UNIVERSITY OF BERGEN

Master Thesis in Information Science

Identifying Dyscalculia Symptoms Related to Magnocellular Reasoning using Smartphones

Author:
Greger Siem Knudsen

Supervisor:
Ankica Babic

2016

I think everybody should learn how to program a computer
because it teaches you how to think

− Steve Jobs

Identifying Dyscalculia Symptoms Related to Magnocellular Reasoning using Smartphones

Greger Siem Knudsen

Abstract

This thesis presents a research that has developed and evaluated a software application for smartphones, called MagnoMath. MagnoMath has been developed for assisting diagnosis of dyscalculia, a learning disability in mathematics, and for identifying dyscalculia symptoms possibly related to magnocellular reasoning. Typical software aids developed for individuals with learning disabilities are focused on both assisting diagnosis and teaching the material. The software developed in this project however maintains a specific focus on the former, while attempting to capture alleged correlations between dyscalculia symptoms and possible underlying causes of the condition. This was achieved by applying k-Nearest Neighbor algorithm classifying five parameters used in evaluating users' mathematical and magnocellular reasoning. Evaluation results were then utilized to support diagnosis and measure correlations between performances across the two task categories. To test MagnoMath's validity, an experiment involving seven dyscalculic and four non-dyscalculic volunteers was conducted. Results show that the software was able to reveal dyscalculic behavior. A strong correlation between deficiencies in arithmetic and spatial reasoning was identified, revealing a possible dependency valuable for detecting early signs of developmental dyscalculia. Learning disability experts at Linköping University, Sweden, University of Oslo, Norway and Dyslexia Norway have found the application to be appropriate and valuable for its intended purposes. Thus, results prove that mobile software is a suitable and valuable tool for assisting dyscalculia diagnosis and identifying possible root causes of developing the condition. Additional evaluation would be necessary to further test mobile softwares' ability to confirm the theory of magnocellular reasoning's involvement in developing mathematical deficiencies.

Acknowledgements

Several people helped and provided encouragement during the writing of this thesis. First and foremost I thank my supervisor Ankica Babic. Your knowledge, guidance and feedback have been invaluable. You always made time for my questions, and worked hard to make sure this thesis would reach its full potential in every aspect possible. For that, I am very grateful.

A big thank you goes to Rickard Östergren, Ulf Träff and Kenny Skagerlund at Linköping University, Caroline Solhem at Dyslexia Norway and Anita Lopez-Pedersen at University of Oslo. Thanks also to the students at Department of Biomedical Engineering, Linköping University and Faculty of Humanities, University of Oslo. Your interest, engagement and knowledge from interviews and evaluations have been valuable.

I would like to thank Solveig Ingersdatter Eriksen for your interest and assistance in the process of recruiting dyscalculic volunteers for user testing.

Gratitude goes to University of Bergen for supporting my travel to the conference EFMI STC 2016 in Paris to present and publish this research by awarding it the Lauritz Meltzer Research Grant. Thank you also for providing an excellent student environment making these last few years a memorable period of my life.

I would also like to thank my family. Especially my twin brother Bjørn for all your support, and my sisters Ingeborg and Victoria for always being only a phone call away. Lastly, I would like to thank my mother Kristin, for your love and for always being there. You mean the world to me.

Greger Siem Knudsen
Oslo, June 2016

Contents

List of Figures

List of Tables

List of Algorithms

Listings

Chapter 1

Introduction

We start to learn topics in mathematics very early in life, being introduced to simple calculations at a young age. However, there is no assurance that the approach applied for teaching the material is suitable and valuable for all students in the classroom. Some of them may be affected by a learning disability (LD), a condition involving difficulties with learning in a typical manner usually caused by an unidentifiable set of factors [61]. More specifically, having difficulties with learning within the field of mathematics characterizes *dyscalculia*, a specific LD. Dyscalculia involves the difficulty of comprehending numerosity including simple arithmetic and identifying quantities of small sets [12].

Helping students affected by dyscalculia is a challenging task. To assist this process, aids in the form of computer software can be applied. The majority of this software is developed to serve several purposes. That is, aiming both at assisting diagnosis of pupils not yet diagnosed and teaching mathematics to those with established diagnosis. An improved approach could be to separate diagnosis and teaching, with the former being conducted before the latter. Furthermore, the majority of available software does not utilize data gathered from user evaluations sufficiently, often being limited to returning the number of correct answers on given tasks.

1.1 Research questions

John Stein, Emeritus Professor of physiology at University of Oxford, is one of several researchers who has studied how information transmission behaves in individuals diagnosed with *dyslexia*, a LD in language reading and writing [49]. He has developed the magnocellular theory of developmental dyslexia claiming there are possible correlations between developing LDs and disruptions in the transmission of information from the eyes to the information processing center in the human brain. The neurons responsible for this transmission are called *magnocellular cells*, which are believed to be involved in developing skills shown to be affected by LDs [49].

If we could apply Stein's research to develop a software for not only assisting the diagnosis of dyscalculia, but also comparing evaluated mathematical reasoning and reasoning performed by these cells, an improved tool for identifying underlying causes of the condition can be implemented. Such a software could facilitate the process of initiating a personalized pedagogical support in a timely manner. It could have the potential of enabling a decrease or even termination of the development of dyscalculia in affected individuals. An additional benefit could include a greater understanding of dyscalculia as diagnosis.

The research questions to be answered by this thesis are presented below.

RQ1. How suitable is mobile software as a tool for assisting diagnosis of dyscalculia?

RQ2. Can mobile software identify dyscalculia symptoms possibly related to magnocellular reasoning?

To answer RQ1 and RQ2, a high-fidelity software prototype for smartphones running Android mobile operating system (OS) has been developed. This was enabled by k-Nearest Neighbor (k-NN) algorithm classifying performances in selected skills in both mathematical and magnocellular reasoning. The reason for developing the application for a mobile OS was due to the low prevalence among mobile applications for identifying underlying causes of dyscalculia at the time of writing. Additionally, exploring the possibilities in utilizing mobile technology was an interesting approach for the author of this thesis. The author's interest in mobile development and back-end software engineering (SE) were contributing factors for choosing smartphone as software platform.

1.2   Outline

The following is an outline of this thesis.

Chapter 2: Literature Review presents relevant theory supporting research questions presented in section 1.1 before discussing related work.

Chapter 3: Method introduces the applied methodological framework *design science* and how it is applied to secure a valid and productive research. Methods supporting software application development and evaluation are also presented.

Chapter 4: Design and Development describes application design and development.

Chapter 5: Evaluation and Results presents expert and user evaluation of the application and results.

Chapter 6: Discussion applies evaluation results in answering research questions presented in section 1.1.

Chapter 7: Conclusion & Future Work summarizes main results and provides suggestions for future work.

Chapter 2

Literature Review

This chapter will present selected relevant theory supporting this research before discussing related work.

2.1 Learning disability

A learning disability (LD) is defined as a classification including several areas of difficulty with learning in a typical manner usually caused by an unknown set of factors [61]. LDs refer to the variety of disorders in a person's ability to obtain, recall, understand, organize or use sensory information. Low intelligence is not a triggering factor as LDs usually affect individuals with average or above average intelligence [22].

2.2 Dyscalculia

Dyscalculia is a specific LD involving difficulties with learning mathematics. The reader might be curious about how an individual who is affected by the diagnosis can be separated from someone who only has normal difficulties with math. A key observation is that there are several known reasons for being weak in mathematics, e.g. the math teacher is not sufficiently qualified or the pupil has missed several lessons. However, the reasons for being dyscalculic are unknown. What is meant by this is that even if a dyscalculic pupil has a great teacher and supportive backgrounds, attends every class and is very motivated, he/she still cannot do the tasks the majority of his/hers fellow students can do. Diagnosing the condition concequently involves several disciplines, including education, psychology and neuroscience [29].

Limited research has been conducted on dyscalculia compared to other LDs. Brian Butterworth, Emeritus Professor of cognitive neurophysiology at University College London, has made a significant contribution to the study of the diagnosis. Butterworth has published several studies on diagnosing dyscalculia [12, 13], especially utilizing time as criteria. The following example illustrates this criteria: Imagine that a dyscalculic ten year old is given the arithmetical problem $5 + 3$. For most of his classmates, the time required to provide the correct answer of eight varies from half a second to a second. However, for the dyscalculic child, it takes more effort for him to respond, requiring five seconds or longer. It does not seem to be too much longer at first glance. Nevertheless, time required represents five to ten times longer than a child not affected by the condition requires. Thus, time is a useful characteristic to apply in differentiating dyscalculic and non-dyscalculic individuals.

2.3   Visual perception

This section will introduce a selection of the literature from neuroscience relevant for this research. In particular, visual perception in the human visual system.

2.3.1   Visual system

The visual system is a part of the central nervous system (CNS). It is centralized in the brain and is responsible for providing the ability to interpret and process information from visible light [10]. This information is used to build a representation of the surrounding environment. The visual system consists of several parts, including the eyes, optic nerve and visual cortex as shown in figure 2.1.
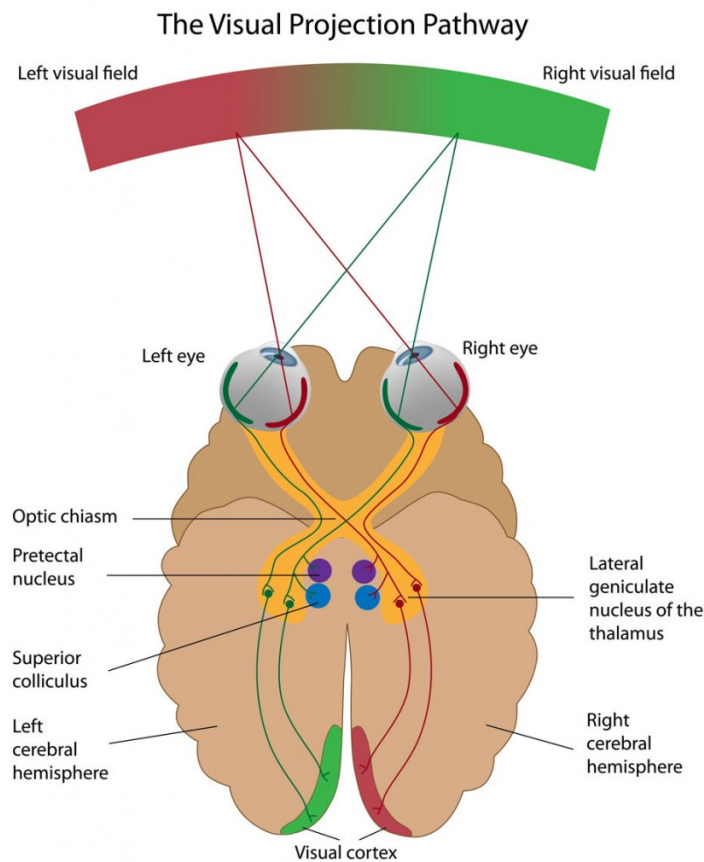
Figure 2.1: The human visual system [25]

The eyes are the visual organs and convert detected light into electrochemical impulses in neurons sent to the brain through the optic nerve connecting the eyes to the visual cortex. The visual cortex is the part of the outer layer of the neural tissue, called *cerebral cortex*, responsible for processing visual information. It is located in the occipital lobe, one of four major lobes of the cerebral cortex located in the lower back of the brain.

2.3.2 Thalamus

The thalamus is a symmetric structure composed of two halves within the human brain and the brain of other vertebrates [48]. Located between the cerebral cortex and the midbrain as shown in figure 2.2, it is the part of the CNS associated with vision, hearing, motor control, alertness and temperature regulation, among others. The thalamus has functions relying on sensory and motoric signals from the cerebral cortex, e.g. signals regulating consciousness.
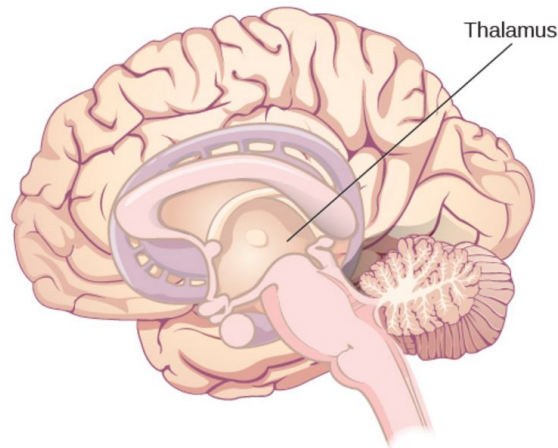
Figure 2.2: Cross-section of the human brain exposing the thalamus [42]

2.3.3 Lateral geniculate nucleus

The lateral geniculate nucleus is a center in the thalamus for the visual pathway of the visual system [10]. It receives large amounts of sensory input information from the retina in the eye and represents the main pathway connecting the optic nerve to the occipital lobe. This connection is analogous to the one between a computer and monitor. Just like a cable is required for obtaining a visual image on a computer monitor, the visual pathway is required for obtaining a visual image in the occipital lobe as we recognize as sight.

2.3.4 Magnocellular cells

Research has shown that there is a relationship between dyslexia and visual perception and processing in the human brain [17, 35, 50]. More specifically, certain cells responsible for transferring information percepted by the visual system in the brain for processing, transfer this information at a slower pace than individuals not diagnosed with a LD. These cells are called *magnocellular cells*, believed to be crucial in visual information processing [49]. They are located within the magnocellular layer of the lateral geniculate nucleus and are responsible for resolving motion and coarse outlines.

Livingstone, Rosen, Drislane and Galaburda at Department of Neurobiology at Harvard Medical School have conducted behavioral studies on a group consisting of five participants diagnosed with dyslexia and seven not affected by the condition [35]. The tests conducted mainly involved contrast sensitivity screening. Each participant was placed in front of a screen providing the stimulus of the experiment represented by a rectangular checkerboard of thirty-six rectangles. These rectangles alternated between different contrast levels. In the evaluation, *evoked response*

was applied on the occipital lobe, an electric potential recorded from the CNS. The method is illustrated in figure 2.3.
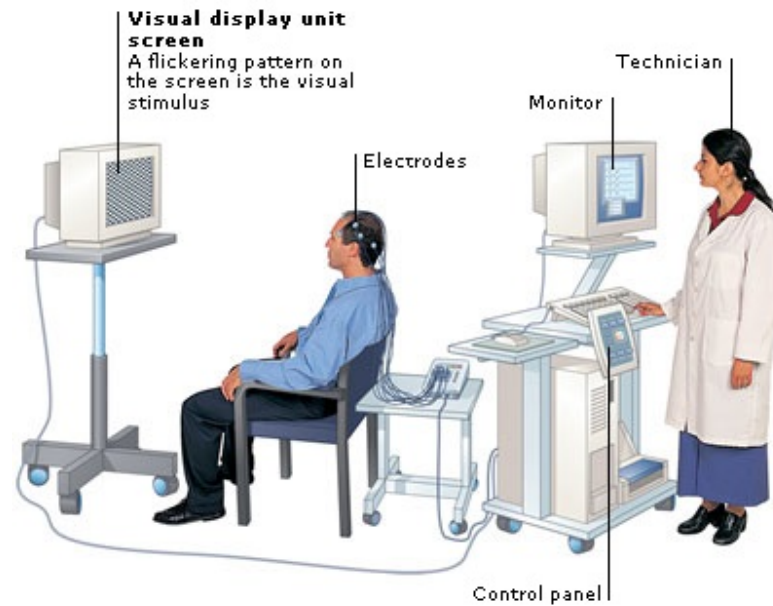
Figure 2.3: Monitoring evoked potential of the human brain following presentation of external stimulus [8]

Evoked responses on 128 different contrast levels from each test subject were logged and averaged. Responses from four dyslexic and six non-dyslexic subjects were then scanned. Finally, generated results from each of the two groups were averaged together as shown in figure 2.4. In the figure, the horizontal axis represents the number of milliseconds required by each participant to recognize the change in contrast. The vertical axis represents the amount of microvolts measuring the activity in the occipital lobe, the evaluated area of the brain. Figure 2.5 illustrates activities in this area marked by colors on a scale where purple is minimum and red is maximum activity level.

After evaluating gathered results, the study concluded that individuals diagnosed with dyslexia require more effort. That is, this group showed higher brain activity to detect changes in contrasts. Consequently, longer time was needed compared to individuals not affected by the LD. This indicated that difficulties in recognizing alternations in the environment are involved in developing deficiencies in visual processing. The study identified a possible correlation between low contrast sensitivity and deficiencies in language reading and writing. As there is a possibility of low contrast sensitivity being related to mathematical deficiencies, it can be valuable to study this possible correlation further to see if it applies to dyscalculia as well.

2.4   Machine learning

Machine learning (ML) is a subfield of computer science stemming from research within artificial intelligence (AI) [37]. It is a scientific discipline exploring the construction, implementation and study of algorithms able to learn automatically from given data. More specifically, these
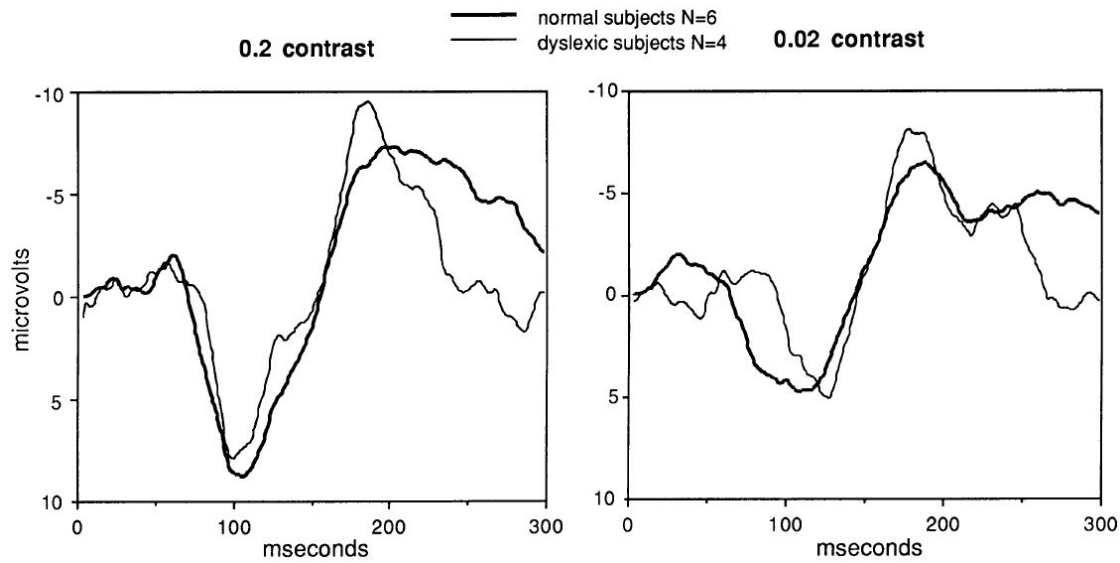
Figure 2.4: Averaged visually evoked responses from dyslexic and non-dyslexic subjects [35]
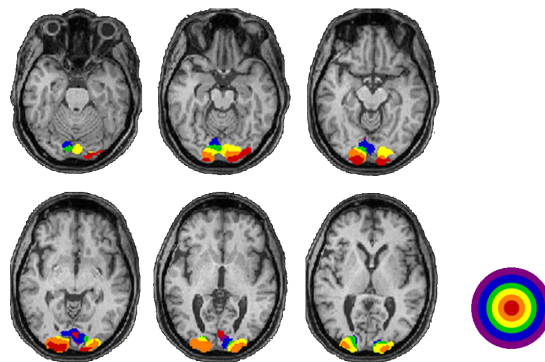
Figure 2.5: MRI section of a human brain exposing the occipital visual areas marking blood flow representing measures of activity during visual pattern simulation [14]

algorithms use data to build a model utilizing information to produce outputs of future data. An advantage of ML techniques is that they are mathematically validated and are capable of performing consistently and reliably.

2.4.1 Classification

Classification is a subfield of ML solving the problem of efficiently and correctly deciding where a new observation belongs among a set of groups called *classes*. That is, based on earlier reviewed data, the classifier needs to identify which group a new observation belongs to, as given in equation 2.1.

$$classifier(o) = c \in C \tag{2.1}$$

In equation 2.1, o represents a received unknown observation. The classifier output c is the class assigned to o from the set of all possible classes C. A popular classification example is separating *spam* from *non-spam* e-mails. Another example is given in figure 2.6 where incoming data is assigned one from a set of two possible classes based on parameters A and B.
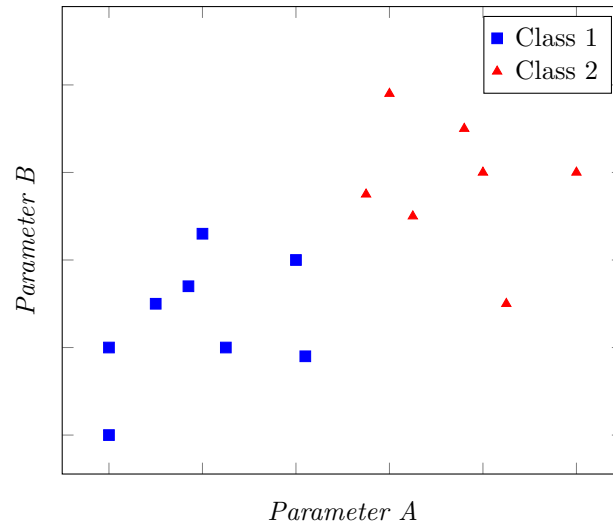
Figure 2.6: Two-parameter classification

2.4.2   Supervised and unsupervised learning

When developing a classifier, the decision on whether to use supervised or unsupervised learning is crucial. Ghahramani presents in his paper regarding ML an in-depth overview of the field [23]. He describes the concept of unsupervised learning as the subfield of ML considering the process of utilizing information retrieved from the surrounding environment to independently produce increasingly better results, eventually learning the machine the optimal method for solving the problem considered. However, in supervised learning, before the machine can begin the learning process, the classifier is required a set of observations and corresponding classes to classify new observations. Therefore, the nature of the problem to be solved is the deciding factor for which approach to select.

2.5   Related work

Software has been developed for supporting diagnosis and remediation of dyscalculia in individuals of all ages. A review of selected available software is presented in subsequent sections.

2.5.1   The Dyscalculia Screener

Butterworth has in his research on dyscalculia developed *The Dyscalculia Screener* [11], a software developed to test children in elementary school for simple number tasks including basic arithmetic and identifying quantities of small sets. However, it does not perform processing of gathered data beyond quantification of mathematical skills. The software's evaluation screen is shown in figure 2.7.
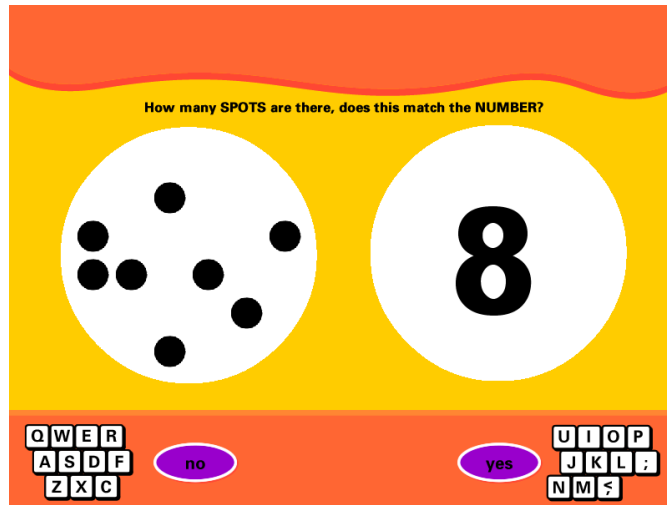
Figure 2.7: The Dyscalculia Screener evaluation screen (Screenshot)

2.5.2 The Number Race

The INSERM-CEA Cognitive Neuroimaging Unit, a leading research institute in mathematical cognition, has developed *The Number Race* [60], a game-based approach for teaching mathematics. It is especially designed to address dyscalculia. This is done by showing numbers in several representations, e.g. digits, words and sets of objects. It focuses on both assisting the diagnosis and teaching mathematical concepts, representing multiple intentions. An improvement would be to focus on one of the two objectives, which is likely to produce more reliable results. The software's evaluation screen is shown in figure 2.8.
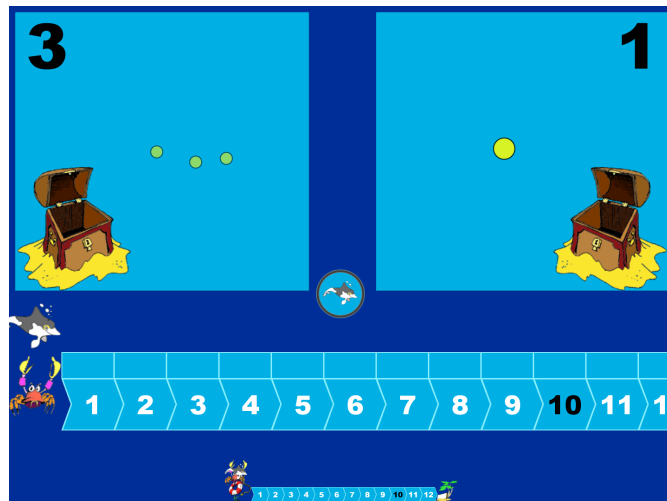
Figure 2.8: The Number Race evaluation screen (Screenshot)

2.5.3   DynamoMath

JellyJames Publishing has developed *Dynamo Math* [19], an online remediation programme for pupils with dyscalculia and limited mathematical skills. It provides teaching plans, comprehensive interactive modules, hundreds of printable worksheets and progress reports. It is aimed both at pupils having symptoms of dyscalculia and pupils already diagnosed, again representing multiple intentions. The nature of the software being hugely complex offering a wide range of functionality and content make the software demanding to use. An improvement would be to decrease the scope of the software to short interactive modules and progress reports, making it more accessible and easy to use for its audience.
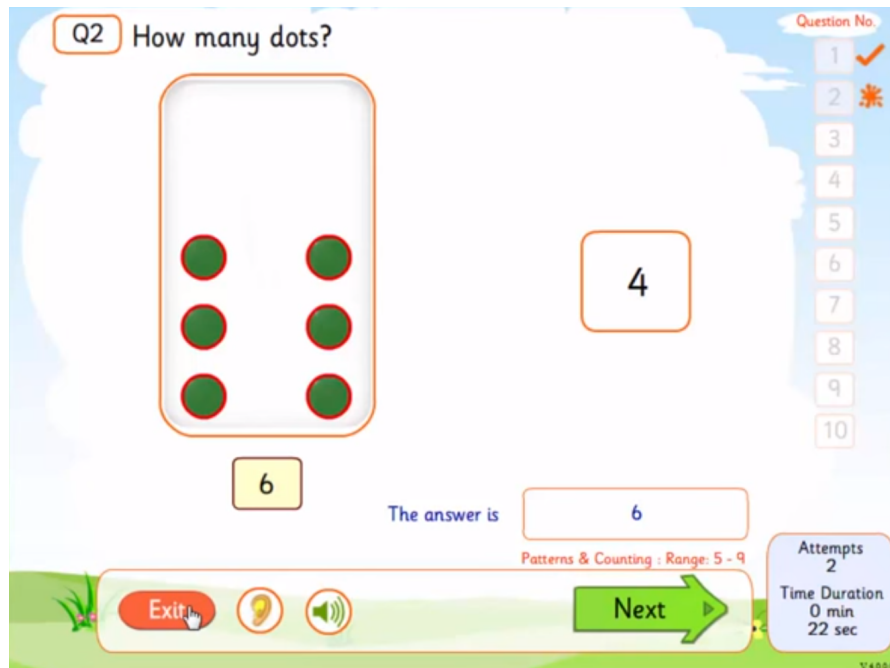
Figure 2.9: DynamoMath evaluation screen (Screenshot)

Correct answers to tasks are presented to users immediately after submitting answers. In the case presented in figure 2.9, the user has submitted 4 before being presented the correct answer of 6. The current task has already been given twice to the user earlier in the evaluation, in figure 2.9 shown by the counter in the bottom right displaying the total number of attempts of solving the task. This approach is not optimal as returning the correct answer after a response is submitted is prone to confuse dyscalculic users. This also suggests that the software is primarily meant for learning, not diagnosing.

2.5.4   DysCalculiUM

Tribal Education in England has developed *DysCalculiUM* [54], a screening tool for evaluating users' understanding of basic number concepts. The software's evaluation screen is shown in figure 2.10. DysCalculiUM's target group consists of adults and students in post-sixteen education. It is developed for screening a large group of users simultaneously. The problem with this approach is that users can be affected by other users' responses and consequently provide feed-

back affected by external influence. Individual evaluation and follow-up could be an improved approach since providing sufficient support to several students at the same time is challenging.
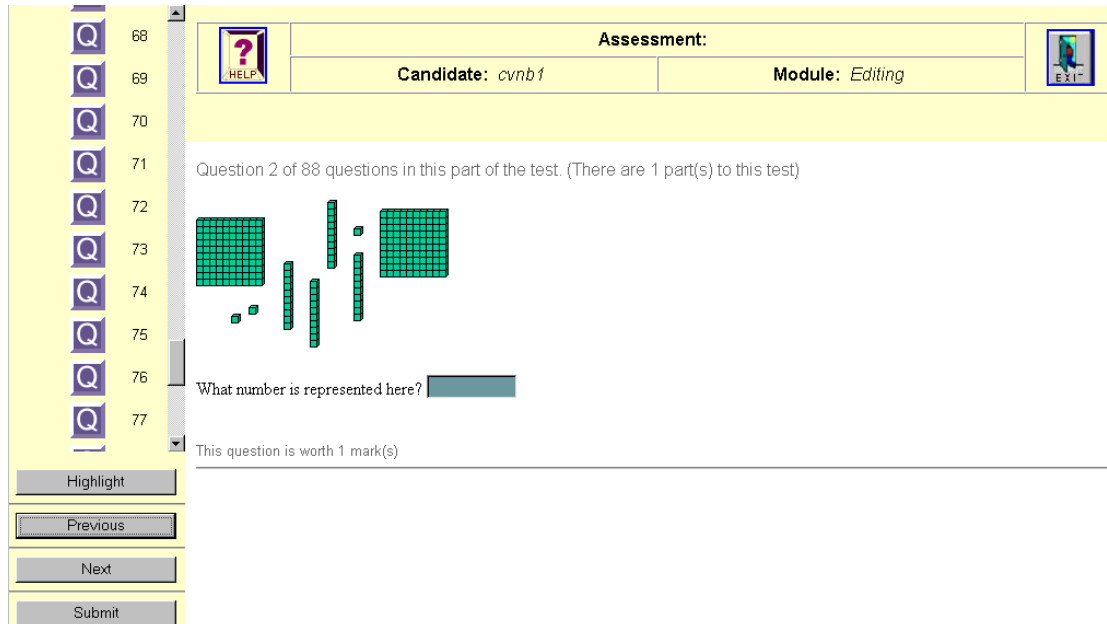
Figure 2.10: DysCalculiUM evaluation screen (Screenshot)

2.5.5 Dyscal

Dyscal shown in figure 2.11 is a mobile application developed for Android OS available at the time of writing at Google's application marketplace, Google Play. It is developed for children with dyscalculia and provides interactive modules containing simple number tasks of counting and identifying quantities.

The interactive modules have several issues. Firstly, they are implemented as infinite sequences of tasks continuing till the user decides to stop. Using this approach, users are not presented any information on when to stop or when an evaluation is completed, which can cause confusion. This behavior can be significantly improved by limiting the scope of the evaluation to a fixed number of tasks, and during the conduct always keeping users updated on the progress.

Secondly, users are required to answer tasks presented correctly before they are allowed to continue. Users are allowed three attempts and must use all three in the attempt of submitting the correct answer before continuing. For dyscalculic users struggling intensively with mathematics, this requirement can be frustrating. A better approach would be to continue the evaluation regardless of the correctness of submitted answers, eliminating the need to exhaust all possibilities in order to continue.

After submitting an answer, users are instantly presented feedback on whether the answer is correct or not. This way of teaching mathematics while diagnosing creates an issue because users' skills are altered along the way decreasing the validity of evaluation results.

(a) Home screen   (b) Evaluation screen   (c) Result screen
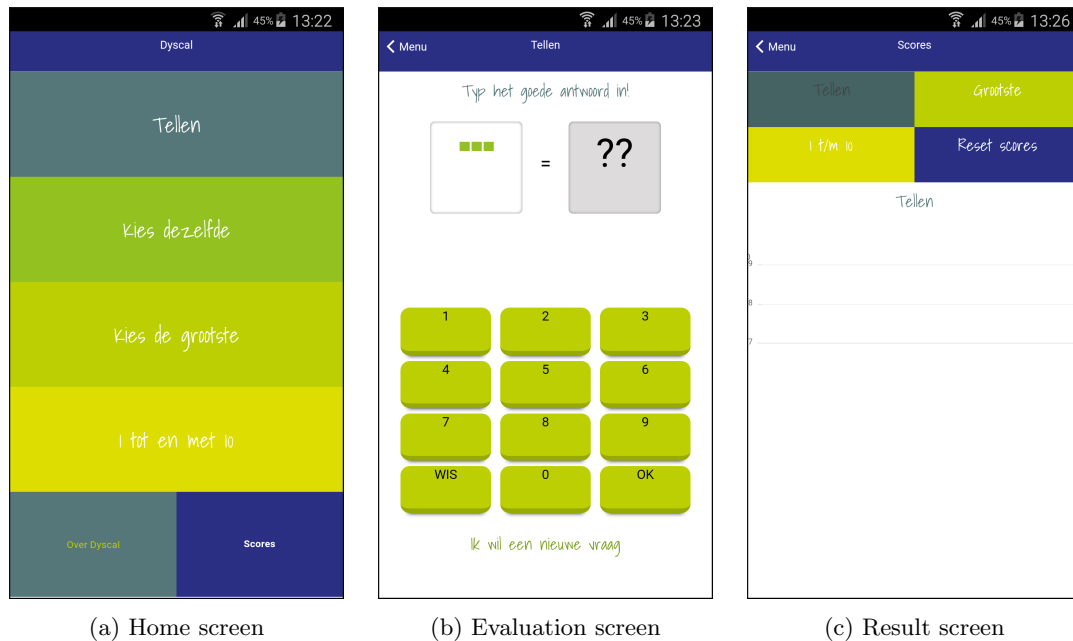
Figure 2.11: Dyscal

2.6   Summary

The majority of available software developed for individuals struggling with mathematics applies
the approach of both identifying symptoms of dyscalculia and at the same time teaching mathe-
matical material. This method is not optimal as this double focus can confuse users as they are
not presented any explanation regarding correctness of answers. This task should be delegated to
the their teacher who has insight into their situation, not to the software. A teacher's guidance
by far exceeds the feedback enabled by software. However, a valuable contribution could be made
by using software to combine LD diagnosis with behavior of magnocellular cells in identifying
possible underlying causes of the condition.

Chapter 3

Method

This chapter will present methods and frameworks applied in this project together with their application.

3.1 Design science

Design science is a research framework involving the design of novel or innovative artifacts and the analysis of the use and/or performance of such artifacts to improve and understand the behavior of aspects of information systems (IS) [55]. Oates defines these artifacts as new IT products [40]. The framework represents a research strategy focusing on developing these artifacts. Mark and Smith present four artifact types given in table 3.1 [36].

Table 3.1: Design science research artifacts [36]

Artifact	Description
Constructs	Concepts or vocabulary used in a particular IT-related domain
Models	Combination of constructs used to aid problem understanding and solution development
Methods	Guidance on the models used to solve problems using information technology
Instantiations	A working system demonstrating that constructs, models, methods, ideas, genres or theories can be implemented in a computer-based system

All artifacts described in table 3.1 represent valuable contributions to the fields of IS and computing. Together they form a composition consisting of both theoretical and technical approaches of contributing to knowledge. According to March and Smith, results of research conducted using the design science framework are often a combination of these artifacts.

Related to the applied approach of conducting research in this project, March and Smith emphasize their criteria for this sort of "development and evaluation" based research projects being considered as research. According to March and Smith, it is not sufficient with a demonstration of technical skills. In order for the development of a technical artifact to qualify as a

research project, the process also has to involve academic qualities such as analysis, argument, explanation, justification and critical evaluation.

March and Smith also stress that the development of the technical artifact is required to contribute to knowledge. How this is carried out depends on the artifact and its role. March and Smith present three roles that an artifact can possess. The role as main focus is the first: achieving contribution to knowledge by developing an IT application using methods not already applied or in a domain not earlier been developed. The role of supporting another aspect is the second: contributing to knowledge by using the developed application to demonstrate or support other methods or findings of the research project. Lastly, contribution to knowledge can be achieved by using the developed application as an end-product for demonstrating a bigger emphasis: the development process. The author of this thesis had a great interest of applying algorithmic techniques within AI in a different manner than what has been earlier conducted within the domain of LDs. This resulted in selecting the first role for the technical artifact developed in this project.

Hevner et al. present seven design science research guidelines given with the purpose of guiding researchers and practitioners in conducting, evaluating and presenting design science research [26]. These guidelines provide an overview of central concepts to consider when utilizing the framework, presented in table 3.2. Subsequent sections will describe how this project applied them.

Table 3.2: Design science research guidelines [26]

#	Guideline	Description
1	Design as an artifact	Design science research must produce a viable artifact in the form of a construct, model, method or instantiation
2	Problem relevance	The objective of design science research is to develop technology-based solutions to important and relevant business problems
3	Design evaluation	The utility, quality and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods
4	Research contributions	Effective design science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations and/or design methodologies
5	Research rigor	Design science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact
6	Design as a search process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment
7	Communication of research	Design science research must be presented effectively to both technology-oriented and management-oriented audiences

3.1.1   Design as an artifact

The first guideline is the requirement that design science research must produce a viable artifact in the form of a construct, model, method or instantiation [26].

The artifact developed in this project is a fully functional high-fidelity software application prototype implemented in Java and eXtensive Markup Language (XML) for smartphones running Android OS. Application development has been the major focus of the first half of the duration of this project. The application was essential for answering research questions presented in section 1.1. Having developed an instantiation using ML techniques within a business domain in a novel way, design guideline one has been fulfilled.

3.1.2 Problem relevance

The second guideline states that the objective of design science research is to develop technology-based solutions to important and relevant business problems [26]. In design science, this is achieved through the construction of innovative artifacts applying novel approaches.

In the process of satisfying this guideline, evaluating research relevant for this project was useful for identifying valuable contributions, as done in section 2.5. However, none of the applications evaluated are developed for identifying underlying causes of the condition. With the new approach of applying k-NN classification on response data, a different method of identifying LDs and underlying causes has been applied. Additionally, the literature review given in chapter two has suggested a low prevalence among mobile applications developed for identifying underlying causes of dyscalculia, making this research relevant at the time of writing, satisfying guideline two.

3.1.3 Design evaluation

The third guideline is the requirement of the utility, quality and efficacy of a design artifact being rigorously demonstrated using well-executed evaluation methods. Hevner et al. [26] further state that the evaluation must be performed through appropriate metrics.

In this project, evaluation constitutes a crucial role in answering research questions presented in section 1.1. Firstly, students, researchers and practitioners evaluated the developed application to test its methods and functionality. Then, its validity, accuracy and usability were evaluated by individuals having symptoms of dyscalculia. Including both experts in underlying fields and potential users of the application has satisfied guideline three.

3.1.4 Research contribution

The fourth guideline states that effective design science research must provide clear and verifiable contributions to the areas of the design artifact, design foundations and/or design methodologies [26]. This guideline can be fulfilled by specifying specific requirements to the software application to be developed. This supports the importance of the artifact contributing to the knowledge base of one or several of its underlying research fields.

In this project, these underlying fields include dyscalculia, LDs, education, psychology, ML and AI. The main contribution to knowledge will be dyscalculia symptoms identified as possibly being related to magnocellular reasoning, a process enabled by the artifact itself. That is, the success the artifact has in fulfilling this purpose can provide new knowledge connecting application output to underlying causes of dyscalculia, providing the required contribution described by guideline four.

3.1.5 Research rigor

The fifth guideline encompasses the rigor of the research to be conducted, explaining that design science relies on the application of rigorous methods in both the construction and evaluation of

the design artifact [26].

Firstly, Personal Scrum was applied for assisting the development process of the artifact as described in section 3.3.2. Personal Scrum was especially useful for organizing the implementation of functional requirements specified for the artifact in section 4.1.1. To assist the development and evaluation of the artifact's user interface (UI), a set of usability heuristics developed by Jacob Nielsen [38] was applied as described in section 3.6.3. Then, to evaluate the usability of the artifact, the System Usability Scale (SUS) developed by John Brooke [9] was applied as described in section 3.6.4. By applying these methods proven to be effective and valuable for developing and evaluating the artifact, guideline five has been fulfilled.

3.1.6   Design as a search process

The sixth guideline emphasizes the fact that the search for an effective artifact requires utilization of available means to reach desired ends while satisfying laws in the problem environment [26]. In this context, *searching* refers to continuous improvement of the artifact through iterative and incremental development. The *ends* are represented by the goals and constraints and how to balance these in order to maximize success. The *means* are the resources applied and actions performed in order to construct the artifact.

In this project, the requirement of continuous improvement is satisfied by the iterative nature of Personal Scrum. *Ends* are represented by functional requirements specified for the artifact given in section 4.1.1. As described in guideline five in section 3.1.5, the artifact was implemented in Java and XML, representing the *means* of the project. Additional resources are given in section 4.3.

By applying these methods for continuous improvement through iterative development, guideline six has been fulfilled.

3.1.7   Communication of research

The last guideline is the requirement that design science must be presented effectively to both technology-oriented and management-oriented audiences [26]. This guideline is essential to take into account as it forms the approach of communicating research questions, their supporting theory and literature as well as the presentation and discussion of results.

This guideline is fulfilled by including in this thesis and related scientific publications both the technology-oriented and the not-so-technology-oriented audience by presenting the material and findings in a manner understandable for non-experts. This is accomplished by providing brief yet effective introductions to the background material in order to draw knowledge and value from results. While software application methodology, technologies and development might be more interesting to technology-oriented readers, application design, functionality, evaluation and findings are presented to be of interest for everyone.

3.1.8   Summary

The design science framework has been applied by developing an instantiation of an artifact in the form of a mobile application for assisting diagnosis of dyscalculia as well as for identifying dyscalculia symptoms possibly related to magnocellular reasoning. The purpose was to contribute to knowledge by examining whether the new approach applied yielded a greater success in the process of identifying underlying causes of developmental dyscalculia. Table 3.3 provides a summary of design science research guidelines as applied in this project.

Table 3.3: Design science research guidelines applications as presented in this thesis

Guideline	Application
Design as an artifact	This thesis describes the development and evaluation of Magno-Math, a mobile application for assisting diagnosis of dyscalculia and identifying dyscalculia symptoms related to magnocellular reasoning. **Chapter 4** presents the design and development of the artifact while **chapter 5** presents artifact evaluation.
Problem relevance	**Chapter 1** presents the business domain and the potential of applying classification algorithms in implementing an improved tool for identifying dyscalculia symptoms related to magnocellular reasoning. Related work reviewed in **chapter 2** suggested that the current situation was in need for improvements. This provides the relevance of this research.
Design evaluation	**Chapter 5** presents artifact evaluation design, conduction and results using interviews, questionnaires and observations.
Research contributions	**Chapters 4 and 5** present and discuss MagnoMath, the designated artifact contributing to the selected business domain. **Chapter 6** presents findings generated by the artifact contributing to the domains of dyscalculia and SE.
Research rigor	**Chapters 3, 4 and 5** present relevant methods for designing, implementing and evaluating the artifact.
Design as a search process	MagnoMath is a result of an incremental development process enabling continuous improvement during implementation and evaluation. **Chapters 4 and 5** together document the process of searching for the best solution for the artifact within the project scope.
Communication of research	The research conducted in this project is communicated through this thesis as a whole and scientific publications.

3.2 Data acquisition

This section will describe methods applied in this project for acquiring data as well as the processing and utilization of this data. Table 3.4 summarizes the data acquisition methods considered for this research.

3.2.1 Qualitative data

3.2.1.1 Interview

The first step of the data acquisition was to obtain an insight into the underlying fields of the study with a focus on dyscalculia and LDs in general. In order to achieve this, method 1 and 2 given in table 3.4 were considered before selecting the former. One-on-one meetings with interviewees yield data of more detail as compared to meetings including several interviewees. Qualitative data acquisition was supported by interview guides given in appendix B applied with the purpose of providing interview sessions structure and consistency.

Table 3.4: Comparison of data acquisition methods [46]

#	Method	Suited for	Types of data	Advantage	Disadvantage
1	Interview	Locating mistakes	Mostly qualitative, some quantitative	The interviewer can assist the interviewee if needed	Time consuming
2	Group interview	Gathering several opinions	Mostly qualitative, some quantitative	Communicating agreements and disagreements	Some participants can dominate and obtain too much attention
3	Questionnaire	Answering specific questions	Quantitative and qualitative	Effective and cost-efficient	Risk of low response rate
4	Direct observation in the field	Obtaining insight into real use in real environment	Qualitative	Generates valuable data not offered by other methods	Time consuming and possibly a huge amount of collected data is unstructured
5	Direct observation in controlled environment	Obtaining insight into real use where the data must be logged efficiently	Qualitative and quantitative	Can focus on details without being interfered by technical issues	Data can be affected by the artificial environment
6	Indirect observation	Obtaining insight in real use in real environment without interrupting the user, data gets collected automatically	Quantitative automatic logging and qualitative notes	The user does not get interrupted by the gathering of data	Possibly the majority of collected data is quantitative

3.2.1.2   Observation

The next part of the data acquisition involves expert and user evaluation of the application. Since participants were not given feedback regarding correctness of responses, they were not affected by the knowledge of their performance being poor or strong. Therefore, the application behaved identically to all participants during evaluation regardless of performance. Additionally, participants were well informed that it is not they who are under evaluation, but the application's ability to acquire and utilize user responses. Because of this, method 5 in table 3.4 was considered most suitable to support application evaluation.

3.2.2 Quantitative data

3.2.2.1 Questionnaire

After completing evaluation sessions, participants were given a set of statements to rank regarding their experience of the application. This method was applied with the purpose of obtaining quantitative feedback covering both application functionality and usability. This type of data was useful due to its quantifiable nature enabling easy comparison and evaluation.

3.3 Software development

Nunamaker, Chen and Purdin provide several frameworks for supporting the use of software development as a methodology in IS research, explaining the nature of system development as a research methodology [39]. SE is especially emphasized in this context. Subsequent sections will present frameworks applied in this project supporting artifact development.

3.3.1 A multimethodological approach to research

According to Nunamaker, Chen and Purdin, a research process involves an understanding of research domains, asking meaningful research questions and applying valid research methodologies to address these questions [39]. Therefore, combining several different methodologies to cover the full aspect of the research process is valuable. The key is to utilize the research life cycle together with the life cycle of SE. This is done by constructing a model combining such methodologies from design science and SE given in figure 3.1. The research performed in this project focused on three of the four IS research methodologies: system development, observation and experimentation. The first was applied in developing the software artifact for assisting dyscalculia diagnosis and identifying possible underlying causes of the condition. The second was applied in the expert and user evaluation of the artifact using both direct observation in the field and indirect observation remotely. The third was applied in post-evaluation conduction where experiments were conducted with the purpose of utilizing collected data in identifying new findings.

3.3.2 Personal Scrum

Scrum software development methodology was applied to support artifact development as presented in figure 3.2. Scrum is a development framework in which cross-functional teams develop products or projects in an iterative and incremental manner [16]. This way, continuous improvement can be maintained. Scrum achieves this by using *sprints* followed by corresponding *sprint backlogs*. A sprint is a period of time allocated to implement a set of specific requirements of the system or software under development. A sprint backlog is a review of the sprint addressing issues occurred and specifying new functionalities to be implemented in the next sprint.

In this project, the Scrum team consisted of one individual only: the author of this thesis. Since the traditional use of Scrum has a minimum of three team members, the methodology had to be adjusted. A version of Scrum suitable for this project is called *Personal Scrum*, an agile methodology adapting Scrum practices to one-person projects [43]. The main difference between Scrum and Personal Scrum is that the scope of sprint backlogs is heavily decreased as compared to traditional Scrum. The individual in Personal Scrum has a broad overview of the project while the Scrum master in traditional Scrum is often limited to client communication.
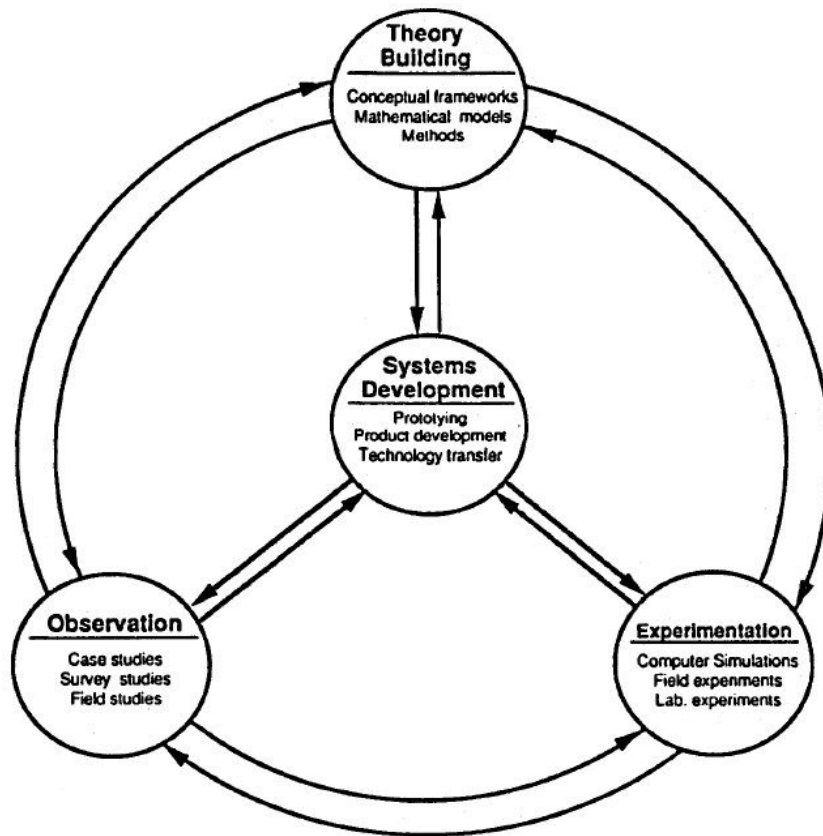
Figure 3.1: A multimethodological approach to research [39]
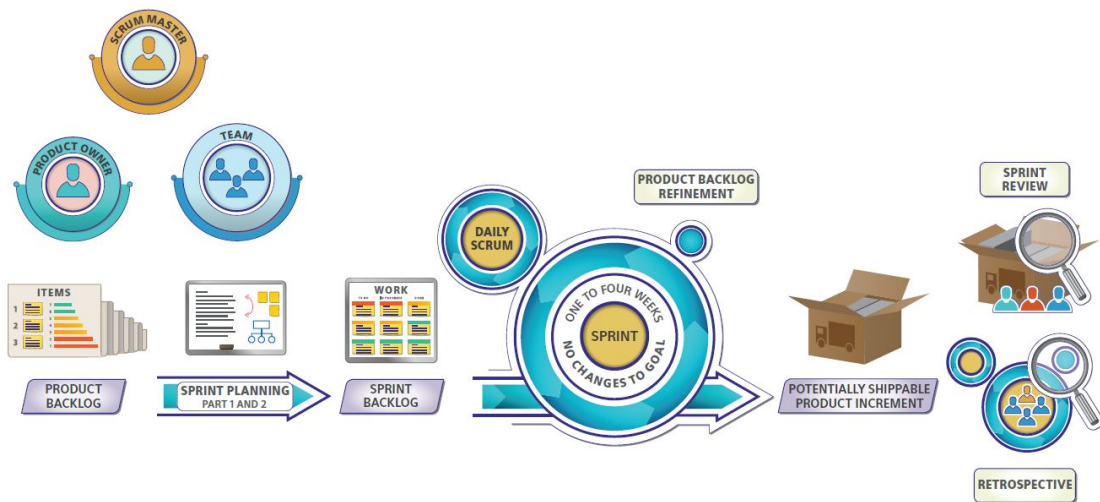
Figure 3.2: Scrum software development model [16]

3.4 Classifying evaluation responses

Most of today's software developed for helping pupils struggling with mathematics perform limited processing on acquired data as discussed in section 2.5. This is unfortunate since this data could be utilized to give feedback regarding their mathematical skills being possibly related to magnocellular reasoning. This has the potential of detecting early signs of developing mathematical deficiencies. The application developed in this project achieves this by classifying evaluation response data into appropriate performance classes. This section will present the methodological background of the classifier implemented and applied in this project.

3.4.1 Parameters

Since ML was the applied approach for classifying input data, this data was required to be discrete. More specifically, each result in the data set had to come from a finite set of possible values. For the classification performed in this project, a data instance is in the form $\alpha \subset \{\mathbb{N} : 0 \leq \alpha_i \leq 100\}$ where α is a data set and α_i is a result in the data set α.

3.4.1.1 Mathematical skills

Since the classifier was to evaluate users' mathematical reasoning, a selection of mathematical skills had to be specified. Since dyscalculia is a LD, specifying a finite set of skills applicable to all having symptoms of the condition is infeasable. Therefore, only a subset of skills known to be affected by dyscalculia [59] are included. Table 3.5 shows skills classified to identify and measure dyscalculia symptoms.

Table 3.5: Mathematical skills classified

Mathematical skill	Elaboration
Counting	Using addition to find the number of a finite set of objects
Subitizing	Recognizing the quantity of patterns of sets without counting elements one by one
Artihmetic	Mentally add, subtract, multiply and divide numbers

Among mathematical skills given in table 3.5, *Subitizing* requires further elaboration. The term was coined by Kaumann et al. and is derived from the Latin verb subitus, meaning "sudden" [30]. It captures the feeling of immediately knowing the number of items when presented all at once, as long as the quantity of elements is within a limited range. Figure 3.3a provides an example of a set of elements where the quantity is assumed to be immediately identifiable by a non-dyscalculic individual. Figure 3.3b however shows a set of elements where the quantity presented is beyond the subitizing range, requiring the elements to be divided into two or more groups in order to determine its quantity.

3.4.1.2 Magnocellular reasoning skills

As specified in functional requirement two given in section 4.1.1, the software application developed needed to be able to measure correlations between dyscalculia symptoms and magnocellular reasoning. Therefore, an additional group of tasks evaluating users' magnocellular reasoning skills

<div align="center">(a) Can be subitized   (b) Cannot be subitized</div>
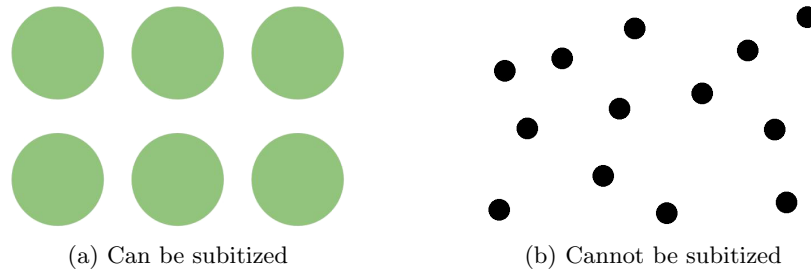
<div align="center">Figure 3.3: Sets of elements</div>

had to be included. These tasks focused on evaluating two skills believed to be associated with magnocellular reasoning: *contrast sensitivity* and *spatial ability* [7, 35, 49, 50], presented in table 3.6.

<div align="center">Table 3.6: Magnocellular reasoning skills classified</div>

Magnocellular reasoning skill	Elaboration
Contrast sensitivity	Identifying and separating contrast levels
Spatial ability	Mental manipulation of three-dimensional figures

The former skill involves distinguishing contrast levels using alternations in color or luminance. The software application developed evaluates this skill by asking users to identify elements in different shades of grey. The latter skill is the ability to understand and remember spatial relationship between objects, a skill also believed to be depending on information transmission in magnocellular cells. In the context of the developed software application in this project, tasks evaluating spatial ability asked users to quantify elements in the three-dimensional space. Mental folding and unfolding of three-dimensional figures were also incuded to evaluate spatial ability.

3.4.2   Algorithm

Classification of skills functions by assigning $\mathbb{N}_{[0,100]}$ to each skill given in tables 3.5 and 3.6. These measures represent a set of discrete values suggesting user performance within each skill. For instance, 0 suggests a poor performance while 100 suggests a strong performance. This approach was inspired by Julie M. David and Kannan Balakrishnan's research where they applied ML in the process of predicting LDs in elementary school-aged children [15].

3.4.2.1   Selection

When the skills to be evaluated had been specified, the parameters they represented could be processed by a classification algorithm. In the context of the application developed, this algorithm had the responsibility of assigning sets of given natural-valued inputs, called *observations*, to appropriate performance classes. Several classification algorithms were evaluated and compared before selecting k-Nearest Neighbor (k-NN). k-NN consists of a group of statistical ML algorithms for learning functions by utilizing information from a given group of input-output pairs called *training set* [47]. It is a non-parametric algorithm within classification and regression and is

among the group of statistical ML algorithms for classifying input data using a training set. Its simplicity, high performance and adaptability were the deciding factors for selecting k-NN [5]. In the context of this project, the classifier did not require a significant amount of training before classification could start. Since the application runs on a smartphone with limited performance compared to a desktop computer, this characteristic suited the selected platform. Additionally, as the set of output classes was known in advance, it was easy to adopt k-NN to the classification problem to be solved.

k-NN, along with several other classification algorithms, utilizes input data to teach the function given in equation 3.1.

$$f(\{x \subset X\}) = c \in C \qquad (3.1)$$

In equation 3.1, x denotes a subset of X which is the set of natural numbers between 0 inclusive and 100 inclusive, represented by $x \subset \{\mathbb{N} : 0 \leq x \leq 100\}$, the set of possible performance measures for each of the five skills presented in tables 3.5 and 3.6. c is the corresponding class from the collection of classes with values of natural numbers from 1 inclusive to 3 inclusive, represented by $c \in \{\mathbb{N} : 1 \leq c \leq 3\}$, the assigned performance class generated from performance values.

k-NN algorithm works by first being trained using a set of k observations, each with n coordinates in the n-dimensional space, one for each parameter to classify. Then, the Euclidean distances between the point represented by a new observation p and each of the existing points q in the space are calculated as given in equation 3.2 [18].

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \qquad (3.2)$$

Euclidean distance is a discrete measurement of the distance between two points in the n-dimensional space as illustrated in figure 3.4. The procedure is given in algorithm 1.
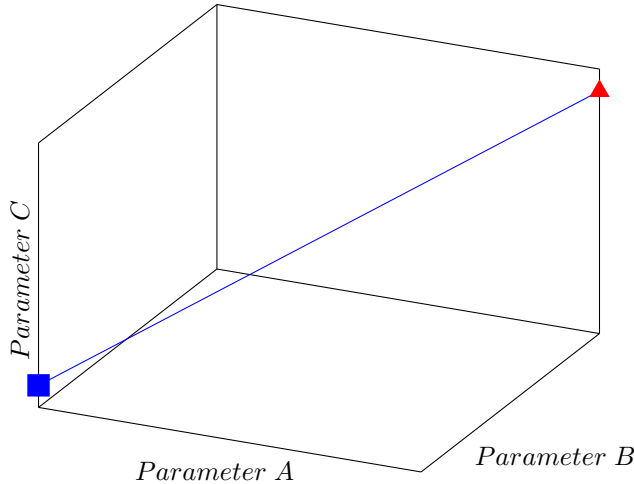
Figure 3.4: Euclidean distance between two points in the three-dimensional space

---

Algorithm 1 Euclidean distance

Input: p and q, n-dimensional cartesian coordinates
Output: α, Euclidean distance between p and q

1: **Initialize** $\alpha \leftarrow 0$
2: **for all** $p_i \in p$ and $q_i \in q$ **do**
3:     $\alpha \leftarrow \alpha + (p_i - q_i)^2$
4: **end for**
5: $\alpha \leftarrow \sqrt{\alpha}$
6: **return** α

3.4.2.2   Initial training

To classify a new observation p, k-NN has to compare p with k or more already classified observations $q \in T$ in order to assign a class c to p. A training set T of size k is required because the k-NN classifier needs a minimum of k training data records in order to start classifying new observations p. With the classifier developed in this project, k-NN receives inputs in the form of points consisting of natural-valued coordinates. The training set used to initiate the k-NN classifier evaluating mathematical skills is given in table 3.7 with performance classes $C = \{1, 2, 3\}$ where 1 is minimum and 3 is maximum performance.

Table 3.7: Initial mathematical performance classifier training set

Counting perf	Subitizing perf	Arithmetic perf	Perf class
0	0	0	1
50	50	50	2
100	100	100	3

The training set given in table 3.7 initiates the k-NN classifier evaluating mathematical skills as shown in figure 3.5. An additional instance of the classifier was trained for classifying magnocellular reasoning skills with the training set given in table 3.8. This training set initiates the k-NN classifier evaluating magnocellular reasoning skills as shown in figure 3.6.

Table 3.8: Initial magnocellular performance classifier training set

Contrast sensitivity perf	Spatial ability perf	Perf class
0	0	1
50	50	2
100	100	3

Euclidean distance is in k-NN applied to calculate the similarity between two observations. New observations p are classified by first calculating the Euclidean distance between p and all existing observation $q \in T$. The class c of q resulting in the lowest Euclidean distance among k nearest neighbors of p is assigned to p as shown in equation 3.3.
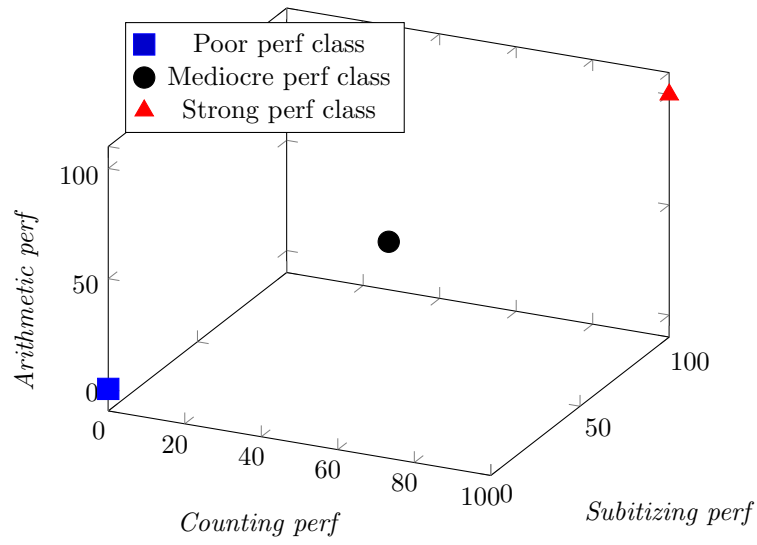
Figure 3.5: Initial mathematical performance classifier state
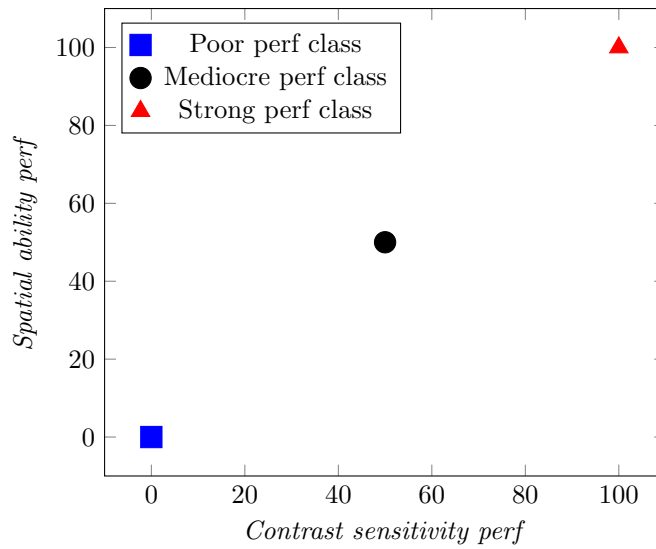
Figure 3.6: Initial magnocellular performance classifier state

$$classification(p) = class \ of \ \underset{q \in T}{\operatorname{argmin}}(eucDis(p,q)) \tag{3.3}$$

The complete k-NN classification procedure is given in algorithm 2.

Algorithm 2 k-NN

Input: T, set of $\geq k$ classified observations q, p, new observation
Output: α, class $c \in C$ of p

  1: **Initialize** shortest distance $d_{min} \leftarrow$ maximum positive value
  2: **Initialize** closest neighbor $q_{min} \leftarrow$ null
  3: **for all** observations $q \in T$ **do**
  4:   $currDist \leftarrow$ eucDist(p,q)
  5:   **if** $currDist < d_{min}$ **then**
  6:     $d_{min} \leftarrow currDist$
  7:     $q_{min} \leftarrow q$
  8:   **end if**
  9: **end for**
10: $\alpha \leftarrow$ class $c \in C$ of q_{min}
11: **return** α

3.5 Correlation computation

Another objective with this project was to evaluate the ability of mobile software to detect and measure correlations between dyscalculia symptoms and magnocellular reasoning abilities as presented in RQ2 in section 1.1. This correlation was partly evaluated by computing the ratio of performances in mathematical and magnocellular reasoning, respectfully. This computation is given in algorithm 3.

Algorithm 3 Mathematical/magnocellular performance correlation

---

Input: R, set of performance classification mappings $classificationRes_i \rightarrow c \in C$
Output: α, correlation in percent

  1: $mathRes \leftarrow$ set of mathematical performance mappings $mathRes_i \in R$
  2: $magnoRes \leftarrow$ set of magnocellular performance mappings $magnoRes_i \in R$
  3: $mathSum \leftarrow \sum mathRes_i \in mathRes$
  4: $magnoCellSum \leftarrow \sum magnoRes_i \in magnoRes$
  5: $mathSum \leftarrow mathSum \times \frac{nrOfMagnoCellParams}{nrOfMathParams}$ {reducing $mathSum$ to result space of $magnoCellSum$}
  6: $\alpha \leftarrow |mathSum - magnoCellSum|$
  7: $\alpha \leftarrow \alpha \times 0.5$ {reducing α to percent}
  8: $\alpha \leftarrow 100 - \alpha$
  9: **return** α

Algorithm 3 returns the similarity in percent between results obtained on mathematical and magnocellular reasoning tasks. This to give users feedback on how their performance in mathematics compares to their performance in magnocellular reasoning. Section 6.2 presents results from methods applied for identifying correlations between task categories. That is, correlations between specific skills rather than between the two groups of tasks.

3.6 Software artifact design

This section will describe methodologies applied for developing the design and UI of the mobile software application developed and applied in answering RQ1 and RQ2 presented in section 1.1.

3.6.1 Prototype

A *prototype* is a manifestation of a design allowing users to explore its suitability [46]. What makes prototypes so useful is that they provide end-users direct interaction with the design of the product under development. This enables developers to get feedback from real users before the final version is released. Receiving this feedback before the final completion of the product is highly cost benefitial and time-saving. Additionally, users participating in this evaluation can assist the development of the final product, thus improve its suitability for its intended users.

There are two types of prototypes: *low-fidelity* and *high-fidelity*. Low-fidelity prototypes use materials very different from the final product, such as paper and cardboard instead of physical devices and program code. On the other hand, high-fidelity prototypes apply materials a user would expect to find in the final release, appearing to be very similar to the real product under development. A comparison is given in table 3.9.

Table 3.9: Comparison of prototype types [46]

Type	Advantage	Disadvantage
Low-fidelity	Low development costs	Limited error checking
	Enables evaluation of multiple design concepts	Cannot test navigation and information flow thoroughly
	Proof-of-concept	Limited usefulness for usability tests
High-fidelity	Complete functionality	Time-consuming to create
	Fully interactive	Not effective for gathering requirements
	Provides look and feel of final product	More expensive to develop

After comparing the different types, it was seen as more useful to base this project on a high-fidelity prototype rather than a low-fidelity one. Additionally, due to the author's interest for SE and experience with programming it was decided that the prototype in this project would be developed to manifest a fully functional product.

3.6.2 Human Computer Interaction

Rogers et al. define Human Computer Interaction (HCI) as the process of designing interactive products to support the way people communicate and interact in their everyday lives [46]. In the context of SE, HCI defines the interface between computers and their human users, hence its name. HCI studies how UIs should be designed in order to appear accessible and easy to use. The UI design of the software application developed in this project was therefore crucial for ensuring a positive experience for users as it is the only part of the software visible to them.

3.6.3 Nielsen's usability heuristics

The Merriam-Webster dictionary defines *heuristic* as an approach serving as an aid for learning or discovering the solution to a problem. Jacob Nielsen defines a set of ten heuristics to guide UI

development for any kind of prototype, may it be a physical object or a software [38]. Nielsen is User Advocate and principal of the Nielsen Norman Group. The majority of his work involves studying usability for making systems and the internet easier to use. His ten heuristics are presented in subsequent sections.

3.6.3.1   Visibility of system status

The first heuristic states that the system developed should always keep users informed about the status of each section of the system or each action the system performs. This is very important, because if an operation requires a significant amount of time to complete and users are not informed about the status, they have a tendency of getting impatient resulting in users leaving the system.

3.6.3.2   Match between system and the real world

The second heuristic states that the language in the system used to communicate with users should be similar to the language used in the real world. That is, it should apply the language used by the users themselves with natural words and phrases rather than technical terms.

3.6.3.3   User control and freedom

The third heuristic emphasizes users' sense of freedom within the system, stating that the system should support undo and redo operations. This way, users can easily exit an unwanted state after having made a mistake.

3.6.3.4   Consistency and standard

The fourth heuristic focuses on avoiding ambiguity of words or actions within the system. That is, the user should not have to wonder whether different words, terms, actions or situations mean the same or result in the same outcome.

3.6.3.5   Error prevention

The fifth heuristic focuses on preventing users performing tasks resulting in system error. Error messages can be used, but optimally errors should be prevented from occurring in the first place. Error-prone sections of the system should therefore apply sufficient communication so that users does not perform actions resulting in system halt.

3.6.3.6   Recognition rather than recall

The sixth heuristic emphasizes that the system should remember repeating choices made by the user, without having to ask for the same information several times. Additionally, this already provided information should be retrievable for the user at all times.

3.6.3.7   Flexibility and efficiency of use

Heuristic seven states that the user should be able to automate frequently performed actions.

3.6.3.8 Aesthetic and minimalistic design

Heuristic eight states that the system should not communicate information which is irrelevant to the current situation or rarely needed in general. The more information users receive, the more cumbersome it is for them to use the system and extract relevant information.

3.6.3.9 Help users recognize, diagnose, and recover from errors

Heuristic nine is concerning the content of error messages. Systems often refer to error codes when displaying an error with limited information about the actual error or suggestion for a working solution. Therefore, to assist the user when an error occurs, the system should precisely indicate and inform about the problem, and constructively offer a solution written in understandable, plain language.

3.6.3.10 Help and documentation

The tenth and last heuristic is regarding the documentation accommodating the system, consisting of information about assistance and elaborations. Although the best is to develop a system so easy to use so it does not require such information, it may be necessary for some users in order to take full advantage of the system's functionality.

3.6.4 Brooke's System Usability Scale

Additional to Nielsens usability heuristics, it was considered useful to obtain singular discrete usability measurements of the developed application as a whole. System Usability Scale (SUS) enables this [9]. Developed by John Brooke, an earlier IT solutions architecture consultant, SUS evaluates a total of ten metrics of a system by presenting a set of statements. Odd-numbered statements are positive while even-numbered are negative as listed below.

1. I think that I would like to use this system frequently.

2. I found the system unnecessary complex.

3. I thought the system was easy to use.

4. I think that I would need the support of a technical person to be able to use this system.

5. I found the various functions in this system were well integrated.

6. I thought there was too much inconsistency in this system.

7. I would imagine that most people would learn to use this system very quickly.

8. I found the system very cumbersome to use.

9. I felt very confident using the system.

10. I needed to learn a lot of things before I could get going with this system.

Each statement is to be ranked on the following scale: *Strongly disagree*, *Disagree*, *Neutral*, *Agree*, *Strongly agree*. Each response is then converted to a number from 1 to 5 before the SUS score is computed as shown in algorithm 4.

Algorithm 4 SUS score
_____
Input: α, set of rankings $\alpha_i \in \{\mathbb{N} : 1 \leq \alpha_i \leq 5\}$
Output: β, SUS score

  1: **Initialize** $\beta \leftarrow 0$
  2: **for all** rankings $\alpha_i \in \alpha$ **do**
  3:    **if** α_i is even **then**
  4:       $\alpha_i \leftarrow 5 - \alpha_i$
  5:    **else**
  6:       $\alpha_i \leftarrow \alpha_i - 1$
  7:    **end if**
  8:    $\beta \leftarrow \beta + \alpha_i$
  9: **end for**
10: $\beta \leftarrow \beta \times 2.5$
11: **return** β
_____

3.7 Evaluation

After an application is developed, a thorough evaluation is crucial as discussed in design science guideline three in section 3.1.3. This will test whether generated results correspond to correct classification of skills.

The evaluation of the developed application represented a significant phase in this project involving a set of *focus groups*. A focus group is an evaluation method which applies qualitative research. It involves questioning a group of people about their perceptions, opinions, beliefs and attitudes towards a product, concept or idea [33]. The data gathered can be in any form, e.g. quantitative data from surveys or qualitative data from observations of participants interacting with the product. This project applied a combination of the two as discussed in section 3.2. The focus groups included in this project are presented in subsequent sections.

3.7.1 Expert group

The first focus group included in this research consisted of individuals having an expertise in dyscalculia, dyslexia and/or LDs in general. This focus group was divided into two subgroups.

The first subgroup consisted of individuals with a strong expertise in SE and mathematics. This group was selected for testing and verifying the algorithmic techniques applied with the purpose of obtaining feedback regarding their accuracy and suitability.

The second subgroup consisted of researchers within the fields of dyscalculia, dyslexia and/or LDs in general. This group was selected to give feedback on methods the software application applied to assist diagnosis and identify dyscalculia symptoms possibly related to magnocellular reasoning.

3.7.2 User group

The second focus group included potential users of the application. This group was divided into two subgroups.

The first subgroup consisted of individuals having symptoms of dyscalculia. This group was crucial to include for answering research questions presented in section 1.1. This group enabled the application to identify possible underlying causes of the condition.

The second subgroup consisted of individuals having no specific struggles with mathematics. More specifically, individuals not having symptoms of dyscalculia. This group was included to test application's ability to separate dyscalculic from non-dyscalculic users.

3.8 Validity

Validity in IS research is regarding the measure of (1) whether an appropriate process has been used, (2) whether the findings of the research do indeed come from the data gathered and (3) whether they answer presented research questions [40].

The processes of implementation and evaluation in this project are supported by techniques and methods widely used and proven to be successful. These include Personal Scrum described in section 3.3.2 and observations followed by associating questionnaires as described in section 3.2.2.1. However, during user evaluations, being observed could affect users' responses resulting in data not corresponding with natural behavior and own opinions. This was taken into account during evaluation data analysis.

Chapter 4

Design and Development

This chapter will present the design and development of MagnoMath, the software application developed in this project, following design science research guideline six presented in section 3.1.

4.1 Establishing requirements

According to Weaver, it is crusial to specify what the software to be developed will do to fulfill its intended purpose [57]. There are two types of requirements useful to specify when developing a system or software: *functional* and *non-functional*. This section will present these requirements set for MagnoMath.

4.1.1 Functional requirements

Weaver defines functional requirements as specifications describing essential system needs to be met [57]. Establishing requirements is an iterative process, which was also the case for MagnoMath before the final requirements listed below were specified.

1. The application must be able to assist diagnosis of dyscalculia.

2. The application must be able to identify and measure possible correlations between dyscalculia symptoms and magnocellular reasoning.

3. Results from conducted tests must be stored for later retrieval.

4. User progress must be visualized.

5. There must be a possibility to sort results by performance, time spent, evaluation date and performance classification.

6. There must be a possibility to send results by e-mail.

7. There must be a possibility to delete result records.

8. The user must be able to choose whether time spent is displayed during evaluation.

9. The user must be able to choose whether evaluation tasks are the same every time or chosen randomly.

10. The user must be able to reset classifiers to initial states.

11. All screens must contain a help-button providing information and assistance.

4.1.2   Non-functional requirements

Weaver defines non-functional requirements as specifications enhancing user experience (UX) involving efficiency, usability and accessibility of a system [57]. The final non-functional requirements of MagnoMath are listed below.

1. The application must be easy to use, also for users without technical backgrounds.

2. The application must be responsive.

3. The application must be well-functioning.

4.2   Design

This section will present the development of MagnoMath's UI together with how the design complies with Nielsen's usability heuristics presented in section 3.6.3.

4.2.1   Complying with requirements

Functional and non-functional requirements given in sections 4.1.1 and 4.1.2 respectfully greatly impacted the development of MagnoMath's UI. More specifically, the graphical user interface (GUI), the interface for computer software taking advantage of computers' graphical capabilities to make the software easier to use [4]. Its purpose was to make it easy for non-technical users to access software functionalities through a visual interface rather than a command prompt.

4.2.2   User interface development

The first GUI design draft was finalized August 30th 2015 using Indigo Studio [27], a wireframe tool for creating interactive conceptual software prototypes. It was selected due to its simplicity, accessibility and ease of use. Resulting was the initial high-level design of the software application to be implemented. Additionally, Indigo Studio enabled interaction beyond a set of static screen sketches as it allowed for button click listeners and dynamic input text boxes, among others.

The draft is shown in figure 4.1 focusing on visualizing initial ideas regarding MagnoMath's main structure, and how the different parts of the application are connected with each other. An initial functional requirement involved that the user was obligated to create an account in order to use the application. The idea was to enable the possibility of several users on the same device having available their personal result records and preferences. A *Logout* button was therefore included in this draft as shown in figure 4.1a. This requirement was later removed to increase application's usability and accessibility.

Once the draft was finalized, programming started. Software and UI development from then on were conducted in parallel. Consequently, during application development, UI was implemented in XML to facilitate a more detailed design specification fully functional with the software source code written.

(a) Home screen   (b) Evaluation screen   (c) Result screen
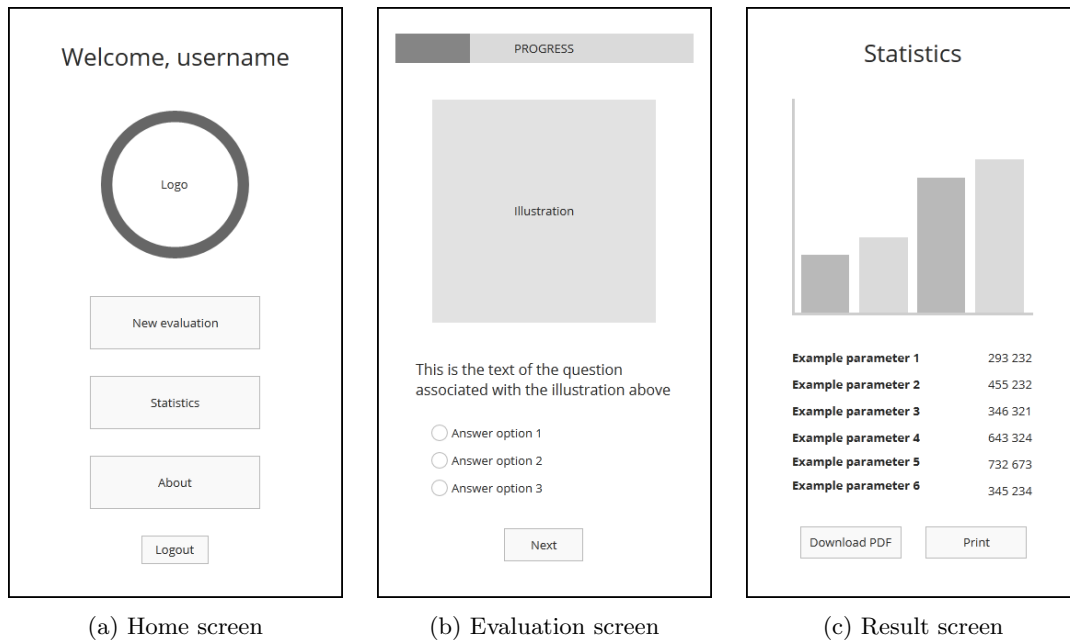
Figure 4.1: MagnoMath GUI draft

4.2.3   Complying with heuristics

This section will present how the design draft shown in section 4.1 complies with Nielsen's usability heuristics presented in section 3.6.3.

Visibility of system status. Status visibility is relevant mainly during evaluation conductionshown in figure 4.1b. This is displayed during the whole evaluation so users are always aware of their progress.

Match between system and the real world. The phrasing of evaluation tasks and other information is adapted to users struggling with mathematics. Consequently, simple wording without technical terms has been applied.

User control and freedom. Users have the ability to exit and return to the home screen at any time during evaluation. However, they are not allowed to return to a previous task to change a response. This behavior was implemented due to the evaluation conduction being developed for users to submit responses quickly and naturally. Additionally, if users know that they have the possibility to re-submit answers, they are likely to act more indifferently as compared to if they know that answers are final.

Consistency and standard. Consistent language is applied for similar evaluation tasks to ensure that users are not confused. Moreover, terms applied across screens use the same phrasing, maintaining consistency during use.

Error prevention & Help user recognize, diagnose, and recover from errors. To ensure that users do not create runtime errors, only keyboards with valid input data are displayed. Additionally, the software does not allow users to continue evaluation before an answer is submitted.

Recognition rather than recall & Flexibility and efficiency of use. Automating data processing is not hugely relevant for MagnoMath. However, two processes available for automation were identified. Instead of inserting an e-mail every time users wish to send a result,

they can submit one which is automatically inserted. The other process concerns sorting of earlier achieved results. By default, results are sorted by date of evaluation. Instead of requiring users to change sorting criteria to the one preferred every time they open the result history screen, this criteria can be automatically selected.

Aesthetic and minimalistic design. To avoid users getting overwhelmed by large amounts of information during use, evaluation tasks are short and concise consisting exclusively of required information for solving the problems. Additionally, the UI is designed so that users are not required to read a lot of documentation in order to use the application and obtain an overview of the functionality it offers.

Help and documentation. As mentioned in the heuristic above, the UI is developed to avoid lots of documentation in order to use the application. In case users need guidance during use, a button is included in each screen for displaying information when clicked.

4.3   Technologies

This section will introduce technologies applied in developing MagnoMath.

4.3.1   Application platform

Firstly, the application platform had to be selected, especially with respect to which OS the software would run on. Several were available, spanning from OSs running on desktop computers to mobile OSs running on portable devices. After a thorough comparison and further consideration, Android was selected for three reasons.

Firstly, it was evaluated as more feasible to have the resources and time to develop a fully functional prototype for a smartphone as compared to an OS running on a desktop computer. Software applications for smartphones are smaller in magnitude so finalizing an application for this platform was more suitable for this project due to its limited timeframe.

Secondly, as discussed in section 1.1, smartphones represent a relatively new technology at the time of writing worth exploring and utilizing.

Lastly, the open source nature of Android made the OS very accessible for development and testing, an advantage valuable in one-person projects.

4.3.2   Programming languages

Software is made possible by one or several programming languages, formal languages designed to communicate instructions to a computer. An application can be developed using a single or a combination of programming languages. This project selected the latter approach, as Android uses the Java programming language for back-end logic and XML for front-end UI structure and design. XML is particularly suited for creating UIs because it gives the developer a great freedom to design and place elements not only according to their position on the screen, but also according to their position in relation to other elements. This is useful when the application developed is to support devices with screens varying in density and size.

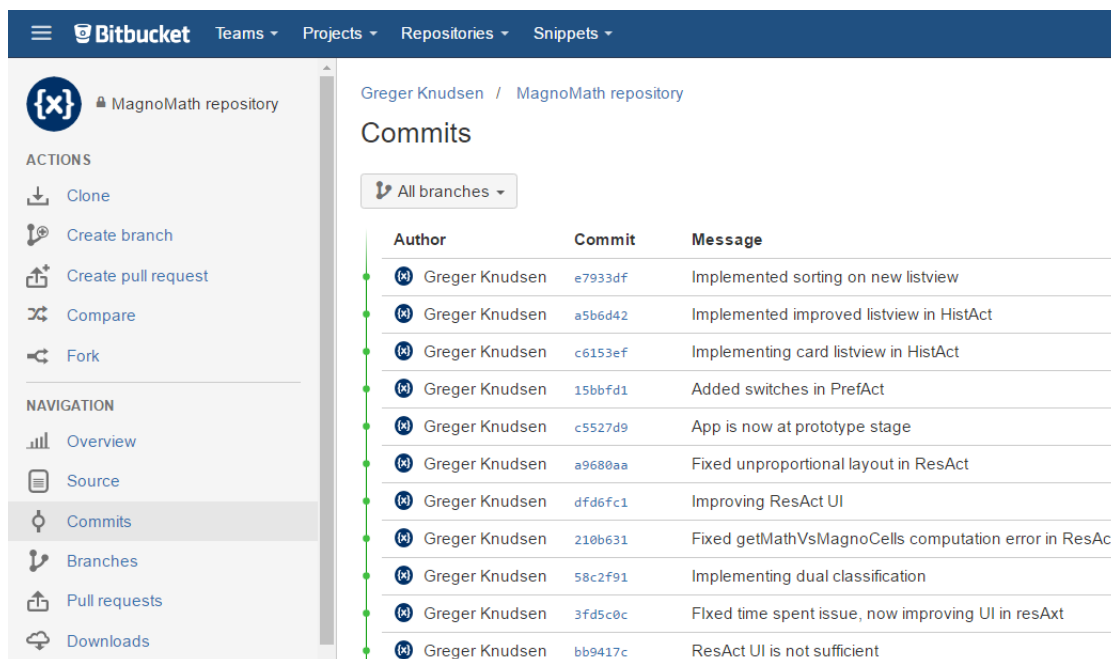### 4.3.3   Integrated development environment

Writing software code using a programming language can be written in any text editor. However, to support development with error-handling, syntax checking, test-running and more, an integrated development environment (IDE) is an invaluable tool. An IDE is a piece of software designed to assist the software development of any application, making the process easier and

more efficient. Although an IDE specifically for Android development is available at the time of writing, called Android Studio [2], the more general Java IDE Eclipse [20] was selected as the IDE for developing MagnoMath. The decision was largely based on the author's experience with Eclipse. Secondly, Android Studio is still in beta stage at the time of writing. In contrast, Eclipse has been used for Android development for years, making its stability a deciding factor. Since Eclipse is a general IDE for Java development, it does not support development for Android when first installed. An Android plugin for Eclipse had to be downloaded and installed, but this was a very manageable process.

4.3.4 Version control

An enormous amount of code is possibly written when developing computer programs. As the amount increases, there is always a possibility of program errors occurring potentially sabotaging development progress. Therefore, tracking changes in the code during development along with a complete version history of every program file spare developers from duplicated work and frustration. This process is called *version control* and was applied in this project. Version control systems are invaluable for not only tracking changes which could be causing program malfunction. They are also effective in large development teams. Git [24] was selected for this project due to its prevalence and author's experience with its use. Together with Git, Bitbucket [6] was used, a web hosting service for software projects. Bitbucket was chosen because it offers a UI providing a full overview of all changes made to the attached program source code. Another advantage of Bitbucket is that it was free of charge for the purposes of this project. Figure 4.2 shows Bitbucket in use during the development of MagnoMath.

Figure 4.2: Bitbucket displaying MagnoMath repository commits (Screenshot)

4.3.5 Project management tool

As presented in section 3.3.2, the software development method Personal Scrum was used. Personal Scum consists of a product backlog divided into several smaller sections, each representing a sprint backlog. Within the timeframe of each sprint backlog, a functionality or a set of functionalities is implemented. To manage and maintain an overview of implemented and not-yet implemented functionalities within each sprint, a project management tool was valuable. The one selected for this project was Trello [53], a simple web based project management tool with a minimalistic UI. The author's own experience with the tool as well as being free of charge were the deciding factors for choosing Trello. It offers the functionality of color-coding each implementation specification by priority, useful when selecting new functionality to implement in a sprint. Figure 4.3 shows one of its states during the development of MagnoMath, where color codes green, yellow and red represent low, medium and high priority, respectfully.
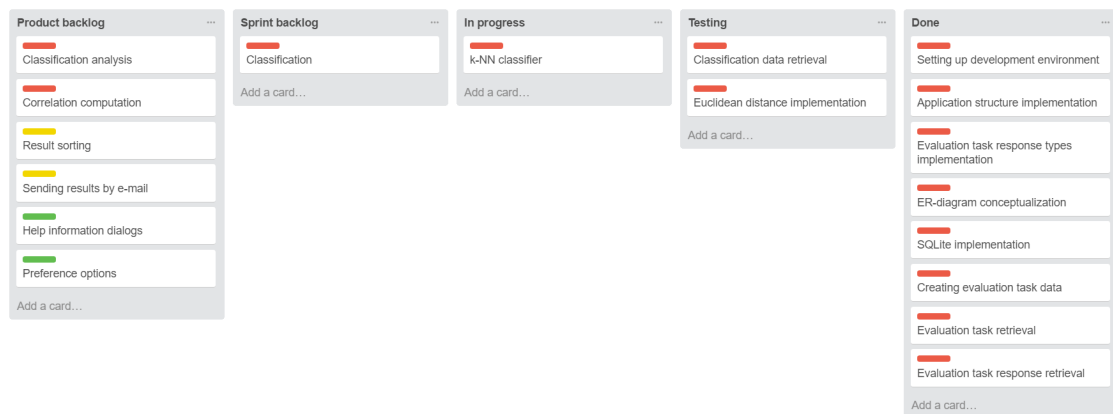
Figure 4.3: Trello board during development iteration five of MagnoMath (Edited screenshot)

First, a task in the *Product backlog* column in figure 4.3 is divided into several subtasks, representing the sprint backlog shown in the *Sprint backlog* column. Each resulting functionality under development was listed in the *In progress* column. Once a functionality in the sprint backlog is implemented, it is moved to the *Testing* column. For each sprint of two weeks, the process repeats itself until all listed functionalities in the product backlog are fully implemented and placed in the *Done* column. In this project, Personal Scrum was useful for planning and structuring the implementation of functional requirements given in section 4.1.1.

4.4 Development

This section will describe the development of MagnoMath applying Personal Scrum methodology presented in section 3.3.2.

4.4.1 Iteration one - Setting up development environment

The first iteration was assigned the timespan of week 36 and 37 of 2015, i.e. August 31st to September 13th. Iteration one of prototype development consisted of setting up the necessary tools to be used throughout the overall development process. Firstly, the Java Runtime Environment (JRE) was installed in order to build and test the application. Then, the newest version of

Eclipse IDE was installed. For developing software for the selected platform, the Android Development Kit (ADK) for Eclipse was installed. Then, the latest version of Git was installed. For managing source code versions tracked by Git throughout development, a designated repository was created on Bitbucket.

4.4.2 Iteration two - Implementing application foundations

The second iteration was assigned the timespan of week 38 and 39 of 2015, i.e. September 14th to September 27th. Iteration two consisted of implementing a high-level foundation of the application and its structure, including designated sections for each main functionality and interactions between these. Sections implemented are shown figure 4.4.

(a) Home screen [52]   (b) Evaluation screen [56]   (c) Result screen
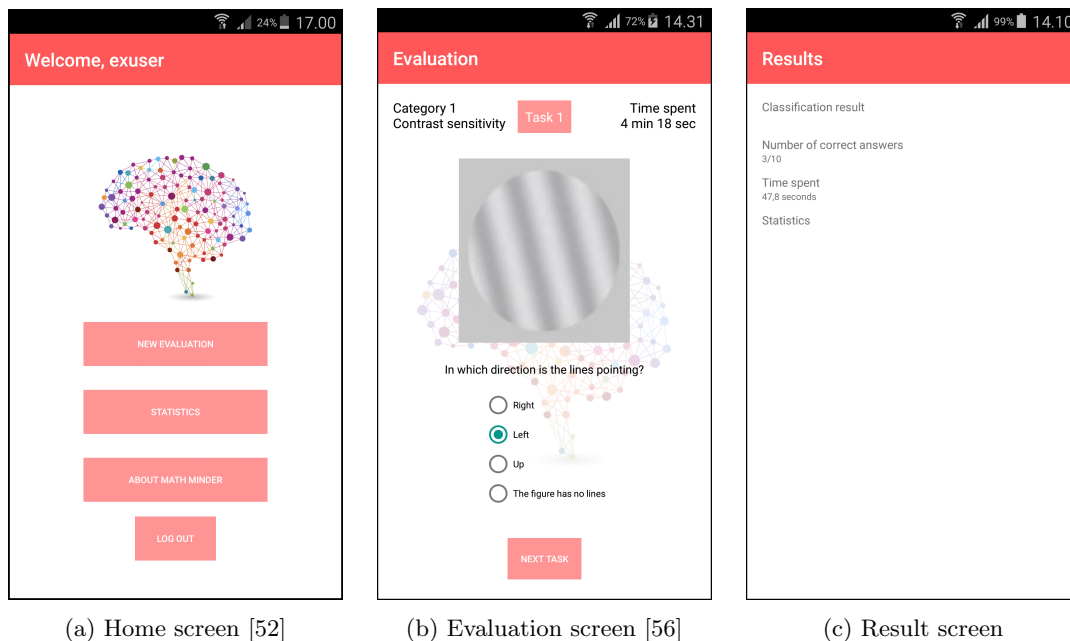
Figure 4.4: MagnoMath GUI after second iteration

4.4.3 Iteration three - Implementing database

The third iteration was assigned the timespan of week 40 and 41 of 2015, i.e. September 28th to October 11th. Iteration three consisted of conceptualizing the data to be stored as described in subsequent sections.

4.4.3.1 Data conceptualization

Before the database could be implemented, its data and structure had to be defined. A useful model for assisting this process is the Entity Relationship (ER) model which visualizes the data and its relationship with other data [21]. Figure 4.5 shows the ER-diagram of MagnoMath's database.
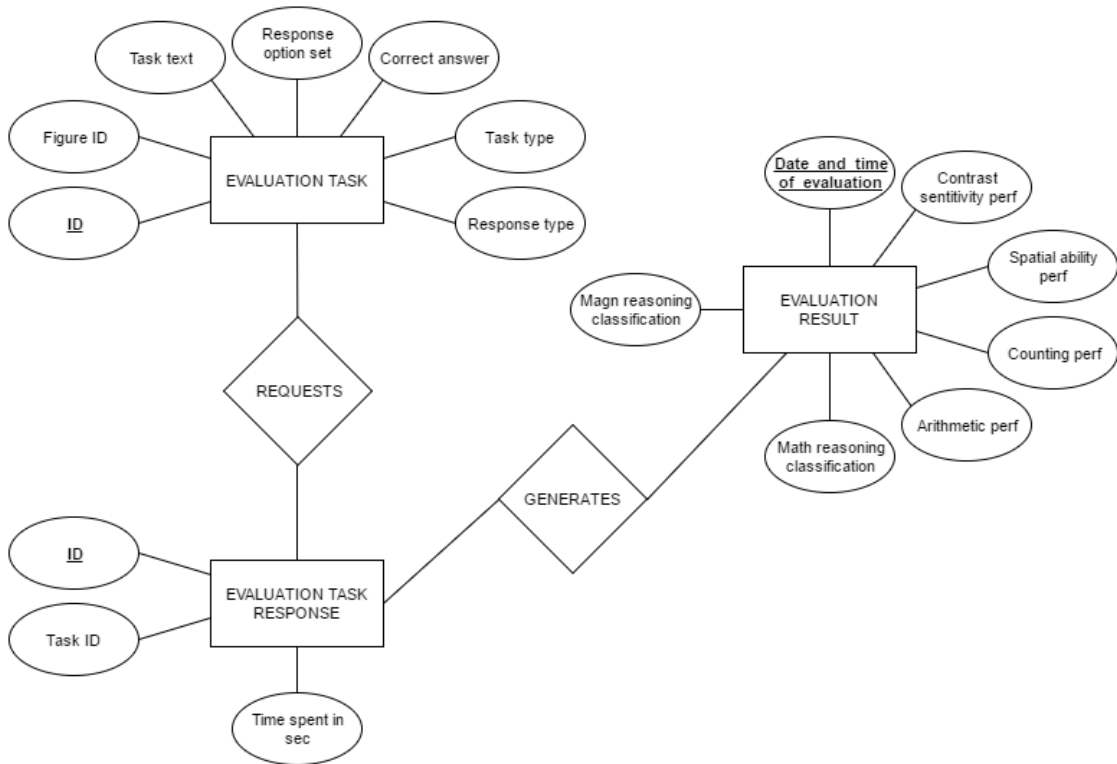
Figure 4.5: MagnoMath database ER diagram

4.4.3.2   Database management system

When data conceptualization was finalized, the next task was to select which system to use for storing, organizing and managing the data. More specifically, a database management system (DBMS), a collection of programs enabling users to create and maintain a database [21]. This includes insertion, manipulation and deletion of data, operations the majority of computer software depends on.

SQLite [51] was selected as the DBMS for the application. SQLite is a library for mobile platforms for creating and managing a database. MagnoMath was from the beginning planned to run without the need of communicating with the World Wide Web (WWW), thus SQLite was particularly suitable due to its localized nature. Additionally, SQLite fully supports Structured Query Language (SQL), hence its name, which provided sufficient control of stored data.

4.4.3.3   Data security

Since the software application was developed for Android, the security and privacy of gathered data was an especially important consideration as the selected OS runs on portable devices. The majority of devices running Android can store data in two possible locations: internal and/or external. The former is located inside the device. The latter is located on a memory card more prone to unautorized access. Therefore, it was crucial to store data internally. However, even if the data is stored inside the device, there is still a possibility that other applications can access it. Fortunately, Android OS by default stores application data on an internal location only

accessible for the specific application. This ensures that no other applications can read the data gathered by MagnoMath. However, it became necessary to create an exception to the isolated nature of MagnoMath's data. In order to comply with the functional requirement stating that users need to be able to send results by e-mail, device's e-mail application had to be explicity granted access. With external access only for sending results, the data is protected while still being available for sending.

4.4.4 Iteration four - Implementing task response acquisition

The fourth iteration was assigned the timespan of week 42 and 43 of 2015, i.e. October 12th to October 25th. Iteration four focused on implementing the acquisition of user response data during evaluation of mathematical and magnocellular skills. More specifically, the data representing discrete performance measurements of selected classification parameters given in tables 3.5 and 3.6. Each task given to the user acquires one of four possible response types listed below.

- $\mathbb{N}_{\geq 0}$

- Radio button

- Image

- Boolean

These response types were implemented to obtain different kinds of responses for different kinds of tasks evaluating mathematical and magnocellular reasoning.

4.4.5 Iteration five - Implementing classification and correlation computation

The fifth iteration was assigned the timespan of week 44 and 45 of 2015, i.e. October 26th to November 8th. Iteration five focused on implementing the core functionality of MagnoMath: the classification algorithm analyzing and classifying evaluation response data. As described in section 3.4.2, k-NNN was selected as classification algorithm. This algorithm was implemented in Java as given in listing 4.1.

Listing 4.1: k-NN algorithm implementation in Java

```java
public ClassificationRes knnClassify (Map<ClassificationParam ,Double>
    newObservation , TrainingSet trainingSet) {
        ClassificationRes classificationRes = null;
        double currMinDistance = Double.MAX_VALUE;
        for(int currObservationIndex = 0; i < trainingSet.getSize();
            currObservationIndex++){
                Observation currOldObservation;
                try{
                        currOldObservation = trainingSet.
                            getObservationByIndex(currObservationIndex); //
                            retrieves observation from local file
                }
                catch(IOException ioe){
                        ioe.printStackTrace();
                }
```

```
12              double distanceToNewObservation = eucDistance(
                    newObservation, currOldObservation);
13              if(distanceToNewObservation < currMinDistance){
14                      classificationRes = currOldObservation.
                            getClassification();
15                      currMinDistance = distanceToNewObservation;
16              }
17          }
18          Utils.addObservationToTrainingSet(newObservation, classificationRes
                , trainingSet);
19          return classificationRes;
20 }
```

In listing 4.1, line 7 retrieves from a local file an observation from the training set. Then, in line 12, the Euclidean distance between this observation and the observation to be classified p is computed. This process is repeated for all observations $q \in T$ until the one with the shortest distance from p is identified. Line 12 calculates the Euclidean distance as shown in algorithm 1, implemented in Java as given in listing 4.2.

Listing 4.2: Euclidean distance computation implementation in Java

```
 1 private double eucDistance(Observation observation1, Observation
       observation2, TrainingSet trainingSet){
 2      double eucDistance = 0.0;
 3      Map<ClassificationParam,Double> observation1Coordinates =
            observation1.getCoordinates();
 4      Map<ClassificationParam,Double> observation2Coordinates =
            observation2.getCoordinates();
 5      List<ClassificationParam> classificationParams = Utils.
            getClassificationParamsByTrainingSetID(trainingSet);
 6      for(ClassificationParam classificationParam : classificationParams)
            {
 7              double currDistance = ((observation1Coordinates.get(
                    classificationParam) - (observation2Coordinates.get(
                    classificationParam))));
 8              eucDistance += currDistance * currDistance;
 9      }
10      return Math.sqrt(eucDistance);
11 }
```

In line 2, the distance is initialized with zero, the lowest possible distance value. Then, in lines 6 through 9, for every coordinate in observation one and equivalent coordinates in observation two, their ratio is squared and added to the total distance value. When all ratios are squared and added, the distance value represents the length of the distance vector in the n-dimensional space, one dimension for each observation parameter retrieved in line 5. The classifier evaluating mathematical skills has observation parameters {*counting*, *subitizing*, *arithmetic*}. The classifier evaluating magnocellular reasoning skills has observation parameters {*contrastSensitivity*, *spatialAbility*}.

When the performance within each task category is classified, we can use this classification to obtain a discrete measurement of the correlation between mathematical and magnocellular

reasoning. This was done to satisfy functional requirement two given in section 4.1.1. To achieve this, algorithm 3 presented in section 3.5 was applied. Its implementation is given in listing 4.3.

Listing 4.3: Mathematical/magnocellular reasoning correlation computation implementation in Java

```java
public static double getMathMagnoCellCorrelation(Map<ClassificationParam,
    Double> results){
        double mathRes = results.get(ClassificationParam.ARITHMETIC) +
            results.get(ClassificationParam.COUNTING) + results.get(
            ClassificationParam.SUBITIZING);
        double magnocellularRes = results.get(ClassificationParam.
            CONTRAST_SENSITIVITY) + results.get(ClassificationParam.
            SPATIAL_ABILITY);
        double mathResReduction = 3.0/2.0; // nrOfMathCategories/
            nrOfMagnoCategories
        mathRes *= mathResReduction;
        double toPercentReduction = .5;
        return 100.0 - ((Math.abs(magnocellularRes - mathRes) *
            toPercentReduction));
}
```

Since the number of parameters evaluating mathematical reasoning differs from the number of parameters evaluating magnocellular reasoning, the performance from the former had to be reduced to the result space of the latter as done in line 5. After their ratio is computed, it is reduced to percent, done in line 7. This percentage is subtracted from the highest correlation possible, which is 100. The implementation then returns the answer representing the correlation measure in percent between the performance within each of the two evaluation task categories.

4.4.5.1 Computational complexity

The computational complexity of an algorithm determines the amount of resources necessary to execute it. Classifying results into corresponding performance classes represents the most resource intensive task performed by MagnoMath. It is therefore useful to evaluate the performance of the classifier as the number of observations $q \in T$ increases. This ensures that the classifier is functional in terms of efficiency as T becomes very large.

To evaluate the computational complexity of the classifier, we have to identify the operations required to classify a new observation p. With equation 3.2 in mind, we have the following operations classifying p using T with n-parametric observations:

1. n subtractions

2. n squares

3. $n - 1$ additions

4. 1 square root

To compute the Euclidean distance between p and another obervation $q \in T$, the number of operations required is $n + n + (n - 1) + 1 = 3n$. Since we want to find q with the shortest distance to p, we have to compute this for all $q \in T$, giving us a total of $3n \times |T|$ operations. Consequently, the running time of the classification algorithm increases as the size of T increases.
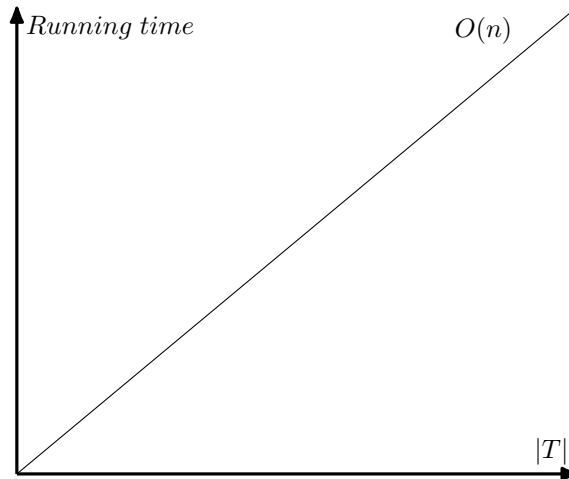
Figure 4.6: Computational complexity of evaluation results classifier

But how much? Since the number of classification parameters n never increases, only the number of observations in the training set $|T|$, the increase in running time is *linear*. The running time is consequently $O(n)$, where $n = |T|$. Figure 4.6 illustrates this behavior. Linear complexity of an algorithm is desirable [31] as the running time increases slowly as T increases in size, thus ensuring a stable classifier performance for the purpose of the software application developed.

4.4.6 Iteration six - Implementing result retrieval and sorting

The sixth iteration was assigned the timespan of week 46 and 47 of 2015, i.e. November 9th to November 22nd. Iteration six focused on implementing the retrieval and presentation of results stored in MagnoMath's database. Additionally, sorting of result records by criteria was implemented to satisfy functional requirement five.

Iteration six also introduced several improvements of the application's GUI eliminating the need for the user to register an account. As discussed in section 4.2.2, this was initially required to use the application. However, it was considered unnecessary to require a login as it would decrease availability and usability. Figure 4.7 shows MagnoMath's GUI after iteration completion.

4.4.7 Iteration seven - Improving graphical user interface

The seventh iteration was assigned the timespan of week 48 and 49 of 2015, i.e. November 23rd to December 6th. Iteration seven focused on improving the GUI so that elements became easier to see and text easier to read. This was achieved by implementing several visual improvements as given in figure 4.8. The most significant changes were implemented in the home screen given in figure 4.8a introducing *Preferences* home menu option.

4.4.8 Iteration eight – Implementing complementary functionality

The eighth iteration was assigned the timespan of week 1 and 2 of 2016, i.e. January 4th to January 17th. Iteration eight focused on implementing remaining functionality specified in functional requirements given in section 4.1.1 including user progress visualization, help dialogs and *Preferences* menu options.
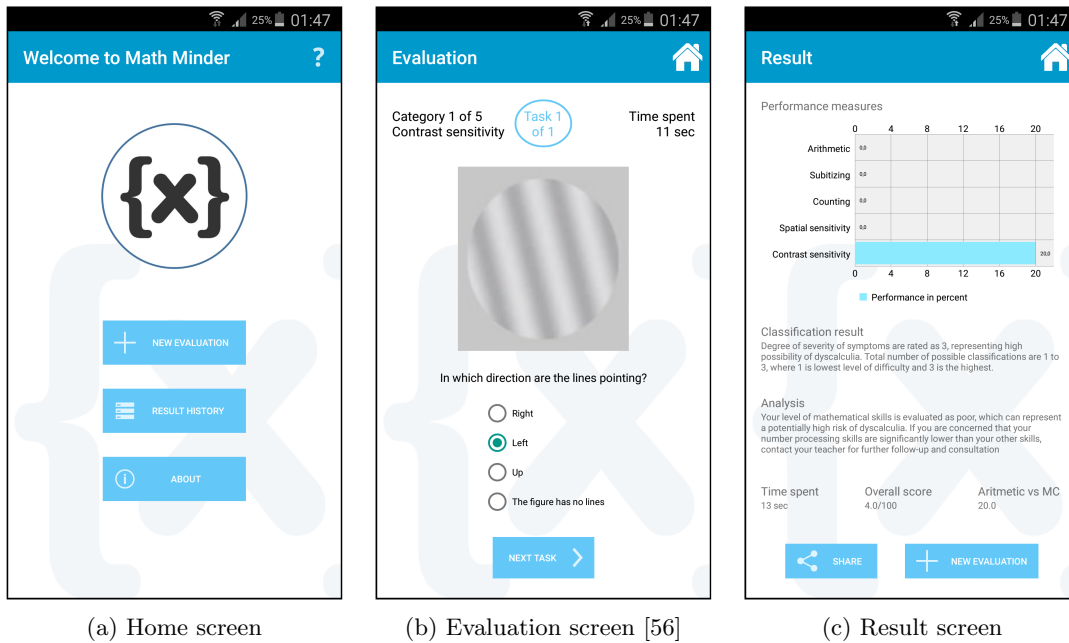
(a) Home screen   (b) Evaluation screen [56]   (c) Result screen

Figure 4.7: MagnoMath GUI after sixth iteration
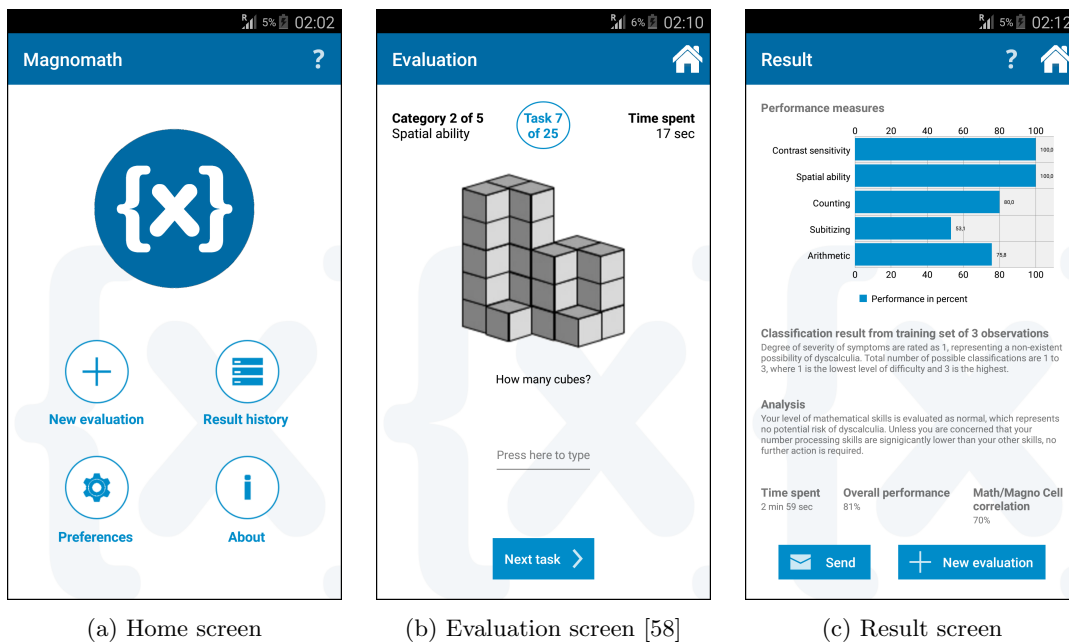
(a) Home screen   (b) Evaluation screen [58]   (c) Result screen

Figure 4.8: MagnoMath GUI after seventh iteration

4.4.9   Iteration nine – Prototype finalization

The ninth iteration was assigned the timespan of week 3 and 4 of 2016, i.e. January 18th to January 31st. Iteration nine focused on implementing remaining pieces of the prototype, as well as performing thorough testing of software behavior making sure it was stable for the upcoming evaluation phase of the project. This iteration expanded evaluation task database as well as improving evaluation result screen structure as shown in figure 4.9c.

(a) Home screen   (b) Evaluation screen   (c) Result screen

Figure 4.9: MagnoMath GUI after ninth iteration

Chapter 5

Evaluation and Results

This chapter will present the expert and user evaluation of MagnoMath and results, following guideline three in design science research framework presented in section 3.1.

5.1   Background

As discussed in section 3.1.3, evaluating the developed artifact is a crucial phase in the research process. Hevner et al. state that the utility, quality and efficacy of the artifact are required to be demonstrated through well-executed evaluation methods [26]. Appropriate metrics has t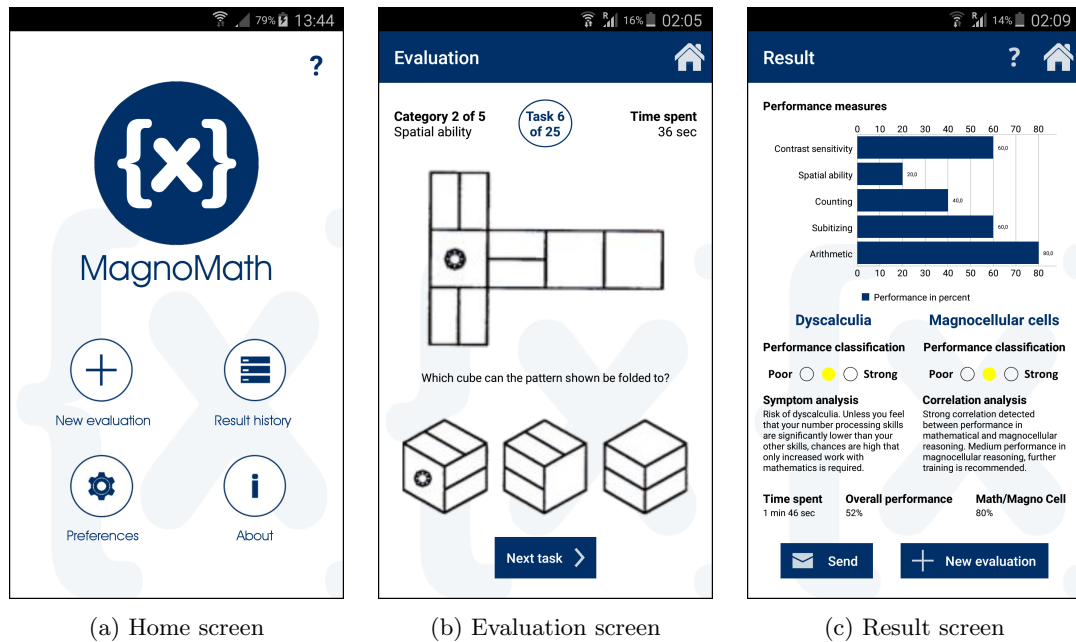o be applied in the retrieval of valuable and relevant data to ensure that the artifact creates value in its business domain. Hevner et al. mention several metrics applicable in evaluation, three of which were used in this research: performance, accuracy and usability. Performance is the most important as it does not matter how well MagnoMath processes data if the user has to wait a significant amount of time before results are presented. Accuracy comes next, crucial for generating correct and reasonable results from evaluation response data. Usability was also included as it is essential for the application to be easy to use in order to retrieve high quality data. Additionally, results had to be presented to users in a way so they could understand them and benefit from them.

5.2   Objectives

The evaluation phase of this project had four main objectives. The first was to study Magno-Math's ability to identify and measure correlations between dyscalculia symptoms and magno-cellular reasoning. The second objective was to study MagnoMath's ability to assist diagnosis of dyscalculia. The third was to get feedback on MagnoMath's functionality and behavior. The last objective was to evaluate application's usability. Together, these objectives cover the research questions of this project given in section 1.1, confirming that the evaluation phase represented a crucial part of this research.

5.3   Eligibility

Participants eligible for enrollment in the evaluation of MagnoMath were divided into three groups. The first consisted of experts in underlying fields with a specific focus on the fields of dyscalcula and LDs. The second consisted of individuals having symptoms of dyscalculia.

Individuals without strong mathematical backgrounds nor having symptoms of dyscalculia were also included to evaluate MagnoMath's ability to differentiate dyscalculic and non-dyscalculic users.

5.4   Expert evaluation

The first group participating in the evaluation of the software prototype was the expert group. This section will present expert evaluation design and results.

5.4.1   Objectives

The expert group evaluation focused on evaluating a set of qualities of MagnoMath given in table 5.1.

Table 5.1: Expert evaluation objectives

Quality	Objective
Responsiveness	Test whether MagnoMath responds quickly to user actions
Stability	Test MagnoMath's error prevention and handling
Functionality	Test whether MagnoMath's functionality works as intended with appropriate methods
Ease of use	Test whether MagnoMath's UI is logical, accessible and easily understandable

5.4.2   Eligibility

The evaluation included only eligible participants meeting all inclusion criteria listed below.

- Undergraduate, graduate or doctoral student in the fields of information science, computer science and/or mathematics *or* researcher/practitioner in the fields of dyscalculia, LDs and/or special needs education.

- Able to provide consent of participation.

The evaluation excluded participants if one or more of the exclusion criteria listed below applied.

- Having symptoms of dyscalculia, dyslexia or other LD.

- Unable to perform given tasks due to visual problems.

- Unable to provide consent of participation.

5.4.3 Participants

5.4.3.1 Software and mathematics experts

Undergraduate, graduate and doctoral students at Department of Biomedical Engineering, Linköping University, Sweden was the selected group for evaluating MagnoMath's behavior, logic and design. This group was selected for two reasons.

Firstly, studies at the department heavily involves computer science, programming and mathematics. Therefore, the students were especially suitable for evaluating the behavior of the application all the way down to algorithms performing calculations and data processing.

Secondly, all participants had strong mathematical foundations from being students across the study programmes offered at the department including applied physics and electronics, computer science, engineering biology and information technology. Therefore, they were suited for evaluating the mathematics behind evaluation results computation.

Another purpose of including the group in the evaluation was to reveal whether evaluation tasks were too difficult. If this was the case, it would cause MagnoMath having too high expectations, representing a design bias.

5.4.3.2 LD experts

The selected group for evaluating the LD aspects of MagnoMath consisted of both researchers and practitioners within the fields of dyscalculia and LDs, presented below.

Rickard Östergren is Junior University Lector at Department of Behavior Studies and Learning at the Institute of Education, Linköping University, Sweden. He has years of research experience in LDs and dyscalculia, as well as experience from the industry as a clinical psychologist. He is a member of the Cognitive Educational Psychology research group and a research group in mathematical didactics, both at Linköping University. He graduated with a Ph.D. in psychology in 2013 with the thesis "Mathematical Learning Disability - Cognitive Conditions, Development and Predictions", making him an uppermost relevant expert for evaluating MagnoMath.

Ulf Träff is Professor at the Department of Behavioural Sciences and Learning, Section of Psychology, Linköping University, Sweden. His chief interest is within the fields of developmental mathematics and mathematical disabilities in children. His work includes treating dyscalculia when diagnosed by applying novel interventions. His research aims at mapping functional disabilities shown in children with mathematical difficulties, as well as identifying underlying causes of developing such disabilities, making him a very relevant expert for evaluating the application.

Kenny Skagerlund is a post-doctoral researcher at Department of Behavior Studies and Learning at the Institute of Education, Linköping University, Sweden. He received his Ph.D. in psychology in 2016 with the thesis "Magnitude Processing in Developmental Dyscalculia - A Heterogeneous Learning Disability with Different Cognitive Profiles". His thesis contributes to our understanding of developmental dyscalculia from the perspective of cognitive psychology and cognitive neuroscience. His study and work made him a relevant expert user.

Caroline Solem is Secretary General at Dyslexia Norway, Oslo. Even if the name of the organization suggests that it exclusively focuses on dyslexia, the organization also has a large focus on dyscalculia as well as specific language impairment (SLI). Dyslexia Norway's work includes online courses for teachers and parents, publications of material for supporting dyscalculic individuals as well as cooperation with schools affected by dyscalculia to support their educational work. Solem is central in all these areas and has years of experience in consulting schools and individuals struggling with mathematics. This made her a great expert user for evaluating MagnoMath.

5.4.4   Design

5.4.4.1   Observation

During the whole course of evaluation sessions, observational data was written down. This data represents the qualitative data of the evaluation session in the form of comments and other feedback given by participants.

5.4.4.2   Usability

Part one: List of tasks to complete. A set of tasks was given to evaluators with the purpose of providing them an insight into functionalities provided by the application to support usability evaluation. Tasks given are listed below.

1. Complete a new evaluation. Select *don't show again* in the dialog.

2. Send results in Excel format to gkn036@student.uib.no

3. Sort all results in the result history by time spent on evaluation.

4. Send all results in the result history to gkn036@student.uib.no in Word format.

5. Send a result record in the result history to gkn036@student.uib.no in plain text.

6. Delete a result record in the result history.

7. Delete all result records in the result history.

8. Set the default result receiver e-mail to gkn036@student.uib.no

9. Reset preferences to default.

10. Disable time display in evaluation.

11. Reactivate the dialog originally showing when a new evaluation starts.

Part two: Nielsen's usability heuristics. For evaluating the usability of MagnoMath, evaluators were asked to rate each usability heuristic presented in section 3.6.3 on a scale of one to ten, where one is *Very poor* and ten is *Very good*.

Part three: System Usability Scale. To obtain a discrete measurement of MagnoMath's usability as a whole, Booke's SUS was applied as described in section 3.6.4. However, statement one saying *I think I would like to use this system frequently* was omitted in this evaluation since experts in mathematics, software and LDs do not need dyscalculia diagnosis assistance. Thus, they were not potential users of MagnoMath. The result space of SUS was reduced to accommodate this modification.

5.4.4.3   Functionality

The session evaluating the functionality of the application was, unlike Nielsen's usability heuristics and Brooke's SUS, focused on behaviors specific to MagnoMath and not general behaviors applicable to all software applications. It applies a selection of relevant parameters not covered by Nielsen's usability heuristics and Brooke's SUS. Each participant was asked to rate a set of statements regarding application functionality on a scale of one to ten, where one is *Strongly disagree* and ten is *Strongly agree*. The statements are given below.

1. The text in the evaluation is easy to understand.

2. The application is responsive.

3. Evaluation tasks are too difficult.

4. Evaluation results correspond to my own perception of skills.

5. The application has useful features.

6. It is easy to understand the purposes of the application.

7. Pre-evaluation introduction was sufficient.

8. Methods applied are suitable for assisting dyscalculia diagnosis.

9. Methods applied are suitable for measuring correlations between dyscalculia symptoms and magnocellular reasoning.

Since the first group of experts consisting of students within the fields of software and mathematics did not have insight into dyscalculia nor magnocellular cells, statements 8 and 9 were only given to LD experts.

5.4.5   Conduction

5.4.5.1   Software and mathematics experts

The evaluation by software and mathematics experts was conducted at Department of Biomedical Engineering, Linköping University, Sweden. Ten participated in the evaluation session, comprised of undergraduate, graduate and doctoral students at the department. Further details regarding the evaluation session are given in table 5.2.

Table 5.2: Software and mathematics expert group evaluation details

Date	February 29$^{\text{th}}$ 2016, 1 PM
Duration	1 hour 30 minutes
Study type	Experimental and observational
Data type	Quantitative and qualitative
Location	Department of Biomedical Engineering, 13$^{\text{th}}$ floor, South Entrance, Linköping University, 581 83 Linköping, Sweden
Population	Undergraduate, graduate and doctoral students at the Department of Biomedical Engineering, Linköping University meeting eligibility criteria
Enrollment	10
Gender distribution	Male: 7 Female: 3

The evaluation session was initiated by a short introduction to the project, MagnoMath and evaluation conduction. Then, a Quick Response (QR) code was given. A QR code is an optical label containing a piece of information readable for smartphones. The code was used in the evaluation session to quickly grant participants access to MagnoMath's installation file. More specifically, the QR code referenced to a Uniform Resource Locator (URL) at the World Wide

Web (WWW) where the installation file was available for download. This enabled participants owning an Android smartphone to install the software application on their personal devices for evaluation. Four participants had an Android smartphone, making a total of five devices used to evaluate the application including the device belonging to the evaluation session instructor, the author of this thesis. Due to fewer devices than participants, they were required to wait for their turn to individually evaluate MagnoMath. This did not represent an issue during evaluation.

After participants owning an Android smartphone had downloaded and installed the application, they were handed out the tasks presented in section 5.4.4.2. Then, they were handed out Nielsen's usability heuristics presented in section 3.6.3. Lastly, they were handed out a set of statements regarding the functionality of the application given in section 5.4.4.3. The evaluator instructor was available for questions during the whole session.

5.4.5.2 LD experts

As described in section 5.4.3.2, participants in the LD expert group were spread across several locations. Therefore, it was natural to conduct each evaluation individually at their respective locations. Details are given in table 5.3.

Table 5.3: LD expert group evaluation details

	Rickard Östergren	Ulf Träff	Kenny Sk- agerlund	Caroline Solem
Date	February 29th 2016, 9 AM	March 22nd 2016, 2 PM		March 9th 2016, 2 PM
Duration	27 min	1 h 5 min		1 h 17 min
Study type	Experimental and observational			
Data type	Quantitative and qualitative			
Location	Department of Behavioral Sciences and Learning, 1st floor, Linköping University, 581 83 Linköping, Sweden			Dyslexia Norway, Storgata 10a, 5th floor, 0155 Oslo, Norway

The session was conducted in the same manner as in the software- and mathematics evaluation session described in section 5.4.5.1. The only difference was that since each LD expert participated in the evaluation individually, they were not affected by other's opinions.

5.4.6 Results

5.4.6.1 Usability

Part one: Tasks to conduct. In conducting task one in section 5.4.4.2 asking the user to complete a new evaluation, time was an important metric in evaluating the difficulty of tasks as discussed in section 2.2. The longer time a participant required to submit an answer, the more difficult he/she experienced the task. Since participants were experts in mathematics and LDs, none should have any difficulties answering correctly to given tasks. However, there were tasks evaluating spatial ability that some evaluators found challenging, examples of which are given in figure 5.1.

(a) Subitizing (1) [45]   (b) Subitizing (2) [1]   (c) Counting
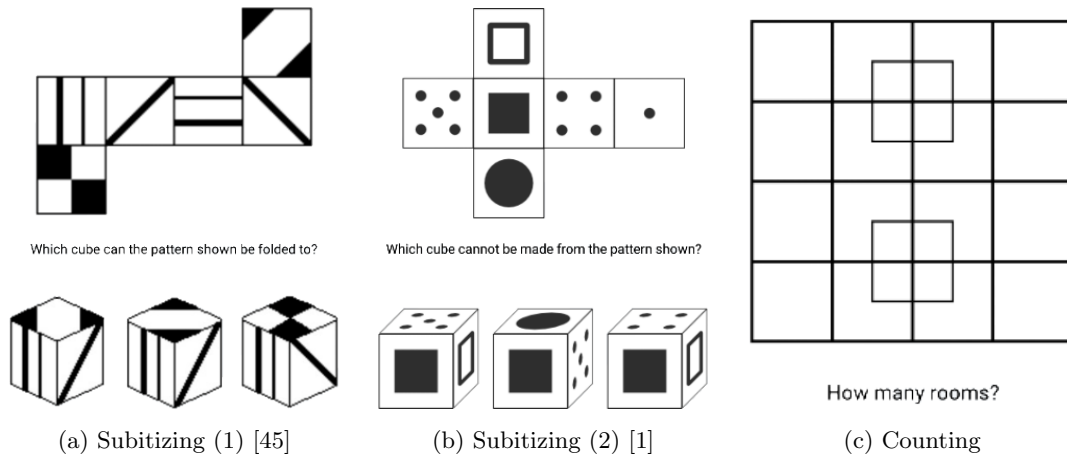
Figure 5.1: Examples of tasks some expert evaluators found difficult (Edited screenshots)

Tasks that are too difficult could seriously threaten the validity of results. This because such tasks are prone to cause false positives. That is, resulting in MagnoMath stating that the user is affected by dyscalculia when this is not the case. Fortunately, this error was easy to correct by replacing featured tasks with ones of decreased difficulty.

Another observational finding concerned MagnoMath's UI. Task three in section 5.4.4.2 for example, asking the user to sort all results by time spent on evaluation, was not trivial for everyone. It was observed that their focus on managing result records was on the button section shown in figure 5.2 instead on the task bar section at the upper part of the screen.
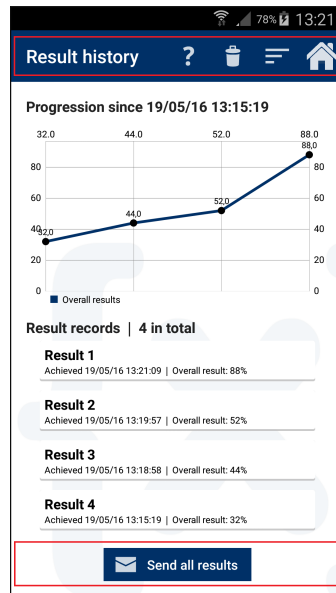

Figure 5.2: Upper and lower sections of result history screen (Edited screenshot)

Part two: Nielsen's usability heuristics. Rankings of Nielsen's usability heuristics were
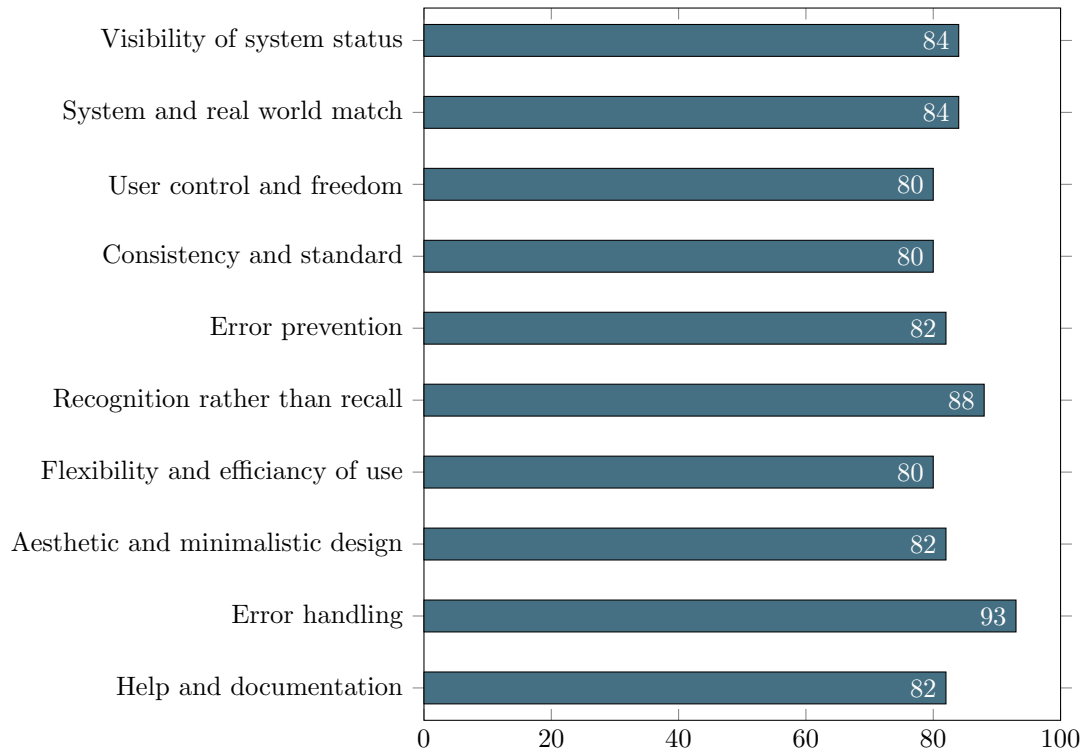
Figure 5.3: Software and mathematics expert users' usability heuristics rankings mean values in percent

averaged and compared as shown in figures 5.3 and 5.4. The heuristic obtaining the highest score among software and mathematics experts was the one regarding MagnoMath's ability to prevent errors. This result corresponds to observational data, as no operational errors or halts occurred during evaluation, proving that the application is as stable as desired during use.

The heuristic obtaining the highest score among LD experts was the one regarding application's ability to remember recurring user choices. Although this was not highly relevant for MagnoMath, it was encouraging to see that evaluators found it useful that result receiver e-mail and result records sorting criteria could be automatically specified.

In the *Preferences* screen, the application did not react to a submitted invalid e-mail address for sending results to. Consequently, an error-check was implemented to eliminate the possibility of sending results to invalid e-mails.

Another heuristic suggested decreased user control and freedom, obtaining a lower score than others in both groups. Some evaluators found it negative that answers cannot be resubmitted. However, this was implemented so that users were required to submit natural and immediate responses without having in mind the possibility of resubmitting answers. This could possibility create a result bias. Therefore, no underlying software behavior regarding evaluation execution was modified.

In the *Preferences* screen, the heuristic focusing on maintaining aesthetic and minimalistic design identified a significant confusion caused by the button for resetting preferences to default given in figure 5.5. Comments included that the button looked too much like a *Go back* button, which is an action really delegated to the *Home* button placed to the right of it. This confusion
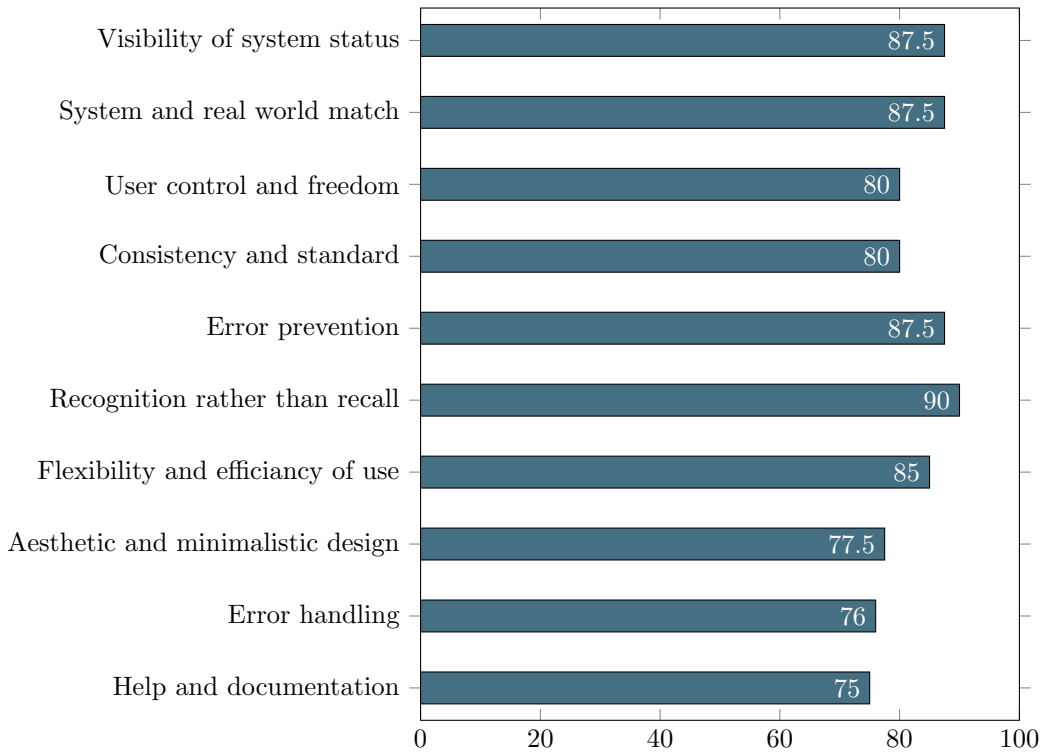
Figure 5.4: LD expert users' usability heuristics rankings mean values in percent
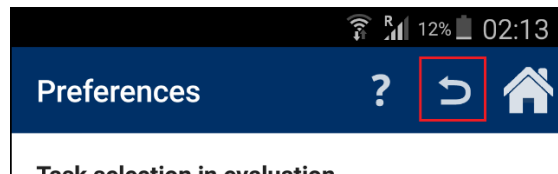

Figure 5.5: Preferences screen action bar with *Reset preferences* button marked in red (Edited screenshot)

was eliminated by replacing the icon with one more suitable and informative for its purpose.

The same heuristic identified another design issue. Some evaluators meant that it was too much text in some evaluation tasks and dialogs. As users should only be presented relevant information without being confused or overwhelmed by large amounts of text appearing on the screen, this issue was resolved by shortening the text in featured tasks and dialogs.

One LD expert commented upon the design of the keyboard for submitting digits in evaluation tasks. The expert meant that the keyboard violated the heuristic stating that the user should not be presented information not relevant to the current situation. As shown in figure 5.6, the keyboard contained buttons for negating the submitted answer as well as retrieving content from the clipboard. Both represent functions not relevant nor necessary for users and was therefore removed.

Part three: System Usability Scale. The SUS score of each participant is given in figures 5.7 and 5.8. A SUS score of 70 is classified as passable. A score of 90 is classified as very good,
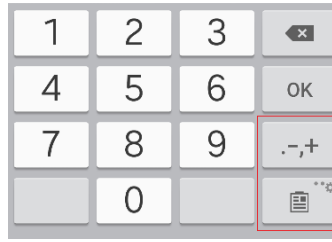
Figure 5.6: Evaluation task digit submission keyboard with unnessecary buttons marked in red (Edited screenshot)
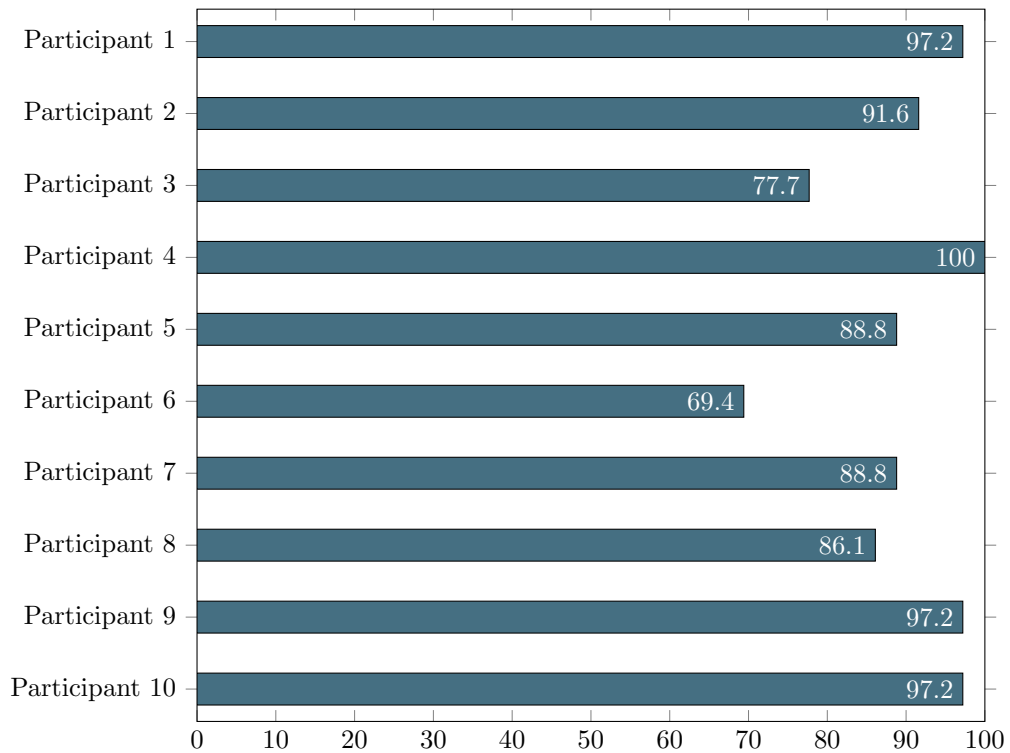

Figure 5.7: Software and mathematics expert users' SUS scores

while a score below 70 should be considered for improvements or changes [3]. The lowest SUS score obtained was 69.4. Comments from the participant generating this score included that some icons were misleading, which corresponded to finds done by Nielsen's usability heuristics.

The general impression among participants was that the application was easy to use. Averaging all scores in figures 5.7 and 5.8 gives a mean score of 88.6, representing a satisfying result.

5.4.6.2 Functionality

Functionality rankings were averaged and compared as shown in figures 5.9 and 5.10. The most interesting is maybe the one regarding difficulty of evaluation tasks obtaining the lowest score
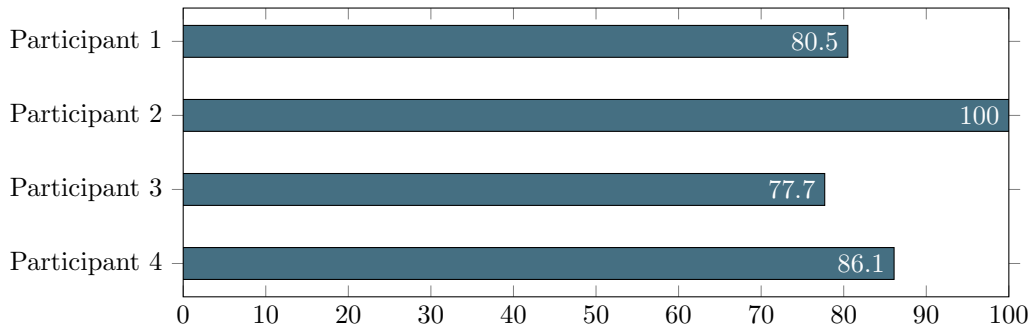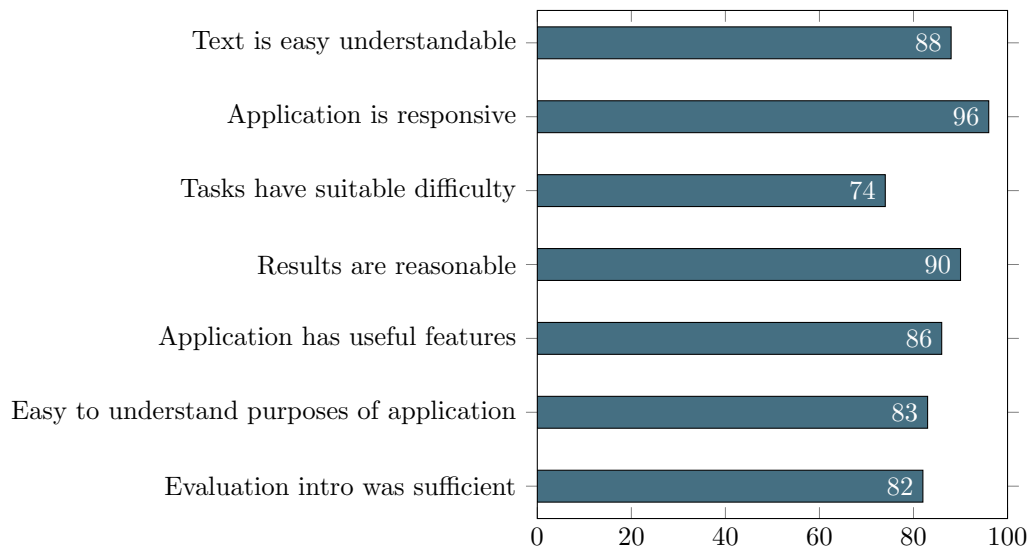
Figure 5.8: LD expert users' SUS scores

Figure 5.9: Software and mathematics expert users' functionality rankings mean values in percent

of all statements in both groups. There were some disagreements among evaluators whether or not evaluation tasks were too difficult. As explained in section 5.4.6.1, there were some tasks evaluators found challenging. However, the majority of participants did eventually submit correct answers. Interestingly, two LD experts had contradicting opinions regarding difficulty of tasks. One of them commented upon the tasks as being too hard, creating a risk for false positives. The individual explained that these featured tasks had to be put into context of other tasks and should not be applied exclusively to assist diagnosis of dyscalculia. On the other hand, the other LD expert commented upon the tasks as challenging, but not too difficult. The person found it positive that tasks given challenge the user. To create a balance between the opposing opinions, the difficulty of featured tasks was slightly decreased since the majority of software and mathematics experts required some time to submit answers on specific tasks. It is however important to emphasize that some tasks were meant to require more time to conduct than others even if the users had strong mathematical backgrounds.

As figure 5.10 shows, LD experts were sceptical to whether MagnoMath was fully suitable for assisting dyscalculia diagnosis. Comments included that tasks were too hard which could result in
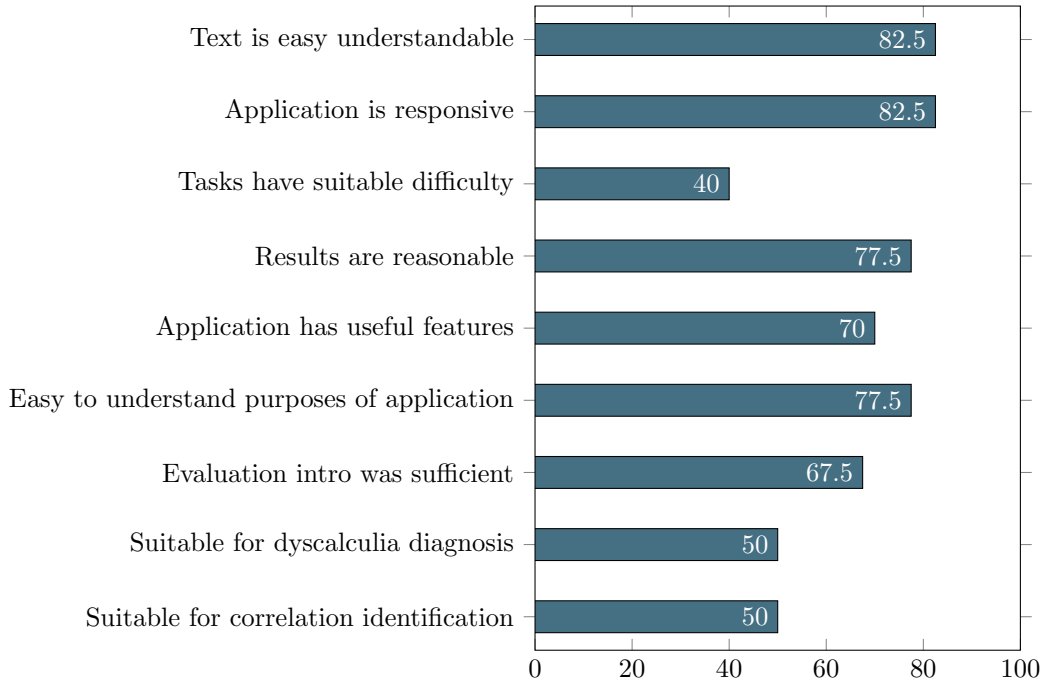
Figure 5.10: LD expert users' functionality rankings mean values in percent

false positives as earlier discussed. MagnoMath's suitability for identifying correlations between mathematical and magnocellular reasoning was also received with some skepticism. Comments from some LD experts included that they were unsure of the success applied methods would have in achieving this due to their limited knowledge regarding magnocellular cells' involvement in developing mathematical deficiencies. Therefore, the majority of participants assigned a neutral ranking on the featured statement.

A significant issue was identified during expert evaluation. Although the application tracked time spent to finish each evaluation task category, this information was not applied in the generation of performance measures to be classified by k-NN. Thus, since performance was not affected by time spent, MagnoMath did not have the ability of differentiating an individual requiring three minutes to finish an evaluation and one requiring twenty minutes. This behavior could result in false negatives, representing a serious result bias. Therefore, generation of results had to be altered to include time spent in each category. Several methods were tested before selecting the solution given in algorithm 5. The algorithm presents a simple approach for including time spent in computing evaluation task category performance. Since time is considered as an extremely important parameter in evaluating dyscalculic behavior as discussed in section 2.2, evaluation category performance was greatly decreased if the user spends more time than a specified threshold. Threshold times were determined by the average time spent by a selection of users without symptoms of dyscalculia nor having strong mathematical backgrounds. Algorithm 5 simply takes the number of correct answers, reduces it to percent and subtract its half if time spent exceeds the default threshold time of the category. This way, the application has an increased potential to differentiate dyscalculic and non-dyscalculic users.

Algorithm 5 Evaluation task category performance

Input: α, number of correct answers, β, time spent, γ, threshold time
Output: δ, performance measure

1: $\delta \leftarrow \alpha \times 20$ {reducing α to percent}
2: **if** $\beta > \gamma$ **then**
3:    $\delta \leftarrow \delta \times 0.5$
4: **end if**
5: **return** δ

5.5 Implementing expert evaluation feedback

The tenth and final software development iteration was assigned the timespan of week 14 of 2016, i.e. April 4th to April 10th. Main improvements, corrections and omissions detected in expert evaluation implemented are listed below.

- Computation of evaluation category performance measures was altered to take time spent into account.

- E-mail validity check in preferences was implemented.

- The icon of the button resetting preferences to default was improved.

- Text in help dialogs were shortened and simplified.

- Some evaluation tasks were rephrased.

- The difficulty of some evaluation tasks was decreased.

- The keyboard for submitting digits in evaluation tasks was replaced by one without unnecessary buttons.

Screenshots of the final prototype are given in appendix D.

5.6 User evaluation

The purpose of conducting expert evaluation was to prepare the application for user evaluation, a crucial project phase. User evaluation focused on MagnoMath's ability to accomplish its intended purposes of assisting dyscalculia diagnosis and identifying dyscalculia symptoms related to magnocellular reasoning. The application's usability was also evaluated. This section will present user evaluation design and results.

5.6.1 Objectives

User evaluation focused on testing a set of qualities regarding MagnoMath given in table 5.4.

Table 5.4: User evaluation objectives

Quality	Objective
Diagnosis assistance	Test Magnolia's ability to assist diagnosis of dyscalculia
Correlation identification	Test MagnoMath's ability to identify dyscalculia symptoms possibly related to magnocellular reasoning
Classification accuracy	Test whether MagnoMath generates too many false positives and false negatives
Usability	Test MagnoMath's ease of use

5.6.2   Eligibility

The evaluation included only participants meeting all inclusion criteria listed below.

- Has symptoms of dyscalculia *or* does not have symptoms of dyscalculia.

- Able to provide consent of participation.

The evaluation excluded participants if one or more of the exclusion criteria listed below applied.

- Having a history of seizures or brain tumor.

- Having history of psychological illnesses.

- Having history of brain surgery.

- Unable to perform given tasks due to visual problems.

- Unable to provide consent of participation.

5.6.3   Participants

5.6.3.1   Individuals having symptoms of dyscalculia

The selected group for evaluating MagnoMath's ability to assist diagnosis and identifying underlying causes of dyscalculia consisted of individuals having symptoms of the condition. Eligible participants were recruited with the help of social media. At the time of writing, the social network Facebook has a Norwegian group for users diagnosed with dyscalculia or having struggles with mathematics. After the author of this thesis published a post on the group informing about the project seeking volunteers for evaluation, the response rate was high. Several volunteered to participate in the study shortly after the post was published. Thanks goes to Anita-Lopez Pedersen at Department of Special Needs Education at University of Oslo, Norway for informing the author about the existence of this group.

5.6.3.2   Individuals not having symptoms of dyscalculia

Students at Faculty of Humanities at University of Oslo, Norway were included to test MagnoMath's ability to separate dyscalculic from non-dyscalculic users. The reason for selecting this group was due to the theoretical-oriented studies at the faculty lacking mathematical and technical topics. Not being significantly strong in mathematics nor having symptoms of dyscalculia made this group suitable.

5.6.4 Design

5.6.4.1 Interview

First, it was valuable to obtain information regarding each participant's background. Questions asked are listed below.

1. Age?

2. Are you diagnosed with dyscalculia?

    (a) If so, how long have you been diagnosed?

    (b) If not, how long have you been struggling with mathematics?

3. Which everyday tasks involving mathematics do you struggle with?

Questions ensuring participation eligibility were also included.

5.6.4.2 Observation

Observation is crucial in studying how dyscalculic individuals reason when presented mathematical problems. Therefore, observations were applied in this evaluation phase. While participants conducted evaluations, their screen was observed and comments reflecting their reasoning were written down. This method of thinking aloud was very useful for obtaining information about how each participant reflected upon and solved a problem presented in front of them. Personal reflections about evaluation results were also written down.

5.6.4.3 Functionality

The most important part of the user evaluation was to test whether MagnoMath's functionality behaved as intended. After participants had completed evaluations and given results, they were asked to rank a set of statements regarding the functionality of the application on the same scale as SUS. They were slightly altered from the set of statements evaluating functionality in the expert group evaluation given in section 5.4.4.3 to be more suitable and relevant for dyscalculic users. The statements are given below.

1. The software is useful for assisting diagnosis of dyscalculia.

2. The software has the potential of identifying underlying causes of developing dyscalculia.

3. Results given on the test are reasonable.

4. Results given on the test present useful information.

5. The software has useful features.

6. It is easy to understand the purposes of the software.

5.6.4.4 Usability

SUS was applied for evaluating MagnoMath's usability as described in section 3.6.4. SUS was especially suitable and valuable to use in this evaluation since it involved potential users.

5.6.5   Conduction

5.6.5.1   Individuals having symptoms of dyscalculia

Since volunteers were recruited through social media, it was natural that they were situated in several different geographical locations, making it time-consuming to arrange one-on-one meetings. Therefore, evaluations involving dyscalculic users were conducted with each participant individually over the internet.

Step one. Date and time for evaluation was agreed upon using e-mail. Prior to the meeting, the screen of an Android smartphone running MagnoMath was mirrored on a computer monitor.

Step two. The author of thesis initiated the meeting by calling the participant over the phone. After conducting a short interview with the participant as described in section 5.6.4.1, the user was sent a link by e-mail. This link mirrored author's screen to participant's screen using the online meeting and screen sharing service *join.me* [28]. When the participant clicked the link and viewed author's screen, mouse control was granted enabling the user to navigate the application freely.

Step three. The participant completed an evaluation before results were explained and elaborated by the observer, the author of this thesis.

Step four. The observer demonstrated additional functionality offered by the application, including sorting and sending of results as well as preference options. This was done with the purpose of facilitating a foundation for acquiring feedback regarding MagnoMath's UI.

Step five. The participant ranked functionality and SUS statements given one-by-one over the phone on the scale *Strongly disagree, Disagree, Neutral, Agree, Strongly agree.*

This evaluation method was applied with the participant in mind. *join.me* does not require any registration nor installation, making evaluation conduction accessible for participants. By having MagnoMath running on author's screen, participants were not required to install the application on their own computers. Additionally, the observer could inspect how the participant navigated and used the application. This represented valuable data together with comments from the participant during use. Evaluation details are given in table 5.5.

Table 5.5: Dyscalculic user group evaluation details

	User 1	**User 2**	**User 3**	**User 4**	**User 5**	**User 6**	**User 7**
Gender	Female						
Age	37	46	14	39	14	62	8
Diagnosed	No						
Date	April 22nd 2016, 1 PM	April 27th 2016, 6 PM		April 28th 2016, 4 PM	April 28th 2016, 6 PM		May 5th 2016, 5 PM
Duration	1 h 24 min	1 h 22 min		1 h 27 min			48 min
Study type	Experimental and observational						
Data type	Quantitative and qualitative						

5.6.5.2   Individuals not having symptoms of dyscalculia

The evaluation consisting of individuals without symptoms of dyscalculia was conducted at Blindern, Oslo, the main campus of University of Oslo holding the Faculty of Humanities. Volunteers eligible for participation were recruited. The sole purpose with this evaluation was to test MagnoMath's ability to differentiate dyscalculic and non-dyscalculic users, not to get feedback on MagnoMath's UI, behavior or functionality.

Firstly, participants were given a short introduction to the project and MagnoMath before conducting evaluations. Observational data was written down and results were registered. Evaluation details are given in table 5.6.

Table 5.6: Non-dyscalculic user group evaluation details

	User 1	**User 2**	**User 3**	**User 4**
Gender	Male	Male	Female	Female
Age	27	20	24	32
Date	May 10$^{\text{th}}$ 2016, Noon			
Duration	1 hour 53 min			
Location	Faculty of Humanities, University of Oslo, Blindern, Oslo, Norway			
Study type	Experimental and observational			
Data type	Quantitative and qualitative			

5.6.6   Results

5.6.6.1   Observation

Observational data was especially useful for obtaining an insight into participants' reasoning and problem solving.

All dyscalculic participants communicated that they had experienced struggles with mathematics since kindergarten, elementary school or junior high, describing their relationship with maths as tense, non-existing or frustrating. All dyscalculic participants could provide examples on how their mathematical struggles affect their everyday life, including telling and estimate time, filling out forms and calculating the price on items on sale.

Non-dyscalculic participants communicated that they did not have very strong mathematical backgrounds, but had not experienced specific struggles with mathematics in everyday life.

Participants' evaluation results are given in tables 5.7 and 5.8. Correlations between math and magnocellular task groups calculated as shown in algorithm 3 are not included as this data was not able to yield any valuable results.

Subsequent sections will present observational data gathered during evaluations performed by MagnoMath.

Contrast sensitivity. All participants submitted correct responses to all tasks evaluating contrast sensitivity. However, the task given in figure 5.11 was time-consuming for two dyscalculic participants, communicating that they became unsure of what is left and what is right. This is interesting behavior showing poor sense of direction, a deficiency possibly related to struggles with mathematics. Eventually, they submitted correct answers. Interestingly, deciding between

Table 5.7: Dyscalculic user group evaluation results

	User 1	User 2	User 3	User 4	User 5	User 6	User 7
Time spent to complete evaluation tasks	24 min 44 sec	10 min 22 sec	6 min 24 sec	6 min 21 sec	8 min 48 sec	4 min 37 sec	5 min 23 sec
Contrast sensitivity performance in percent	100	50	100	100	50	100	100
Spatial ability performance in percent	30	40	40	10	30	40	20
Counting performance in percent	100	50	50	50	50	100	50
Subitizing performance in percent	100	50	50	100	50	50	50
Arithmetic performance in percent	50	50	50	50	50	40	20
Math performance classification result	3/3	2/3	2/3	2/3	2/3	2/3	2/3
Magnocellular performance classification result	2/3	2/3	2/3	2/3	2/3	2/3	2/3
Overall result in percent	76	48	58	62	46	66	48

Table 5.8: Non-dyscalculic user group evaluation results

	User 1	User 2	User 3	User 4
Time spent to complete evaluation tasks	3 min 20 sec	2 min 42 sec	4 min 1 sec	2 min 50 sec
Contrast sensitivity performance in percent	100	100	100	100
Spatial ability performance in percent	50	100	50	100
Counting performance in percent	100	100	100	100
Subitizing performance in percent	100	100	100	100
Arithmetic performance in percent	100	100	100	100
Math performance classification result	3/3	3/3	3/3	3/3
Magnocellular performance classification result	3/3	3/3	3/3	3/3
Overall result in percent	90	100	90	100

the directions up and down did not initiate a similar behavior, answered correctly virtually instantly by all participants.

In summary, none of the non-dyscalculic participants while a few dyscalculic participants found contrast sensitivity tasks challenging.
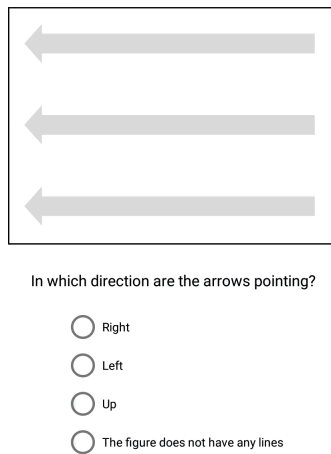
In which direction are the arrows pointing?

○ Right

○ Left

○ Up

○ The figure does not have any lines

Figure 5.11: Task evaluating contrast sensitivity some dyscalculic participants found challenging (Edited screenshot)

Spatial ability. As intended, all participants required time to reflect in order to submit answers to tasks evaluating spatial ability, making it the most difficult category overall. This is shown in tables 5.7 and 5.8, showing the lowest performance measures among all task categories. Some dyscalculic users answered quickly rather than correctly, resulting in submitting incorrect answers.

None of the dyscalculic participants were able to solve the task given in figure 5.12. Answers varied from 7 to 9 without any submitting the correct answer of 11. Interestingly, all non-dyscalculic participants were able to solve the task.

How many cubes?
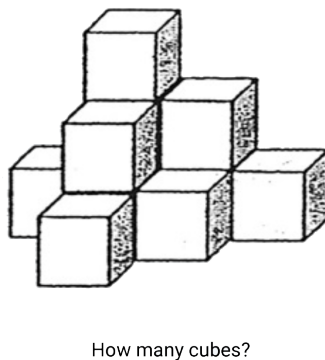
Figure 5.12: Task evaluating spatial ability all dyscalculic participants answered incorrectly (Edited screenshot) [41]

The task given in figure 5.13 also triggered interesting behavior. The majority of dyscalculic participants selected the incorrect left option. This shows poor reasoning as this answer option was selected by many due to it being the only one showing the star shaped symbol shown in the unfolded figure. Only one of the dyscalculic participants spent enough time to reflect to discover the correct middle option, while all non-dyscalculic participants submitted the correct answer
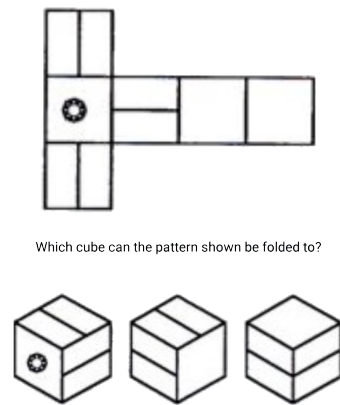
Figure 5.13: Task evaluating spatial ability most dyscalculic participants answered incorrectly (Edited screenshot)

after spending some time.

In summary, dyscalculic participants found spatial ability tasks challenging while non-dyscalculic participants were able to submit more correct answers as compared to dyscalculic ones.

Figure 5.14: Task evaluating counting skills some dyscalculic participants found challenging (Edited screenshot) [34]

Counting. As with tasks evaluating contrast sensitivity, all participants submitted correct answers to all tasks evaluating counting skills. However, dyscalculic participants required a significant amount of time to respond as shown by category performance measures given in table 5.7. In particular, the task given in figure 5.14 was challenging for some dyscalculic participants. For instance, while dyscalculic user 1 evaluated the task, she communicated that she did see patterns of two windows, but could not count them two by two. Instead, she had to count them one by one. This is interesting behavior, indicating that while she did see that the positioning of windows grouped in patterns of two, she could not add these groups together in order to count the windows more efficiently. User 1 also communicated that the crowded positioning of the windows gave her an overwhelming feeling, commenting that it was a large amount to count at once. Even if she could not add sets of two windows, she had no difficulties in quickly responding correctly to tasks evaluating subitizing skills. This can indicate that subitizing small

stand-alone sets of elements is significantly easier than subitizing such sets when mixed with other elements. Other dyscalculic participants communicated that they found solving the task "overwhelming" before eventually submitting the correct answer. Some dyscalculic participants were required to count the windows several times using their fingers, representing dyscalculic behavior. Non-dyscalculic participants did not show any signs of struggling with counting tasks.

Subitizing. Tasks evaluating subitizing skills triggered interesting dyscalculic behavior. Four dyscalculic participants were required to count each element individually in order to determine the quantity of elements shown. This resulted in mediocre results within the category as shown in table 5.7. Even if they did not spend a significant amount of time to submit correct answers, it still represented dyscalculic behavior as correct answers to the same tasks were given virtually instantly by all non-dyscalculic participants.

Arithmetic. While tasks evaluating spatial ability produced the highest amount of incorrect answers among dyscalculic participants, the category evaluating skills in arithmetic by far exceeds all others in terms of time required to complete it. All dyscalculic participants found one or more tasks within this category challenging. This revealed several interesting dyscalculic behaviors.

For instance, time required by dyscalculic user 1 for solving the task $\frac{24}{3}$ exceeded all other tasks in all categories. When presented the task, she communicated that she became stressed and started to break a sweat. She spent a couple minutes on finding a method for solving the problem before selecting the following approach: First, she used a pen to draw vertical lines of three dots counting up to twenty-four as shown in figure 5.15.
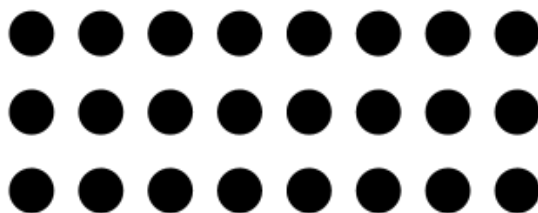

Figure 5.15: Quantity visualization applied by dyscalculic user 1 for solving the problem $\frac{24}{3}$

Then, user 1 counted the number of dots in the top horizontal row, eventually providing the correct answer of 8. In total, she spent five minutes and seven seconds to solve the problem, which by far exceeded the average time needed among non-dyscalculic participants. This is an interesting observation, proving that MagnoMath had the ability to reveal dyscalculic behavior.

Dyscalculic user 3 also struggled with the problem $\frac{24}{3}$. She first spent half a minute reflecting upon the task before pressing the *Next task* button without first submitting an answer. When she realized that MagnoMath would not allow her to continue before answering, she further reflected upon the task before saying to the observer, the author of this thesis, that she simply does not know the answer. She then asked the observer for guidance on how to proceed. Interestingly, after the observer suggested her to submit a guess, she submitted the correct answer of 8. After she finished the evaluation, she communicated that she had to use her fingers in nearly all tasks, demonstrating dyscalculic behavior. Dyscalculic user 7 was unable to solve the task, submitting the incorrect guess of 7. As the problem was given by MagnoMath in the form 24/3 instead of $\frac{24}{3}$, she explained that she saw it as a date rather than a ratio, and could therefore not solve the problem.

When dyscalculic user 1 was presented the problem of deciding whether $9 \times 3 = 28$ is correct, she solved it using the following method: First, she placed her hands in front of her. Then, in her mind, she labeled each finger left to right the numbers from one to ten. Then, she put down

the finger which was to be multiplied by nine, which in this case is 3. For finding the answer, she counted each finger positioned left of the finger down as tens, and each finger positioned right of the finger down as ones. Then, she simply added the tens and ones together before saying 27, which is the correct answer of the multiplication. Figure 5.16 illustrates the method. She then submitted that the given solution to the multiplication is incorrect, eventually solving the task after spending 1 minute and 16 seconds.
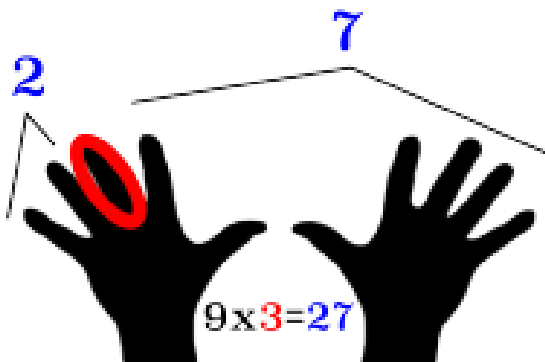
Figure 5.16: Solving the problem 9×3 using hands [32]

Even if dyscalculic participants applied inefficient methods, the majority submitted correct answers on arithmetic tasks. Non-dyscalculic participants did not express any struggles in solving tasks within this category in a timely manner.

Summary. Observations proved that MagnoMath was able to detect dyscalculia symptoms. Dyscalculic participants found categories evaluating spatial ability and arithmetic most challenging. Within the former, especially three-dimensional mental manipulation of figures proved to be challenging. In the latter category, division was most difficult among all dyscalculic participants. Non-dyscalculic users found the category evaluating spatial ability the only one challenging before all submitting correct answers. Comparing results in tables 5.7 and 5.8 proves that MagnoMath was able to differentiate dyscalculic and non-dyscalculic users. This especially applied to classifications of mathematical performances given in the row *Math performance classification result* in both tables, which is a sought outcome.

5.6.6.2 Functionality

Results from functionality rankings provided by dyscalculic users are given in figure 5.17. As the graph shows, MagnoMath scores high on most criteria. The most interesting and important among them is maybe the one regarding their own perception of evaluation results. As figure 5.17 shows, participants found results reasonable indicating high accuracy. Some participants expressed that they acknowledged deficiencies indicated by the application. As shown in table 5.7, dyscalculic user 5 received mediocre scores in counting, subitizing and arithmetic. Interestingly, she acknowledged that results were accurate as she communicated that she needed to use her fingers for nearly all evaluation tasks. This comment further supported the accuracy of MagnoMath. It was reassuring that users found results descriptive for their situation. This together with generated results given in tables 5.7 and 5.8 prove that MagnoMath did not generate too many false positives and false negatives.

The functionality statement obtaining the highest mean ranking value was the one regarding usefulness of application features, as shown in figure 5.17. Assisting diagnosis of dyscalculia as

well as identifying mathematical deficiencies possibly related to magnocellular reasoning are the main purposes of the application. However, it was also valuable to include additional functionalities in the evaluation. As figure 5.17 shows, participants found accommodating functionalities including sending and sorting of results useful, expressing that they were natural and valuable complements to MagnoMaths main functionality of evaluating and classifying user skills. This feedback is useful for future work potentially implementing MagnoMath in schools. Knowing that all functionalities were considered valuable indicated that these could be developed further for pedagogical settings, which is an encouraging outcome.

As it is not optimal to ask an individual struggling heavily with mathematics whether the application is suitable for assisting diagnosis of dyscalculia, it was natural that this functionality statement obtained the lowest score as shown in figure 5.17. It was nevertheless included to obtain information whether users felt that MagnoMath identified struggles with mathematics possibly related to dyscalculia. The majority of participants expressed that the software was useful for assisting diagnosis. Some participants seemed to be unsure of what constitutes dyscalculia symptoms. Therefore, it would be advisable to include special education teachers in future evaluations.

Another interesting functionality statement was the one regarding usefulness of results, ranked as *Neither agree or disagree*, *Agree* or *Strongly agree* by all participants. The former ranking was due to some participants being unsure whether evaluation results actually corresponded to their own abilities. This is natural, as it is challenging for an individual struggling intensively with mathematics to know exactly what constitutes dyscalculic behaviors. However, it was useful to include the statement to get an insight into how each participant perceived their own skills and what MagnoMath could bring in order to increase this introspection. Certainty, special education teachers could in future work contribute to the evaluation of usefulness of evaluation results.
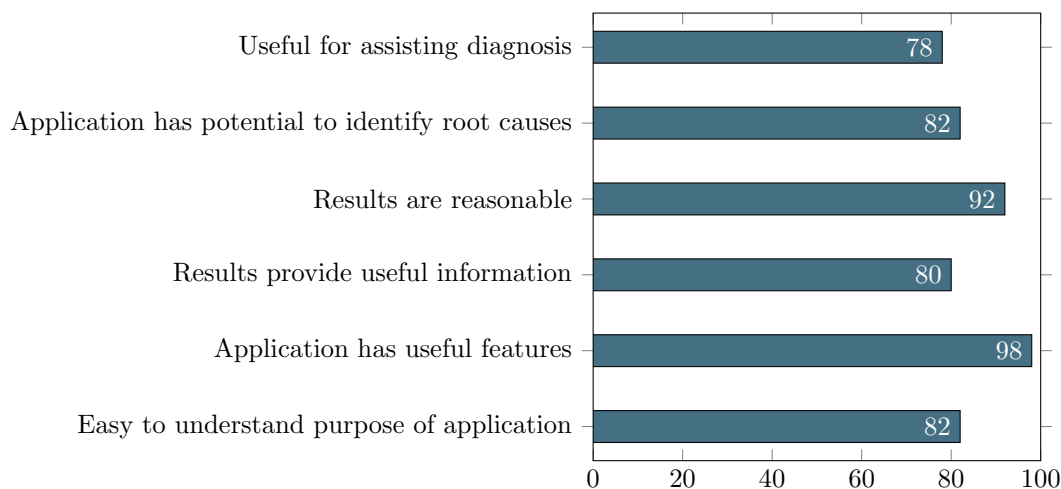
Figure 5.17: User evaluation functionality rankings mean values in percent

5.6.6.3 Usability

After completing evaluations and given results, the majority of participants expressed that MagnoMath is easy to use and understand. Comments included that the UI was clean and logical, creating a positive user experience.

An interesting observation during use was done. Some dyscalculic participants simply did not know the answer to some evaluation tasks presented as discussed in section 5.6.6.1. When they found themselves in this situation, they asked the observer how to proceed before submitting a guess. This requirement of submitting an answer can be removed in future development of MagnoMath by including a *Pass* answer option.

More detailed usability feedback generated by SUS scores is given in figure 5.18. As described in section 5.4.6.1, a SUS score of 70 is passable and 90 is very good. Averaging all scores given in figure 5.18 gives a mean score of 74.6, representing a score that can be improved. The reason for obtaining this score could be partly due to participants' focusing more on their experience during evaluation of mathematical and magnocellular reasoning instead of their experience of the application as a whole. Furthermore, participants' insecurity regarding own mathematical and magnocellular performance could be a contributing factor. As this is expected for individuals struggling intensively with mathematics, it was considered natural that the average SUS score would be affected by this. This outcome could be resolved in future work by including teachers giving users of MagnoMath a greater insight into their own skills.
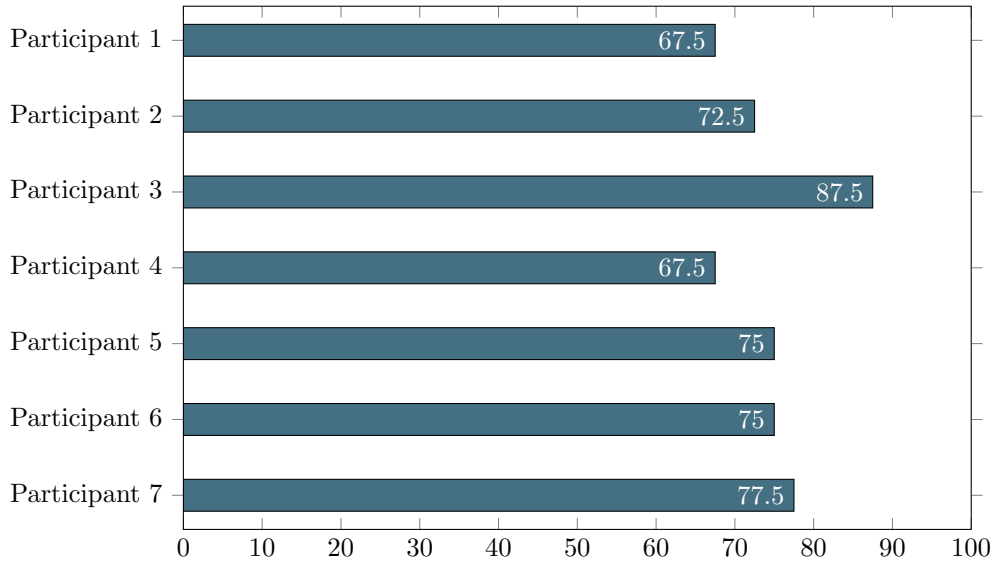
Figure 5.18: User evaluation SUS scores

Chapter 6

Discussion

This chapter will discuss the research questions of this project, following design science research guideline four presented in section 3.1.

6.1   Research question 1

The first research question to be answered by this thesis is given in section 1.1 and below for reference.

RQ1. How suitable is mobile software as a tool for assisting diagnosis of dyscalculia?

The software artifact in this project was developed for smartphones. This decision was made prior to designing and developing the classifier evaluating mathematical and magnocellular reasoning. This project has consequently tested smartphones' ability to classify data efficiently. After prototype finalization and evaluation, observational data proved that the platform has sufficient performance and memory for necessary data processing. This enabled efficient generation of accurate evaluation results. In addition of being efficient, evaluations did not reveal a single operational error such as halt or crash, thus proving that MagnoMath is suited for its platform.

During expert evaluations, researchers and practitioners within the fields of dyscalculia and LDs were positive to the idea of developing the tool for mobile devices. At the time of writing, the prevalence of similar applications was very low making MagnoMath a valuable addition to mobile software.

One expert evaluator was skeptical about the quantity of tasks used in evaluation, pointing out that a greater number of tasks would be required to obtain a sufficient insight into users' mathematical and magnocellular reasoning. This could be resolved in future development by increasing the number of tasks. However, the objective within the scope of this project was to implement an efficient tool not covering the entire spectrum of users' mathematical skills. This would have put a great demand on both MagnoMath and its users. Task categories could however be added to identify additional possible correlations between mathematical and magnocellular reasoning. The average smartphone at the time of writing should not raise any obstacles for achieving this.

A possible constraint taken into account when selecting smartphones as platform was the small screen size. During evaluations, participants communicated that smartphone's screen was large enough to display figures and text so that content is easy to see and text easy to read. Responses

included that the screen presented content clearly, proving that screen size and sharpness are sufficient for presenting tasks evaluating mathematical skills.

During user evaluations, MagnoMath was able to reveal symptoms of dyscalculia. However, programming the software application to actually detect these symptoms was a more challenging task. To achieve this, that is, to differentiate dyscalculic and non-dyscalculic users, it was essential to take time required to complete tasks into account. The applied method yielded a success in identifying dyscalculic behavior without the need of a teacher being present, introducing a sought advantage. User evaluations proved that MagnoMath differentiates dyscalculic and non-dyscalculic users, which is one of the most appreciated outcomes.

In summary, the average smartphone running Android OS has the sufficient performance and memory capacity to process necessary data gathered from evaluating users' mathematical skills. Screen size and sharpness were sufficient to communicate tasks and acquire responses. User evaluations proved that MagnoMath was able to reveal dyscalculia symptoms while separating dyscalculic from non-dyscalculic users. Dyscalculic evaluators found results both reasonable and valuable. Some expert evaluators were a bit sceptical of using mobile software, while user evaluators were in general very positive to the approach. These results together prove that mobile software represents a tool highly suitable for assisting diagnosis of dyscalculia. To obtain further insight into mobile software's suitability of assisting dyscalculic individuals, further evaluation including a larger number of volunteers having symptoms of dyscalculia would be necessary.

6.2   Research question 2

The second research question to be answered by this thesis is given in section 1.1 and below for reference.

RQ2. Can mobile software identify dyscalculia symptoms possibly related to magnocellular reasoning?

MagnoMath applies two types of tasks: tasks evaluating mathematical and magnocellular reasoning, respectfully. The former type is divided into categories (1) contrast sensitivity and (2) spatial ability. The latter type is divided into categories (3) subitizing, (4) counting and (5) arithmetic. Responses submitted to tasks within these categories provide crucial information applicable for identifying dyscalculia symptoms related to magnocellular reasoning. After processing response data, MagnoMath calculates and presents performance measures within each category.

Results generated from dyscalculic users' responses are given in table 5.7. We can study this data to measure correlations across categories. We have $3 \times 2 = 6$ possible correlations. To calculate the correlation in percent between two categories in a result record, represented by a column in table 5.7, we first calculate the ratio of performances obtained in each category. Then, we subtract this ratio from the highest correlation percentage possible, which is 100. This is done since the correlation is maximum when there is no difference between two category performance measures. This correlation is calculated for each of the six ($mathTaskCategory$, $magnocellTaskCategory$) pairs in each of the seven result records from dyscalculic users, given in table 6.1. In the table, the interesting results are those showing high correlation percentage when category performances were low. Since users 6 and 7 were both weak in arithmetic and spatial reasoning, their correlation percentage of 100 is interesting, as being weak in arithmetics may be partially caused by poor spatial reasoning. Another interesting result was the high correlation between counting and spatial ability in users 2 and 3, suggesting a possible dependency.

To obtain an overview of all six possible correlations between the five categories across the two groups, the mean values of all correlation measures in table 6.1 were computed, given in figure 6.1.

Table 6.1: Dyscalculic users' mathematical/magnocellular evaluation categories correlation measures in percent

	User 1	User 2	User 3	User 4	User 5	User 6	User 7
Arithmetic & Spatial ability	80	90	90	60	80	100	100
Arithmetic & Contrast sensitivity	50	100	50	50	100	40	20
Subitizing & Spatial ability	30	90	90	10	80	90	70
Subitizing & Contrast sensitivity	100	100	50	100	100	50	50
Counting & Spatial ability	30	90	90	60	80	40	70
Counting & Contrast sensitivity	100	100	50	50	100	100	50
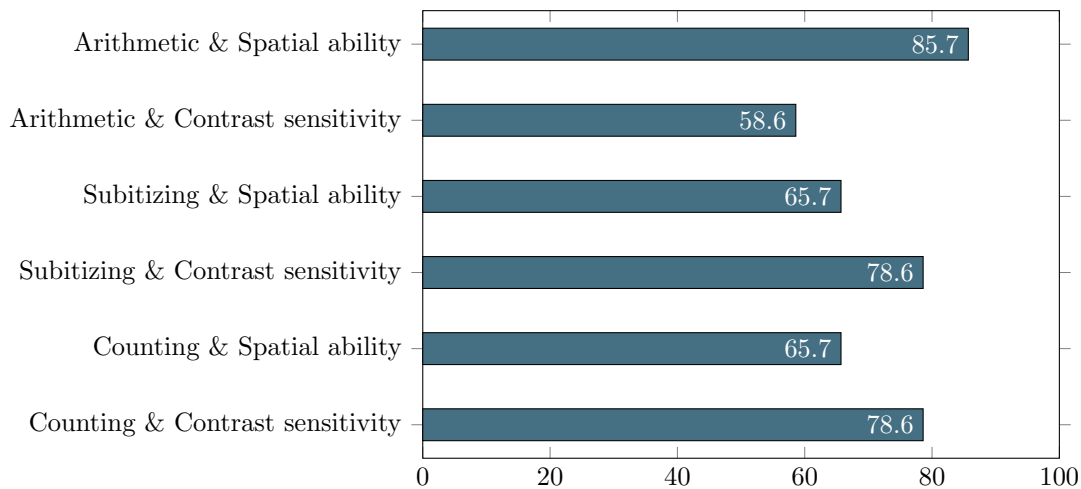
Figure 6.1: Dyscalculic users' mathematical/magnocellular evaluation categories correlation measures mean values in percent

As figure 6.1 shows, the correlation between deficiencies in arithmetic and spatial ability was the strongest among dyscalculic participants. This find corresponds to observational data where the majority of dyscalculic participants who struggled with tasks evaluating arithmetic also struggled with those evaluating spatial reasoning. This is an interesting outcome, especially

as this correlation was derived from results generated by MagnoMath. Figure 6.1 also shows strong correlations between the groups subitizing and contrast sensitivity, as well as counting and contrast sensitivity. However, these are not as interesting as the performances within these groups were strong.

In summary, evaluation data proves that mobile software is, to a certain extent, able to identify dyscalculia symptoms possibly related to magnocellular reasoning. The current version of MagnoMath utilizes only smartphone's touch screen. Further development could include additional features of smartphones, such as camera, microphone or gyro sensor. Future evaluations involving a larger number of professionals and volunteers could improve MagnoMath's accuracy and validity.

Chapter 7

Conclusion and Future Work

The research presented in this thesis has studied the suitability of mobile software as a tool for assisting diagnosis of dyscalculia as well as whether mobile software is able to identify dyscalculia symptoms possibly related to magnocellular reasoning. This research was motivated by the low prevalence among mobile software applications developed for identifying underlying causes of developing the condition.

The design science research framework was applied to support this research. The project has developed MagnoMath, a mobile software application for smartphones for assisting diagnosis of dyscalculia and detected symptoms' relation to magnocellular reasoning. The application was implemented in Java and XML for Android OS and applies k-NN algorithm for classifying user evaluation responses into appropriate performance classes. To support software development, Personal Scrum was utilized, applying incremental development enabling continuous integration and improvement spread across ten iterations.

To test MagnoMath's functionality and applied methods, the software application was evaluated by researchers and practitioners in the fields of dyscalculia and learning disabilities from Linköping University, Sweden and Dyslexia Norway, Oslo. After implementing expert evaluation feedback, an experiment involving seven dyscalculic and four non-dyscalculic individuals was conducted to test MagnoMath's accuracy, validity and usability. Findings showed that MagnoMath was able to reveal dyscalculia symptoms by presenting a set of tasks in mathematics evaluating counting, subitizing and arithmetic. Additionally, the software activated magnocellular reasoning by presenting tasks evaluating contrast sensitivity and spatial ability. After evaluation conduction, dyscalculic users found the application useful and easy to use. Results indicated a strong correlation between deficiencies in arithmetic and spatial ability, representing a possible correlation valuable for identifying early signs of developmental dyscalculia. Both experts and dyscalculic volunteers were positive to the application and found it valuable, representing a great potential for future development. Further evaluation involving a larger number of professionals and dyscalculic individuals would be necessary to secure MagnoMath's accuracy and validity.

7.1   Future work

The next step for this project would be to conduct more thorough evaluation of MagnoMath. This would be valuable for further testing the application's ability to assist the diagnosis of dyscalculia as well as identifying possible additional dependencies between mathematical and magnocellular reasoning. As conducted evaluations only included female volunteers due to availability, male volunteers would be required to include in future studies. Due to availability and the limited

timespan of this project, user evaluations were conducted over the internet, which was sufficient for this project. One-on-one meetings have advantages making it easier to obtain more feedback and should therefore be considered.

MagnoMath is currently available for Android OS only. To accommodate additional users, it is natural that further work will develop versions for additional platforms.

This research has found a great need and motivation for implementing a mobile software in schools as a tool for assisting dyscalculic pupils. Further testing and development would enable teachers to perform early diagnosis by studying early underlying signs of dyscalculia. This has the potential of providing crucial help early enough.

Bibliography

[1] 123test. (2016). Spatial reasoning test. Available from: `https://www.123test.com/spatial-reasoning-test`

[2] Android Studio. Available from: `https://developer.android.com/studio/index.html`

[3] Bangor, A., Kortum, P. T., Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. Published in: International Journal of Human-Computer Interaction, 24 (6), 574–594.

[4] Beal, V. (2016). GUI – graphical user interface. Available from: `http://www.webopedia.com/TERM/G/Graphical_User_Interface_GUI.html`

[5] Bhatia, N. & Vandana, A. (2010). Survey of Nearest Neighbor Techniques. Published in: International Journal of Computer Science and Information Security, 8 (2), 302–305.

[6] Bitbucket. Available from: `https://www.bitbucket.org`

[7] Brightstar Learning. (2016). The Magnocellular visual pathway. Available from: `http://www.brightstar-learning.com/the-science/magnocellular-visual-pathway/`

[8] British Medical Association. (2010). BMA Complete Medical Guide – The essential reference for every family (1st ed.). United Kingdom: Dorling Kindersley.

[9] Brooke, J. (1986). SUS – A quick and dirty usability scale. Available from: `http://www.usabilitynet.org/trump/documents/Suschapt.doc`

[10] Bull, R.D. (2014). Communicating pictures – A Course in Image and Video Coding (1st ed.). United States of America: Academic Press.

[11] Butterworth, B. (2003). The Dyscalculia Screener Manual. Available from: `http://sebastien.brunekreef.com/dyscalculie/Dyscalculia_Screener_Manual.pdf`

[12] Butterworth, B. (2005). The development of arithmetical disabilities. Published in: Journal of Child Psychiatry, 46 (1), 3–18.

[13] Butterworth, B., Landerl, K., Bevan, A. (2003). Developmental dyscalculia and basic numerical capacities: a study of 8-9-year-old students. Published in: Cognition, 93 (2), 99–125.

[14] Creel, D.J. (2015). Visually Evoked Potentials. Available from: `http://webvision.med.utah.edu/book/electrophysiology/visually-evoked-potentials/`

[15] David, J.M. & Balakrishnan, K. (2010). Machine Learning Approach for Prediction of Learning Disabilities in School-Age Children. Published in: International Journal of Computer Applications, 9 (11), 7–14.

[16] Deemer, P., Benefield, G., Larman, C., Vodde, B. (2012). The Scrum Primer.

[17] Demb, J.B., Geoffrey, M.B., Heeger, D.J. (1997). Brain activity in visual cortex predicts individual differences in reading performance. Published in: Proc. Natl. Acad. Sci. USA, 94 (24), 13363–13366.

[18] Deza, M.M. & Deza, E. (2009). Encyclopedia of Distances (1st ed.). Germany: Springer.

[19] Dynamo Maths. (2014). Dynamo Maths – A three-stage programme. Available from: `http://www.dynamomaths.co.uk/index.html`

[20] Eclipse IDE. Available from: `https://www.eclipse.org/ide/`

[21] Elmasri, R. & Navathe, S.B. (2011). Database Systems – Models, Languages, Design, and Application Programming (6th ed.). United States of America: Pearson Education, Incorporation.

[22] Geary, D.C., Hamson, C.O., Hoard, M.K. (2000). Numerical and Arithmetical Cognition: A Longitudinal Study of Process and Concept Deficits in Children with Learning Disability. Published in: Journal of Experimental Child Psychology, 77 (3), 236–263.

[23] Ghahramani, Z. (2004). Unsupervised Learning. Published in: Lecture Notes in Computer Science, 3176, 72-112.

[24] Git. Available from: `https://www.git-scm.com`

[25] Glaucoma Associates of Texas. (2016). Glaucoma. Available from: `http://glaucomaassociates.com/glaucoma/`

[26] Hevner, A.R., March, S.T., Park, J., Ram, S. (2004). Design Science in Information Systems Research. Published in: MIS Quarterly, 28 (1), 75–105.

[27] Indigo Studio. Available from: `http://www.infragistics.com/products/indigo-studio`

[28] join.me. Available from: `https://www.join.me`

[29] Kaufman, E.L. & Aster, M.V. (2012). The diagnosis and Management of Dyscalculia. Published in: Deutsces Ärzteblatt International, 109 (45), 767–777.

[30] Kaufman, E.L., Lord, M.W., Reese, T.W., Volkmann, J. (1949). The Discrimination of Visual Number. Published in: The American Journal of Psychology, 62 (4), 498–525.

[31] Kleinberg, J. & Tardos, É. (2005). Algorithm Design. (1st ed.). England: Pearson Education Limited.

[32] Knowles, E. (2015). Use Your Fingers To Multiply As Fast As You Can Count. Available from: `http://thescienceexplorer.com/universe/use-your-fingers-multiply-fast-you-can-count`

[33] Kontio, J., Lehtola, L., Bragge, J. (2004). Using the focus group method in software engineering: obtaining practitioner and user experiences. Published in: Empirical Software Engineering, 271–280.

[34] Lee, R. (2012). So many windows. Available from: `https://www.flickr.com/photos/roboppy/7127794749`

[35] Livingstone, M.S., Rosen, G.D., Drislane, F.W., Galaburda, A.M. (1991). Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. Published in: Proc. Natl. Acad. Sci. USA, 88 (18), 7943–7947.

[36] Mark, S.T. & Smith, G.T. (1995). Design and natural science research on information technology. Published in: Decision Support Systems, 15 (4), 251–266.

[37] Mitchell, T.M. (1997). Machine learning (1st ed.). United States of America: McGraw-Hill.

[38] Nielsen, J. (1995). 10 Usability Heuristics for User Interface Design. Available from: `https://www.nngroup.com/articles/ten-usability-heuristics/`

[39] Nunamaker, Jr, J.F., Chen, M., Purdin, T.D.M. (1990). Systems Development in Information Systems Research. Published in: Journal of Management Information Systems, 6 (4), 89–106.

[40] Oates, B.J. (2006). Researching Information Systems and Computing (1st ed.). England: SAGE Publications Limited.

[41] Old School. (2008). Continual Assessment 2 (#4). Available from: `http://www.oldschool.com.sg/index.php/module/PublicAccess/action/Wrapper/sid/ea5f3768e0dae62aa6d01beee8123931/cmbn_id/2660/qba/1`

[42] OpenStax. (2015). The Brain and Spinal Cord. Available from: `https://cnx.org/contents/_Io4zP0c@7/The-Brain-and-Spinal-Cord`

[43] Pruitt, J. (2011). Personal Scrum. Available from: `http://blog.jgpruitt.com/2011/04/10/personal-scrum/`

[44] Psychometric Success. (2016). Sample Spatial Ability Questions. Available from: `http://www.psychometric-success.com/faq/faq-sample-spatial-ability-questions.htm`

[45] Psychometric Success. (2016). Spatial Ability Tests. Available from: `http://www.psychometric-success.com/aptitude-tests/spatial-ability-tests-cubes.htm`

[46] Rogers, Y., Sharp, H., Preece, J. (2011). Interaction Design: Beyond Human-Computer Interaction (3rd ed.). United Kingdom: John Wiley & Sons Limited.

[47] Russel, S. & Norvig, P. (2010). Artificial Intelligence – A Modern Approach (3rd edition). United States of America: Pearson Education.

[48] Sherman, S.M. & Guillery, R.W. (2001). Exploring the Thalamus (1st ed.). United States of America: Academic Press.

[49] Stein, J (2001). The magnocellular theory of developmental dyslexia. Published in: Dyslexia, 7 (1), 12–36.

[50] Stein, J. (2014). Dyslexia: the Role of Vision and Visual Attention. Published in: Current Developmental Disorders Reports, 1 (4), 267–280.

[51] SQLite. Available from: `https://www.sqlite.org`

[52] The Healers Journal. Brain hemisphere balance pineal gland 1. Available from: `http://www.thehealersjournal.com/wp-content/uploads/2013/05/brain-hemisphere-balance-pineal-gland1.png`

[53] Trello. Available from: `http://www.trello.com`

[54] Tribal Education. (2015). DysCalculiUM. Available from: `https://dyscalculia.advancelearningzone.com/index.php?option=com_content&view=article&id=2&Itemid=2`

[55] Vaishnavi, V. & Kuechler, B. (2004). Design Science Research in Information Systems. Available from: `http://desrist.org/desrist/content/design-science-research-in-information-systems.pdf`

[56] VistaVision DMD Medtech. (2016). Visual Acuity Tester. Available from: `http://www.dmdmedtech.it/doc_free/VistaVision-brochure-2.4.2-en.pdf`

[57] Weaver, P. (2004). Success in your project: A guide to student system development projects (1st ed.). England: Pearson Education Limited.

[58] WikiJob. (2016). Spatial Awareness Test. Available from: `https://www.wikijob.co.uk/content/aptitude-tests/test-types/spatial-awareness-test`

[59] Wilson, A. (2012). What are the symptoms of dyscalculia? – Symptoms established by research. Available from `http://www.aboutdyscalculia.org/symptoms.html`

[60] Wilson, A. J. & Dehaene, S. (2004). The Number Race. Available from `http://www.thenumberrace.com/nr/home.php?lang=en`

[61] Woolfolk, A. (2013). Educational Psychology: Active Learning Edition (12th ed.). United States of America: Pearson Education Limited.

Appendix A

Participation information sheets and concent forms in Norwegian

A.1   Interview

Forespørsel om deltakelse i forskningsprosjektet

Identifisering av dyskalkulisymptomer relatert til magnocellulær resonnement ved bruk av smarttelefoner

Bakgrunn og formål

Mitt navn er Greger Siem Knudsen (greger.knudsen@student.uib.no) og jobber med et mastergradsprosjekt ved Institutt for Informasjons- og Medievitenskap ved Universitetet i Bergen. Veileder for dette mastergradsprosjektet er professor ved Linköpings Universitet og førsteamanuensis ved Universitetet i Bergen, Ankica Babic (ankica.babic@uib.no).

Målet med denne studien er å utvikle en prototype på en programvare som er ment til å benyttes til å assistere diagnostiseringen av dyskalkuli. Dette ved å inkludere en alternativ metode sammen med den metoden som vanligvis brukes til dette formålet av programvare i dag. Denne alternative metoden har ikke hovedfokus på å evaluere regneferdigheter og tallprosessering direkte, men istedenfor evaluere mulige underliggende årsaker til vanskeligheter med disse emnene. Mer spesifikt innebærer dette å undersøke oppførselen til en spesifikk type nevroner som har ansvar for blant annet å oppdage raske endringer i omgivelsene og å flytte informasjon fra synet til senteret for informasjonsprosessering i hjernen, kalt magnocellulære celler. Disse cellene vil ikke bli undersøkt direkte, men gjennom responsen avgitt på et sett med oppgaver gitt av programvaren.

Utvalget er basert på relevante ekspert- og brukergrupper av programvaren som vil bli utviklet gjennom dette mastergradsprosjektet, og er derfor ikke tilfeldig utplukket.

Hva innebærer deltagelse i studien?

Studiet vil bli gjennomført i form av et semi-strukturert personlig intervju. Intervjuet vil ta omtrent 30 minutter og vil bli gjennomført på avtalt sted og tidspunkt. Ettersom det er vanskelig å skrive ned alt som sies underveis under et intervju vil det, kun dersom det er i orden for deg som deltaker selvfølgelig, bli foretatt lydopptak av intervjuet.

Hva skjer med informasjonen om deg?

Alle personopplysninger vil bli behandlet konfidensielt og som forsker har jeg **taushetsplikt**. Intervjuet vil som tidligere beskrevet bli tatt opp som lydopptak og senere bli skrevet ned. Lydopptaket vil bli slettet så fort dette er gjort. Det er kun meg og min veileder som vil ha tilgang til dataene og alt som samles inn vil bli anonymisert. Ved publisering og presentasjon av prosjektet vil personlig informasjon som for eksempel navn bli anonymisert. Studien er meldt inn til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS. Prosjektet skal etter planen avsluttes 1. juni 2016 og alle opptak og innsamlet data som ikke er anonymisert vil da være slettet.

Frivillig deltakelse

Det er helt frivillig å delta i studien, og du kan når som helst trekke ditt samtykke om å være med, uten å måtte oppgi noen grunn. Dersom du trekker deg, vil alle opplysningene om deg bli anonymisert. Det er også helt valgfritt å svare på spørsmålene som blir gitt underveis.

Dersom du har spørsmål til studien, vennligst kontakt Greger Siem Knudsen (greger.knudsen@student.uib.no) eller veileder Ankica Babic (ankica.babic@uib.no).

Studien er, som beskrevet tidligere, meldt inn til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Samtykke til deltakelse i studien

Erklæring fra deltaker

Jeg har mottatt informasjon om studien, og er villig til å delta.

☐ *Jeg har lest og forstått hva det innebærer å delta i studiet «Identifisering av dyskalkulisymptomer relatert til magnocellulær resonnement ved bruk av smarttelefoner» og samtykker til å delta i intervju*

_____

(Signatur av prosjektdeltaker, dato)

Erklæring fra forsker

☐ *Jeg bekrefter å ha gitt den potensielle deltakeren konsis og riktig informasjon, og har etter beste evne sørget for at han/hun forstår hva vi skal gjennom i dette intervjuet. Jeg bekrefter at deltakeren har blitt gitt mulighet til å stille spørsmål om dette intervjuet, og at alle spørsmålene har blitt riktig besvart av undertegnede. Jeg bekrefter at deltakeren selv har valgt å delta i dette intervjuet, samt at personen ikke har blitt overtalt til å gi sitt samtykke.*

_____

(Signatur av forsker, dato)

En kopi til dette samtykkeskjemaet har blitt gitt til deltakeren.

A.2 User testing

A.2.1 Adults

Forespørsel om deltakelse i forskningsprosjektet - Myndige

Identifisering av dyskalkulisymptomer relatert til magnocellulær resonnement ved bruk av smarttelefoner

Bakgrunn og formål

Mitt navn er Greger Siem Knudsen (e-post: greger.knudsen@student.uib.no) og jobber med et mastergradsprosjekt ved Institutt for Informasjons- og Medievitenskap ved Universitetet i Bergen. Veileder for dette mastergradsprosjektet er professor ved Linköpings Universitet og førsteamanuensis ved Universitetet i Bergen, Ankica Babic (e-post: ankica.babic@uib.no).

Målet med denne studien er å utvikle en prototype på en programvare som er ment til å benyttes til å assistere diagnostiseringen av dyskalkuli. Dette ved å inkludere en alternativ metode sammen med den metoden som vanligvis brukes til dette formålet av programvare i dag. Denne alternative metoden har ikke hovedfokus på å evaluere regneferdigheter og tallprosessering direkte, men istedenfor evaluere mulige underliggende årsaker til vanskeligheter med disse emnene. Mer spesifikt innebærer dette å undersøke oppførselen til en spesifikk type nevroner som har ansvar for blant annet å oppdage raske endringer i omgivelsene og å flytte informasjon fra synet til senteret for informasjonsprosessering i hjernen, kalt magnocellulære celler. Disse cellene vil ikke bli undersøkt direkte, men gjennom responsen gitt på et sett med oppgaver gitt av programvaren.

Prototypen er nå ferdigstilt, og er klar for brukerevaluering. Du er forespurt om å delta i denne studien grunnet din rolle som potensiell bruker i prototypens målgruppe.

Hva innebærer deltagelse i studien?

Studiet vil bli gjennomført i form av en evaluering av prototypen. Først vil du som deltager fullføre en test gitt av prototypen bestående av et sett med oppgaver som tester utvalgte matematiske ferdigheter og magnocellulær resonnementsevne. Deretter vil du bli stilt noen enkle spørsmål om dine meninger om prototypen. Kun ditt kjønn, din alder, dine oppnådde resultater på testen og svar på spørsmål stilt etter denne vil bli samlet inn. Navn eller annen data som på noen måte kan identifisere deg blir ikke samlet inn eller publisert. Deltagelse er derfor helt anonymt. Deltagelse vil bli gjennomført på avtalt sted og tidspunkt. Deltagelse kan også foregå over internett ved hjelp av videosamtale, ved å bruke for eksempel Skype. Totalt vil deltagelse ta omtrent en halv time.

Hva skjer med informasjonen om deg?

Alle innsamlede data vil bli behandlet konfidensielt og som forsker har jeg **taushetsplikt**. Det er kun meg og min veileder som vil ha tilgang til dataene og alt som samles inn vil bli anonymisert.

Prosjektet skal etter planen avsluttes 1. juni 2016.

Frivillig deltakelse

Det er helt frivillig å delta i studien, og du kan når som helst trekke ditt samtykke om å være med, uten å måtte oppgi noen grunn. Dersom du trekker deg, vil alle opplysningene om deg bli anonymisert. Det er også helt valgfritt å svare på spørsmålene som blir gitt.

Dersom du har spørsmål til studien, vennligst kontakt Greger Siem Knudsen (e-post: greger.knudsen@student.uib.no) eller veileder Ankica Babic (e-post: ankica.babic@uib.no).

Studien er meldt inn til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Samtykke til deltakelse i studien

Erklæring fra deltaker

Jeg har mottatt informasjon om studien, og er villig til å delta.

☐ *Jeg har lest og forstått hva det innebærer å delta i studiet «Identifisering av dyskalkulisymptomer relatert til magnocellulær resonnement ved bruk av smarttelefoner» og samtykker til å delta i brukerevaluering*

_____
(Signatur av studiedeltaker, dato)

Erklæring fra forsker

☐ *Jeg bekrefter å ha gitt den potensielle deltakeren konsis og riktig informasjon, og har etter beste evne sørget for at deltakeren forstår hva deltakelsen innebærer. Jeg bekrefter at deltakeren har blitt gitt mulighet til å stille spørsmål om denne studien, og at alle spørsmålene har blitt riktig besvart av undertegnede. Jeg bekrefter at deltakeren selv har valgt å delta, samt at personen ikke har blitt overtalt til å gi sitt samtykke.*

_____
(Signatur av forsker, dato)

En kopi av dette samtykkeskjemaet vil blitt gitt til deltakeren.

A.2.2 Minors

Forespørsel om deltakelse i forskningsprosjektet - Umyndige

Identifisering av dyskalkulisymptomer relatert til magnocellulær resonnement ved bruk av smarttelefoner

Bakgrunn og formål

Mitt navn er Greger Siem Knudsen (e-post: greger.knudsen@student.uib.no) og jobber med et mastergradsprosjekt ved Institutt for Informasjons- og Medievitenskap ved Universitetet i Bergen. Veileder for dette mastergradsprosjektet er professor ved Linköpings Universitet og førsteamanuensis ved Universitetet i Bergen, Ankica Babic (e-post: ankica.babic@uib.no).

Målet med denne studien er å utvikle og evaluere en prototype på en programvare som er ment til å benyttes til å assistere diagnostiseringen av dyskalkuli. Dette ved å inkludere en alternativ metode sammen med den metoden som vanligvis brukes til dette formålet av programvare i dag. Denne alternative metoden har ikke hovedfokus på å evaluere regneferdigheter og tallprosessering direkte, men istedenfor evaluere mulige underliggende årsaker til å utvikle vanskeligheter med disse emnene. Mer spesifikt innebærer dette å undersøke oppførselen til en spesifikk type nevroner som har ansvar for blant annet å oppdage raske endringer i omgivelsene og å flytte informasjon fra synet til senteret for informasjonsprosessering i hjernen, kalt magnocellulære celler. Disse cellene vil ikke bli undersøkt direkte, men gjennom responsen gitt på et sett med oppgaver gitt av programvaren.

Prototypen er nå ferdigstilt, og er klar for brukerevaluering. Ditt barn er gjennom deg forespurt om å delta i denne studien grunnet barnets rolle som potensiell bruker i prototypens målgruppe.

Hva innebærer deltagelse i studien?

Studiet vil bli gjennomført i form av en evaluering av prototypen. Først vil barnet fullføre en test gitt av prototypen bestående av et sett med oppgaver som tester utvalgte matematiske ferdigheter og magnocellulær resonnementsevne. Deretter vil det bli stilt noen enkle spørsmål om barnets meninger om prototypen. Kun barnets kjønn, alder, oppnådde resultater på testen og svar på spørsmål stilt etter denne vil bli samlet inn. Navn eller annen data som på noen måte kan identifisere barnet blir ikke samlet inn eller publisert. Deltagelse er derfor helt anonymt. Deltagelse vil bli gjennomført på avtalt sted og tidspunkt. Deltagelse kan også gjennomføres over internett ved hjelp av videosamtale, ved å bruke for eksempel Skype. Totalt vil deltagelse ta omtrent en halv time.

Hva skjer med informasjonen om barnet?

Alle innsamlede data vil bli behandlet konfidensielt og som forsker har jeg **taushetsplikt**. Det er kun meg og min veileder som vil ha tilgang til dataene og alt som samles inn vil bli anonymisert.

Prosjektet skal etter planen avsluttes 1. juni 2016.

Frivillig deltakelse

Det er helt frivillig å delta i studien, og du kan når som helst trekke ditt samtykke om å være med, uten å måtte oppgi noen grunn. Dersom barnet velger å trekke seg, vil alle opplysningene om barnet bli anonymisert. Det er også helt valgfritt å svare på spørsmålene som blir gitt.

Dersom dere har spørsmål til studien, vennligst kontakt Greger Siem Knudsen (e-post: greger.knudsen@student.uib.no) eller veileder Ankica Babic (e-post: ankica.babic@uib.no).

Studien er meldt inn til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

Samtykke til deltakelse i studien

Erklæring fra deltaker

Jeg har mottatt informasjon om studien, og bekrefter herved at mitt barn kan delta.

☐   *Jeg som foresatt har lest og forstått hva det innebærer at mitt barn deltar i studiet «Identifisering av dyskalkulisymptomer relatert til magnocellulær resonnement ved bruk av smarttelefoner». Jeg samtykker til at mitt barn kan delta og bekrefter at mitt barn deltar av fri vilje.*

_____

(Signatur av foresatt, dato)

Erklæring fra forsker

☐   *Jeg bekrefter å ha gitt den foresatte konsis og riktig informasjon, og har etter beste evne sørget for at den foresatte forstår hva deltakelsen innebærer. Jeg bekrefter at deltakeren har blitt gitt mulighet til å stille spørsmål om denne studien, og at alle spørsmålene har blitt riktig besvart av undertegnede. Jeg bekrefter at deltakeren selv har valgt å delta, samt at personen ikke har blitt overtalt til å gi sitt samtykke.*

_____

(Signatur av forsker, dato)

En kopi av dette samtykkeskjemaet vil blitt gitt til den foresatte.

Appendix B

Interview guides in Norwegian

B.1 Researchers and practitioners within the fields of dyscalculia and learning disabilities

Introduksjon

Hei og takk for at du ville delta på dette intervjuet. Før vi begynner tenkte jeg vi kunne raskt oppsummere det som stod i informasjonsskrivet og samtykkeskjemaet. Mitt navn er Greger Siem Knudsen og er mastergradsstudent ved Institutt for Informasjons- og Medievitenskap ved Universitet i Bergen. Kort oppsummert vil det i min masteroppgave bli utviklet en prototype på en programvare for å tidlig detektere symptomer på lærevansker, mer spesifikt matematiske lærevansker kjent som dyskalkuli, ved å inkludere en alternativ metode enn den som vanligvis brukes av programvare i dag. Programvaren vil ha et sett med funksjoner som inkluderer blant annet gjennomføring av tester, loggføring av resultater og visualisering av progresjonen til brukeren.

Du ble valgt ut til å delta i denne undersøkelsen grunnet din rolle her. Det er helt frivillig å delta i studien, du svarer kun på spørsmålene du selv ønsker og du kan når som helst trekke ditt samtykke uten å oppgi grunn. Dataen som samles inn vil bli anonymisert.

Spørsmål

1. Introduksjon
   (a) Hva er din stilling her?
   (b) Hvor lenge har du vært aktiv i denne stillingen?

2. Egen forskning/eget arbeid
   (a) Kan du fortelle litt om hva du forsker på/jobber med?
   (b) Er det noen forskningsfelt/arbeidsområder innen lærevansker, assistert læring, dysleksi og/eller dyskalkuli du er aktiv i eller har vært aktiv i tidligere?
      i. I så fall, kan du fortelle om din forskning/ditt arbeid rundt disse feltene?
      ii. Dersom din forskning/ditt arbeid har omhandlet dyskalkuli spesielt, kan du fortelle om din oppfattelse av denne diagnosen og om din forskning/ditt arbeid rundt denne?

    (c) Hvor langt har forskningen generelt kommet i prosessen av å forstå de nevrologiske prosessene som forårsaker lærevansker?

        i. Hva mener du er årsakene til dette?

3. Dyskalkuli

    (a) Hva mener du er årsakene til at noen personer sliter spesielt med å lære materiale som andre lærer på normal måte?

    (b) Hvilke mulige sammenhenger og likheter mener du det er det mellom dysleksi og dyskalkuli?

    (c) Har personer med lærevansker ofte andre psykologiske diagnoser?

        i. I så fall, hvilke?

        ii. Hva mener du kan være årsakene til dette?

    (d) Hva anser du som de største forskjellene mellom en person som er berørt av dyskalkuli og en person som bare har normale mattevansker?

    (e) Hva vil du si er prognosen til et barn i 10-12 års alderen som har sterke symptomer på lærevansker?

        i. Er denne prognosen uavhengig av type lærevanske?

        ii. Er denne prognosen uavhengig av andre mulige psykologiske tilstander hos den enkelte?

4. Programvare

    (a) Hva er dine holdninger til å benytte programvare til å diagnostisere lærevansker?

    (b) Har du erfaringer med programvare for å detektere symptomer på lærevansker?

        i. Omfatter dine erfaringer i hovedsak benyttelse av programvare til diagnostisering eller læring til dyskalkuliske personer?

        ii. Anser du det som nyttig og verdiskapende å benytte programvare til å detektere symptomer på dyskalkuli? Hvorfor, hvorfor ikke?

           A. Hva er dine etiske synspunkter rundt dette?

    (c) Er det noen funksjoner du etterspør/anser som nyttige i programvare for å assistere diagnostiseringen av dyskalkuli?

    (d) Mener du det er nyttig å loggføre progresjonen til regneferdighetene til dyskalkuliske elever elektronisk? Hvorfor/hvorfor ikke?

    (e) Hva er dine etiske synspunkter rundt dette?

5. Prototypen

    (a) Hva er ditt førsteinntrykk av beskrivelsen av min prototype?

    (b) Majoriteten av programvaren tilgjengelig for dyskaluliske personer i dag er ment til å bli benyttet i samråd med lærer eller annen veileder. Prototypen dette prosjektet utvikler derimot kan også benyttes alene uten ytre påvirkninger fra andre. Syntes du denne muligheten representerer en fordel eller ulempe sammenlignet med annen programvare? Hvorfor, hvorfor ikke?

(c) Prototypen min benytter metoder for å evaluere kontrastsensitivitet, spatial sensitivitet (evnen til å identifisere posisjoner av objekter i rommet), subitizing (evnen til å gjenkjenne mengder uten å telle elementer én og én), telling og grunnleggende aritmetikk for å identifisere mulige sammenhenger mellom dyskalkuli og hvordan informasjonsoverføringen i magnocellulære celler er hos den enkelte. Hensikten er å finne underliggende årsaker til lærevansken slik at den kan oppdages tidligere. Hva mener du om denne tilnærmingen? Mener du denne tilnærmingen skaper verdi?

B.2   Individuals having symptoms of dyscalculia

Introduksjon

Hei og takk for at du ville delta på dette intervjuet. Før vi begynner tenkte jeg vi kunne raskt oppsummere det som stod i informasjonsskrivet og samtykkeskjemaet. Mitt navn er Greger Siem Knudsen og er mastergradsstudent ved Institutt for Informasjons og Medievitenskap ved Universitet i Bergen. Kort oppsummert vil det i min masteroppgave bli utviklet en prototype på en programvare for å tidlig detektere symptomer på lærevansker, mer spesifikt matematiske lærevansker kjent som dyskalkuli, ved å inkludere en alternativ metode enn den som vanligvis brukes av programvare i dag. Programvaren vil ha et sett med funksjoner som inkluderer blant annet gjennomføring av tester, loggføring av resultater og visualisering av progresjonen til brukeren.

Du ble valgt ut til å delta i denne undersøkelsen grunnet din rolle som relevant bruker av programvaren. Det er helt frivillig å delta i studien, du svarer kun på spørsmålene du selv ønsker og du kan når som helst trekke ditt samtykke uten å oppgi grunn. Dataen som samles inn vil bli anonymisert.

Spørsmål

1. Introduksjon

    (a) Alder?

2. Egen undervisning

    (a) Hvordan opplevde du matematikkundervisningen du fikk i grunnskolen?

    (b) Hvordan løste matematikklæreren din på barneskolen situasjonen der elever hang etter i undervisningen?

    (c) Hvilke erfaringer har du med tilpasset undervisning i matematikk?

        i. Kan du fortelle litt om metoden(e) som ble benyttet i denne undervisningen?

        ii. Var det noen av metodene som ble brukt som fungerte godt for deg? I så fall, kan beskrive disse?

3. Dyskalkuli

    (a) Hvor gammel var du da du først opplevde at du hadde vanskeligheter med matematikk?

    (b) Opplever du at dine vanskeligheter med matematikk forbedrer seg eller forverrer seg ettersom du blir eldre? Hvorfor, hvorfor ikke?

    (c) Er det noen oppgaver som involverer tallprosessering du finner spesielt vanskelig?

        i. Kan du gi noen konkrete eksempler?

    (d) Føler du at dine matematikkvansker hemmer deg i hverdagen?

        i. Kan du beskrive noen konkrete eksempler?

    (e) Hva anser du som de viktigste forskjellene mellom en person som er berørt av dyskalkuli og en person som bare har normale vansker med matematikk?

    (f) Syntes du at dyskalkuli blir vektlagt i tilstrekkelig grad i skolen? Hvorfor, hvorfor ikke?

4. Programvare

   (a) Hva er dine holdninger til å benytte programvare til å detektere symptomer på lære-vansker?

   (b) Har du erfaringer med programvare for å detektere symptomer på lærevansker?

       i. Omfatter dine erfaringer i hovedsak benyttelse av programvare til diagnostisering eller læring til dyskalkuliske personer?

   (c) Anser du det som nyttig og verdiskapende å benytte programvare til å detektere symptomer på dyskalkuli?

   (d) Er det noen funksjoner du etterspør/anser som nyttige i programvare for å assistere diagnostiseringen av dyskalkuli?

   (e) Mener du det er nyttig å loggføre progresjonen til regneferdighetene til dyskalkuliske elever? Hvorfor, hvorfor ikke?

       i. Dersom du kunne benyttet min programvare prototype til å loggføre progresjonen din, ville du ha brukt denne funksjonen ofte? Hvorfor, hvorfor ikke?

5. Prototypen

   (a) Hva er ditt førsteinntrykk av beskrivelsen av min prototype?

   (b) Tenker du at du ville ha benyttet deg av denne programvaren?

       i. I hvilke situasjoner kunne du tenkt deg å benytte applikasjonen?
       ii. Hvordan kunne du tenkt deg å benytte applikasjonen?

   (c) Majoriteten av programvaren tilgjengelig for dyskalkuliske personer i dag er ment til å bli benyttet i samråd med lærer eller annen veileder. Prototypen dette prosjektet utvikler derimot kan også benyttes alene uten ytre påvirkninger fra andre. Syntes du denne muligheten representerer en fordel eller ulempe sammenlignet med annen programvare? Hvorfor, hvorfor ikke?

   (d) Applikasjonen er, på dette tidspunktet, under utvikling. Det er estimert at prototypen vil være klar i februar/mars 2016. Det vil da bli gjennomført brukertesting. Du representerer en aktuell kandidat for denne brukertestingen. Er denne brukertestingen noe du kunne vært interessert i å delta på?

Appendix C

Java™ listings

Listing C.1: Evaluation generation

```
 1  private void updateTask() throws IllegalAccessException,
        IllegalArgumentException, NoSuchFieldException, IOException{
 2          if(currTaskIndex < taskSet.size()){
 3                  currEvalData = taskSet.get(currTaskIndex);
 4                  selectedTasks[currTaskIndex] = Integer.parseInt(
                        currEvalData.getId());
 5                  updateTaskType();
 6                  taskStopwatch = new Stopwatch();
 7                  taskHeaderContainer.setText(Html.fromHtml(Utils.getBoldHTML
                        (TextResources.EVAL_TASK_HEADER_TEXT + (currTaskIndex
                        +1))));
 8                  taskHeaderContainer.append(TextResources.NEW_LINE);
 9                  taskHeaderContainer.append(Html.fromHtml(Utils.getBoldHTML(
                        "of " + taskSet.size())));
10                  imageContainer.setImageDrawable(Utils.getImageByFileName(
                        this,currEvalData.getFigureId()));
11                  taskTextContainer.setText(currEvalData.getTaskText());
12                  addResponseInput(currEvalData);
13                  currTaskIndex++;
14                  if(enableTextToSpeech){
15                          Utils.textToSpeech(taskTextContainer.getText().
                                toString(),this);
16                  }
17          }
18          else{
19                  if(updateRepeatingTasksFile){
20                          updateRepeatingTasksFile();
21                          Utils.showMessage(this, Utils.getTxtFileContent(
                                repeatingTasksFilepath), Toast.LENGTH_LONG);
22                  }               Double timeSpentInEval = HomeActivity.db.
                        getTimeSpentFromRepData();
23                  IntentInformation[] information = {new IntentInformation("
                        timeSpent",timeSpentInEval.toString())};
24                  Utils.switchActivity(this, ResultActivity.class,
                        information);
25          }
26  }
```

Listing C.2: Evaluation results database retrieval

```
1  public void populateResHistoryCardList(){
2          listView = (ListView) findViewById(R.id.card_listView);
3          cardArrayAdapter = new CardArrayAdapter(getApplicationContext(), R.
               layout.list_item_card);
4          for(int i = 0; i < HomeActivity.db.getTableSize(AppDatabase.
               RES_DATA_TABLE_NAME); i++) {
5                  String resultID = resHistoryCardListLabelsIDs[i];
6                  Result currResultData = HomeActivity.db.getResultRecordByID
                       (resultID);
7                  String currLabel = "Logged at " + resultID + "  |   Overall
                       performance: " +
8                                              Utils.getDoubleConversion(
                                                  currResultData.getOverallScore
                                                  ()) + "\%";
9                  Card card = new Card("Result " + (i+1),currLabel);
10                 cardArrayAdapter.add(card);
11                 listView.setAdapter(cardArrayAdapter);
12         }
13 }
```

Listing C.3: Evaluation results infographics generation

```java
public static void generateResultChart(Map<ClassificationParam,Double>
    results, HorizontalBarChart chart){
        ClassificationParam[] classificationParams = ClassificationParam.
            class.getEnumConstants();
        List<BarEntry> dataSet = new ArrayList<>();
        List<String> YAxisValues = new ArrayList<>();
        for(int i = classificationParams.length-1; i >= 0; i--){
                ClassificationParam currClassificationParam =
                    classificationParams[i];
                YAxisValues.add(Utils.classParamToString(
                    currClassificationParam));
                dataSet.add(new BarEntry(results.get(
                    currClassificationParam).floatValue(),(
                    classificationParams.length-1) - i));
        }
        BarDataSet barDataSet = new BarDataSet(dataSet, TextResources.
            RES_PERFORMANCE_LABEL);
        XAxis xAxis = chart.getXAxis();
        xAxis.setPosition(XAxisPosition.BOTTOM);
        xAxis.setDrawAxisLine(true);
        BarData data = new BarData(YAxisValues, barDataSet);
        chart.setData(data);
                chart.setDescription(TextResources.CLEAR);
        chart.animateY(1500);
                chart.setBackgroundColor(Color.TRANSPARENT);
                chart.setDrawGridBackground(false);
                chart.getXAxis().setDrawGridLines(false);
        barDataSet.setColor(Color.parseColor(TextResources.APP_THEME_COLOR)
            );
        chart.invalidate();
}
```

Appendix D

MagnoMath screenshots

Home screen
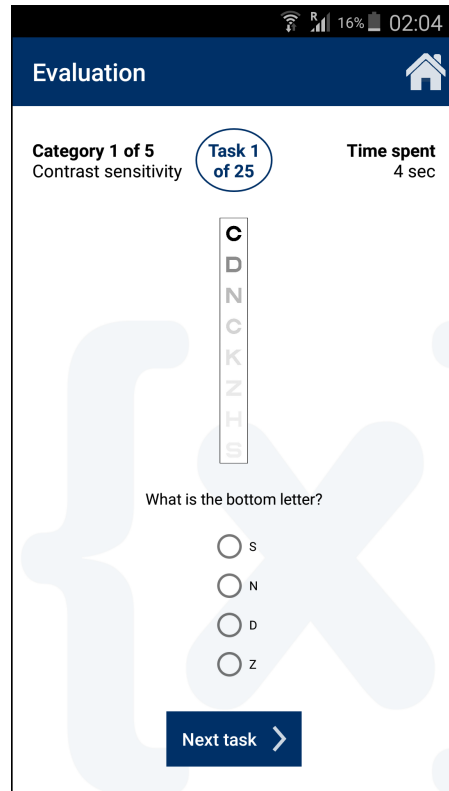
Screen evaluating contrast sensitivity
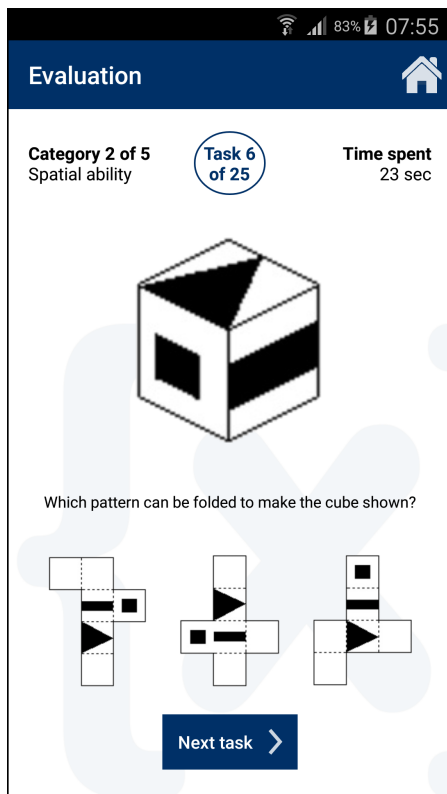
Screen evaluating spatial ability (1) [44]
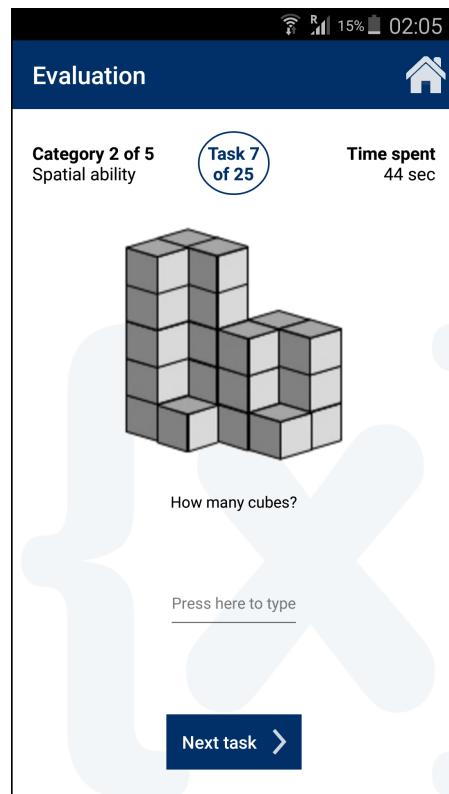
Screen evaluating spatial ability (2) [58]
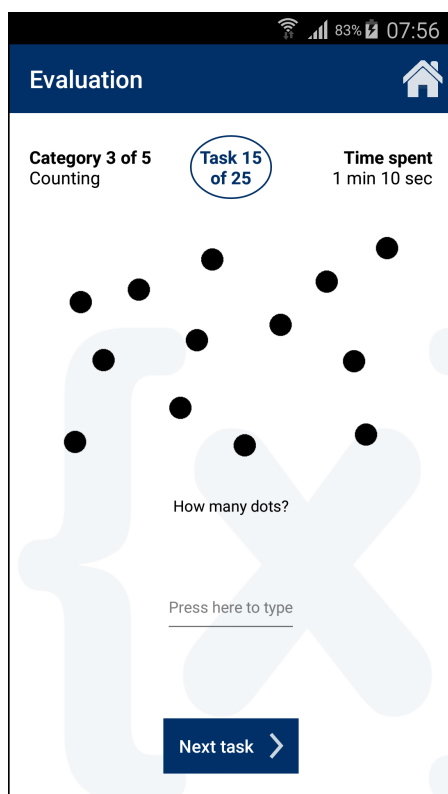
Screen evaluating counting skills
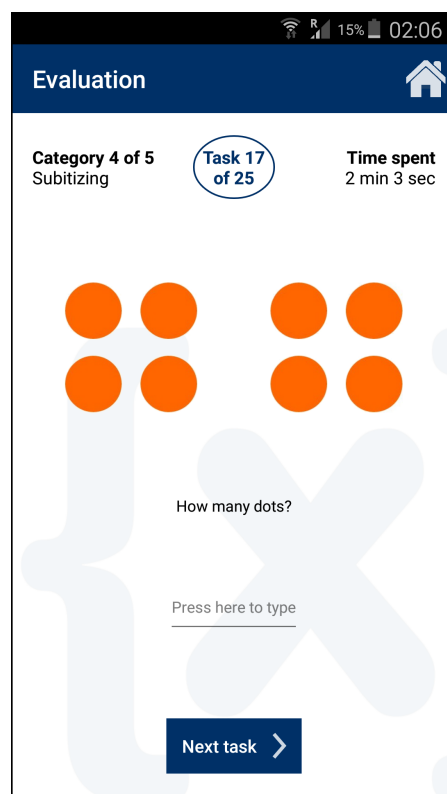
Screen evaluating subitizing skills (1)
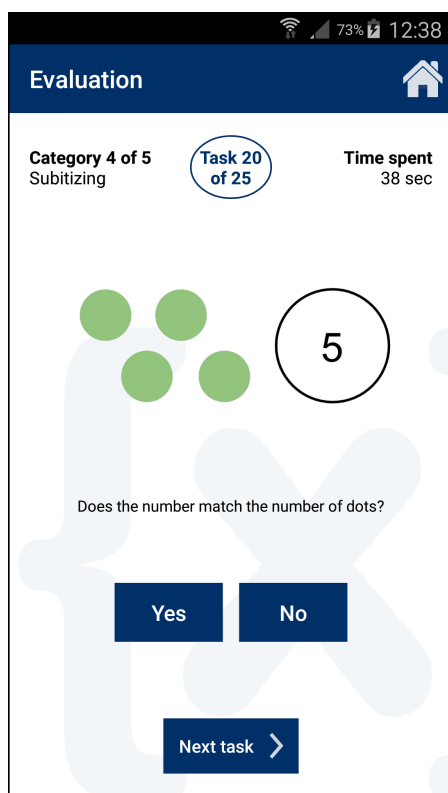
Screen evaluating subitizing skills (2)
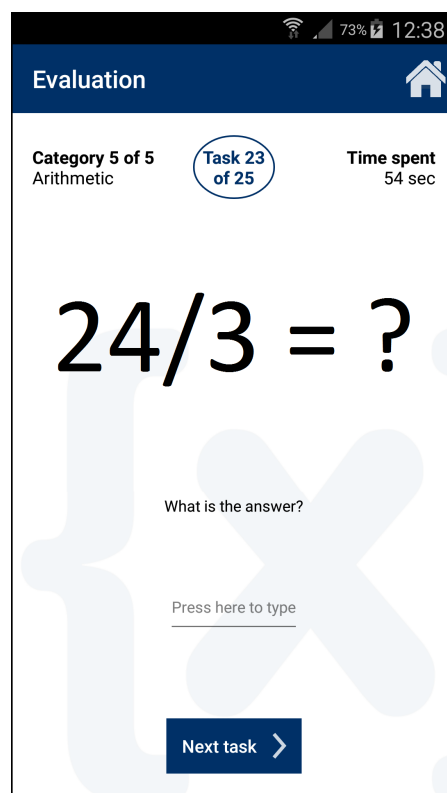
Screen evaluating arithmetic skills
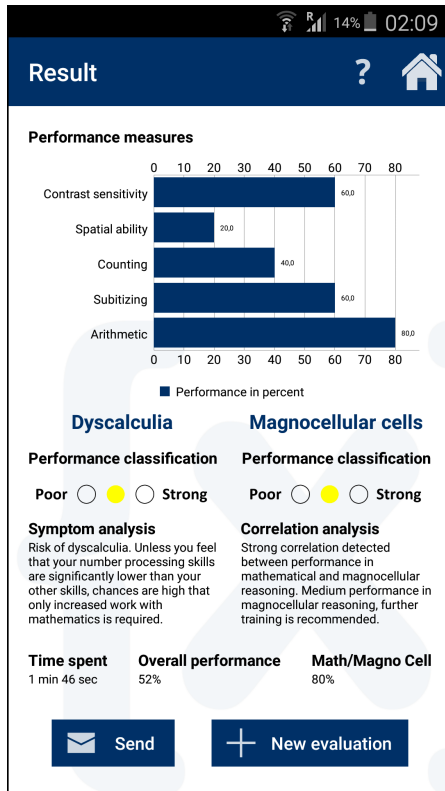
Result screen
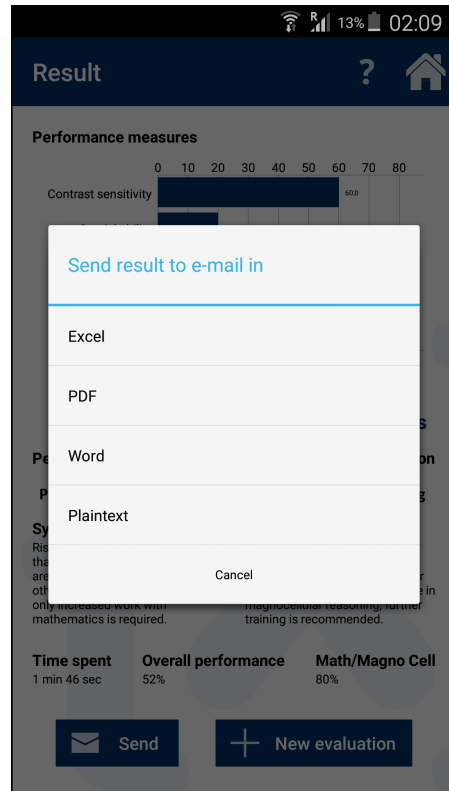
Send results by e-mail screen
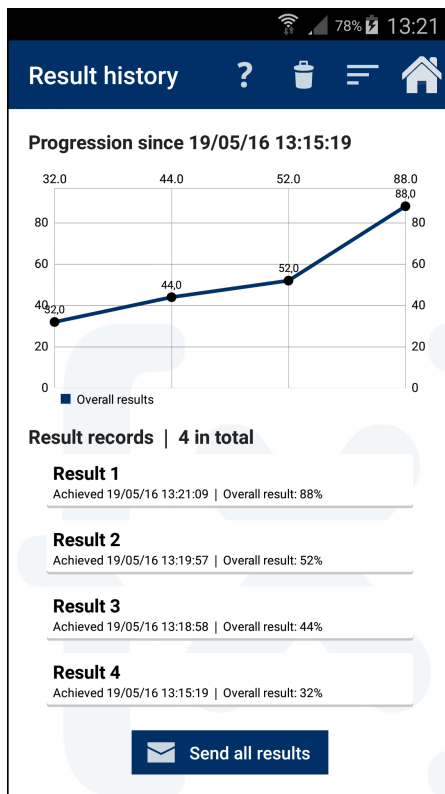
Result history screen
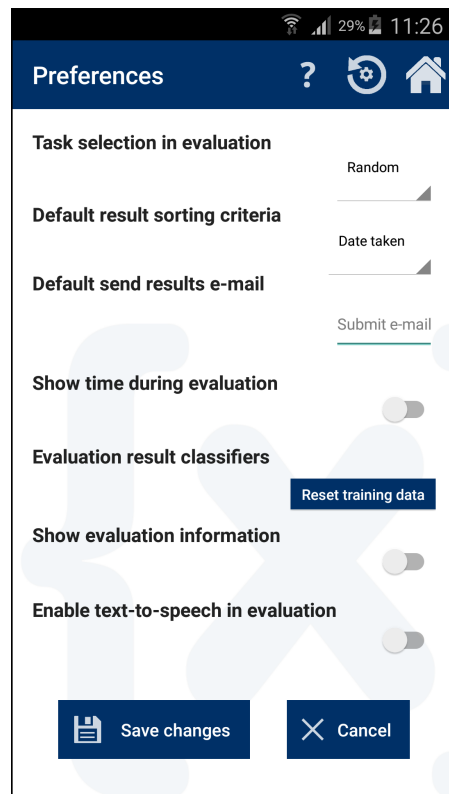
Preferences screen

Appendix E

Ethical approval from Norwegian Centre for Research Data

NSD

Ankica Babic
Institutt for informasjons- og medievitenskap Universitetet i Bergen
Fosswinckelsgate 6
5007 BERGEN

Vår dato: 19.04.2016   Vår ref: 45448 / 3 / HIT   Deres dato: Deres ref:

TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 01.11.2015. All nødvendig informasjon om prosjektet forelå i sin helhet 16.04.2016. Meldingen gjelder prosjektet:

45448	*Statistisk Klassifikasjon av Symptomer på Diskalkuli ved bruk av Smarttelefoner og Oppførselen til Magnocellulære Celler*
Behandlingsansvarlig	*Universitetet i Bergen, ved institusjonens øverste leder*
Daglig ansvarlig	*Ankica Babic*
Student	*Greger Siem Knudsen*

Etter gjennomgang av opplysninger gitt i meldeskjemaet og øvrig dokumentasjon, finner vi at prosjektet ikke medfører meldeplikt eller konsesjonsplikt etter personopplysningslovens §§ 31 og 33.

Dersom prosjektopplegget endres i forhold til de opplysninger som ligger til grunn for vår vurdering, skal prosjektet meldes på nytt. Endringsmeldinger gis via et eget skjema, http://www.nsd.uib.no/personvern/meldeplikt/skjema.html.

Vedlagt følger vår begrunnelse for hvorfor prosjektet ikke er meldepliktig.

Vennlig hilsen

Kjersti Haugstvedt

Hildur Thorarensen

Kontaktperson: Hildur Thorarensen tlf: 55 58 26 54
Vedlegg: Prosjektvurdering
Kopi: Greger Siem Knudsen greger_knudsen@hotmail.com

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.

Personvernombudet for forskning

Prosjektvurdering - Kommentar

Vi kan ikke se at det behandles personopplysninger med elektroniske hjelpemidler, eller at det opprettes manuelt personregister som inneholder sensitive personopplysninger. Prosjektet vil dermed ikke omfattes av meldeplikten etter personopplysningsloven.

Det ligger til grunn for vår vurdering at alle opplysninger som behandles elektronisk i forbindelse med prosjektet er anonyme.

Med anonyme opplysninger forstås opplysninger som ikke på noe vis kan identifisere enkeltpersoner i et datamateriale, verken:
- direkte via personentydige kjennetegn (som navn, personnummer, epostadresse el.)
- indirekte via kombinasjon av bakgrunnsvariabler (som bosted/institusjon, kjønn, alder osv.)
- via kode og koblingsnøkkel som viser til personopplysninger (f.eks. en navneliste)
- eller via gjenkjennelige ansikter e.l. på bilde eller videoopptak.

Personvernombudet legger videre til grunn at navn/samtykkeerklæringer ikke knyttes til sensitive opplysninger.

Appendix F

Related publications

F.1   Knudsen, 2016

Statistical Classification of Dyscalculia Symptoms and Behaviour of Magnocellular Cells

Greger Siem Knudsen

Presented at

European Federation for Medical Informatics Special Topic Conference

in

Paris, France
April 17−19, 2016

Published by

IOS Press

Indexed in

MEDLINE

F.1.1   Abstract

Statistical Classification of Dyscalculia Symptoms using Smartphones and Behaviour of Magnocellular Cells

Greger SIEM KNUDSEN[a,1]

[a] *Department of Information Science and Media Studies, University of Bergen, Norway*

Keywords. Dyscalculia, Classification, Diagnostics Assistance, Application

Having the sufficient expertise to assist pupils diagnosed with dyscalculia, a disability in learning arithmetic, can be challenging for teachers and parents. Therefore, with the ongoing technological development, several forms of computer software are developed for this group. However, the majority of this software is exclusively focused on learning mathematics, not assisting establishing diagnosis. The current project is developing a mobile application for smartphones for assisting diagnosing of dyscalculia, using the magnocellular theory of developmental dyslexia. Users will have constant access to self-evaluation and overview of their own progress.

Figure 1. Spatial ability evaluation task example

The application is designed to satisfy user needs by utilizing a set of key functionalities. On the user's side, these are represented as evaluating performance, applied in diagnosis classification using machine learning, and infographics generation of results. On the educator's side, results are received directly from the application through e-mail. The solution is implemented as a high fidelity prototype developed for the Android operating system, targeting the age group of 14-30 years. Figure 1 shows an example of a task evaluation design for spatial ability, one of the parameters for the symptom classifier. Contrast sensitivity, counting, subtilizing and arithmetic are examples of other evaluated skills. The data gathered will be stored and assessed automatically, providing a full overview of the progress.

Expected outcomes includes a greater, more accurate insight into the presence or absence of a correlation between information transmission in magnocellular cells and dyscalculia. For the users, the application is expected to increase access to evaluation and provide more direct feedback, which will not require visit at the educator's office. Educators can also receive evaluation results automatically. An advantage of this approach is improved validity of response data, which is not affected by any confounding factors, such as the educator's or parents' presence. User interface is simple and easy to use to accommodate all age groups.

---

[1] Corresponding Author.

F.1.2 Poster

F.2   Knudsen & Babic, 2016

Identifying Dyscalculia Symptoms Related to Magnocellular Reasoning using Smartphones

Greger Siem Knudsen & Ankica Babic

Will be presented at

International Conference on Informatics, Management and Technology in Healthcare

in

Athens, Greece
July 1−3, 2016

Will be published within July 3rd 2016 by

IOS Press

Will be indexed within July 3rd 2016 in

MEDLINE

F.2.1   Full paper

Identifying dyscalculia symptoms related to magnocellular reasoning using smartphones

Greger Siem KNUDSEN, M.Sc[a,b,1] and Ankica BABIC, PhD[a,b]

[a]*Department of Information Science and Media Studies, University of Bergen, Norway*
[b]*Department of Biomedical Engineering, Linköping University, Sweden*

Abstract. This paper presents a study that has developed a mobile software application for assisting diagnosis of learning disabilities in mathematics, called dyscalculia, and measuring correlations between dyscalculia symptoms and magnocellular reasoning. Usually, software aids for dyscalculic individuals are focused on both assisting diagnosis and teaching the material. The software developed in this study however maintains a specific focus on the former, and in the process attempts to capture alleged correlations between dyscalculia symptoms and possible underlying causes of the condition. Classification of symptoms is performed by *k*-Nearest Neighbor algorithm classifying five parameters evaluating user's skills, returning calculated performance in each category as well as correlation strength between detected symptoms and magnocellular reasoning abilities. Expert evaluations has found the application to be appropriate and productive for its intended purpose, proving that mobile software is a suitable and valuable tool for assisting dyscalculia diagnosis and identifying root causes of developing the condition.

Keywords. Dyscalculia, magnocellular cells, k-Nearest Neighbor, mobile application

Introduction

Software has been developed for the purpose of diagnosing and remediation of dyscalculia in individuals of all ages. However, the majority of this software focus on teaching mathematical principles. Further insight regarding the causes is seldom given. Research conducted at University of Oxford [1] has studied how information transmission behaves in individuals diagnosed with dyslexia, a learning disability in language reading and writing. This work has developed the magnocellular theory of developmental dyslexia claiming there are possible correlations between development of learning disabilities and disruptions in the transmission of information from the eyes to the visual system in the human brain. The neurons responsible for this transmission are called *magnocellular cells*. If we could utilize the data collected by software to not only score the performance, but also to compare the evaluated mathematical reasoning and reasoning performed by these cells, it will be easier to establish diagnosis and develop a personalized pedagogical support in timely manner.

---

[1] Corresponding Author: Greger Siem Knudsen, e-mail: greger_knudsen@hotmail.com

Eventually, this would help to obtain a greater understanding of the diagnosis. The purpose of this study was to develop a software that could achieve this purpose.

The following paragraphs will present a selection of available software for diagnosed individuals and/or individuals having symptoms of the condition.

Institute of Cognitive Neuroscience at University College London has developed *The Dyscalculia Screener* [2] to test elementary school children for dyscalculia using very simple number tasks, such as performing basic calculations and identifying small quantities.

The INSERM-CEA Cognitive Neuroimaging Unit, a leading research institute in mathematical cognition located in Paris, has developed *The Number Race* [3], a game-based approach for teaching mathematics. It is especially designed to address mathematical learning disabilities, achieved by presenting numbers in several representations, e.g. digits, words and sets of objects.

Dyscal [4] is a mobile application developed for the Android operation system, available in Dutch only. It provides interactive modules containing simple mathematical tasks of counting and identifying quantities.

The purpose of these software solutions is either to assist diagnosis by nearly exclusively scoring performance, or to apply alternative methods for teaching the material. None of them utilizes calculated results to identify underlying causes of the condition, which is the main purpose of the application presented in this paper.

Method

The software application, named *MagnoMath*, is developed for the Android operating system for smartphones, using the Java programming language and eXtensive Markup Language (XML). Firstly, MagnoMath evaluates mathematical skills with the intention of identifying and measuring dyscalculia symptoms. Selected tasks for evaluating these skills are divided into three groups: skills in (1) counting, (2) subitizing and (3) arithmetic. Subitizing is the process of recognizing patterns of objects and immediately determining the quantity without counting them one by one, a useful metric in revealing dyscalculic behavior.

Secondly, MagnoMath identifies and measures correlations between these symptoms and magnocellular reasoning. This was achieved by designing a set of tasks with the intension of activating this reasoning, by evaluating (1) contrast sensitivity and (2) spatial ability. The latter is the process of mentally manipulating three-dimensional figures.

The two groups of parameters were classified separately by k-Nearest Neighbor algorithm [5], using two classifiers applied on the data gathered from evaluation tasks. The key evaluation result is the computation of the correlation between dyscalculia symptoms and magnocellular reasoning. This computation applies mathematical reduction to directly compare the two evaluation task groups, which otherwise differs in metric space, enabling identification of dependencies.

The evaluation of the application included experts in mathematics and software engineering, as well as researchers and practitioners in learning disabilities. The usability of the application was evaluated by using Nielsen's usability heuristics [6] and Brooke's System Usability Scale [7].

Results and discussion

A prototype of the software application was finalized for the Android operating system. Figures 1 and 2 show the application's user interface during evaluation, testing spatial ability and counting skills, respectfully. Figure 3 presents the result screen after evaluation, showing the performance within each task category, as well as a computed correlation measure between mathematical skills and magnocellular reasoning ability.
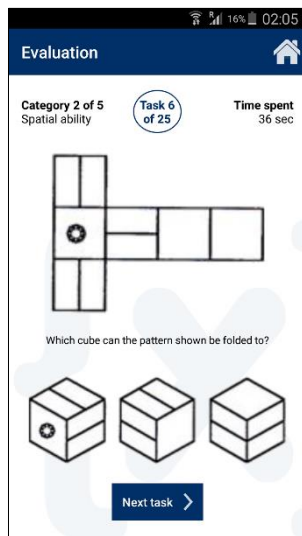

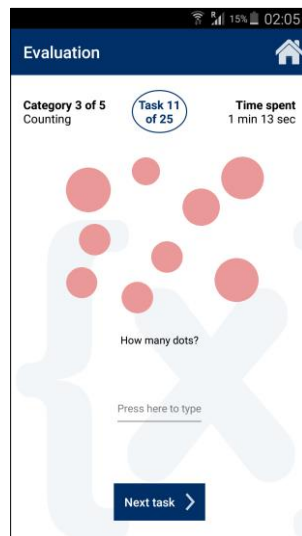**Figure 1:** Spatial ability task example (Screenshot).

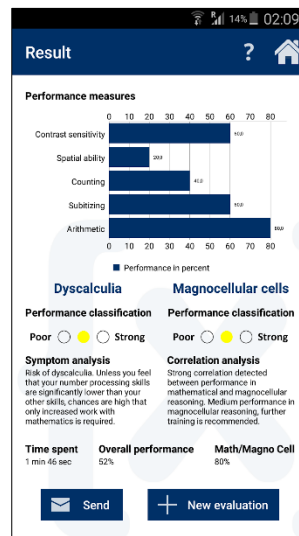**Figure 2:** Counting task example (Screenshot).

Figure 3: Result screen (Screenshot).

To test the application's validity and usability, two evaluations were conducted. Firstly, ten software and mathematics experts from the Department of Biomedical Engineering, Linköping University assessed the application's usability and behavior. Then, four learning disability experts from the Department of Behavioral Sciences and Learning, Linköping University, Sweden and Dyslexia Norway, Oslo, Norway evaluated the application's validity and accuracy.
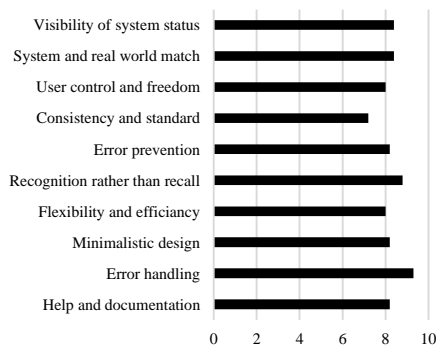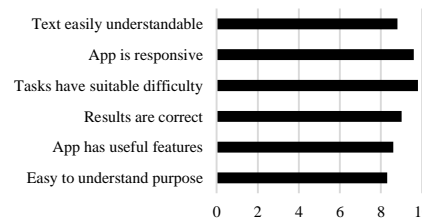

Figure 4: Usability test results.

Figure 5: Functionality test results.

Overall usability performance was satisfactory, as shown in Figure 4. The heuristic obtaining the highest score was the one regarding the application's ability to handle errors. Observational data also indicated that the application was stable and well-functioning, without creating a single operational error during evaluation. The heuristic obtaining the lowest score was the one regarding consistency and standards during evaluation. The cause of this was partly the usage of mathematical symbols as illustrations on some tasks which confused some users. These were later removed.

Figure 5 shows results obtained from evaluation testing the functionality of the application, again representing a satisfactory result. This especially applies to the prompt response times during evaluation, which is a sought quality. The characteristic obtaining the highest ranking was the one regarding the difficulty of evaluation tasks, indicating that the difficulty was appropriate; not too easy or too hard. The heuristic regarding the correctness of results also received a high ranking. This suggest that the k-NN classification algorithm performed well, showing a high degree of efficiency and accuracy.

Conclusion

This paper has presented a research focused on both evaluating software's ability to accurately assist diagnosing of dyscalculia and its ability to identify and measure correlations between the diagnosis and magnocellular reasoning. The mobile application developed in this study and its evaluation proves that using software applying k-Nearest Neighbor algorithm is suitable and useful for its purposes, achieving results of high accuracy and value. Future development will include evaluation conducted by users affected by dyscalculia, to compare the developed software's performance to existing software applications based on already established tests. Additionally, further work involves looking into the means of establishing a more precise correlation measurement between the physiological processes related to mathematical skills.

References

[1] J. Stein, The magnocellular theory of developmental dyslexia, *Dyslexia: An International Journal of Research and Practice* **7** (2001), 12-36.
[2] B. Butterworth, *The Dyscalculia Screener Manual*, nferNelson Publishing Company Limited, United Kingdom, 2003.
[3] A.J. Wilson, S.K. Revkin, D. Cohen, L. Cohen, S. Dehaene, An open trial assessment of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions* **2** (2006), 20.
[4] Dyscal. Available from: http://www.dyscal.org
[5] S. Russel & P. Norvig, *Artificial Intelligence - A Modern Approach*, Pearson Education, United States of America, 2010.
[6] J. Nielsen, 10 Usability Heuristics for User Interface Design (1995). Available from: https://www.nngroup.com/articles/ten-usability-heuristics/
[7] J. Brooke, SUS - A quick and dirty usability scale (1986). In P.W. Jordan, B.A. Thomas (editors), *Usability evaluation in industry*, Taylor & Francis Ltd, Great Britain, 1996.