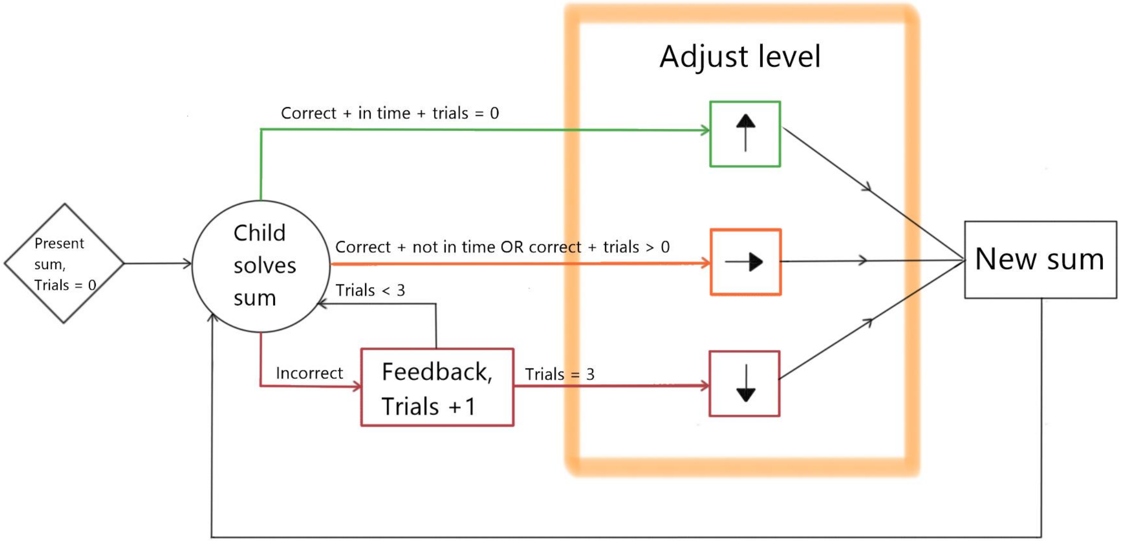


Figure 4.1: Interaction between student and the application

4.3 Extensive explanation of goals and requirements

The goal of the application will be to practice addition and subtraction at speed and give the students some guidance when they have trouble solving the exercise. All sums that are given should be adapted toward the level of the user, hence in the zone of proximal development. The goal is to improve the level of the user by giving more challenging sums. However, the application should also be able to decrease the level as users can stagnate or even deteriorate. Therefore, the application can not follow a simple route for increasing the level. The application should be able to give feedback according to the amount of interaction with the application. This feedback should help the student better understand the given equation and provide some hints or tips on how to solve it or how to address any mistakes. One equation at a time will be given and the user can proceed after solving it correctly or after making too many mistakes. The interaction that leads to a level adjustment is made visible in figure 4.1. The prototype focuses on correct and incorrect answers. When the level is adjusted, the chances of getting a specific equation are changed. The program internally uses a distribution that corresponds with the chance of a certain type of category being picked. One category consists of similar exercises, and an exercise is randomly generated. How much distribution is moved per interaction depends on the learning rate and the amount of distribution that was already present in the corresponding level. More details on how this algorithm works can be found in section 4.4.7.

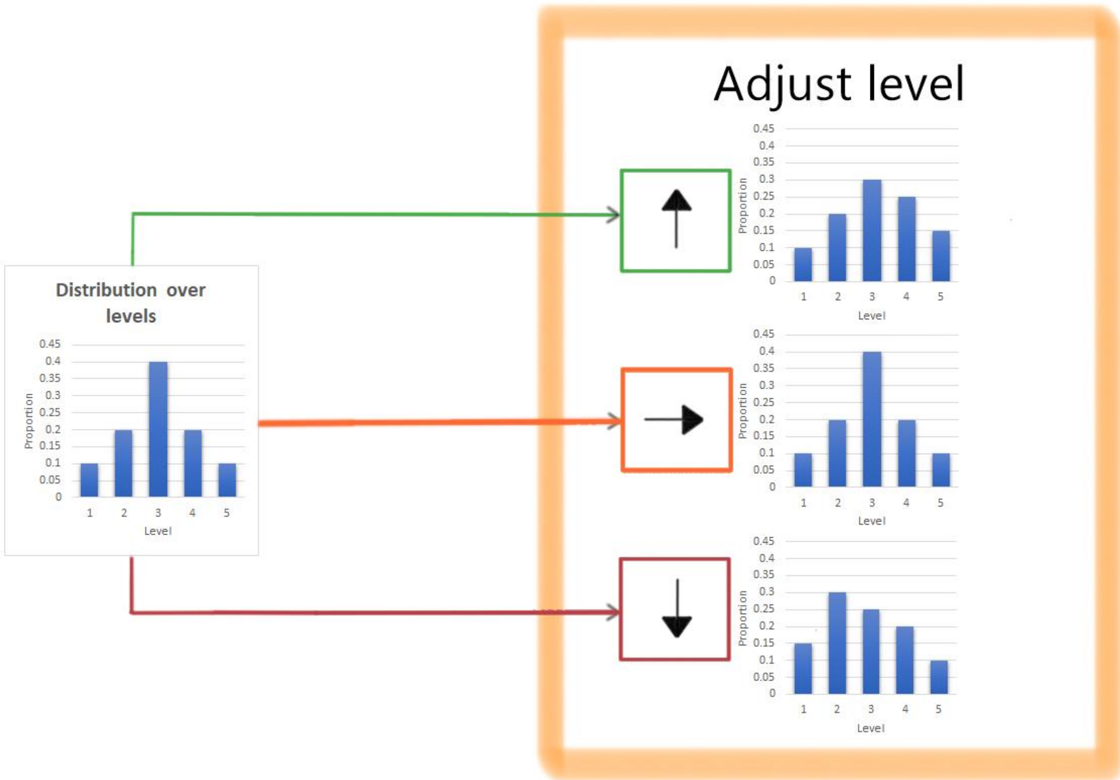


Figure 4.2: Example of a level adjustment

Figure 4.2 gives an example of a level adjustment. Every level has a chance to get picked, in figure 4.2 level 2 has a 2 out of 10 chance to become the next equation. Based on this interaction, the chance either increases, stays the same or decreases. An example of a possible adjustment is given in figure 4.2 To display the equation to the student, the application needs to have an interface that can interact with the student. Equations should be displayed and input will need to be analyzed and stored. The information will need to be stored per student. An adaptive algorithm will decide upon which equations will be displayed and given to the student. The application will need to be intuitive and simple, as young students will work with the application. This is also why the application should look attractive and modern as students need to be motivated to work with the application.

4.4 Application

In this section, the application is discussed that will be a vessel for the adaptive algorithm. The development of this application is an ongoing process and will be explained in detail below. All design choices are described with corresponding argumentation. This section is written top-down where the most general choices are explained and thereafter more details are given. Eventually, the technical details will be discussed.

4.4.1 Type of application

In order to build an application that can be used in schools, the application should be easily accessible. There are multiple options to launch an application. The application can be either online on a website or offline as an application on a device. Combinations are also possible, as applications can be installed but still communicate with a server through the internet. Figure 4.1 can be used as a reference for the envisioned behaviour of the application.

A purely online application has multiple advantages. The data is collectively stored in the cloud and can be retrieved from any location. This makes it possible to collect, store and process data in real-time as the application is data-driven. The data is available from every location through the online portal. This makes sure that personal progress is accessible on every device, even if the device is not connected to a certain network, making it possible to practice at home as part of homework or as extra practice. The online application can work in different ways depending on where the heavy lifting is done. This means that the server can process most of the data or the browser can process the data upfront before sending it to the server. The latter means that it will consume more of the student's computing power. For a standard computer this would not be a problem, but for tablets or phones, it could cause delays in the application. Therefore, the heavy lifting is done at the server-side. This also means that the source code is not completely visible to the user, making it harder to copy the whole application for malicious Internet businesses. As the application is rather small compared to video games or other resource heavy programs As a result, there are no performance problems expected when the heavy lifting is done by the server and accessed through an internet browser. Servers can also easily and quickly be scaled up. The disadvantage is that it is reliable on an internet connection. If the connection is down, the application can not be accessed anymore. If the school has a stable and reliable connection it should not be an issue.

Another option would be to build an application that can be installed and sends data while connected to the internet. This application can function even without an internet connection. The drawback is that one would need to install the application on the device and old versions of the application may still be active when updates are not honoured. Also, different operating systems require different application versions. To compare, online browser applications run on nearly every operating system. Considering the advantages and the disadvantages, the application would ideally be completely online as it is a data-driven application that should be easily accessible from different devices. For this research, the application should be accessible by the computer, laptop, chrome book or tablet, though the application can still be accessed on the mobile phone, due to the smaller form factor it would not be a priority to optimise the design for use on mobile phones. In addition, most schools work with computers, laptops, chrome books or tablets and not with (personal) phones during school hours.

Figure 4.3: Login screen

4.4.2 Log in

For an online application, it is important that the application can keep track of the progress of a specific student by using secure credentials. Therefore, the application needs a personal log in. As the application is targeting young children (ages range from 4-12) the login should not be too complicated but still safe to work with.

This can be done in different ways. Every student can get their own credentials to log in. As this login portal is available for everyone with an internet connection, it is important to use strong passwords. Preferably a generated password as they are in general harder to guess than when people pick their own passwords. This could give some trouble for the youngsters as they should remember and type long and strong passwords. The consequence will be that the passwords will be written down. This is not a problem in a school setting. The teacher can keep a list and make cards with the credentials for the students. These cards can also be distributed to the student's parent or guardian. If the credentials leak, no real harm can be done with a generated password. For the login, a username needs to be created, this can be an email address or a username. Usernames are normally shorter than email addresses and are less harmful when leaked.

Another option would be to make a login for the school or teacher and a second login for the children. This makes it possible to create a secure way of logging in within the school environment. Children can thereafter log in by clicking on their own icon or inserting a short PIN. The drawbacks are that the children can easily log in to other accounts and it is not possible to log in outside the school.

Considering the pros and cons of different login systems, the login system will create individual credentials. As it makes it possible to register classes and individuals by a register page. Leaving the option open to also work with the application outside the classroom setting. Many online applications that are available nowadays use individual credentials to log in for young people. Therefore, it will be assumed that even younger students will be able to log in.

As mentioned before, safety is important. Therefore, the passwords are encrypted and saved in the database, this means that even if the database was hacked, the real passwords are not acquired. The details about these security protocols can be found in the technical background section. When someone successfully logs in, the session will be valid for a fixed amount of time, resetting after each interaction. Students that do not log out will be automatically logged out after a certain amount of time. As the computer they work on would be used by many students, leaving them logged in is not preferred. The application does not save any personal information except the username, email address for resetting the password, hashed password and information about the practice sessions. The email address will not be encrypted as it is needed to contact the users once in a while. The real safety risk is the leakage of the email addresses that can be used for phishing and sending spam.

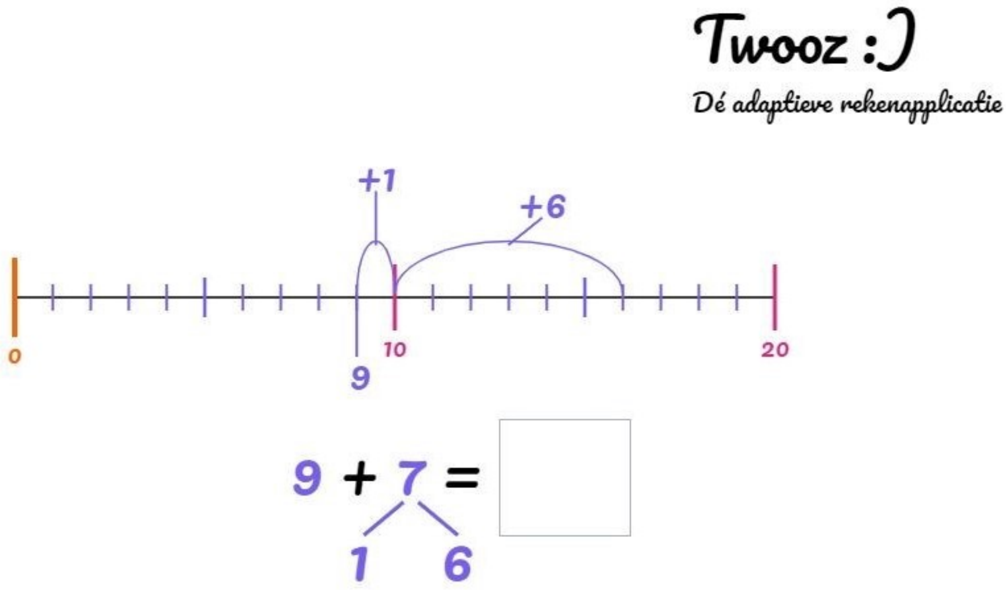


Figure 4.4: Appearance of the application

4.4.3 The appearance of the application

The login page will have a simple appearance with two text areas for the username and the password, as can be seen in Figure 4.3. After a successful login at the login page, the main application will appear. The application can be divided into two elements. The first displays the exercise that needs to be solved. The second is used to help the user solve the exercise at hand, depending on the level of the user and the exercise this element is not always completely visible. These elements are arranged in the following way: figure 4.4. The interface in general is kept plain on purpose, to draw attention toward the exercise and cause no extra distraction for the user. The idea behind the numeric line and division will be explained in section 4.4.4.

When looking at Figure 4.4, the use of colours is prominent. The idea behind the colours is the following. The numbers are coloured based on the position of the digits. Single digits are coloured purple and tens are coloured pink for example. To help the child understand the position based system behind the numbers. For example, 147 is built based on 100, 40 and 7. To keep everything consistent, the colour scheme will be used in every element that consists of numbers. Therefore it will return in the second element as well; the numeric line and the division of numbers. The colours will always be the same throughout different training sessions.

To choose the colour scheme, a few considerations were made. First, the colours red and green were excluded from the colour scheme as they are bound to be correct and incorrect, or positive and negative. These colours will come back when giving feedback to the user. The three colours that were picked for the scheme should be easy to differentiate between one another. The second constraint was that the colour scheme should also be functional for people that suffer from colour blindness. This constraint is important as the colour scheme is helping the user understand how the numbers are built up and making it easier to solve the sum. IBM [54] presented different colour schemes that were developed to be inclusive and one of these schemes was chosen. The colour scheme is presented in image 4.5, also a representation of how the colours are perceived by someone suffering from colour blindness.

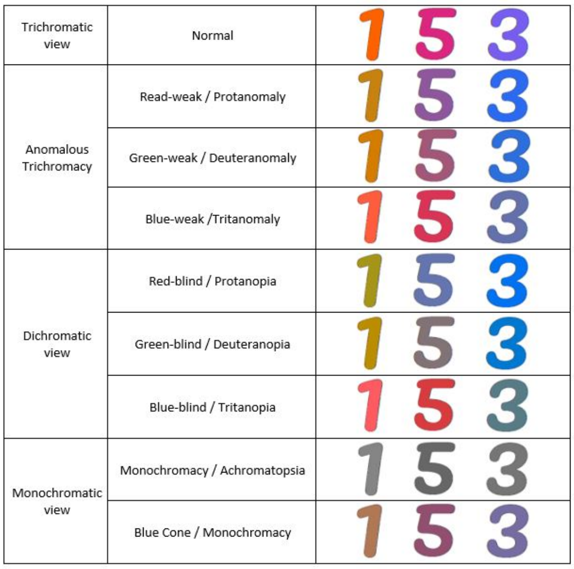


Figure 4.5: Colour scheme in combination with color blindness

4.4.4 Content: arithmetic

The application's goal is to administer addition and subtraction equations. A description on the content and the structuring is given in this section. How this content is created can be found in section 4.4.6 Specific for this content, the numeric line and division of numbers is used as guidance for the students. This section gives a general explanation about the numeric line and the division. How this numeric line is generated in the code can be found in section 4.4.6.

Category	Answer smaller than	Second number consist out of			Crossing		Additional rules	Example
		Units	Tens	Hundreds	Tens	Hundreds		
1	10	X						3 + 2
2	20	X						14 + 8
3	20	X			X			9 + 7
4	100		X				First number = tens	20 + 10
5	100	X						32 + 5
6	100		X					32 + 10
7	100			X				32 + 15
8	100	X			X			32 + 9
9	100	X	X		X			32 + 19
10	1000			X			First number = hundreds	200 + 100
11	1000	X						521 + 300
12	1000		X					321 + 50
13	1000			X				321 + 500
14	1000	X	X					321 + 501
15	1000		X	X				321 + 530
16	1000	X		X				321 + 501
17	1000	X	X	X				321 + 531
18	1000	X			X			325+ 9
19	1000	X	X		X			325 + 19
20	1000	X		X	X			325 + 209
21	1000	X	X	X	X			325 + 219
22	1000		X			X		725 + 90
23	1000	X	X			X		725 + 92
24	1000		X	X		X		725 + 190
25	1000	X	X	X		X		725 + 192
26	1000	X			X	X		199 + 5
27	1000	X	X		X	X		125 + 99
28	1000	X		X	X	X		195 + 109
29	1000	X	X	X	X	X		125 + 129

Figure 4.6: Different categories for addition and subtraction until 1000

Categories

The application is built around the usages of categories. The different categories for arithmetic are displayed in figure 4.6 These categories are ordered by difficulty. This is not set in stone but to develop this prototype, a hierarchical structure is needed. Some categories might be arguable or are almost the same level. For example category 4 might be easier than category 3 if a student discovers the principle that the sums are similar to category 1 except there is an extra 0. For the sake of simplicity a linear hierarchical structure is chosen for the prototype. However, the adaptive algorithm is written in such a way that it also supports different hierarchical structures.

This would mean that when doing category 1 correct, category 2 and 4 will appear more in the practice session.

Discovering the best hierarchy of the categories is out of the scope of this thesis. The categories are linearly connected in the prototype and the simulation focuses on only 9 categories and different hierarchies. This in order to make the results accessible.

Numeric line and division

As mentioned in the section above, the numeric line [29], [28], [27] can be used as a tool to solve the exercise and as help. This tool is also used in most primary schools in the Netherlands and study books. The numeric line works as follows. First, the child will look at the equation at hand. Thereafter they smartly divide the equation. For example, when tackling 9+7, the 7 will be divided into 1 and 6, first, the 1 will be added, creating a round number (10), thereafter the 6 is added. The equation 10 + 6 is easier to solve. This division is given to the user below the sum. Corresponding to the division below the equation, the steps on the numeric line are made, this can also be seen in image 4.4. The numeric line is a visualization of the numeric system. It starts on the left with 0 and every horizontal space between the vertical bars is 1 part. Towards the right, the parts are totaled. Going from left to right is chosen as this is the reading direction in the western world. When using the numeric line, the user will start at the first number of the equation, in the example that number is 9. And jump one bar towards the right ending at 10. Thereafter, 6 is added in the same way, represented by a jump of six bars. The jump is visualized by using half of an oval.

Ideally, the user would be able to make these steps on the numeric line by dragging and dropping. The current state of the application gives the numeric line with the jumps to the user as a form of help.

4.4.5 User interaction with the application

In this section, the basic behaviour of the application will be described after logging in. The user will be asked to solve an exercise. Depending on the level of the user in combination with the level of the given exercises, the user will see the exercise and some support. This support knows 3 levels. The highest level is only the exercise with an empty numeric line. The numeric line is adjusted to the current exercise, meaning that when the exercise is below 20 the numeric line will be given up to 20 and vice versa for an exercise above 20 but below 100. The middle level is when the division is given beneath the exercise. The lowest level is the same as the middle level however with the steps on the numeric line. The numeric line is always visible, even if there is no support given. This is to encourage the student to use the numeric line to solve the exercise at hand.

The student has three attempts per exercise as can be seen in figure 4.1, when failed three times a new exercise will pop up. When making a mistake, the support level will automatically decrease leading to more support until the lowest level is reached. The background colour of the answer box will turn red, and a pictogram like a cross indicates that the answer is incorrect. The pictogram is added as the application should be inclusive for users with color blindness, as red and green are not well distinguishable. The same goes for a correct answer.

When a correct answer or three incorrect answers are given the next exercise will begin. Which exercises and which level of support will be determined by the adaptive algorithm that tends to achieve the best learning outcome.

4.4.6 Technical background

In this section, the technical background of the application will be discussed. To start with which programming language and framework could be used. Thereafter, general security measures that are taken will be presented. The section will be concluded with a general description of how the application is built and specifically the development of the numeric line.

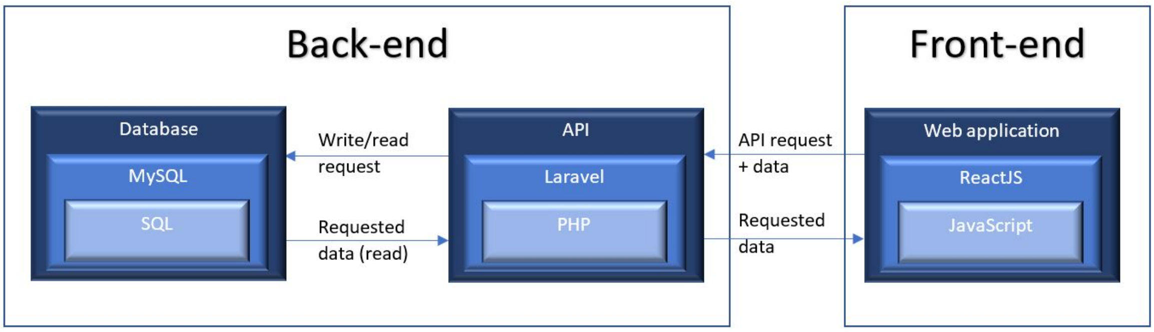


Figure 4.7: Technical overview application

Programming languages and frameworks

At first, the application was built in python with Django as the framework. The front-end used JavaScript. However, the JavaScript code was completely visible and would have been easy to reuse. It would be preferred to keep this code hidden. Furthermore, the heavy load was directed towards the user through JavaScript. As the application can be expanded in the future, it might not run smoothly on every device, especially on devices with limited RAM. Therefore, a change of programming language and frameworks was proposed and the existing code was partially re-written. To exploit the application, NodeJS is picked. The framework React is used to build the front-end. Furthermore, an API was written in PHP with the Laravel framework to make a connection with the MySQL database and the front-end. Figure 4.7 gives a broad overview of the application and the different elements of the framework and the programming languages.

As the implementation of the algorithm was not heavy, the calculations are done in real-time by the server and the results are returned by NodeJS. The primary code is not sent to the user making it possible to hide most of the code. This also gives the advantage that less code needs to be sent to the user, making the website lighter to load. The reverse side of the coin is that it will ask more about the computational power of the server. The database is filled with information directly from the application (question, amount of trials, time spent on a sum, performance etc.). The machine learning or adaptive algorithm component uses these results from the database and puts the results back into the database. These results are used by the application to predict the next best sums and thereby generate information again to put in the database.

Security

To make the application secure, different considerations were made. The server that hosts the application is protected by the use of SSH keys and two passwords. The SSH key can only be used in combination with a password. When logged in to the server, there is a second password check with a different password to make changes as a root user (Sudo command). The server will be protected by a firewall that only gives access to installed applications on the server, for example, the web server. The web server is configured with the standard settings. The API to access the database will only consider requests coming from the domain name, at the same server and containing a valid CSRF token (cookie). This protects from unauthorized requests targeted at the database. As mentioned before, the password is saved by hashing the password with Bcrypt. This hash function incorporates a salt that protects against rainbow table attacks and has a work factor that makes the hash ”slow”. This is a protection measurement towards brute force attacks and can be increased when the computational power increases. Therefore, it is expected to be secure in the future. The content of the website is only available while being logged in, and this is checked for every interaction with the application. The moment the user is not considered to be logged in, the login page will time out and the user is brought back to the login page. Overall, many security measures are taken to protect the application. The application itself does not form a high risk of being attacked to collect personal information as the personal information that is stored is kept at a minimum level.

Building the application

The application is built in the Web Framework React and the following languages are used: HTML, CSS, JavaScript. The basic idea is that based on the sum, the whole page is dynamically generated. This means that even though there are returning elements, the page is specific for that sum. This has the advantage that the application can display every sum possible without the need for pre-made images, this makes the application more flexible and efficient. Also, style changes can be adjusted with a few lines of code. React is well suited for this purpose as it regenerates on the server and looks at the differences with the load before. This means that it will only change what is new. This makes the website light and it updates without refreshing the pages, which is user-friendly behaviour. The application relies on a given sum and generates content based on the sum.

There are some general rules for the layout. First, the application looks if the answer of the equation is crossing the 100, in that case, there is a need for a number line until 1000. If it stays under 100 one line is sufficient. The answer above 100 can be displayed on the screen and therefore need a higher order of number line. The consequence of a number line till 1000 is that the scale is 10 times smaller and units are not well displayed anymore. Therefore the application will display two number lines. One displaying the 1000 and one displaying a zoomed-in version of the same line. The zoomed line is cropped around the area where the tens and units are added or subtracted. Another important layout rule comes into place when the sum is crossing tens or hundreds (for example 7 + 5 is crossing the ten). Normally the number is divided into units, tens and hundreds, however, when crossing tens or hundreds, there is another division made. This is done to first jump to the whole tens or hundreds and after that add the remaining amount (for example 7 + 5, 5 is divided into 3 and 2). This division is shown underneath the sum but also affects the jumps made on the number line.

Generation of the equations

The equations are generated per category. This is done by using mathematical principles. First two numbers are generated with a pseudo random algorithm based on the max value, in this case 10, 20, 100 or 1000 and some additional constraints. These constraints focus on the second number that only consists of hundreds, tens or units depending on the specific category. Thereafter, the sum is checked based on rules specific for the category. The pseudo-code is given in the algorithm 1. The code might give the impression that it is specific for addition, however to create a subtraction, the addition is taken and flipped to a subtraction. For example, 7 + 5 = 12 belongs to category 3 as it crosses the ten, when flipped 12 - 7 = 5 or 12 - 5 = 7 is created and still belongs to category 3 as it still crosses ten. From the algorithm 1 it can be seen that multiple sums might be generated if the right conditions are not met. However this is not a problem as the generation and checking the sum is rather fast and the application starts generating when the user still receives feedback from the application. The user will not experience any lag. The algorithm might not be the most optimised algorithm as it is possible that there are techniques to generate the sums faster and more efficiently. However to meet the requirements of this prototype, the developed algorithm is sufficient.

Algorithm 1 Generate sums

```
for Every interaction do
    Number1 = 0
    Number2 = 0
    while (Number1 == 0 & Number2 == 0) or specific conditions* do
        Number1 = Random number**
        Number2 = Random number**
    end while
end for
```

* Possible conditions (can be combined):
Number1 + Number2 $>$ 10
Number1 + Number2 $>$ 20
Number1 + Number2 $>$ 100
Number1 + Number2 $>$ 1000
Sum crosses tens
Sum crosses hundreds
** Random numbers are generated based on the rules of a category. For example, category 1 a number is generated between 0 and 10, and category 16 number1 and number2 are generated between 0 and 1000, however, number2 only consists of hundreds and units, not tens.

Generation of the numeric line

The numeric line is generated live. This has some advantages over using existing images. First, the numeric line can be customized completely; colour schemes, fonts or size can easily be adapted. This gives schools the possibility to adjust the application to their needs. For example, schools use different methods and these methods use different colours to point at single digits, tens or hundreds. This means that the application can meet the current method that the school is using, which makes it easier for a child to switch between the application and the textbook. It can be confusing if the book uses the colour purple for tens and the application uses purple for hundreds. This customization will also help to create a mobile-friendly application as font size for example can be connected to the screen width or height. Second, the application does not have to save all possible numeric lines images. Especially when every equation has its own numeric line with steps on it, this amount of image would grow exponentially.

Creating an image can be done in different ways. There is a canvas element in HTML5 that allows you to draw based on a coordinate system. It also supports text and images. As a numeric line is based on lines and text, canvas elements can be used. The canvas element can be influenced by JavaScript, making it possible to create real-time changes without refreshing the page.

Quite similar to canvas elements are Scalable Vector Graphics (SVG) and vector-based graphics in XML format. It also works with a coordinate system and supports text and vectors. Images can be created by software that draws vectors or by creating the XML. Overall, Canvas is more suitable for interactive designs and uses JavaScript by default. The differences are rather small making both elements suited for drawing the numeric line, choosing between the two is a personal preference. In this case, the canvas elements from HTML5 are chosen. To conclude the comparison, the numeric line can also be created by making div elements and colour the borders. As this is not the main purpose of divs and some properties about positioning differ per browser and update, this will not be used.

4.4.7 The algorithm in detail

In the algorithm 2 pseudo-code is given for the developed algorithm. To move the distribution around the use of integers is necessary. The consequence of not using integers is that the decimals will keep growing. Rounding would cause variations in the total distribution and with repeating this algorithm these variations could grow quite large. Therefore, integers are chosen to work with. This does lead to the question of how to handle the distribution evenly based on a changing percentage and a changing learning rate. To tackle this problem, the following solution is used: the distribution is rounded down and summed. The total is subtracted from the total amount of shifting distribution, leading to a rest amount. This amount is distributed over the categories by ordering the categories based on the proportion and handing out the remaining distribution. For example, category 1 was correct and will shift

Algorithm 2 Adaptation algorithm

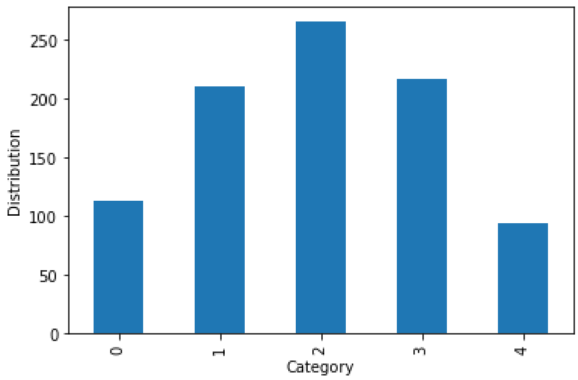
```
for Every interaction do
    Choose sum based on distribution over categories
    User inputs a correct or incorrect answer
    Calculate weight shift with learning rate
    if User input is correct then
        Weight shift divided over more challenging categories
    end if
    if User input is incorrect then
        Weight shift is divided over less challenging categories
    end if
end for
```

19 towards the other categories. Category 2, 3, 4 and 5 will take respectively 30%, 50%, 20% and 0%. They will take 5, 9, 3 and 0. Leading to a rest of 3 that will be distributed over 2, 3 and 4 in the following order: 3, 2, 4. The amount of distribution that is shifted depends on the learning rate. This heavily influences the adaptive behavior of the application. The learning rate can be changed per student as some students would be advancing faster than others.

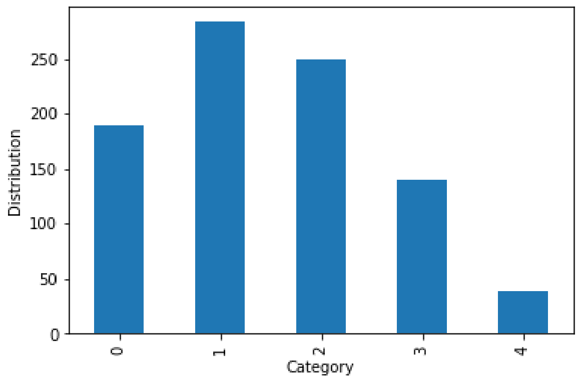
The idea behind the algorithm is that when a child almost masters the category, it will move on to the next category. This is done by creating a normal distribution around the category that is almost mastered. When the algorithm is used as proposed the distribution centers around 50% of the time correct. Meaning that a child will get most exercises about the category that was only correct half of the time. If the student only gets half of the exercises correct while practising, it could be quite demotivating. Therefore, it is important to be able to move the centre towards a category that the student has nearly mastered, for example, around a category that is correct 75% of the time. This is done by penalizing mistakes more than rewarding correct responses. The distribution is shifted to the left every time and only to the right half of the time. The effects can be viewed in figure 4.8, the 50% rule leads to centralising around 75%. To show that this shift can be modified to serve different needs, 25% and 10% is added. This effect could also be established by using 2 separate learning rates, for example, 0.1 for correct and 0.2 for incorrect.

4.4.8 Connectivity of categories

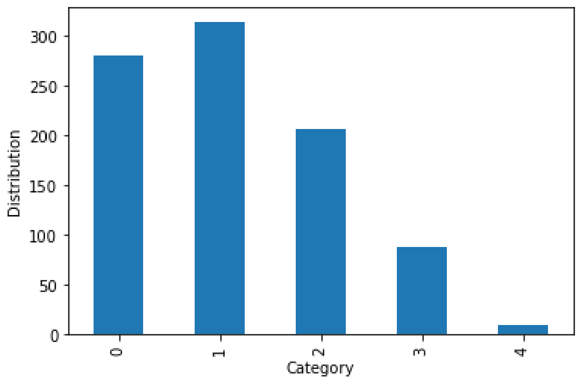
There are 3 types of connections within the categories. For an overview of these different types, three examples are given in figure 4.9. The first type is the simplest one, linear, where the categories are lined up. Category 1 influences category 2,



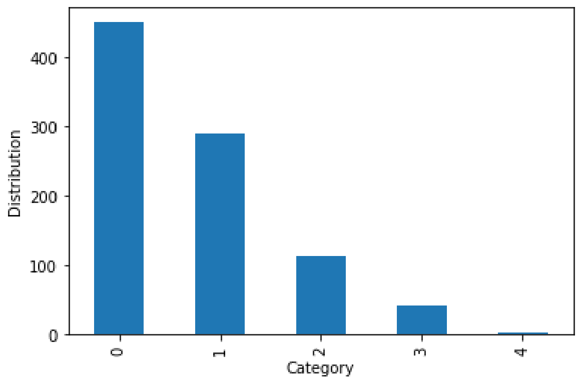
(a) Without rule



(b) With 50% rule



(c) With 25% rule



(d) With 10% rule

Figure 4.8: Shifting the average distribution by adding an extra rule. The 5 categories have a respective skill level of 1.00, 0.75, 0.50, 0.25 and 0”

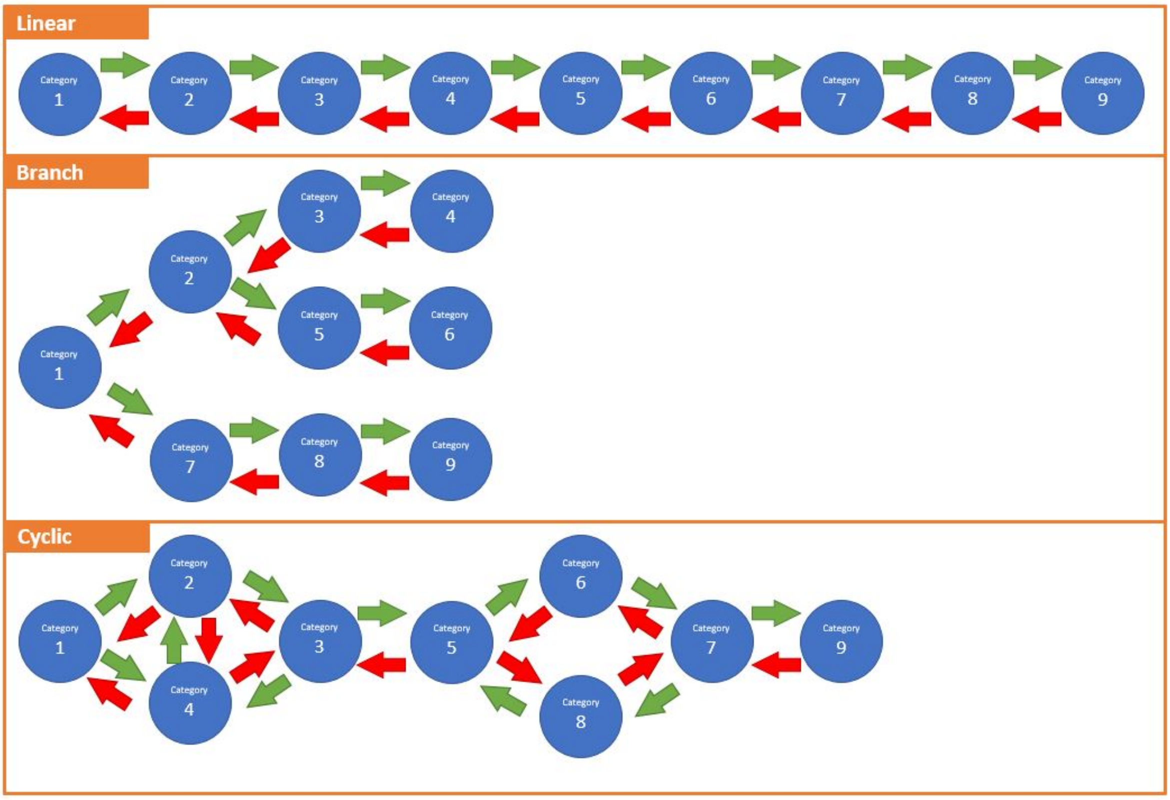


Figure 4.9: Three different hierarchical connections

category 2 influences category 3 etc.. This reflects when the categories are completely building upon each other. The second type, branching, is where category 1 influences category 2 but also category 3 and might even influence category 6. However, the proportion will always move to the right, never backwards. This reflects the situation where learning one category will give you the basis to solve multiple other categories. The third type, cyclic, is where category 3 can influence category 1 and category 4. This reflects when the learning process doesn't have a complete hierarchy. Cycles are occurring in the structure. The application will use the first type of connections in the first version. However, it can be assumed that when the application is developed even further, the second type of connections would be used as some aspects of categories are the basis for other categories.

4.4.9 Summary

In this section a prototype is developed that is a specific implementation of the global design that is discussed in chapter 3. This prototype is discussed in much detail and is also carried out. The prototype can be found on the following website: www.twooz.nl, please contact the author of this thesis for a demonstration or a demo account. The prototype focuses on the interface of the students and the implementation of the adaptive algorithm. The next steps to take with the prototype are to develop the teacher interface and to add more content, for example, multiplication and division could be added. The current prototype is functioning well enough to be used in a field study. Multiple children can work simultaneously on the website in their own account and interactions are recorded into a database for further research.

4.5 Validation

Due to the Covid-19 pandemic, the planned field study was not possible. The set-up for this field study and the use of a neural network is described in appendix A. Instead, two other experiments were conducted, a simulation and a group discussion. The simulation had the goal to test, improve and validate the adaptive component of the prototype. The method, results and discussion are described in section 4.5.1. As no humans were involved during the simulation, a group discussion was organised to validate the need for an AI assistant and to gather feedback on the design of the prototype. Details on this group discussion can be found in section 4.5.2.

4.5.1 Simulation

In this section, the method is discussed to test and validate the algorithm that makes the prototype adaptive by using simulations. More details on the developed algorithm can be found in section 4.4.7. The simulation is divided into two phases, the testing and the validation phase. The results of both phases are presented in the result section and are concluded by a discussion on the results.

4.5.1.1 Method

The simulation makes use of agents that resemble the users of the application, different agents were built as students differ as well. The algorithm makes use of a schema of how the categories of the content are connected. Different schemes are tested as the algorithm should work for different content. Thenceforth, an overview of the different agents that are used is given. Thereafter, a description is given of how the testing and validation phase is conducted.

Simulation built

The simulation consists of the AI system powered by the adaptive algorithm and the agents that interact with this AI system. Each interaction and outcome is recorded as well as the change that it will bring to the distribution over the categories. The AI presents a category to the agent, the agent responds with correct or incorrect. After

one interaction the AI adapts based on the new information. Depending on the type of agent, they will have some internal learning process.
The simulation is built in such a way that different settings can be tested and the interaction can be repeated many times. The algorithm works with an internal distribution and the agent has a certain skill set. In the following experiments the goal is to understand if and how well the algorithm can evolve the internal distribution toward the skill set of the agent, especially when the skill set of the agent is developing over time.
The following variables are included in the comparison: learning (including different learning rates) vs non-learning agents, different skill sets from agents, different learning rates from the algorithm, initial distribution and different connections between categories, and the use of the kick-start.

The following learning rates for the agents are used: 0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.5, 0.75, 1.0. The different skill sets that are tested are displayed in table 4.1. Every row is an agent and the columns present the skill per category. For example agent 3 masters category 1 up to and including 5 completely, hence answers always correct but fails at category 6 up to and including 9, hence answers always incorrect. Agent 5 answers 60Finally, different connections between categories are tested. As one set of connections between categories can be configured in many different ways, for the sake of simplicity, one configuration per connection type is chosen for the simulation. This will still give enough information about the behaviour of this type of connection in general. In figure 4.9 the three types of connections are shown. Note that except for linear, an increased number is not directly related to an increase in level.

Agents

The agent is responding to the category that is presented. How they react depends on their skill set and learning capacity. Different agents are built to observe the behaviour of the algorithm. The agents used can be categorized based on the type of agent. The first type of agents are the simple agents. These agents do not learn from the interaction. Per category, it is decided if the agent will answer correctly or incorrectly.
The second type of agent also doesn't learn from the interaction. However, they will respond per category through a stochastic distribution. This means that for example when category 1 is presented, in 90% of the interactions, the agent will be correct and in 10% of the interactions incorrect. Each category can have a different chance of being correct. The third type of agents does learn from the interaction.

	Skill per category								
Agent	1	2	3	4	5	6	7	8	9
1	0%	0%	0%	0%	0%	0%	0%	0%	0%
2	100%	100%	100%	100%	100%	100%	100%	100%	100%
3	100%	100%	100%	100%	100%	0%	0%	0%	0%
4	0%	0%	0%	0%	100%	100%	100%	100%	100%
5	100%	80%	60%	40%	20%	0%	0%	0%	0%
6	0%	0%	0%	0%	20%	40%	60%	80%	100%
7	100%	100%	100%	100%	100%	80%	60%	40%	20%
8	20%	40%	60%	80%	100%	100%	100%	100%	100%
9	20%	40%	60%	80%	100%	80%	60%	40%	20%
10	100%	90%	80%	70%	60%	50%	40%	30%	20%
11	20%	30%	40%	50%	60%	70%	80%	90%	100%
12	0%	100%	0%	100%	0%	100%	0%	100%	0%
13	100%	0%	100%	0%	100%	0%	100%	0%	100%
14	0%	0%	0%	100%	100%	100%	0%	0%	0%
15	0%	0%	0%	50%	100%	50%	0%	0%	0%
16	90%	90%	70%	30%	50%	70%	50%	40%	80%
17	50%	20%	60%	40%	50%	30%	70%	90%	70%
18	70%	40%	10%	30%	90%	60%	20%	20%	10%
19	100%	50%	50%	20%	20%	20%	0%	0%	0%
20	0%	0%	0%	20%	20%	20%	50%	50%	100%
21	100%	70%	30%	20%	20%	20%	0%	0%	0%
22	0%	0%	0%	20%	20%	20%	30%	70%	100%
23	100%	100%	100%	60%	60%	30%	0%	0%	0%
24	0%	0%	0%	30%	60%	60%	100%	100%	100%
25	100%	70%	30%	20%	80%	20%	0%	0%	0%
26	0%	0%	0%	20%	80%	20%	30%	70%	100%
27	100%	0%	50%	50%	50%	50%	0%	0%	0%
28	0%	0%	0%	50%	50%	50%	50%	0%	100%

Table 4.1: Skill set of different agents

Initial distribution over categories								
1	2	3	4	5	6	7	8	9
900	0	0	0	0	0	0	0	0
100	100	100	100	100	100	100	100	100
180	180	180	180	180	0	0	0	0
0	0	0	0	0	180	180	180	180
300	240	180	120	120	60	0	0	0
0	0	0	0	60	120	180	240	300
127	129	129	129	129	103	77	51	26
26	51	77	103	127	129	129	129	129
36	72	108	144	180	144	108	72	36
167	150	133	117	100	83	67	50	33
33	50	67	83	100	177	133	150	167
0	225	0	225	0	225	0	225	0
180	0	180	0	180	0	180	0	180
0	0	0	300	300	300	0	0	0
0	0	0	225	450	225	0	0	0
142	142	110	47	79	111	79	63	126
94	37	112	75	94	56	131	170	131
180	103	26	77	232	154	51	51	26
347	173	173	69	69	69	0	0	0
0	0	0	75	75	75	187	187	376
347	242	104	69	69	69	0	0	0
0	0	0	69	69	69	104	242	347
200	200	200	120	120	60	0	0	0
0	0	0	60	120	120	200	200	200
282	197	84	56	225	56	0	0	0
0	0	0	56	225	56	84	197	282
300	0	150	150	150	150	0	0	0
0	0	0	150	150	150	150	0	300

Table 4.2: Initial distribution of algorithm

The stochastic distribution will still decide if the category is correct or not, based on this outcome the agent will learn from this interaction. This learning is captured by increasing the chances of the category being correct if the agent's response was correct.

Kick-start

As imagined, the initial level is not known to the application, so a kick-start is developed to make the algorithm converge faster towards the level of the student. The algorithm would start with a high learning rate that decays over time. This makes it possible to guess the level of the student faster. This could give the application a kick-start. This kick-start is defined in the following way: the learning rate starts at 1 and decays in a linear way. For example, every interaction the learning rate decreases with 0.05 until it reaches the intended learning rate of 0.2 after 16 interactions. Based on the configurations the amount of interactions necessary to reach the intended learning rate differs, this is subject and purpose dependent. The effect of the kick-start can be seen in the results.

Testing phase

In the testing phase, all combinations of different variables are simulated as this gives around 900.000 different plots. Only the most important results will be displayed in the results section. All simulations are run to give a good impression of the behaviour in different situations. The testing phase is conducted with the first two types of agents.

The first type is used to understand if the algorithm will become stable and how long that will take. The desirable output is that the distribution will shift toward the correct categories. The three different types of connectivity of the algorithm will be tested.

The second type is used to understand how the algorithm is dealing with categories that are not completely mastered or with random mistakes. What if a child is only correct 50% of the time? The algorithm has to find the zone of proximal development. This is the point where a child is still learning from the interaction but is not lost.

Validation phase

The validation phase will be conducted with the last type of agent. In this phase, the focus lies on creating an agent that resembles a child and understanding how the algorithm responds to this type of interaction. As children differ from each other, so do the agents. Some agents will learn fast and some will learn slow. In which cases is the algorithm responding as it was envisioned? The expected output is that some of the agents will not always be served well, for example, the algorithm might not keep up with a really fast learning agent. To make this more concrete, three different learning agents are used. These resemble the difference in learning speeds between children. The agents are named, slow, fast and random. The slow agents learned from each interaction between the 0 and the 5%. The fast agent learned from each interaction between the 5% and the 10%. The random agent learned from each interaction between the 0 and the 10%. These percentages were randomly picked to get less predictive learning results. This is to model a child as they might not learn from each interaction. The opposite is also possible, a child could learn a lot from one specific interaction. The percentage was added to the skill until it reaches 100%.

4.5.1.2 Results and interpretation

First, the results of the testing phase are described. The focus lies on the effect of the separate variables. The effect of the learning rate is described first. By narrowing the learning rate down to a small range, the effect of different skills in combination with initial distributions becomes more clear. The effect of the kick-start is also shown. Thereafter, different types of agents were simulated. Last, the effect or the connections are tested. Thereafter, the results from the validation phase are described. The focus lies on analyzing the behaviour of the algorithm during the interactions. This will be translated to what this behaviour means with practical examples.

Testing phase

As many simulations have been conducted, the results are given per feature. All experiments are conducted with 50% rule, with the consequence that the distribution will centre around a category with a skill level of 75%. The figures are created by using 5 categories with a respective skill level of 1.00, 0.75, 0.50, 0.25 and 0. The categories were linearly connected, category 1 was followed by category 2 and category 2 by category 3. The learning rate of 0.1 was used and the agent did not learn for interaction. The average distribution of over 1000 interactions was taken.

Learning rate The learning rate influences the moment that the algorithm converges. This is visible when using a non-learning agent. The following scenario is used: the whole distribution starts at the first category and the student answers every question correctly.

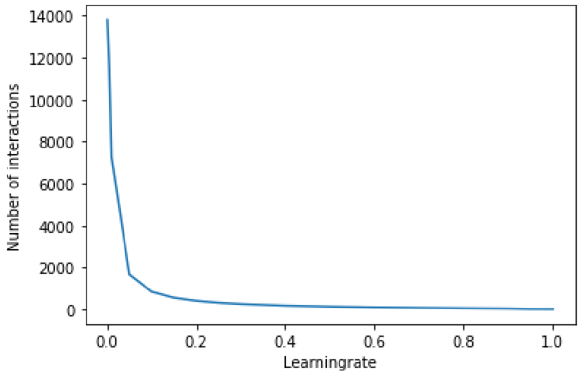


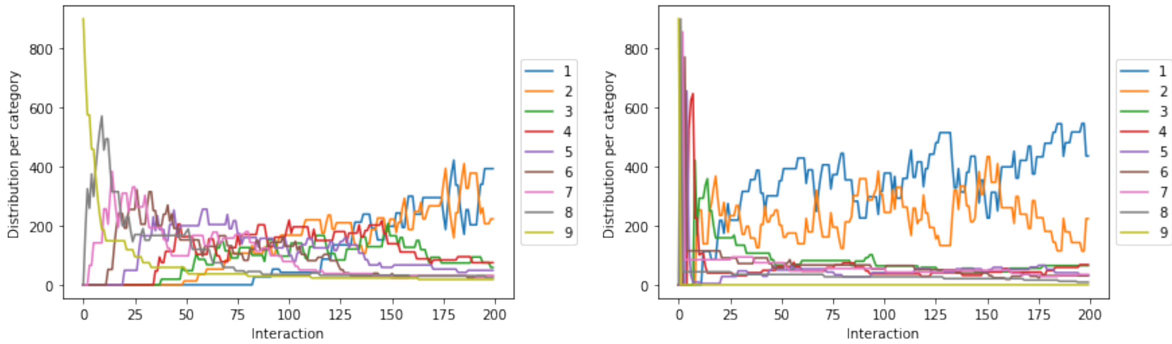
Figure 4.10: Effect of the learning rate, number of interaction needed before algorithm converges from category 1 to category 9

From figure 4.10, it can be seen that the learning rate influences heavily how fast a student can proceed through the application. It would take around 850 interactions to get 50% of the distribution to end up in the 9th category with a learning rate of 0.1. Depending on the educational activity this might be too much or not enough practice. How fast a student has to proceed through the application depends on many factors: those factors include the individual student, his or her capacities, learning speed and also the content that needs to be practised. For example, if a category has only two questions, the student could advance faster than when a category that has many different questions testing the same skill. Future research is needed to figure out which learning rate is efficient for which educational activity. For the simulations learning rates between 0.05 and 0.2 are used as it seems to give a reasonable amount of exercises before ending up in the last category.

Furthermore, the initial distribution is also important. When this is more closely towards the level of the agent, the algorithm converges earlier. This means that it would be helpful to get a good initial distribution per student. This can be done by using information about the student like age, grade and information about the curriculum. A teacher can also help with this distribution. Another option would be to start with a high learning rate and decrease this learning rate over time to the preferred learning rate, using a kick-start. The effects of the kick-start are shown in figure [ADD].

The learning rate is the same for each category, however, it might be possible that every category has its own learning rate as some categories need more practice time than others. The simulations did not take this scenario into account, however, the algorithm can easily use a learning rate per category. Therefore, it would also be possible to climb faster in level and a slower drop in level, depending on the needs of the user working with the application.

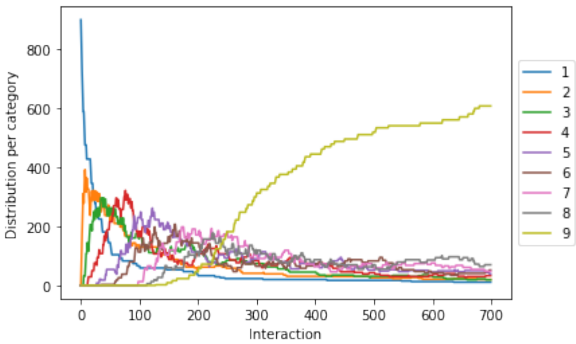
To conclude, the learning rate is behaving as envisioned. It is a parameter to change the adapting speed of the algorithm making the algorithm converge faster or slower. High learning rates like 0.5 and above are moving the distribution around quickly and causing unwanted behavior. The student will not stay in the zone of proximal development but will encounter many exercises that the student cannot solve and when making one small mistake it will directly get many easier exercises as much distribution is moved to the lower category. This is good to keep in mind when choosing a learning rate. Because the student is also learning, a slightly higher learning rate might be wise to choose. The algorithm works with a fixed learning rate, this can be customized per student. However, in the future, a learning rate that is adapting as well might be interesting as the learning rate for a student can vary in time.



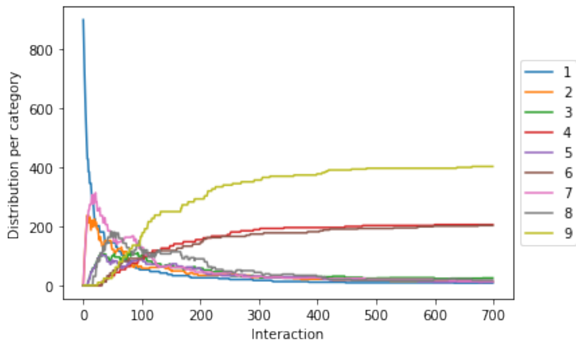
(a) Without kickstart *(b)* With kickstart

Figure 4.11: Skills and distribution mismatch. Learning rate of 0.2, initial distribution starting at category 9 and skill set of agent only category 1 correct.

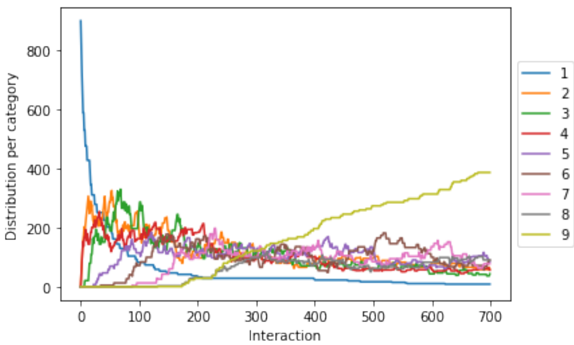
Skills and initial distribution When using categories that are linearly linked, the importance of the initial distribution is shown. In the extreme case that the whole initial distribution was set at category 9 and the agents could only answer category 1 correctly, the distribution has to move through all categories to end up in category 1. In figure 4.11a is an example of this behaviour given. The learning rate was set to 0.2. It is visible that first category 9 is fast decaying. After 100 interactions this category is rarely asked anymore. When 9 starts to decay, category 8 is increasing. When the chances are growing of picking category 8, the distribution is also moved to category 7. Category 7 is still growing when 8 is starting to decay. This behaviour is repeated for the other categories until category 1 is reached. The categories are following each other up. This would be quite frustrating for a child to get many exercises above their level and have to make many errors before ending up at their level. The same goes when starting in category 1 and the student is already at the level of category 9. It has to answer many easy questions before getting to the right level. This shows the importance of a good initial distribution. To counter this, the kick-start could be used. The effects of this kick-start are visible in figure 4.11b. Showing that the categories 1 and 2 are reached much faster than without the kick-start. Making it less frustrating if the level of the initial distribution was too high. When the initial distribution and the skill level are slightly off, the application will converge fast towards the correct skill level with the use of the kick-start. To conclude, the algorithm is behaving as envisioned as it can find the zone of proximal development. However, one important constraint has to be made. The categories have to be connected to each other, meaning that mastering category 1 will increase the chances of mastering category 2 and thereafter category 3. If this is not the case, for example category 2 will never be mastered. Exercises from category 3 will never be asked. This means that how the categories are chosen and connected towards each other is influencing if the algorithm can work as envisioned.



(a) Linear



(b) Branching



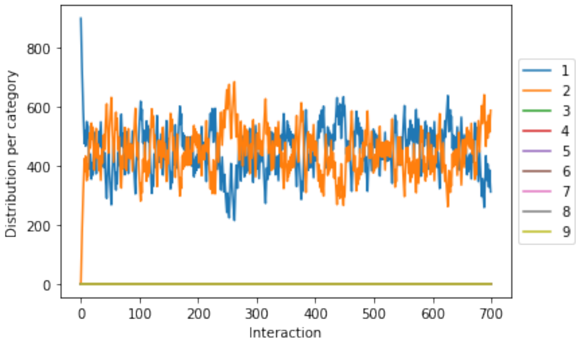
(c) Cyclic

Figure 4.12: Different connections

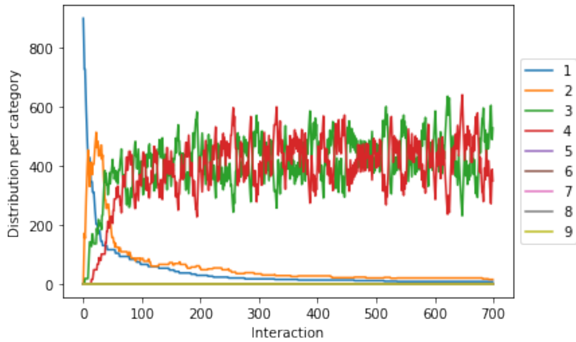
Connections between categories The content can be structured in three different ways, namely linear, branch and cyclic. These types are visualised in 4.9. Figure 4.12 shows the results of the simulation with different connections. For these graphs the following settings were used: skill level was 100% for each category, the learning rate was 0.1 and the distribution started in category 1. In the linear plot, it is visible that all categories follow each other up. Start with category 1, thereafter 2 grows and 1 decreases. In the middle, the distribution is more distributed over multiple categories. Eventually, category 9 starts growing until all distribution is moved to this category. In the branching plot, it is visible that the distribution is moving to categories 2 and 7 as they are both connected with category 1 as can be seen in figure 4.9. The consequence of the branching is that it converges with around half of the total distribution in category 9, a quarter will end up in category 4 and the other quarter in category 6. This distribution is kind of stuck in these categories. In the cyclic plot, it is visible that some categories will have an increased distribution compared to the linear plot. This is because category 4 receives the distribution from category 2 and category 3. Eventually, the distribution from categories 2 and 3 will come partially back to category 4. Some will leak to category 5 and move on. The consequence is that the cyclic connections take more time to reach category 9 as the distribution literally moves in circles until all is leaked out of the cycle.

These plots show how the distribution is moving through the system. The graphs in figure 4.13 are made with the same settings, however, the skill for category 2 or 4 was fixed at 0%, thus failing this category completely. The consequence for the linear connection is clear. The distribution stayed between levels 2 and 3 or 3 and 4. As the next category is fully dependent on the category before, the distribution will not proceed to higher categories. This mechanism can be used for categories that need to be mastered till a certain level before proceeding with more difficult exercises.

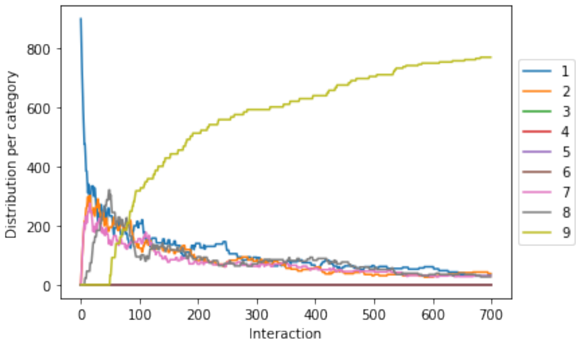
The branching connection shows two different results. If category 2 is set to 0%, all distributions will move to the branch that ends up in category 9. Categories 2 up to 6 are completely cut off and will receive any distribution. However when category 4 has a skill level of 0% the distribution will act as in figure 4.12b with one difference, a quarter of the distribution will stay between category 3 and 4. The cyclic connection also shows two different plots. Depending on which category is cut off. Category 2 had a key role in the proceeding and therefore the distribution is stuck between 1 and 2. Category 4 on the other hand is not necessary to master to get to category 9.' To conclude, the way the categories are connected influence the behavior of the algorithm. Depending on the subject the right graph needs to be chosen as they are converging in different ways. For the prototype the linear structure is chosen as the categories as there is a certain hierarchy in these categories. However a branching or even cyclic graph could also work as one category might have different categories as basis, this does not have to be a linear hierarchy. However this needs to be discovered per subject and exercise set.



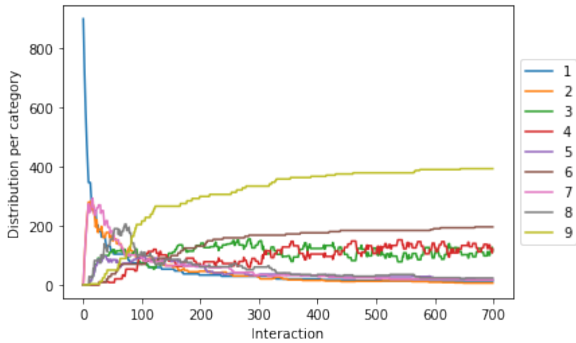
(a) Linear with skill 0% for category 2



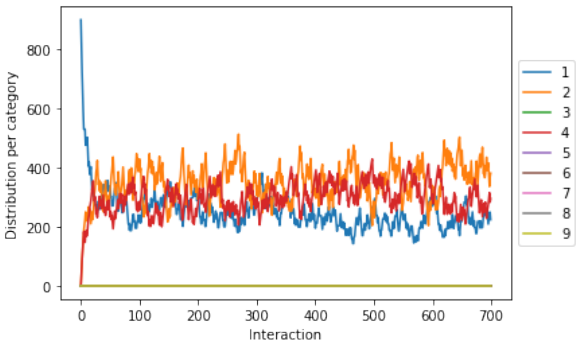
(b) Linear with skill 0% for category 4



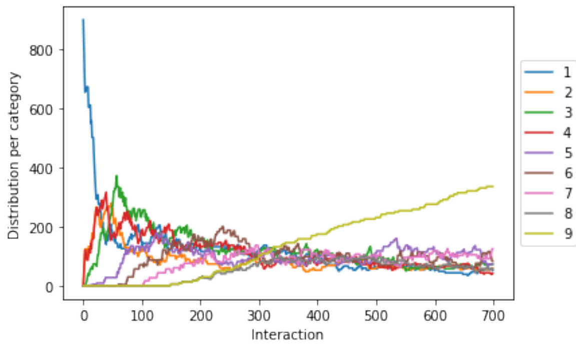
(c) Branch with skill 0% for category 2



(d) Branch with skill 0% for category 4



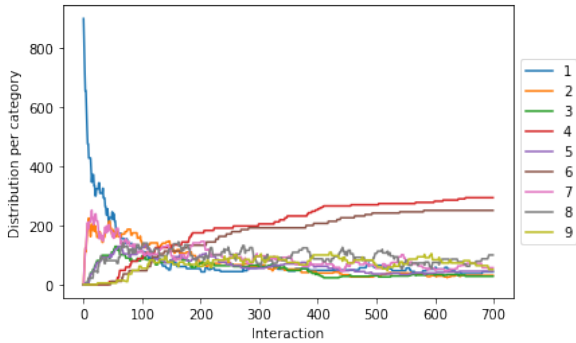
(e) Cyclic with skill 0% for category 2



(f) Cyclic with skill 0% for category 4

Figure 4.13: Development of distribution for different connection with skills fixed at 0% for respectively category 2 and 4

Summary of testing phase The different variables are working as envisioned. The main takeaway is the importance of a well-designed content structure and the expected learning rate of the user. Based on the structure, the application responds differently. When a category is really important to master, it would be smart to use a more linear approach. When multiple concepts are practised from one basic understanding, branching could be useful. Although it is important to think about how much proportion should end up in which category. This needs to be done carefully, as can be shown in figure 4.14. Because categories 7, 8 and 9 are not completely mastered, the distribution moves to the branch that is mastered. Resulting in a sort of shortcut in the application. The cyclic would be suitable for content that has 'many ways to Rome' and the goal is to master the higher category and not all categories before. The expected learning rate of the user is not known in this thesis, however it heavily influences which learning rate needs to be chosen for the algorithm. The speed of the adaptive behavior depends on this learning rate. Therefore, it needs to be chosen carefully and further research is needed.



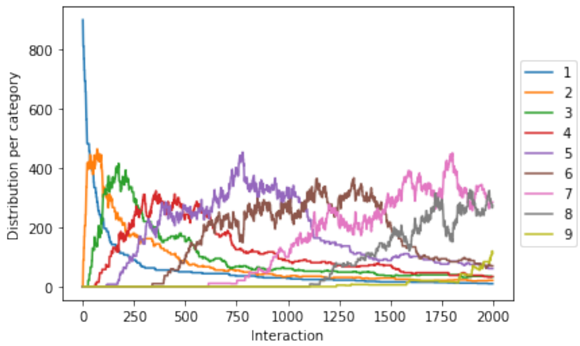
(a) Branch leakage

Figure 4.14: Unwanted behaviour

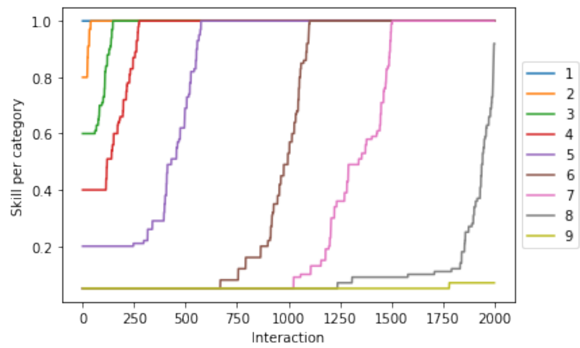
Validation phase

Similar to the testing phase, many different settings were tested. The most important results are shown here. The focus is on how the different agents are developing their skill and how the distribution of the algorithm is moving in response to the agent skill level. Three different agents are used that resemble a slow, fast and random learning child. Random indicates there is a combination of fast and slow learning, but randomly configured. From the testing phase it becomes clear that the learning rate influences the adaptive behavior of the algorithm. Therefore, multiple graphs will be shown with different three learning rates: 0.05, 0.1 and 0.2. As the linear connected categories give the best overview of the different response of the algorithm to the slow, fast and random learning agent, all graphs in this paragraph are created with linear connected categories.

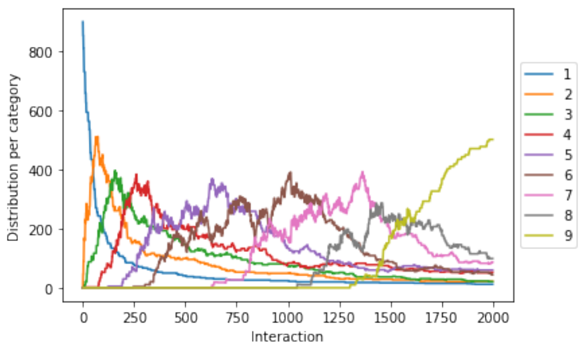
The results are displayed in figure 4.15, 4.16 and 4.17. The three different figures correspond to the different learning rates that were used, respectively 0.05, 0.1 and 0.2. Within the figure the three different agents are shown horizontally. The left graphs display the distribution of the algorithm over time and the right graphs display the development of the skill level of the agent. They are placed horizontally together to make the comparison easier as they influence each other. The skill of the agent that goes up depends on which category was given by the algorithm and if the agent



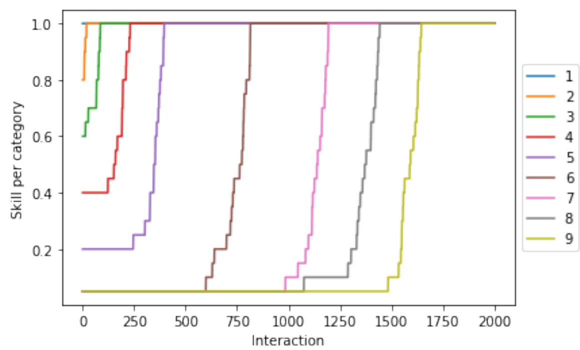
(a) Distribution slow agent



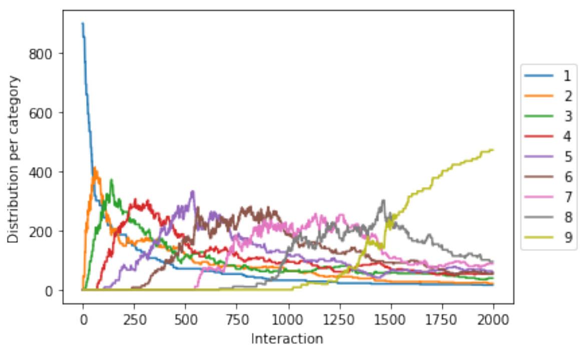
(b) Development skills slow agent



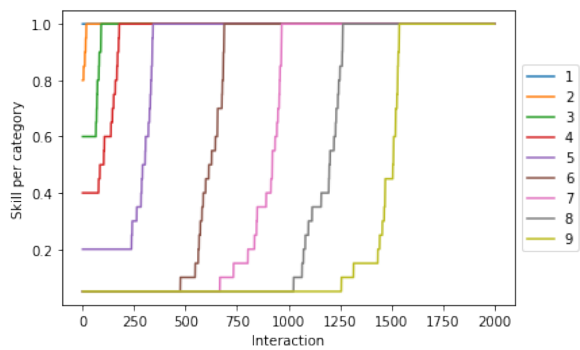
(c) Distribution fast agent



(d) Development skills fast agent



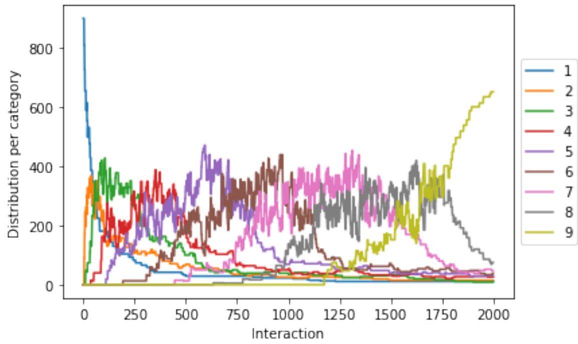
(e) Distribution random agent



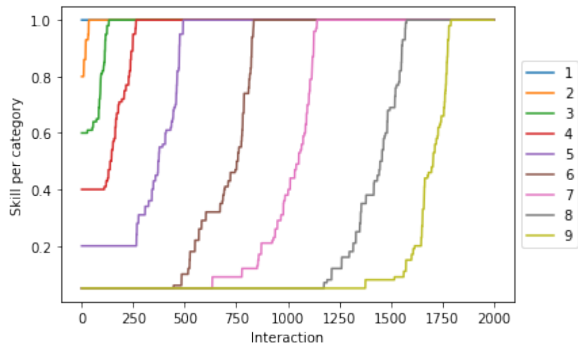
(f) Development skills random agent

Figure 4.15: Learning agent with a learning rate of 0.05

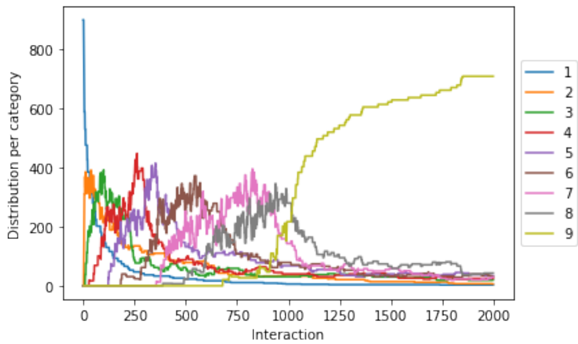
was correct. The algorithm is also changing the distribution based on the same interaction. The algorithm works with a learning rate that is fixed, the skill level of the agent is more randomly configured as this might resemble a child better. Not all practice leads to the same learning outcome. It is important to note that these graphs are of one simulation, meaning that when the graphs would be recreated, it will be different graphs. This is because of the random component in choosing the next category by the stochastic distribution and the random growth of the skill level. The fast agent and the random agent are quite similar in the graphs. Based on intuition, it was expected to be more a mixture of the slow and fast agent. The probable cause



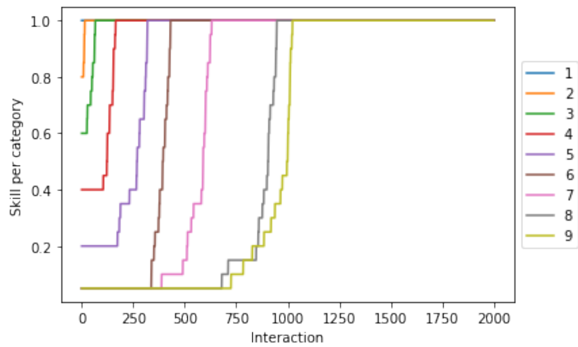
(a) Distribution slow agent



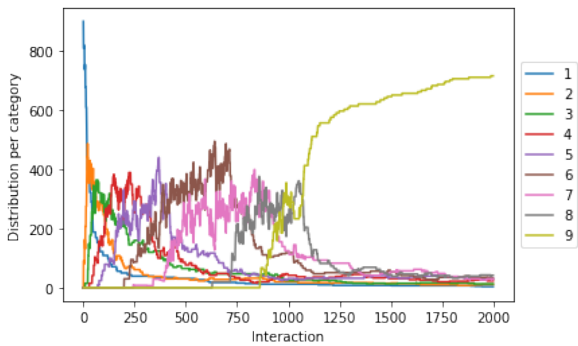
(b) Development skills slow agent



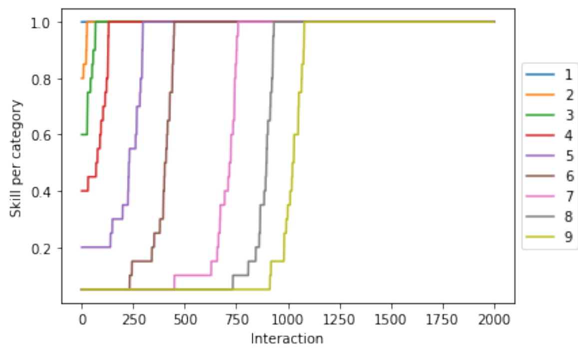
(c) Distribution fast agent



(d) Development skills fast agent



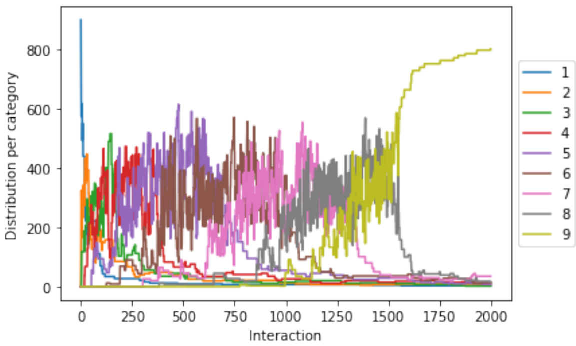
(e) Distribution random agent



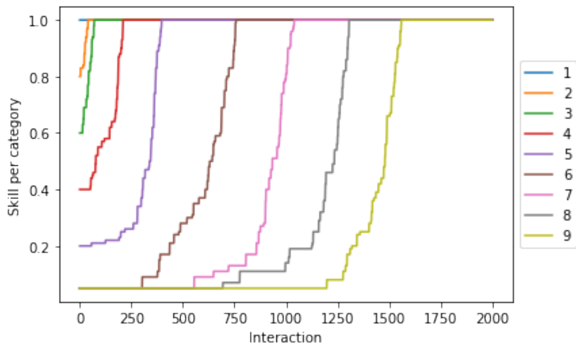
(f) Development skills random agent

Figure 4.16: Learning agent with a learning rate of 0.1

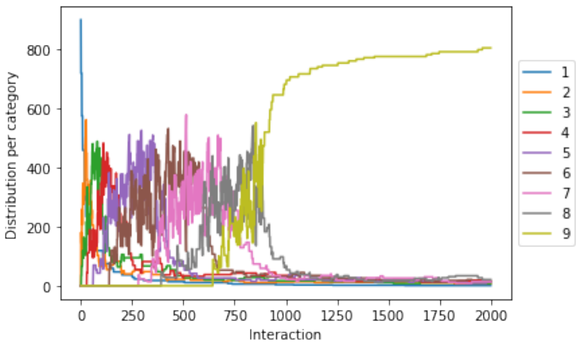
lies in the randomness of the graphs. The graphs are not regenerated and cherry picked but a random run is included as this also displays the variation. When looking at the individual figures, it becomes clear that the distribution of the algorithm and the skill level of the agent are connected as expected. When the distribution of a certain category is relatively high, the skill level is also increasing around that point in time. This is the intended behavior. The application is focusing on certain categories to get the skill level of the user up. When this is increasing, it will automatically start with the next category. The distribution of learned categories continues to increase after the skill level reaches the maximum level as there is still distribution



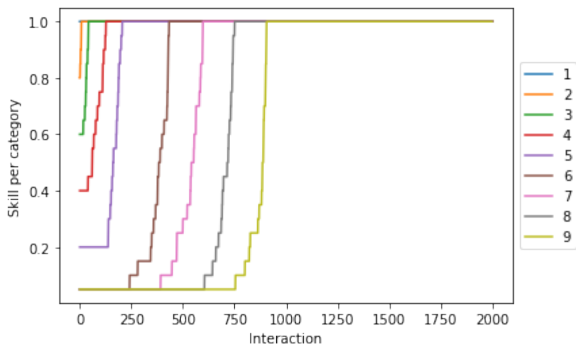
(a) Distribution slow agent



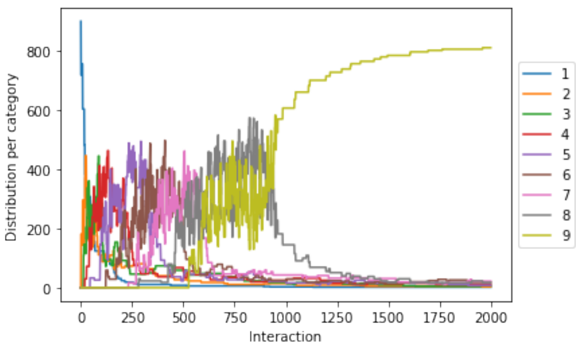
(b) Development skills slow agent



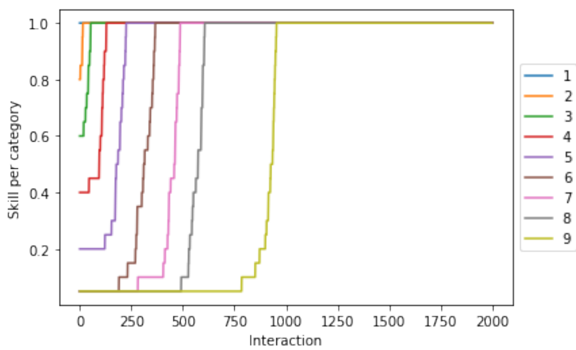
(c) Distribution fast agent



(d) Development skills fast agent



(e) Distribution random agent



(f) Development skills random agent

Figure 4.17: Learning agent with a learning rate of 0.2

from lower categories that needs to move through a specific category. Thereafter, it starts reducing. When it starts reducing, it is influenced by the learning rate. High learning rates lead to an earlier reduction moment. Also the total amount of distribution that sits in one category differs per learning rate. This could be explained by the fact that the distribution is more clutter per category with a high learning rate. The lower learning rate leads to more scattered distribution of the different categories, hence more variation in categories that are asked. Which learning rate is better to use depends on the specific case where it will be used. A relatively higher learning rate means that more questions of fewer categories are asked, a lower learning rate means that more categories will be asked. For the prototype, the latter may work better as it helps repeating previous learned content and still gives more difficult categories that are in the zone of proximal development. For faster learning students, this might not be true as they could benefit from a higher learning rate, meaning that they will have fewer repeating materials and continue faster with more difficult categories.

To conclude, the figures show that the algorithm is responding towards the development of the skill level of the agent. It could be questioned if it might be slightly out of the zone of proximal development as quite some categories are asked for a long time after the skill level reaches the maximum. When looking at the real world, this might not be a problem as the skill level is connected to the correctness. When a student masters a category in a practice session, it is not a bad thing to repeat these previous mastered categories in the next practice sessions as the skill level of the child may drop over time. This behavior is not tested in the simulation; the skill level cannot drop therefore still touching the zone of proximal development. However, this hypothesis needs to be tested using real participants.

4.5.1.3 Summary

Overall, the simulation showed that the algorithm can be used for an adaptive application. It gives the possibility to adapt based on the results of the student. When the student is correct, the level increases slightly. When the student is incorrect the level is decreasing slightly. Furthermore, the algorithm can find the level of the student when this is not known. With the help of the kick-start this goes even faster. The validation showed that there are indications that the algorithm can find the zone of proximal development, however to make a strong claim, more extensive research is needed. For now it can be said that the algorithm gives challenging categories that the user cannot always solve and repeats recently mastered categories.

The results and discussion show the complexity of using the right settings as this depends on the environment where the algorithm is put in. Still, some insights can be drawn. To elaborate, the learning rate influences the speed of the application, how fast new categories are introduced, how fast the level drops when mistakes are made, and also which categories are given in which proportion. This learning rate needs to be determined per subject by experts or by experimental research. The simulation can help to understand how fast the distribution converges to a category in combination with an initial distribution and an assumed skill level. Therefore, the simulation can be used as a tool for further development of the application. Even though the simulations give many insights into the behaviour of the algorithm, there is a need for experiments in the classroom, as there is much diversity in the learning process of children. The insights will help the teacher and content experts to start with basic settings and from here different settings can be tested to increase efficiency. It was shown that the speed that the user will go through the content can be influenced by changing the learning rate. It is important to think carefully about the connections between the nodes. Linear is the simplest form, however not all learning content has this structure. When categories are connected toward more than one category, a branch is created. The simulation showed that the distribution is stuck at the end of the individual branches. This could lead to unwanted behaviour. For example, distribution can leak away from end branches that are not yet mastered and this distribution will never return. This problem also emerges with the cyclic connections. When looking at figure 4.9 the cyclic connection, when distribution flows back from category 4 to category 1, half of this distribution will go to category 2 and thereafter to category 3. Category 2 will bring a quarter of this distribution back to category 4, the other quarter is leaking away to category 5. This means that a user never has to master category 4 to get to the other categories. The question is if it is necessary to master all categories, something that is implied in the linear structure. If so, additional rules are needed in the branch and the cyclic structure. For example, in the cyclic structure a rule could be that when category 2 is correct and 4 is incorrect, more distribution should flow from 1 to 4 instead of from 1 to 2. This could be done by an additional rule that says for category 4 needs at least a certain percentage of correct exercises be measured over the last X practices as assurance that the category has been mastered.

It is assumed that a student has a kind of internal skill level that can be modeled by a stochastic distribution. For example, a student might be confused and make the same mistake repeatedly. When the student understands what they are doing incorrectly, they may then proceed to answer all equations correctly. The model has a more gradient skill level. By using a skill level that is influenced by a stochastic distribution, it is still possible that a category is incorrect 10 times and correct 10 times after that. Even though this specific behavior might be rare when the chances of getting correct is 90% versus incorrect 10% it still might happen due to the stochastic modelling. This ensures that the simulation also includes certain rare, unexpected or rather random behavior patterns. This will make the simulations emulate more realistic environments.

The validation results indicate that the algorithm can perform as envisioned. However good design is necessary to get this correct performance. The learning rate must be based on the content and the student for example. If the content is hard, a lower learning rate could be better than a high learning rate. The connections of the categories are also heavily influencing the behaviour. If not well designed, the algorithm can easily be stuck or skipping a few categories. All these things will be discovered when using this algorithm in the prototype and test this prototype with real students. However, this is future work.

4.5.2 Group discussion

To validate the need for an AI assistant for teachers, a group discussion was organised with six people from different backgrounds. The backgrounds differed from education, technology, law, research and pedagogic as the design is also multidisciplinary, hence input from different fields is preferred. Three people have studied Pedagogical and Education Science at the university. One of them also received a degree as a teacher and is currently working as a teacher at a primary school as well as a care coordinator. The other had experience in kindergarten teaching and has also begun studying for the position of school director. The last participant's expertise is more in the field of policy. One participant has a background in law and is interested in privacy issues and regulations around the use of AI in Europe. Two other participants are professors associated with two different universities. One specialised in educational science and the use of technology, the other in technology in general. Both with teaching experience to university students.

First, the method is described, thereafter the results are presented. Concluding with a short discussion.

4.5.2.1 Method

The group discussion is led by the researcher and featured the following topics: AI in general and the math application.

The first part focuses on the existing problems in education and the opportunities for AI to (partly) solve those problems. In the introduction of this part, a general description of AI is given to the participants. The participants are asked to think about possible solutions or innovations that would improve the education system.

One of the participants is asked to make a mind map during the conversation. This mind map is visible for all participants and organises the topics from the discussion. The second part focuses on the developed application. The application is introduced and the prototype is shown. Questions about the functionality of the application are answered by the researcher. The goal is to get feedback on the design, what kind of improvements can be made and how they would use it. Thereafter, the discussion is directed towards a more general discussion about these types of applications in the education system and how these innovations could be implemented.

The participants were asked to discuss the topics that were introduced amongst themselves while the researcher remained in the background. The researcher kept a couple of questions that could be asked when the discussion came to a natural end and there was still time left. The role of the researcher was more active in the second part as the participants could ask questions about the prototype and the design in general.

The discussion was planned for 2 hours. The first half-hour is used to establish the online connection and introduction of the participants to one another so the participants could have an opportunity to learn about one another's respective backgrounds. Part one and two are both half an hour. The remaining 30 minutes could be used for overtime.

The discussion is recorded and analysed by the researcher. A brief summary is provided about the results and contents of the discussion. The focus lies on the different problems that the participants named and the possible solutions. Furthermore, feedback about the application is gathered.

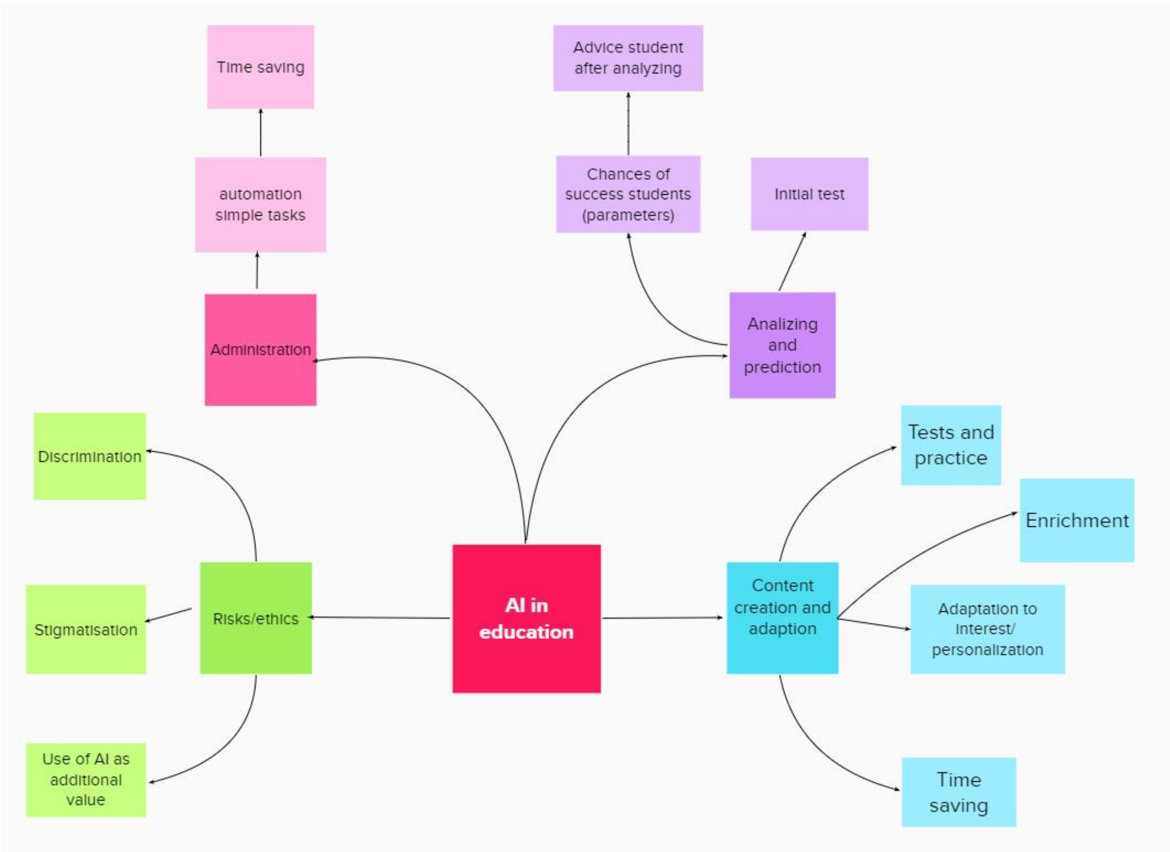


Figure 4.18: Mind map made by the participants

4.5.2.2 Results

In figure 4.18 the mind map is given from the conversation. The participants mentioned different projects that are related to AI and technology in education, such as formative testing and study results prediction. The overall topics that were revisited many times were the purpose of AI in education. The purpose focused on that it should be efficient and lower the workload of a teacher in order to have more hands-on time with the students.

The definition of AI leads to some questions. One of the participants asked if a tracking system is also AI. The researcher clarified that if it is only logging information it is not AI, however it can be used by an AI. Another participant asked if advice that comes out of the tracking system is AI. The research argued that it depends on how this advice came to be. If the advice is only based on specific rules that are set by an education specialist it is not considered AI. However the system has to learn how to interpret it based on data, it is indeed AI. This led to the remark of one of the participants that AI and the amount of data are connected. The more data you have, the more the AI can do. One participant noted that the EU created a definition for AI and that it was focused on the techniques that were used like deep learning. They agreed with each other that a complete definition is hard to give. One of the participants elaborated on the definition: it has to be a system that has to learn from data and make decisions based on that. This means that not all decisions of the system are designed and programmed by a rule set. Later on the question was asked: 'is the application at hand AI?'. Some of the elements would not be considered AI if viewed alone, like simple administration tasks. However the application as a whole does become smart as it can adapt to the level of the students in a smart and in a self-sufficient way. However this is not yet done by the use of data.

The feedback on the application was overall positive. They liked that it was adaptive towards the level of the student so that children can work at their own speed. One of the participants even mentioned that she could see herself using the application in the classroom. Questions about the application were directed towards elements that were already included in the design but not yet communicated. For example, the participants asked about how the initial level is determined. The answer was that the teacher could include an initial level or that the application can find the level by starting with a high learning rate that decays over time. The participants agreed that these two options would work well. The participants liked that there was a lot of customization, however, they warned of overwhelming the teachers. Customization should be easy to handle. The participants hypothesized that for some teachers, it works better to have a standard application to start with and not to give many customization options.

One of the participants critically mentioned that AI should not be used to replace the core tasks of being a teacher but rather support the teacher and take over administration and partially analyzing tasks. The more boring and time consuming tasks. The example was given that it would be useful if you could tell your computer that a certain child is sick and that would automatically be placed in the correct administration system by the computer. Or that it would be useful if you could just scan documents with your phone and they get automatically checked. The participants agreed with each other that taking over the administration and some analyzing tasks would lead to time savings that can be used to focus more on teaching.

While talking about customization, personalisation was mentioned by a participant. When a child has a certain interest it would be good to use this interest as a motivation when generating content, instead of only focusing on adapting towards the level of the student. Another participant mentioned that this technology already exists and is used for advertising on the internet.

One of the participants asks if the numeric line is the only strategy that is presented to the student. The answer was that this can easily be changed into a different strategy and even set per student, but for the sake of simplicity the prototype only offers one strategy. The participant mentioned that it would be helpful if the strategy can be changed based on the strategy used by the school or teacher. For some children a mix of strategies might even work, although for the younger and weaker math students one strategy would be preferred.

The question was raised if the application is only for practising or if it can also be used as instruction. The prototype focuses on practising. However, there are possibilities to also include some instructional material like videos and animations. The participant said that practising would be the most important function of the application as the teacher can still give the instruction and the application can give feedback that is based on this instruction, like showing the numeric line.

4.5.2.3 Discussion

From the group discussion, it became clear that everyone has some ideas about AI in general and were curious to learn from each other. All participants had some expertise and knowledge from AI and from education. The participants gave a lot of anecdotes about encounters with AI, ethics and education. A returning topic for most participants was whether the use is ethical or not, this could mean that people do think about the risks of AI.

Customization is a positive point of the application. However, how this customization is presented to the teachers is very important. It should be an application that is plug and play and does not have a long extensive manual before working with the application. This is also something that is seen and implemented in other technologies. For example, mobile phones do not come with an extensive manual anymore. You start your phone and the phone helps you with the set-up of the basic settings. Thereafter, the basic settings can be changed if necessary.

A limitation of the group discussion was that the format was quite open, therefore an interesting discussion took place with a lot of information. However this makes it hard to draw conclusions from this discussion. For future work, a usability study with concrete questions about the design where concrete examples are shown such as a video that shows a child interacting with the system or images from the interface, would be preferred. This might lead to more concrete feedback on the application. Even so, it was interesting to hear that the participants mentioned certain ideas such as customization and supporting multiple learning strategies to the application that were already included or part of future work and not mentioned in the discussion before. This could be seen as a validation for these elements. The conclusion from the participants was that it would be worth exploring further development of the design in the future.

Chapter 5

General discussion

This thesis started with the research question: “How to design and develop an AI assistant that adapts towards the zone of proximal development of the students, while reducing the workload of the teacher?”

To answer this question a problem analysis was made with different literature studies that were conducted in many different fields. As the question is related to different fields such as education, learning, development, AI and machine learning. This has led to a design of such an AI system. A returning topic in this thesis are three specific tasks: administration, analyzing and content creation. They are individually discussed in sections 5.1, 5.2 and 5.3 respectively. The discussion on how to use such a system can be found in section 5.4. Thereafter the developed math application is discussed in more detail in section 5.5. As the usages of AI systems are always tangled with privacy issues, a dedicated section is written about this topic in section 5.6. Thereafter, the possibilities of emotion recognition are discussed as this might be an add-on to the application in the future. This is discussed in section 5.7. As emotion recognition is an additional data source, it is wise to think about what the reasoning is behind adding new data sources. A start of this discussion is made in section 5.8. This chapter is concluded with recommendations for future work that was not mentioned before in section 5.9. To return to the research question, all elements described in this thesis are a piece of the answer to this research question. However, this is just one example of how this question can be answered. Future research must be conducted to understand how effective the proposed design is and what kind of improvements can be made.

5.1 Administration

The administration of the teacher consists of gathering data that can be analyzed and that helps to keep track of the progress of the students. Since administration can take a lot of time, the possibilities of using an AI are discussed in section 2.4.1, to reduce the time spent on this task. There are many different techniques available that could help with the administration. Therefore, a few techniques that can be used in the application will be discussed with the advantages and the disadvantages, concluding with the technique that is implemented in the prototype. As an additional remark, administration and analysis are presented as two different tasks, however, there is much overlap. When checking answers with natural language processing, data is still analyzed but on a micro level. When results from multiple training sessions are analyzed and stored in the administration, the scope is bigger. Eventually, the administration of the teacher does not focus on every single result but results per day, week or per test. In this section, both micro and macro levels are taken into account.

The proposed application supports different subjects from a technical point of view, as it is built in a modular way. This makes it easier to implement new techniques to support a different subject. For example when supporting the subject history, natural language processing might be needed to check the answers. Natural language processing can be used in different ways. One of these ways is creating fill-the-gap exercises based on a corpus [55] [56]. Another way would be to analyze sentences on content and meaning, however this is more complex. More research would be needed to develop these techniques even further and to implement them in the application. The application is designed in a modular way, meaning that it supports different techniques. An advantage is that when a new technique is developed, it can be easily implemented in the application. Even though the application could support different subjects, this does not mean that the implementation of a new subject is not time consuming as many considerations have to be made: how to present the new content, which type of support needs to be given and when and how the results are processed to the teacher interface. Also the internal hierarchy of the content needs to be designed, and the content itself needs to be created.

The prototype offered math exercises, these are easy to check by the system as simple math rules are standard of the programming languages. The administration of the application consists of logging the interaction of the user with the system. The data is stored in a simple database where the interaction is stored per user and timestamp. This makes it possible to use SQL queries for the more complex analyzing tasks. The system is a stand alone application, however, it is possible to create a connection between a tracking system of the school with a simple API. The prototype kept a current state of the level of the student and updated this after each interaction. This makes it possible to analyze all steps separately. This is done for research purposes, however, when the application is available for schools, not all information would be stored as this would be a huge amount of data that may not even have a purpose. Therefore, it is important to carefully design the administration in such a way that it is effective at achieving its goal task. The prototype focuses for example on logging results and response time. It would be interesting to add more features that could help the adaptive algorithm. For example, emotion and engagement might influence the learning process. This could be measured by using data about the practice session. Mouse movement and keyboard strokes might hold information about the emotion and stress level of the user [45], [46]. Machine learning might be suitable to find those hidden patterns in the data. However, there is a need for research that focuses on how to capture and use this data to enhance the learning process.

One of the goals that is central in this thesis is to reduce the workload of a teacher. However, not all data sources are suitable for using an AI system. One of these sources might be exercises on paper. If the teacher has to scan all documents from the children after school hours, checking by hand during the lesson might be more efficient. The result is that it might not even be worth using an AI for this specific administration task. There are alternative solutions possible, for example, to have an application on the phone to scan the documents when the children are finished. The teacher could even receive the score and previous scores in real time, giving opportunities to give feedback directly after finishing the exercises. To take it even further, high resolution cameras in different angles might be able to scan the work in real time and analyze it. However the question is always how to reach a goal in the most efficient and least intrusive way. The proposed application has a strong advantage here as the system is already on an electronic device making it rather easy to log information and the devices are already available in the schools.

The administration is always tangled with privacy issues. The design should always keep in mind why the data needs to be collected and the risks of collecting that data. Summarizing, there are many possibilities to use AI for administration. It is important to think about what data needs to be logged and what is the most (time-)efficient way of doing that. The prototype showed a basic but efficient way of logging. The application in general is suitable for multiple data sources and makes it possible to push the data to a tracking system through an API.

5.2 Analyzing

A teacher has the task to analyze results and observations with the goal to keep track of the progress and to interfere when needed. An AI can help the teacher by monitoring progress, analyzing relations between different data and for prediction as can be read in section 2.4.2. These three processes will be discussed below in more detail. Thereafter, it is discussed how a teacher can use this analysis of the AI in the classroom by taking the proposed application as an example. First, monitoring progress can be done with the use of statistical methods, for example to look for outliers in results from different training sessions. When the performance from a certain training session is significantly lower than from the training sessions before, something is going on with the student. It would be the task of the teacher to figure out if the student is just having a bad day, or the student is missing an important concept that is causing the drop in performance. The teacher can take measurements like talking with the student or explaining a concept. Second, analyzing relations between data can be done using statistical methods, for example, by looking at the correlation between different results. The correlation could be used to make underlying misconceptions clear. One misconception might cause multiple results to be low. If the teacher can discover what is causing the correlation, they could tackle this problem at the root. This could be the case with English and history. If they are both low, it might be caused by a lack of reading skills. Correlation can also be used to understand how much practice time is needed for a certain result, this could differ per student. In the future AI might be able to discover what is causing the correlation to support the teacher even more. Finally, correlation can also be used for prediction purposes together with machine learning. The advantages of using machine learning would be that patterns that are not yet discovered may come to light.

The teacher can use the analysis in different ways. The teacher can get notifications on a device that helps monitor the progress of the children. When a child drops significantly in level, the teacher will be notified and can interfere based on this information. If necessary the teacher can get more information about the notification,such as graphs that provide more information as to why the notification was raised. With this detailed information the teacher can better help the student as more of the initial situation of the student is known.

The AI also provides analysis about the relationships between data. This can be between students, between subjects or even branching subjects. The goal of providing this information is so the teacher can adapt their curricula based on these results. For example, if the whole class is struggling with a specific branch of a subject the teacher can focus the next lesson on this specific topic. The analysis that focuses on prediction might be the most complex of the application, as machine learning is involved. Ideally prediction can help the teacher with choosing a good focus for the next lessons as information about how well the students probably will be performing is given. A drawback of this technique is that it requires a lot of data and the data that is gathered afterwards is probably biased as the teacher is adapting based on the prediction. For example when the system predicts that students have issues with a certain type of exercise, the teacher will spend more time on it and therefore the student performs better. If the system is trained with new data, the intervention should also be included otherwise the system will think that the prediction was wrong, as the student did not have any issues with that exercise. All these different ways of implementing the AI for analysis in the classroom could lead to reducing the workload of the teacher and making the lessons more efficient as the teacher has more initial knowledge and can monitor the students in real time. This means that the teacher may be able to give the student education that is in the zone of proximal development of the individual students. The teacher does not have to put a lot of effort in the analysis tasks as the system presents information in a structured way and with the use of notifications. An example of what kind of teacher's interface with the results of the analysis is given in figure 3.2.

5.3 Content creation

Ideally, primary school teachers would adapt their education based on the progress of the students. As the method that they follow might not always be sufficient, it could be most efficient for the students to make custom adaptations or even to create some content. However, adaptations and content creation can cost the teacher a lot of time as goals need to be set and transferred into exercises and instruction materials. Especially if the teacher wants to adapt the lesson to the individual student(s). Individual adaptation by creating content manually is not time-efficient nor possible. This is where the AI can help as the AI can generate content, adapt this content to the class or individual student and do this in real time. In section 2.4.3 a general description is given of the possibilities. This section focuses on how the content creation can be implemented in the proposed application. Thereafter, the prototype is discussed and this gives a good example of adaptive content creation.

The proposed application can support multiple subjects, therefore different techniques are needed to support these subjects. For example Natural Language Processing (NLP) can be used to generate history, grammar or spelling exercises. More specifically with the use of a corpus, 'fill the gap' exercises can be created [55]. Or a structured database can be used to change small elements of a question. An example of this principle was given in section 3.3.2 with the corresponding image 3.3. One drawback of using NLP or structured databases may be that it depends heavily on a corpus or data set. If it is not available, it needs to be created by content creators and this is a time consuming task. However, when entered once, it can be used by all teachers that are working with the application.

The prototype created math exercises on the fly based on mathematical rules. The exercises were grouped per category. These categories were hierarchically structured based on level. It includes the possibility to adapt the exercises based on the level of the student. When the level of the student increased, the level of the categories that were given also increased and vice versa. It is important to have an internal structure of the data to be able to adapt based on the level of the student. Therefore, content creation is not only capable of creating different exercises, it should also be able to have a sense of the level of the exercises and know which exercises are easier or harder than the current exercise. In the future, knowledge on common misconceptions could also be used to create exercises that challenge or test a certain misconception. The challenge will be to create such a system that also corresponds with the natural learning flow of the student. This needs to be set by educational experts from the subjects, or experimental research could be used to figure out this structure. Pre-defined curriculum might be a good starting point. Thereafter, research about the efficiency is necessary to make adjustments. This would be a circle that repeats itself. Based on research, adjustments are made and implemented generating new data to research again.

This creates the possibility to make a system that can be used in every educational setting. It is not limited to the use in primary schools. For example, mathematics at high school level could also be added. An advantage would be that if the program senses that the basics of mathematics are not well established, it could lower the level and start practising the basics first before the more challenging questions. When the student is ready the more challenging questions will come automatically while practising and when the student shows progress.

One of the challenges of content creation is that it needs to be validated and be highly accurate, as mistakes in the generated content can lead to checking mistakes and misconceptions. If the system is telling the student that 1+1=2 is wrong, the student might believe that. Or if the grammar is incorrect, a student might take over that mistake and believe that it is correct. There are many techniques available to counter this problem. The most essential technique would be to use testers: Experts that interact with the system and check generated content. However this would be the most time consuming option. If the system has proven itself within the test environment, the quality can be tested by the users. The data of the user can be analyzed to find outliers in the generated questions. These might be incorrect or the level might not correspond to the category it is in. Instead of a more statistical approach, a self-report button could also help. If a student is convinced that the system made an error, it could report the question. However reported questions need to be checked by hand, this could be done by the teacher or by the development team of the application. This validation process would take some iterations before the application can be widely used. However designing, implementing and validating generated exercises would be future research.

Summarizing, when using an AI for content creation, the AI adapts the content to the current level of the student in real time. This current level of the student is aimed at the zone of proximal development to accelerate the learning process. However, if the teacher wants to use the application to practice a specific exercise, there is an option to temporarily override the adaptive behavior. For example, if the teacher just explained a new concept and wants the children to practice this concept right away.

5.4 Usages in the classroom

In section 3.3.4 an extensive example is given about how an AI assistant can be used in the classroom. This example is focused on primary schools, however, it can also be used in higher education. Even at university level, it would be interesting to follow the development of the students while they are engaged with practice materials. Based on the group results, the professor could adapt their next lecture. In university the number of students in one class can be quite high, making it impossible for a teacher to keep track of the individual progress of the students. An application like the one proposed can help with this. The professor can add practice materials as homework and this can give a good overview if the students understood the lecture. The application is checking the answers and the professor only needs to keep an eye on the analyzed results in his own teacher interface. It should be noted that not all subjects might be suitable for the application. For example, it doesn't support writing papers, it is more focused on skill and knowledge acquisition.

The focus lies on using AI as an assistant of the teacher. In a more extreme case, could the application take over the role of the teacher? One important thing is missing in the prototype: instruction. The design of the application does give the possibility to add instruction in the form of multimedia such as animations and videos. However, would a student learn enough without the interference of a human? However, there are certain cases where interference from a teacher is not necessary. For example, many people already follow online prerecorded courses or use written tutorials on the internet to acquire a skill. If they could use a program that adapts based on their progress and their current level, that would be a huge advantage. The program can skip content that is already mastered or take more time for content that the user finds difficult, making the course more adaptive to the student.

One of the challenges of this application is that it may be difficult to implement in the educational setting. This is not because it is not possible to implement but because it is innovative and will ask the users to let go of some of the ideas that they have about education. Also to work with this digital environment, some basic skills are needed. Not all teachers are comfortable working with these types of innovations. Future research about how these types of innovations can efficiently and smoothly be added to the classroom would be preferred. When using an AI as an assistant, the role of the teacher might also change. The assistant can help the teacher with administration, analysis and content creation. Therefore, the teacher can more efficiently adapt the education towards the student and become more of a coach that guides the students through the materials. With the help of the AI, the teacher may even get an even better overview of the students and their progress. Making it easier to help the students with their specific learning needs. For nuance, this is still a difficult task for a teacher to adjust the instruction based on the needs of the individual students. The application does not directly help with this task but provides information that the teacher can use.

5.5 Adaptive Math Application

The adaptive component of the application seems to be working as envisioned. The simulations showed that with changing the parameters, the behaviour also changes, making it flexible to use in many situations.

From the simulations, it becomes clear that the design needs to be well thought out. The development of the application needs a multidisciplinary team. For example, there is a need for an expert that structures the content in such a way that the application can work with it. The easiest structure is content that is linearly connected. Level 2 is harder than level 1 and easier than level 3. And if you mastered level 1 you have a better chance at mastering level 2 thereafter. However, when teaching history, it could be preferred to first learn the more general concepts before diving into one specific topic. A branching structure might be a better option for this content and learning goal. Preferably, an additional rule should be implemented to make sure that first the general topics should be correct before going into depth. This shows that depending on the content, the program needs to be configured differently. Which configuration works best in which setting is a topic for future research.

The simulation assumed that the children can be modelled by the agent. However, the model used is not validated, therefore no real statements can be made about the effectiveness of the program. The envisioned behaviour is based on learning theory and instructional principles [52] [14] [53], increasing the chances of being effective, however future research needs to be conducted to figure out if the application is indeed effective. A first step is made with the formulation of an usability and field study.

Important to note that effectiveness is not only whether or not the children are learning faster, but also if the teacher can reduce the time spent on administration, analyzing and content creation and therefore spend more time on teaching.

To conclude, the adaptive math application showed promising results, however, the next step would be to test this prototype with students in a classroom setting, preferably in an experimental setting with the use of a control group.

5.6 Ethics and privacy issues

As mentioned in section 2.3.3 rules and regulations around data and the usages of AI are created by the EU. Even if the application complies to the regulations, there are still some risks to keep in mind. The most important risk is the implications that occur when the application is not adapting in the right way. When the zone of proximal development is not found, the student either gets too easy exercises or too hard. With the consequence that the student might not even be learning at all but does spend time on practicing. Maybe even more important it probably would be really frustrating for the student to practice below or above their level. To counter this risk, it is important that the teacher should always have a good overview of the progress of the students but also needs to have insights in the assumed level of the students. The proposed adaptive algorithm makes this rather easy, as it is not a black box and the assumed level of the student can be drawn from the distribution. The program can give the teacher an overview of which categories are asked often and even the results per category can be viewed in the teacher's interface.

5.7 Emotion recognition

The current state of the art makes it possible to detect emotion to some extent, however emotion is still a complex concept to capture. Emotion detection can be drawn from different sources like speech, image, video or text [57]. This field, named Affective Computing, is rather new and still in development. Many studies have been conducted towards elements like emotion and satisfaction on academic performance [58] [59] [60] [61]. Emotion and results could be linked together for example and might help to explain disappointing results. Yun et al. [62] conducted a pilot study to explore emotion detection in a learning context. Emotions do not always have to be collected through sensors, surveys are also a powerful data source. Self surveys can give a lot of information and can easily be logged. Small questions about how children are feeling before, during or after exercises could also lead to useful information. These questions can simply be a couple of emoticons to express their emotions. Therefore, the child would be actively helping the data collection in the form of self-reported emotions. Emotion detection was not included in the global design and the prototype, however it might be a useful source of information as it could possibly hold information about the efficiency of the practice session. This could be used as an additional data source for the algorithm that makes application adaptive.

5.8 Adding data sources

One question that arises apart from the privacy issues is, what is the amount of data that needs to be logged? For someone from the machine learning, deep learning or neural network field there is no such thing as too much data, except in terms of computational complexity. Data that is not useful can always be dropped, but data that is not logged can't be retrieved afterwards. This feeling of missing out on data could be quite harmful. Students are more than a collection of data on how they behave and achieve in schools. To take it to the extreme, children from before the computer age also grew up to be fine adults. So the questions should always be, why do you need it, what data do you want to retrieve and what do you want to do with it. The question of why should be backed up with valid argumentation.

Inspired by Pawson and Tilley [6] with their division between method- and theory-driven research, this division is also visible when looking at strategies for data collection. Method-driven strategies collect as much data as possible and find useful data afterwards compared to theory-driven strategies that collect only specific data based on theory and existing frameworks on education.

For example, it is possible to log when children go to the restroom. You probably wonder, why would you want to collect that data when talking about study achievements. In my experience, children that go to the restroom more often than necessary could be avoiding something, for example, a certain lesson or activity. If you do go to the restroom during gymnastics, you might miss the warming-up activity. Avoidance could be related to disappointing results and therefore it could be useful to log. However, with this type of reasoning, where do you stop? With theory-driven research, the focus lays on creating a framework of possible connections with the data before actually collecting the data. These connections are hypothesised based on literature and learning theory. The drawback of the theory-drive approach is that it is building upon existing theories that can result in tunnel-vision and potentially missing out on certain not yet discovered connections or formalized theories.

Based on the European law, the theory-driven strategies are preferred over the method-driven strategies as they state a clear goal of the data that is collected. The design application is also only gathering data that is necessary to make the adaptive behavior possible and to give the teachers an overview of the progress.

For future work, it is a goal to add more data sources to make the application adapt better to the individual. The ones mentioned before are keyboard strokes and mouse behavior and the section above mentioned emotion detection. Data sources that were not named before are the background characteristics such as age, gender or education level of the parents. These background characteristics have to be used carefully as the risk of discrimination increases. For example if the algorithm would give a slow learning rate based on the education level of the parent, this deprives the child of showing his or her capabilities. It could even be argued that no background characteristics should be included and that the application only should adapt based on the encountered interaction.

5.9 Other future work

From the group discussion, it became clear that the application should be easy to handle and have room for customization. Ideas about this customization are presented in this thesis like changing colours and background images, content creation by teachers, changing feedback strategies. Future work should focus on the user interaction between the teacher's interface and the teacher. This interface should be easy to understand, give a good overview of how the students are doing, make it easy to customize the application, and most importantly it should decrease the workload of the teacher.

The prototype only works with arithmetic at the moment. In the thesis, ideas are presented to also use other content and the simulation showed that as long as there can be multiple questions per category and the categories are linked together, the algorithm can be used to make an adaptive practice application. One of the ideas is to extend the content towards language, more specific verb conjugations. The same structure can be used to generate sentences as mentioned in section 3.3.2. The same database with links can be used to generate word problems. However, more research is necessary to build such a system efficiently. The idea is that this can be built on top of the prototype that has already been developed. This prototype is built in a modular way with the potential to extend it to other areas of content as well. For example, a new category can be added easily and if linked to other categories, it will automatically pop up in the application as an exercise.

Were it not for the Covid-19 pandemic, the prototype would have been tested in the classroom. This would still be advisable to do in the future. This is to understand how the children are responding to the adaptive behaviour of the application and to measure the effectiveness of this system as well as to get more feedback from students and teachers to improve the application.

One specific topic that could be developed even further, is the numeric line. The number line presented in the application is static but could be made interactive. Research even suggests displaying an empty numeric line on screen might be less efficient than displaying it on paper although the main issue was that the numeric line was rarely used [63]. To counter this, it might be useful to add interaction to the numeric line as they could make the jumps on their own with their finger or mouse. The application can give feedback based on the steps that they make and it would give insight into the thinking process of the child. This can be used to track down misconceptions.

Chapter 6

Conclusion

Overall, this thesis showed that there are many ways to implement an AI in the classroom to assist instructors. An AI can help with core tasks: administration, analysis and content creation. There are multiple advantages of using such a system.

First, when the adaptive application is correctly implemented the children will have the greatest advantage as the children can get education that is in their zone of proximal development at all times. This could help them progress at their own speed without holding back or going too fast. Second, the AI provides the teacher with information about the progress of the students through the teachers interface, it can be seen as having an extra eye in the classroom. This could assist the teacher with the complex task of making their curricula more adaptive towards the needs of the students. It still asks a lot from the teacher to truly adapt to the individual needs. Third, the AI gives the teacher the possibility to provide generated exercises that are adapted based on the level of the student. And finally, the adaptive application can assist with administration, analysis and content creation. This would ideally reduce the time spent on these tasks and therefore reduce the workload of the teacher.

The prototype that was built was centred around the core tasks and led to an adaptive application that responds to the interaction with the child and finds the zone of proximal development. The application is transparent, the teacher should have a good overview of the progress of the children and how the system is adapting. The adaptive algorithm is also not a black box, the influence of changing parameters is studied with the simulation and a good overview of this influence is provided.

The design of the application combines the current state of the art of technology with a human-centred approach making the application “high tech, human touch”.

Bibliography

[1] L. Hummel, W. Hooftman, and R. Schelvis, “Werkdruk in het onderwijs,” *Rapportage voor Ministerie van OCW*, 2019.

[2] L. S. Vygotsky, *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.

[3] C. Lacey, *Hightown Grammar: the School as a Social System [by] C. Lacey*, 1970.

[4] S. Braster, I. Grosvenor, and M. del Mar del Pozo Andres, *The Black Box of Schooling: A Cultural History of the Classroom.* ERIC, 2011.

[5] J. Karabel and A. Halsey, *Power and Ideology in Education*. Oxford University Press, 1977.

[6] R. Pawson and N. Tilley, *Realistic evaluation*. sage, 1997.

[7] C. e. W. Ministerie van Onderwijs, “Leerplicht en kwalificatieplicht,” Jun 2021. [Online]. Available: https://www.rijksoverheid.nl/onderwerpen/leerplicht/leerplicht-en-kwalificatieplicht

[8] ——, “Hoe regelen basisscholen de schooltijden en lesuren?” Jul 2021. [Online]. Available: https://www.rijksoverheid.nl/onderwerpen/schooltijden-en-onderwijstijd/vraag-en-antwoord/hoe-regelen-basisscholen-de-schooltijden-en-lesuren

[9] “Scholen met een speciale onderwijsmethode,” Aug 2021. [Online]. Available: https://oudersenonderwijs.nl/kennisbank/schoolkeuze/soorten-scholen/scholen-met-een-speciale-onderwijsmethode/

[10] G. W. McDiarmid, “Tilting at webs of belief–field experiences as a means of breaking with experience. research report 89-8.” 1990.

[11] D. Holt-Reynolds, “What does the teacher do?: Constructivist pedagogies and prospective teachers’ beliefs about the role of a teacher,” *Teaching and teacher education*, vol. 16, no. 1, pp. 21–32, 2000.

[12] D. L. Ball, “Breaking with experience in learning to teach mathematics: The role of a preservice methods course; issue paper 89-10.” 1989.

[13] C. Wood, *Yardsticks: Children in the classroom, ages 4-14*. Northeast Foundation for Children Turners Falls, MA, 2007.

[14] S. Alan and B. Gavin, “An introduction to developmental psychology,” *BPS Textbooks in Psychology*, 2003.

[15] R. Sikkes, “Weer meer burn-outklachten in het onderwijs,” Sep 2020. [Online]. Available: https://www.aob.nl/nieuws/weer-meer-burn-outklachten-in-het-onderwijs/

[16] P. A. Ertmer and T. J. Newby, “Behaviorism, cognitivism, constructivism,” *Foundations of Learning and Instructional Design Technology*, 2017.

[17] L. J. Levine and D. A. Pizarro, “Emotion and memory research: A grumpy overview,” *Social cognition*, vol. 22, no. 5: Special issue, pp. 530–554, 2004.

[18] R. Garris, R. Ahlers, and J. E. Driskell, “Games, motivation, and learning: A research and practice model,” *Simulation & gaming*, vol. 33, no. 4, pp. 441–467, 2002.

[19] A. Azizah, “The use of students’recap strategy to improve writing skill (a classroom action research of seventh class of mts yasinta salatiga in the academic year 2014/2015),” Ph.D. dissertation, IAIN Salatiga, 2015.

[20] R. C. Atkinson, “Optimizing the learning of a second-language vocabulary.” *Journal of experimental psychology*, vol. 96, no. 1, p. 124, 1972.

[21] P. I. Pavlik and J. R. Anderson, “Using a model to compute the optimal schedule of practice.” *Journal of Experimental Psychology: Applied*, vol. 14, no. 2, p. 101, 2008.

[22] P. A. Wozniak and E. J. Gorzelanczyk, “Optimization of repetition spacing in the practice of learning,” *Acta neurobiologiae experimentalis*, vol. 54, pp. 59–59, 1994.

[23] E. Mettler, C. M. Massey, and P. J. Kellman, “Improving adaptive learning technology through the use of response times.” *Grantee Submission*, 2011.

[24] C. A. Walkington, “Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes.” *Journal of Educational Psychology*, vol. 105, no. 4, p. 932, 2013.

[25] B. Williamson, “Digital education governance: data visualization, predictive analytics, and ‘real-time’policy instruments,” *Journal of Education Policy*, vol. 31, no. 2, pp. 123–141, 2016.

[26] M. Van Groenestijn, C. Borghouts, and C. Janssen, “Protocol ernstige rekenwiskunde-problemen en dyscalculie [protocol severe mathematics difficulties and dyscalculia],” *Assen, the Netherlands: Van Gorcum*, 2011.

[27] W. Oonk, R. Keijzer, S. Lit, and F. Barth, “Rekenen-wiskunde in de praktijk-verschillen in de klas.[mathematics in practice: Differences in class],” *Groningen, the Netherlands: Noordhoff Uitgevers bv*, 2013.

[28] W. Oonk, R. Keijzer, and S. Lit, *Rekenen-wiskunde in de praktijk*. Noordhoff, 2011.

[29] ——, *Rekenen-wiskunde in de praktijk Onderbouw*. Noordhoff, 2010.

[30] M. J. Brinkhuis, A. O. Savi, A. D. Hofman, F. Coomans, H. L. van Der Maas, and G. Maris, “Learning as it happens: A decade of analyzing and shaping a large-scale online learning system.” *Journal of Learning Analytics*, vol. 5, no. 2, pp. 29–46, 2018.

[31] P. McCorduck and C. Cfe, *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. CRC Press, 2004.

[32] P. R. Norvig and S. A. Intelligence, *A modern approach*. Prentice Hall Upper Saddle River, NJ, USA:, 2002.

[33] J. McCarthy, “What is artificial intelligence?” 2004.

[34] W. Sarle, “Neural networks and statistical models: Proceedings of the 19th annual sas users group international conference,” 1994.

[35] F. Okubo, T. Yamashita, A. Shimada, and S. Konomi, “Students’ performance prediction using data of multiple courses by recurrent neural network,” in *25th International Conference on Computers in Education, ICCE 2017*. Asia-Pacific Society for Computers in Education, 2017, pp. 439–444.

[36] F. Okubo, T. Yamashita, A. Shimada, and H. Ogata, “A neural network approach for students’ performance prediction,” in *Proceedings of the seventh international learning analytics & knowledge conference*, 2017, pp. 598–599.

[37] F. Okubo, T. Yamashita, A. Shimada, Y. Taniguchi, and S. Konomi, “On the prediction of students’ quiz score by recurrent neural network.” in *CrossMMLA@ LAK*, 2018.

[38] T. Saito and Y. Watanobe, "Learning path recommender system based on recurrent neural network," in *2018 9th International Conference on Awareness Science and Technology (iCAST)*. IEEE, 2018, pp. 324–329.

[39] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra *et al.*, "Matching networks for one shot learning," in *Advances in neural information processing systems*, 2016, pp. 3630–3638.

[40] M. Wiering and M. Van Otterlo, "Reinforcement learning," *Adaptation, learning, and optimization*, vol. 12, no. 3, 2012.

[41] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[42] "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)," Apr 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A02016R0679-20160504&from=EN

[43] "Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts," Apr 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A52021PC0206&from=EN

[44] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science Robotics*, vol. 3, no. 21, 2018. [Online]. Available: https://robotics.sciencemag.org/content/3/21/eaat5954

[45] L. Pepa, A. Sabatelli, L. Ciabattoni, A. Monteriù, F. Lamberti, and L. Morra, "Stress detection in computer users from keyboard and mouse dynamics," *IEEE Transactions on Consumer Electronics*, vol. 67, no. 1, pp. 12–19, 2020.

[46] P. Khanna and M. Sasikumar, "Recognising emotions from keyboard stroke pattern," *International journal of computer applications*, vol. 11, no. 9, pp. 1–5, 2010.

[47] S. Colognesi, C. Vassart, B. Blondeau, and L. Coertjens, "Formative peer assessment to enhance primary school pupils' oral skills: Comparison of written feedback without discussion or oral feedback during a discussion," *Studies in educational evaluation*, vol. 67, p. 100917, 2020.

[48] S. An and Z. Wu, "Enhancing mathematics teachers'knowledge of students'thinking from assessing and analyzing misconceptions in homework," *International Journal of Science and Mathematics Education*, vol. 10, no. 3, pp. 717–753, 2012.

[49] K. Thyagharajan and R. Nayak, "Adaptive content creation for personalized e-learning using web services," *Journal of Applied Sciences Research*, vol. 3, no. 9, pp. 828–836, 2007.

[50] E. Holohan, M. Melia, D. McMullen, and C. Pahl, "Adaptive e-learning content generation based on semantic web technology," 2005.

[51] R. E. Clark, P. A. Kirschner, and J. Sweller, "Putting students on the path to learning: The case for fully guided instruction." *American Educator*, vol. 36, no. 1, pp. 6–11, 2012.

[52] B. Rosenshine, "Principles of instruction: Research-based strategies that all teachers should know." *American educator*, vol. 36, no. 1, p. 12, 2012.

[53] J. R. Hollingsworth and S. E. Ybarra, *Explicit direct instruction (EDI): The power of the well-crafted, well-taught lesson*. Corwin Press, 2017.

[54] "Resources: Color library." [Online]. Available: https://ibm-design-language.eu-de.mybluemix.net/design/language/resources/color-library/

[55] I. A. Arregik, "Automatic exercise generation based on corpora and natural language processing techniques," Ph.D. dissertation, Ph. D. thesis, Universidad del Paıs Vasco, 2011.

[56] E. Baglaeva, "Natural language processing for the case of the lexical text adaptation and generating grammar exercises," in , 2014, pp. 285–286.

[57] J. M. Garcia-Garcia, V. M. Penichet, and M. D. Lozano, "Emotion detection: a technology review," in *Proceedings of the XVIII international conference on human computer interaction*, 2017, pp. 1–8.

[58] M.-J. Sanchez-Ruiz and J. El Khoury, "A model of academic, personality, and emotion-related predictors of university academic performance," *Frontiers in psychology*, vol. 10, p. 2435, 2019.

[59] P. A. Graziano, R. D. Reavis, S. P. Keane, and S. D. Calkins, "The role of emotion regulation in children's early academic success," *Journal of school psychology*, vol. 45, no. 1, pp. 3–19, 2007.

[60] G. Gumora and W. F. Arsenio, “Emotionality, emotion regulation, and school performance in middle school children,” *Journal of school psychology*, vol. 40, no. 5, pp. 395–413, 2002.

[61] J. De la Fuente and M. Cardelle-Elawar, “Research on action–emotion style and study habits: Effects of individual differences on learning and academic performance of undergraduate students,” *Learning and Individual Differences*, vol. 19, no. 4, pp. 567–576, 2009.

[62] H. Yun, A. Fortenbacher, N. Pinkwart, T. Bisson, F. Moukayed, C. Ullrich, and M. Wessner, “A pilot study of emotion detection using sensors in a learning context: Towards an affective learning companion.” in *DeLFI/GMW Workshops*, 2017, pp. 1–11.

[63] J. Vermeulen, F. Scheltens, and T. Eggen, “Strategy identification using the empty number line: a comparison between paper-and-pencil and tablets,” *Pedagogische Studiën*, vol. 92, no. 1, pp. 39–54, 2015.

Appendix A

Field study and use of a neural network

Due to the COVID-19 pandemic no field study was conducted, however future plans were made to gather data and to use this data to train a neural network together with a small usability study. The developed method of this field study and usability study can be found below.

A.1 Building a neural network

The goal of the neural network is to predict the level of the student on different categories by outputting a distribution over these categories. In image A.1, an example is given of a possible neural network structure. The neural network will take the following input: every category will get two nodes, the first node is an accuracy score. The second node is the average response time.

The accuracy score is based on a minimum of 5 samples of the category. If the category was not used for fewer than 5 samples, the accuracy score will be 0 or 100 based on the position of the category. Categories that the student has already mastered and therefore do not appear as often are assumed to have a high accuracy score. The categories that the student has not yet mastered will get a score of 0.

The second node, response time, will be 0 or 120 when there is not enough data for that category, otherwise, it will be the average response time. Similar to the accuracy score, the response time will be 0 for categories that are already mastered and 120 for categories that are not yet mastered. This number 120 will also be the maximum input of this node.

The architecture of the neural network will consist of a minimum of two hidden layers, as the categories are interconnected with each other. Based on the training and testing phase the layers can be adjusted. The hidden layers will use a differential nonlinear activation function, for example, ReLu, Sigmoid or Tanh, instead of a linear

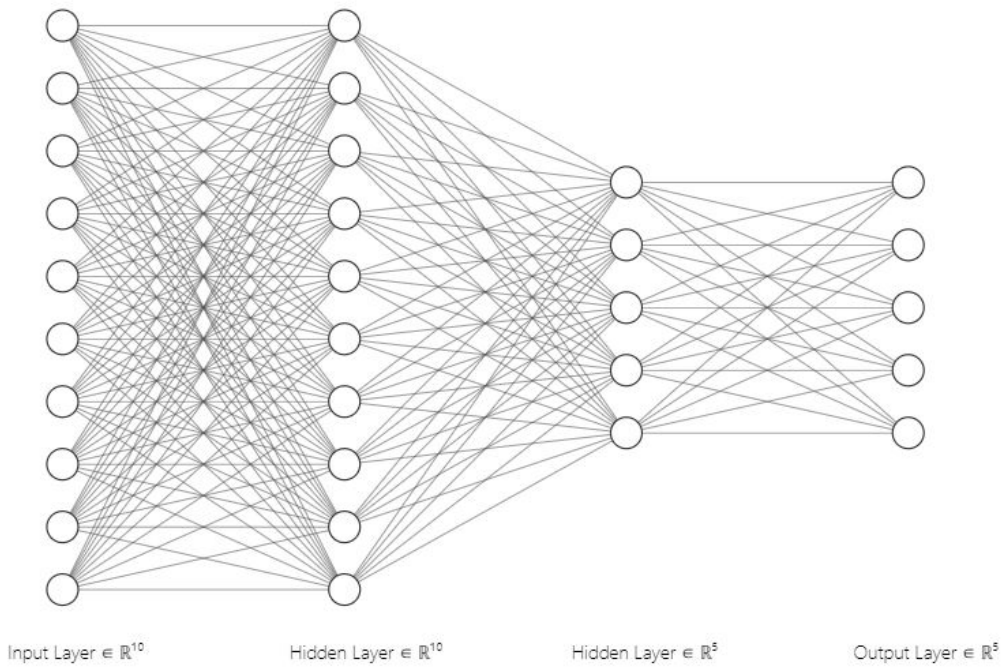


Figure A.1: Example of a neural network

activation function. This has the advantage that the network can learn more complex patterns.

The output of the neural network consists of the number of categories and the individual nodes are a number ranging from 0 to 1000. This resembles the distribution of the categories for a specific child. The neural network is trained with a set that always sums up to 1000, however, the output of the neural network might differ from 1000. Therefore, the output will be scaled back to 1000 based on the distribution of the output.

The neural network will be trained with 70% of the collected data. The validation set and the test set will consist out of 15% each. In the training phase, the parameters will be tuned. Eventually, the best model will be chosen to use in the application.

A.1.1 Collecting ground truth

To train the neural network, a ground truth should be acquired. This can be done by asking a teacher to provide adaptations for their students. As the extra workload of the teacher should not be too high, the teacher only adapts the level after the children have practised for some time with the application. Furthermore, the teacher will see an interface with the different categories, the current distribution and the performance per child. With a slider, the distribution over the different categories can be changed. When one category is increased, others will decrease automatically. To further help the teacher even more, a route map will be available for the different categories. This route map gives an indication of which categories are a good follow-up category as they are not sequential. Making a ground truth per update would give the teacher an immense workload and it would be difficult to make an update in distribution based on one update, as the information one update holds is limited. If a child did one equation correctly or incorrectly, it would be hard to tell if the student has mastered the category based on the one update. In the category, easier and more difficult equations coexist, making it more useful to look at the performance after a series of different updates. When working with teachers in a classroom, a teacher needs to make a ground truth for 20-30 accounts. This ground truth should be updated a couple of times to collect data that belongs to one teacher and one child. As for every neural network, the more data that is collected the better. However, the quality of the data is also really important. Most of the data is recorded by the application (response time, mouse movement, performance etc.), despite that, the ground truth can be noisy. It can happen that a teacher doesn't have the time to update, updates only partially, or makes mistakes in the ground truth. The more data collected, the less this noise will influence the neural network.

A.1.2 Usability study

Together with the data collection for the neural network, a user study is conducted. This focuses on how the children are interacting with the system and how the teacher is experiencing the usage of the application. For the students an small experimental user study is conducted where the interaction with the system is recorded on video and analysed thereafter. The level of engagement with the program is measured. Furthermore the children are asked to rate if they find the activity engaging,if they think they have learned something or made progress and if they think that the exercises are challenging but not too hard. For the teacher a questionnaire will be designed with open-ended questions in order to get qualitative data as there is only one or two teachers participating per class. Questions focus on how the teacher is interacting with the system and if it helps the teacher make the curricula more adaptive.

A.1.3 Time plan

To correctly train a neural network, the collection of the ground truth is of utmost importance. This means that the schools should be contacted early on. Due to the Covid-19 pandemic, visiting schools is out of the question. Therefore, all contact should be done by email, telephone or video call. To contact different schools, an email will be sent to many schools explaining the application and the research. A list of all primary schools in the Netherlands from the government will be used. Sadly no email addresses were given, only the name, address and sometimes the website. Based on this information the website was found and the contact email on this website was used to send an email. To make the email more personal, if the recipient was known, that salutation was used, otherwise, the name of the school was used. Also, the login for the demo application was given. After the email is sent, a follow-up call is used to make the contact more personal. Schools that sign up for the application will test the application during a month for data collection and after building the neural network the school can use the application for another month. In the first month, the teacher will fill in the ground truth every week, bringing the ground truth to 4 updates. The schools can decide to start one week later if that fits their schedule better. The neural network can be developed before or during the data collection as the data format will be known and test data can be generated.

A.1.4 Finding and testing the application with the users

This is an important part of the thesis. The idea is to have a dummy website available that can be shown to schools. From there, it will be necessary to ask the school if they want to participate in this research. Upon approval, they can freely work with the application during the school year 2020-2021. It will be expected of them to use the application with all the children from their class every week, preferably twice a week. More time with the application is encouraged as extra practice for certain students as part of homework. It is important that children with different levels participate. The primary focus will be data collection, and when enough data is collected the adaptive and predictive behaviour will be activated.