

Jakob Westermoen & Martin Lunde

# SmartU: Investigating the effects of visualizations in adaptive self-assessment systems

Master's thesis in Informatics: Interaction Design, Game and Learning Technology

Supervisor: Michail Giannakos & Zacharoula Papamitsiou

June 2020





Jakob Westermoen & Martin Lunde

# **SmartU: Investigating the effects of visualizations in adaptive self-assessment systems**

Master's thesis in Informatics: Interaction Design, Game and Learning Technology

Supervisor: Michail Giannakos & Zacharoula Papamitsiou

June 2020

Norwegian University of Science and Technology

Faculty of Information Technology and Electrical Engineering

Department of Computer Science



Kunnskap for en bedre verden





## Abstract

The increased focus on digitizing learning in recent years has created new opportunities for how learning in digital environments is perceived. Learning Analytics Dashboards (LADs) is one of the areas receiving increased attention, and they mainly visualize information about a student's learning habits and personal learning history. This information is expected to increase the student's self-awareness and to support decision-making and self-regulation accordingly. How to build a LAD, how to support the underlying learning theory and how to present different types of information in an easy-to-understand and efficient manner, are all challenges that must be considered when designing and creating such a solution.

The present thesis investigates whether different types of visualized statistics can affect students' motivation, their performance in self-assessments and their overall learning outcomes. To address this objective, a new LAD called SmartU was developed and an experimental study was conducted. During the study, data was collected through questionnaires, interviews and usage-data from students' interactions in SmartU. Results from the data analysis indicate that several components of the system are considered useful by the users and that they contribute to increasing motivation and learning effort, but there was no statistically significant increase in performance.

## Sammendrag

Det økte fokuset på digitalisering av læring de siste årene har åpnet nye dører for hvordan man tenker om læring på digitale flater. Dashbord for læringsanalyse (Learning Analytics Dashboards, LADs) er et av områdene som har fått økt oppmerksomhet, og består i hovedsak av visualisert informasjon om en elevs lærevaner og personlige læringshistorikk. Denne informasjonen er forventet å øke selvbevissthet til eleven og som en støtte til beslutningstaking og selvregulering. Hvordan et dashbord for læringsanalyse skal bygges opp, hvordan det skal støtte underliggende læringsteori og på hvilken måte det skal vise frem forskjellige typer informasjon på en lettfattelig og effektiv måte, er alle utfordringer som må tas hensyn til når en slik løsning designes og lages.

Denne forskningen ser nærmere på om forskjellige typer visualisert statistikk kan påvirke elevers motivasjon, deres prestasjoner i selvevalueringer og deres overordnede læringsutbytte. For å se nærmere på dette ble det lagd et nytt dashbord for læringsanalyse kalt SmartU og utført en eksperimentel studie. Under studien ble data hentet inn gjennom en spørreundersøkelse, intervjuer og brukerdata fra elevenes interaksjon med SmartU. Resultatene fra data analysen indikerer at flere komponenter i systemet blir betraktet som nyttige av brukerne og at de bidrar til økt motivasjon og læringsinnsats, men ingen statistisk signifikant økning i prestasjon.

## Acknowledgments

This thesis was written for the Department of Computer Science (IDI) at the Norwegian University of Science and Technology (NTNU) from autumn 2019 to spring 2020.

We would like to specially thank our great supervisors, Michail Giannakos and Zacharoula Papamitsiou for their excellent guidance, support and feedback throughout our work on the thesis. We really appreciate the effort put onto helping us with everything, from bi-weekly meetings to discussions and to other practicalities of the research.

Last but not least, we want to thank our families and loved ones, as the thesis would not be the same without their support and motivation throughout the past year.

Jakob Westermoen & Martin Lunde  
Trondheim, May 28, 2020



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem Description . . . . .	1
1.2	Goals and Research Questions . . . . .	2
1.3	Research Design & Research Methods . . . . .	2
1.4	Contributions . . . . .	5
1.5	Thesis Structure . . . . .	6
<b>2</b>	<b>Background Theory and Related Work</b>	<b>7</b>
2.1	Definitions . . . . .	7
2.1.1	Learning Analytics . . . . .	7
2.1.2	Learning Analytics Dashboard . . . . .	8
2.1.3	Adaptive assessment system . . . . .	9
2.2	Literature Review . . . . .	9
2.2.1	Iteration 1: Databases search for articles . . . . .	9
2.2.2	Iteration 2: Generic Coding of the articles . . . . .	10
2.2.3	Iteration 3: Coding articles marked “ <i>maybe</i> ” . . . . .	10
2.2.4	Iteration 4: Full review of the relevant articles . . . . .	11
2.3	Previous Studies . . . . .	11
2.3.1	Customizable dashboards . . . . .	11
2.3.2	Features students expect from Learning Analytics . . . . .	12
2.3.3	Peer comparison . . . . .	13
2.3.4	Linking learning analytics and learning science . . . . .	14
2.3.5	Usage patterns of metacognitive feedback . . . . .	15
2.3.6	Clarity, context and mobility . . . . .	16
2.4	Insights Coming From the Literature . . . . .	17
<b>3</b>	<b>Design &amp; Implementation</b>	<b>19</b>
3.1	From LAERS to SmartU . . . . .	19
3.2	Functional Requirements . . . . .	19
3.3	Development Tools . . . . .	21
3.3.1	Figma . . . . .	21

3.3.2	Google Drive . . . . .	21
3.3.3	SharePoint . . . . .	21
3.3.4	Github . . . . .	22
3.3.5	Docker . . . . .	22
3.4	Iterative Design Process . . . . .	22
3.4.1	Iteration 1 . . . . .	22
3.4.2	Iteration 2 . . . . .	23
3.4.3	Iteration 3 . . . . .	24
3.4.4	Iteration 4 . . . . .	26
3.4.5	Final user interface . . . . .	29
3.5	The Current SmartU System . . . . .	34
3.5.1	Architecture . . . . .	34
3.5.2	Mobile implementation . . . . .	35
3.5.3	Challenges during development . . . . .	36
<b>4</b>	<b>Methods</b>	<b>37</b>
4.1	Study Design . . . . .	37
4.1.1	Participants . . . . .	37
4.1.2	Physical test-setup . . . . .	38
4.1.3	Procedure . . . . .	38
4.2	Data Collection . . . . .	40
4.2.1	Usage data collected from SmartU . . . . .	41
4.2.2	User-test & observation . . . . .	42
4.2.3	Questionnaire . . . . .	42
4.2.4	Interviews . . . . .	44
4.3	Data Analysis . . . . .	45
4.3.1	Qualitative analysis . . . . .	45
4.3.2	Quantitative analysis . . . . .	45
<b>5</b>	<b>Results</b>	<b>47</b>
5.1	Participants & Demographics . . . . .	47
5.2	Descriptives . . . . .	48
5.2.1	Questionnaire mean variables . . . . .	49
5.2.2	SmartU aggregated descriptives . . . . .	50
5.3	Correlations . . . . .	51
5.4	Independent Samples T-test . . . . .	53
5.5	System Usability Score . . . . .	55
5.6	Results From Interviews . . . . .	55
5.6.1	Participants insights towards motivation of use . . . . .	56
5.6.2	Perceptions towards intuitiveness of features . . . . .	58
5.6.3	Features perceived as extra useful . . . . .	62

5.6.4	Opinions on Learning-tools and gamification . . . . .	64
<b>6</b>	<b>Discussion</b>	<b>67</b>
6.1	Differences in Performance . . . . .	67
6.2	Usefulness of Visualizations . . . . .	69
6.3	Factors Affecting Attitude and Motivation . . . . .	70
6.4	Research Limitations . . . . .	72
6.4.1	Covid-19 and its implications . . . . .	72
6.4.2	Diversity in participants and test activities . . . . .	74
6.4.3	Inexperienced researchers . . . . .	74
<b>7</b>	<b>Conclusions &amp; Future Work</b>	<b>77</b>
	<b>Bibliography</b>	<b>79</b>
	<b>Appendices</b>	<b>85</b>
<b>A</b>	<b>Study</b>	<b>87</b>
A.1	Form of Consent . . . . .	88
A.2	Introduction To Adaptiveness and Task Related Analytics . . . . .	91
A.3	SmartU Post-Experiment Questionnaire . . . . .	93
A.4	Observation Guide . . . . .	110
A.5	Interview Guide . . . . .	113
A.6	Interviews . . . . .	115
<b>B</b>	<b>Design, Implementation &amp; Product</b>	<b>153</b>
B.1	User Tests . . . . .	153
B.2	Final User Interface . . . . .	161
B.3	Mobile User Interface . . . . .	172
<b>C</b>	<b>Results</b>	<b>177</b>
C.1	Questionnaire & Descriptives . . . . .	177
C.2	SUS-Scores . . . . .	183





# List of Figures

- 1.1 Oates’ research model, highlighted with the adopted research strategies . . . . . 3
- 1.2 Steps in the research method . . . . . 4
- 1.3 Data generated from the strategies / methods . . . . . 5
  
- 2.1 Growth of the search-terms LA, LAD and EDM acc. Google Trends 8
- 2.2 Stages of the literature review . . . . . 10
  
- 3.1 Iteration 1 - Pen & Paper draft . . . . . 23
- 3.2 Iteration 2 - High Fidelity Drafts . . . . . 23
- 3.3 Iteration 3 - Activity View . . . . . 25
- 3.4 Iteration 4 - Activity View . . . . . 26
- 3.5 Iteration 4 - Prototype Network . . . . . 28
- 3.6 Final user interface wireframes . . . . . 29
- 3.7 Color palette . . . . . 30
- 3.8 Login and user creation . . . . . 31
- 3.9 Dashboard interfaces . . . . . 32
- 3.10 Quiz interfaces . . . . . 33
- 3.11 Application architecture . . . . . 34
- 3.12 Mobile screen scaling . . . . . 35
  
- 4.1 Test-setup without person . . . . . 38
  
- 5.1 Gender and year of study of the participants . . . . . 48
- 5.2 Graphical representation of Independent Samples T-tests between the control and experimental group. The dots represent the mean value while the black bars represent the standard deviation. Statistical significance is not marked due to lack of significant results. . . 54



# List of Tables

2.1	List of tentative features . . . . .	18
4.1	Usage data collected from SmartU . . . . .	41
4.2	Examples of events/actions to note during observation . . . . .	42
4.3	Overall categories in the questionnaire and their constructs . . . . .	43
5.1	Sample of total participants . . . . .	47
5.2	Sample of participants in groups . . . . .	49
5.3	Questionnaire mean variables of categories . . . . .	49
5.4	Questionnaire mean variables between groups . . . . .	50
5.5	Descriptive of usage data for the groups . . . . .	50
5.6	Correlation table for questionnaire . . . . .	52
5.7	Independent Samples T-test for analyzing the effect of learners performance using statistics . . . . .	53
5.8	Independent Samples T-test for the difference of OEUS, AGV and UGV with and without statistics. . . . .	53
5.9	Average SUS-score of both groups for the variables in the SUS-schema	55
5.10	Nodes making up the "Motivation"-group, with number of files referenced and number of statements. . . . .	56
5.11	Nodes making up the "Intuitiveness"-group, with number of files referenced and number of statements. . . . .	59
5.12	Nodes making up the "Usefulness"-group, with number of files referenced and number of statements. . . . .	62
5.13	Nodes making up the "Learning-tools and gamification"-group, with number of files referenced and number of statements. . . . .	64



# Acronyms

**AGV** Attitude towards Graphs and Visualizations. xi, 44

**EDM** Educational Data Mining. ix, 8

**GVA** Graphs and visualizations in an assessment. 44

**LA** Learning Analytics. ix, 1

**LAD** Learning Analytics Dashboard. ix, 1

**LMS** Learning Management System. 7

**MVP** Minimum Viable Product. 20

**NSD** Norwegian Centre for Research Data. 43

**NTNU** Norwegian University of Science and Technology. iii

**OEUS** Overall evaluation of the usability of SmartU. xi, 43

**SPSS** Statistical Product and Service Solutions. 5

**SUS** System Usability Score. 43

**UGV** Usability of Graphs and Visualizations. xi, 44



# Chapter 1

## Introduction

This thesis describes the planning, development, and testing of the Learning Analytics Dashboard (LAD) for the Learning Analytics (LA) application SmartU.

Designing, developing, and evaluating a learning analytics dashboard is a challenging task, commonly raised in the field of learning analytics. Here, we describe the problem and motivation behind conducting the research, the goals and research questions, the methods used to reach these goals, followed by the contributions of this research. Lastly, an outline of how the thesis is organized is provided as an overview to facilitate the readability of the text.

### 1.1 Problem Description

Understanding how students learn and acquire knowledge is a field of research that, in a historical context, came together in the late 1800s to the mid-1900s under the name *learning theories* [Illeris, 2018]. Prosser and Trigwell [1999] investigated how to understand learning and teaching in higher education. Their research indicates that there exist better and worse ways for students to learn, and suggests to university teachers that “*by altering the learning context it may be possible to improve learning by encouraging that approach*” [Prosser and Trigwell, 1999, p. 3].

The emergence of new technologies and the availability in processable data have seen new research-areas such as LA and LADs appearing, trying to solve problems in new ways, previously thought impossible [Schumacher and Ifenthaler, 2018]. However, social sciences tend to slowly adopt new technologies, and, up-to this point, the use of visualizations in classroom settings is often characterized by uninterpretable and basic implementations, low interactivity, and little grounding in pedagogical practices [Vieira et al., 2018]. Few previous studies have looked at bringing visual LA into a classroom setting, thus, more insights of the long-term effect and affordances of LADs are needed for the technology to be adopted by

teachers and learners [Vieira et al., 2018; Schwendimann et al., 2016].

Previous research on providing feedback to learners using LAD visualizations, shows that learners perceive feedback differently based on their respective performance levels [Jivet et al., 2018; Papamitsiou et al., 2019]. To design and deliver visual feedback that is perceived as motivating and useful by all groups of students regardless of their performance levels, additional research is required. According to Papamitsiou et al. [2019], most of the students participating in their study were aware that they needed support (feedback) to deal with the problems at hand, but it was unclear to the students how to properly use the visual feedback they received when they asked for it.

The hypothesis for the present thesis is that visualized statistics can boost learners' motivation and thus improve their performance in self-assessments and enhance the learning outcome. However, to motivate the learner and be perceived as useful, the visualizations will have to be adapted to the learner's level of performance. This implies that students of different performance levels will perceive the usefulness of the provided visualized statistics differently.

The thesis will provide insights into how learners respond to/perceive the different visualizations, in terms of usefulness, how the visualizations affect their attitude, and how they affect their motivation. These differences will be explored with respect to the effect the visualizations have on learning performance.

## 1.2 Goals and Research Questions

With the hypothesis in mind, this research aims to investigate the effects the visualizations have on learners' motivation, attitudes and performance in a learning activity, when integrated into an adaptive learning system. To verify the hypothesis this study attempts to answer the following three questions:

**Research question 1** *What is the effect of different visualizations on learners' performance in the context of an adaptive assessment system?*

**Research question 2** *What is the learners' perceived usefulness of visualizations in an adaptive interface?*

**Research question 3** *What are the differences in learners' motivations/attitudes towards systems that provide visualizations?*

## 1.3 Research Design & Research Methods

In order to answer the research questions in Section 1.2 with *rigorous* and *relevant* research, appropriate research methods are necessary [Oates, 2005]. The selected



methods used in this research can be found in Figure 1.1, where Oates' research model has been used to outline the utilized methods.

The review of relevant literature was conducted to provide a pool of relevant articles, which the body of knowledge is made up of (Section 2.2). The research questions and the conceptual framework are based on the accumulated body of knowledge, as well as the previous experiences and motivations of the researchers of the thesis.

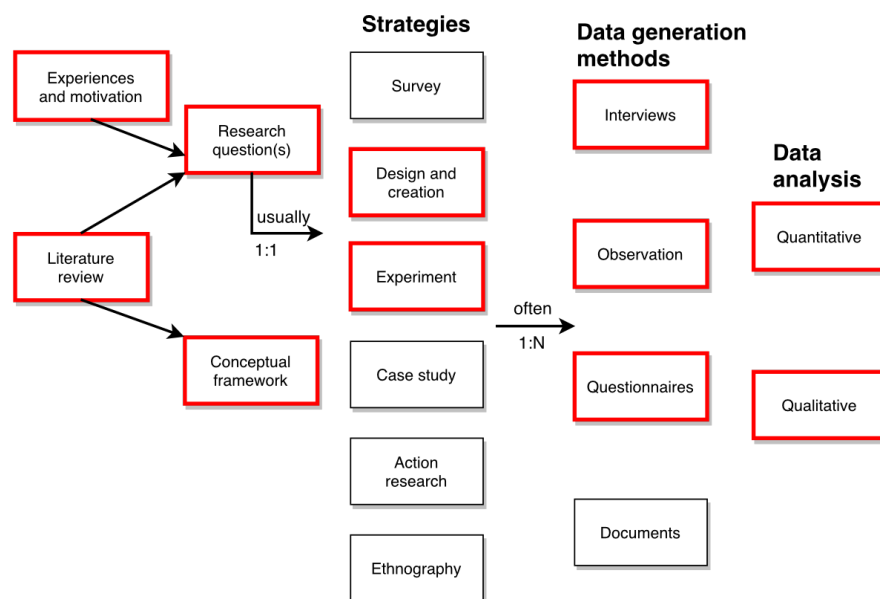


Figure 1.1: Oates' research model, highlighted with the adopted research strategies

### Research strategies

The research reported in this thesis used a mixed-methods design, utilizing a triangulation of research strategies, meaning that two or more strategies were adopted [Johnson et al., 2007]. Utilizing a triangulation of strategies provides multiple ways of responding to the research questions by bringing different perspectives together [Oates, 2005]. The strategies adopted in this thesis were the *Design and Creation* and the *Experiment*, as presented in Figure 1.2.

The *Design and Creation* strategy was used to create a software artifact corresponding to the definition of an instantiation, i.e., “a working system that demonstrates that constructs models, methods, ideas, genres or theories can be implemented in a computer-based system” [Oates, 2005, p. 108]. This strategy was applied by following the five-step plan created by Vaishnavi et al. [2004], which Oates describes in her book [Oates, 2005]. This plan consists of the steps awareness, suggestion, development, evaluation, and conclusion, and revolves around understanding the problem and gathering insight through literature and iteratively

designing, testing and evaluating the created artifact. The process of designing the artifact, i.e., the SmartU system, is covered in detail in Section 3.4 and the results of the development are presented in Section 3.5.

The *Experiment* strategy was used to perform the testing of the SmartU system in a controlled environment, while observing the participants using the system [Oates, 2005]. By conducting the experiment in a controlled environment as Oates suggests, variables could be controlled, allowing the research to focus on the relationship between the factors of the hypothesis. The experiment followed a static group comparison design, where the participants were divided into two random groups, with the groups being provided two different versions of the system. This allowed for measurements of the performance of the two groups and thus detecting the differences between them [Oates, 2005]. Information regarding the participants of the study and how they were recruited can be found in Section 4.1.1, while the description of the experiment itself can be found in Section 4.1.

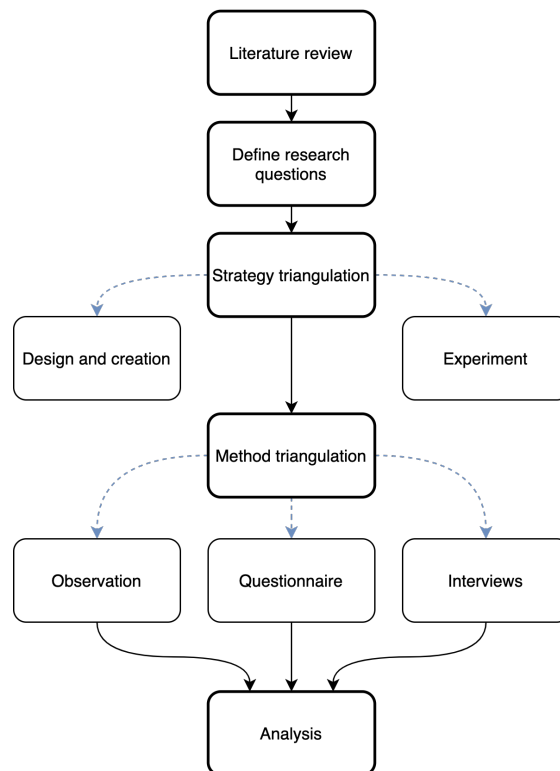


Figure 1.2: Steps in the research method

### Data generation methods

As the research followed a mixed-methods design, a triangulation of data generation methods was used. As shown in Figure 1.2, this triangulation consisted of

three different methods, generating both qualitative and quantitative data. *Observation* was used during the iterative design phase of the Design and Creation strategy and during the testing of the Experiment strategy. *Questionnaire* was used immediately after the testing in the Experiment strategy to collect data regarding the usability of the system (Design and Creation strategy) and to capture the participants' attitudes towards and perception of the usefulness of different features of the interface (Experiment strategy). *Interviews* were conducted three to four days after the testing, using a semi-structured interview approach, to probe for qualitative data not captured by the questionnaire. By utilizing several data generation methods it was possible to cross-check the findings from one method with the findings from the other methods, and thus, to validate the findings [Oates, 2005]. Figure 1.3 illustrates the different methods utilized in this research. Detailed information about the data collection process and how it was conducted can be found in Section 4.2.

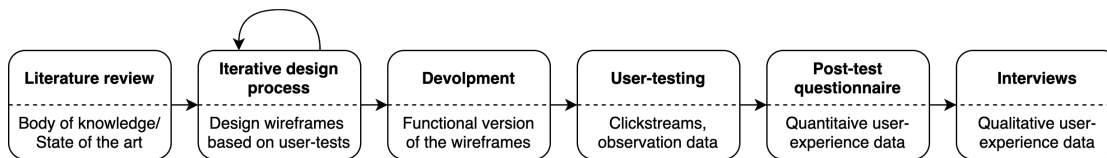


Figure 1.3: Data generated from the strategies / methods

### Data analysis

As the data generation triangulation provided both quantitative and qualitative data, the data analysis employed both qualitative and quantitative approaches. The interviews were transcribed from recording to text and iteratively coded into categories using a combination of a deductive and inductive approach [Oates, 2005]. For the quantitative data analysis, we used IBM Statistical Product and Service Solutions (SPSS). Thus, it was possible to detect correlations in the dataset, as the Person correlation is “*a measure of the strength and direction of association that exists between two variables measured on at least an interval scale*” [Laerd Statistics, 2020b]. Furthermore, IBM SPSS was used to test whether there were statistically significant differences between the groups participating in the study, using Independent Samples T-test to “*compare the means between two unrelated groups on the same continuous, dependent variable*” [Laerd Statistics, 2020a]. The qualitative and quantitative data analysis methods are presented in Section 4.3.

## 1.4 Contributions

This thesis contributes with a mixed-methods study that combines qualitative and quantitative empirical results, and with the evaluation of the usability of a new

interface, which highlights the strengths and weaknesses of the new system. This thesis also provides preliminary insights into how participants respond to different visualizations, in terms of perceived usefulness, how these visualizations affect their attitudes, and how they affect their motivation. These insights point out the features and aspects of the system that can potentially improve the learning outcomes and the effort put into the tasks, and open up new directions and discussions for future work.

## 1.5 Thesis Structure

The thesis contains seven chapters and is structured as followed:

- **Chapter 1: Introduction** – presents the problem description, the research goals and questions, the research methods used to address the questions, and the contributions of this thesis.
- **Chapter 2: Background Theory and Related Work** – provides definitions of relevant terms, describes the steps and findings of the literature review and provides a list of tentative features for a LAD based on literature.
- **Chapter 3: Design & Implementation** – describes the iterative design and the agile development processes followed for creating the interface of the SmartU system.
- **Chapter 4: Methods** – demonstrates the methods employed in the study, from the design of the controlled experiment, to the data collection and analysis methods.
- **Chapter 5: Results** – presents the results from the quantitative and qualitative data analysis.
- **Chapter 6: Discussion** – elaborates on the findings and their interconnections, and compares the results of this study to previous and related studies.
- **Chapter 7: Conclusions & Future Work** – presents the conclusions of this thesis and the main points of the discussion, also listing interesting areas for future work.

# Chapter 2

## Background Theory and Related Work

This chapter gives an overview of the existing research in the field of LA. This is done by looking closer at the previously conducted research and provides an insight into the motivation behind this thesis.

Section 2.1 Definitions provides definitions of important terms in the thesis, followed by Section 2.2 Literature Review, which describes how the current body of knowledge was established. Section 2.3 Previous Studies, contains background theory based on previous work from the articles retrieved from literature. Finally, Section 2.4 Insights Coming From the Literature, presents a list of proposed features for a LAD prototype, based on the findings from literature.

### 2.1 Definitions

#### 2.1.1 Learning Analytics

Learning Analytics (LA) is a relatively new field of research, established in 2011, exploring how to best collect, process and present data, in order to facilitate the needs of teaching staff and/or learners [Brown, 2012; Jivet et al., 2018]. LA is defined by Siemens and Gasevic [2012, p. 1] as “*the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs*”. According to Google Trends, the search-term Learning Analytics experienced a substantial growth after its establishment in 2011, as can be seen in Figure 2.1.

There are several ways to collect the data used in LA. Many schools and universities heavily utilize and rely on a Learning Management System (LMS) to enhance the distribution of information, assignments and learning material. Thus,

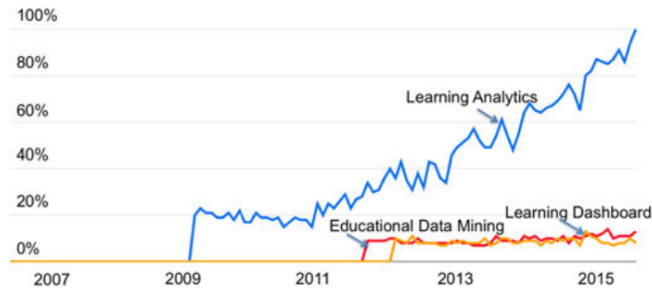


Figure 2.1: Growth of the search-terms LA, LAD and EDM acc. Google Trends

fetching personal information and collecting data from the usage of the LMS is one of the methods used for populating LA with data [Yoo et al., 2015]. Another common way of collecting data in LA is through an integrated, devoted assessment system, collecting data from conducted assessments and presenting the processed data within the same system [Schwendimann et al., 2016].

LA has benefited from the growth of big data and its many applications, as it has introduced new ways of collecting and analyzing data [Schumacher and Ifenthaler, 2018]. Meanwhile, the application of data mining on datasets made up of student-data, known as Educational Data Mining (EDM), provides new ways of discovering previously hidden information in the vast amount of accessible data. This has increased the interest in LA and the opportunities for analysis of big datasets of student data.

### 2.1.2 Learning Analytics Dashboard

The emergence of Learning Analytics has also caused the emergence of Learning Analytics Dashboard [Schwendimann et al., 2016]. With all the available student data, the different ways of presenting statistics and information have created a field of research on its own. Different approaches are being explored, such as textual feedback and visualizations in the form of e.g. graphs. Research conducted on LADs is often linked with other established research areas such as learning theory, design and how to provide feedback, making it a complex challenge to handle.

As defined by Bodily and Verbert [2017, p. 406], LADs should “*support users in collecting personal information about various aspects of their life, behavior habits, thoughts and interests. They also help users to improve self-knowledge by providing tools for the review and analysis of their personal history.*” In this thesis, the term LAD is used to describe a control panel supporting learner’s decision-making and self-regulation, by displaying processed and visualized data collected from all learners’ interactions with an integrated system.

### 2.1.3 Adaptive assessment system

Adaptive assessment systems, as a sub-category of computerized adaptive testing systems, became popular in the early 2000's according to Wainer et al. [2000]. In 2017, the National Council on Measurement in Education (NCME) defined adaptive testing as “*A computer-administered test in which the next item or set of items selected to be administered depends on the correctness of the test taker's responses to the most recent items administered*” [National Council on Measurement in Education (NCME), 2017]. This means that an adaptive assessment system will try to determine the ability level of the person conducting the assessment, and will provide items (i.e., questions) tailored to the given ability level.

## 2.2 Literature Review

Before any research is conducted in an academic setting, an understanding of the current body of knowledge should be established first [Oates, 2005]. This will ensure that the research will contribute positively to the community in terms of providing *new* and *relevant* knowledge, with respect to the State-of-the-Art. Thus, an initial literature scan was performed to review previous work in the field of visualizations and LADs. To locate and retrieve previous work and studies, the following scientific databases and search terms were used:

**Scientific databases used:** *ACM Digital Library, ScienceDirect, ERIC DB, SCOPUS, Google Scholar*

**Search terms used:** *Learning Analytics Dashboard, Learning Analytics Visualisation / Visualization, Educational dashboards, Educational data Visualisation / Visualization, Visual Analytics and interaction design, (Customised) UI for education, Dashboard UI and UX for education, Design for education*

The search terms were used in order to cover the research areas of visualizations and LADs catering to the needs of a learner in an educational context; they were also used to identify best practices for designing visualizations that are easy to understand, yet expressive and correct for the given context.

The review was conducted iteratively over four iterations, with each iteration narrowing down the scope, reducing the number of relevant articles, and increasing the overall relevance of the pool of remaining articles. The four different iterations are illustrated in Figure 2.2 and further elaborated on in the following subsections.

### 2.2.1 Iteration 1: Databases search for articles

The scientific databases were queried systematically using the above search terms. To reduce the number of articles provided by each query, advanced search settings

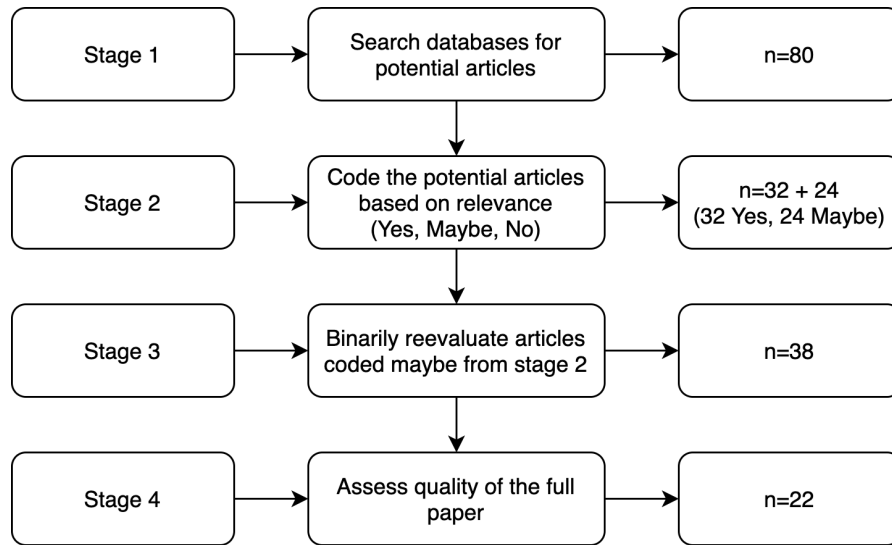


Figure 2.2: Stages of the literature review

were used, providing the option of using query strings and searching directly for keywords. The first 20 - 30 articles appearing when sorting by “most relevant” in each database were briefly scanned and put into a list of potential articles. For one of the databases, only 12 articles appeared when searching. After the first iteration of the search, the generated list contained 80 articles.

### 2.2.2 Iteration 2: Generic Coding of the articles

The list of articles from iteration 1 was the starting point in this iteration. Each article in the list went through a thorough reading, with the abstract and results/analysis being the most interesting sections. Based on the 11 criteria created by Dybå and Dingsøy [2008] to ensure inclusion of rigorous, credible and relevant papers, and considering the general topic and discussion in regard to the research to be conducted, the articles were coded with one of three codes: *yes*, *maybe* and *no*. As a result of the coding, 32 articles were marked *yes*, 24 articles were marked *maybe* and 24 articles were marked *no*.

### 2.2.3 Iteration 3: Coding articles marked “*maybe*”

The 24 articles marked *no* from iteration 2 were discarded from the list prior to iteration 3, thus only the 56 articles coded *yes* or *maybe* were included. The third iteration aimed at ending up with only relevant articles, thus the articles marked *maybe* had to be coded either *yes* or *no*. The articles went through more strict criteria than in the previous phases, and the potential end-users were taken into



consideration. Articles focusing on LA for teaching staff were coded *no*, as these articles were not focusing on learners' self-regulation. From the 24 articles marked *maybe*, 17 articles were coded *no* and 6 articles were coded *yes*.

### 2.2.4 Iteration 4: Full review of the relevant articles

Following the filtering of articles from the previous iterations, the final list included 38 relevant articles. These articles were read carefully and had their coding re-evaluated. The re-evaluation was based on the 11 criteria provided by Dybå and Dingsøy [2008], the scope of the research and the researchers' growing knowledge on the field of research. When articles in this iteration were coded *no*, an explanation was provided in order to create a base for comparison when meeting edge-cases. After the re-evaluation, the concluding list contained 22 relevant articles, within which six of the articles were literature reviews. The pool of articles following this phase made up the foundation the current research builds upon.

## 2.3 Previous Studies

This section introduces the relevant previous studies from the literature review within the field of LA, focusing on the visual parts of LADs and how to display information in the most effective way.

### 2.3.1 Customizable dashboards

In a study from 2017, Roberts et al. [2017] investigated students' opinions on LADs and what students would like to have incorporated in such a system. A qualitative study with focus groups consisting of 41 students was conducted. The focus groups were introduced to the topic of LA through two short videos. Then they went through 3 scenarios discussing their perceptions of dashboards and how they may or may not be useful. Other features the participants perceived as relevant for a dashboard were also discussed. As a part of the results, it was uncovered that a wide range of features was requested, and, as stated by Roberts et al. [2017, p. 324], "*students indicated that they would like to be able to customize their own dashboard to include only information that they perceived to be useful to them*". However, the analysis revealed that there was a common understanding between the students that some features were more important than others. The study culminated in a list, ranking features students perceived as important to facilitate a meaningful LAD. The following list summarizes the percentage of the participants' endorsement towards the different features from the study:

- 93.3% - A reminder of when upcoming assignments are due

- 88.6% - Links to additional resources
- 88.1% - My grades
- 85.5% - Program guidance (i.e., suggested readings)
- 81.4% - Early alerts about performance
- 80.3% - Support services and their contact details
- 80.3% - A calendar
- 72.0% - My attendance
- 71.5% - A link to provide feedback to teaching staff
- 71.0% - A color system that – at a glance – indicates if I am doing well or not
- 69.4% - My grades in comparison to my peers
- 63.7% - My class participation

### 2.3.2 Features students expect from Learning Analytics

Similarly to Roberts et al. [2017], Schumacher and Ifenthaler [2018] conducted a qualitative exploratory study with 20 university students, focusing on students' expectations towards features of LA systems, especially considering aspects of self-regulated learning. The results revealed that students find LA useful as an additional resource to, but not a replacement for, traditional learning. The option of customizing ones' own dashboard was welcomed, to make it as personally relevant as possible. It was also emphasized that the use of LA should be voluntary.

In addition to their qualitative study, a quantitative study with 216 participating students was conducted to complement the findings from the initial exploratory study. This complementary study consisted of a four-part online study, lasting a period of three weeks. The combined findings from the studies reveal that students expect a multitude of features from LA:

- Reminder function - The most requested feature was a reminder function, allowing one to set reminders for deadlines of assignments or enrollments. This feature would facilitate and contribute to the time management of activities.
- Revision of learning content - A feature for revising corresponding learning content from previous semesters was expected, as it would help students learn the content better and thus contribute to an increased learning outcome.

- Self-assessment prompts - Being able to receive relevant self-assessment questions with just-in-time feedback enables the students to review their current level of knowledge. It also enables the students to make alterations and plan their future learning activities. Thus, a self-assessment prompt would contribute to the self-monitoring of knowledge and learning processes.
- Receiving feedback - Receiving feedback for assignments taken in the online learning environment was also seen as important for the students. This would allow students to compare the feedback from the assignments to external feedback or the just-in-time feedback provided by the self-assessment prompts.
- Adaptive recommendations - Students expected a feature providing personal recommendations for learning activities or content based on individual performance. This feature would help students successfully complete their courses by providing a way of modifying learning activities to reach learning objectives.

### 2.3.3 Peer comparison

In a systematic literature review, Jivet et al. [2018, p. 32] wanted to “*explore how the educational concepts and learning theories are used through the whole development cycle of the dashboard, focusing on how they are integrated into the evaluation of learning dashboards*”. The review revealed that previous studies on dashboards trying to target self-regulation, emotional and/or cognitive competences, have not been able to evaluate all aspects of a LAD.

During the analysis of the results, a connection between different reference frames and students’ motivation was discovered. The connection showed that different levels of academic performance affected students’ perceptions of dashboards and specifically the social comparisons. Jivet et al. reported that there was a correlation between a student’s level of academic performance and how the same student perceived the comparison with peers. Based on these differences, it was possible to divide the students into performance-groups (i.e., high-, above-average-, under-average- and low-performers) based on how they were affected differently, in terms of motivation, by the comparisons shown in the dashboards. While high-performers, who were already highly motivated, had a split opinion on peer comparison and would rather be able to set their own goals, the under-average performers appreciated the peer comparison. Low-performers on the other hand, expressed the same opinions as the high-performers, as they would not be motivated by the comparison, but rather demoralized by the knowledge of being on the bottom of their class. The last group, i.e., the above-average performers, had

an unexpected behavior of being quite relaxed when realizing they were slightly above the class average, thus losing their motivation to increase their performance towards better results. One of the outcomes of the research was a set of recommendations regarding the design of LADs for learners, including the following:

- D1 - LA dashboards should be designed as pedagogical tools that enhance awareness and reflection to catalyze changes in the cognitive, behavioral and emotional competences.
- D2 - Educational concepts from learning sciences should be used to motivate design decisions.
- D3 - Comparison with peers should be used cautiously.
- D4 - It should not be assumed that the dashboard will have the same effect on all users. Instead, it should be determined which group of learners benefit the most and how to customize the dashboard to provide the same support to all users.
- D5 - The dashboard should be seamlessly integrated into the online learning environment and into the usual learning activities of the learner.

### 2.3.4 Linking learning analytics and learning science

In another study, Sedrakyan et al. [2018a] looked closer at how to link LA with learning science, much like what Jivet et al. [2018] proposed in D2 in the set of recommendations. Sedrakyan et al. [2018a] stated that a common denominator between all LADs containing feedback, is the lack of grounding in learning sciences and research on feedback and its underlying mechanisms. The study provides preliminary answers to which concepts are needed to allow LAD feedback to have theoretical support grounded in learning sciences, and what is needed to enable the capture of data that will allow measurement of the concepts in a learning process.

According to Sedrakyan et al. [2018a], the learning process can be positively influenced by dashboard feedback, but only if it considers the underlying mechanics of learning processes. For instance, it could be based on a learner's regulatory processes to detect inefficient processes in learning. Two examples of feedback-types to provide include (a) cognitive feedback to support regulation of learning in line with task-specific goals, which can improve immediate learning outcomes, and (b) behavioral feedback that could inform the learner of the need for a behavioral change, by improving the learner's awareness of their learning progress. In addition to this, Sedrakyan et al. [2018a] support that learning goals should be considered in virtual-learning environments such as LADs, in order to expand the scope of

feedback to include mastery orientation along with performance orientation. By combining such feedback with the effectiveness and efficiency of learning processes, Sedrakyan et al. [2018a] claim that it would be possible to better track learning progress and even detect ineffective or inefficient processes during learning.

### 2.3.5 Usage patterns of metacognitive feedback

Papamitsiou et al. [2019] performed a group study aiming to “*broaden current knowledge on the patterns of on-demand metacognitive feedback usage, with respect to learners’ performance*” [Papamitsiou et al., 2019, p. 1]. This was done by dividing the 174 participating students in two equally sized groups, with one group being the experimental group and the other being the control group. The experimental group had access to task-related visual analytics during the learning tasks, showing statistics about the collective previous performance on the specific tasks. The data collected from the study was time-spent on viewing visual analytics, frequency of visual analytics request and learning performance. In order to find the effect that the visual analytics had on learning performance, independent t-test samples were conducted between the two participating groups.

The results revealed that there were statistically significant differences between the groups in terms of performance, and the task-related visual analytics had an effect on the learning performance of the individuals in the experimental group. However, it was also revealed that there were significant differences in the patterns of usage of the visual analytics within the experimental group based on the students’ performance on the tasks. Similarly to Jivet et al. [2018], the experimental group in this study could also be divided into performance-groups based on how the visual analytics was used during the tasks. Papamitsiou et al. [2019] provide an example of this saying that “*as the difficulty of the tasks increases, low-performers tend to gradually use less the metacognitive feedback, both in terms of the average requests for on-demand metacognitive information and of the average time allocated to view and study this information*” [Papamitsiou et al., 2019, p. 8].

The general trend found in the results was that most students start the first task by asking for visual analytics. Next, the experimental group was split into three categories of performance: high, mid and low. High-performers mostly accessed the visual analytics on the hard tasks, while low-performers were sequentially not accessing it on any task; mid-performers had steady access rate on the visual analytics, as it is used on most of the tasks, regardless of their difficulty. However, the time mid-performers spent processing the information was low. Thus, Papamitsiou et al. imply that most of the participating students were aware that they needed support, as they utilized the available visual analytics. However, it was unclear for the students what actions they should take after using it.

### 2.3.6 Clarity, context and mobility

Klein et al. [2019] investigated how undergraduates make use of LADs by conducting a case study of focus groups with 81 participants. Even though students found LADs to be useful in general, it was stated that “*while interventions have the potential to be useful, often issues of relevancy, accuracy, clarity, and context limit their ability to make meaning of and respond to those interventions*” [Klein et al., 2019, p. 139]. Students expressed that having “everything in one place” was a desire, as they would like to have all academic records in the LAD, which would facilitate easier decision-making. It was noted that the lack of alignment to student needs, made it difficult to benefit and access the platforms, as the LADs often were not functioning on the devices preferred by the users and tailored to their usage, being frustrating for students and making them lose interest.

In addition to making sense of the overall impression and perceptions of students using LADs, specific components of LADs were also investigated. LADs mostly consist of data visualizations and feedback, which makes clarity and context important aspects for students to be able to extract meaningful information from dashboards. For instance, the traffic light component, featured in several LADs, is most commonly used to visualize the prediction of a student’s performance or ability to complete a course. However, according to Klein et al. [2019] not all students perceive the traffic light as the best way to do so. “*While many students liked the idea of receiving information on their progress, there were mixed reactions to the uses of color signals, alone, to indicate their current status*” [Klein et al., 2019, p. 142]. Most students liked the idea of easily understanding visualizations of their performance, but they would also like to have some accompanying information to inform how those visualizations have been aggregated. This was highlighted by Klein et al. [2019, p. 144] stating that “*without context students felt that LAD data were being provided in a vacuum or were irrelevant to their perspectives, which disconnected them from understanding how that data connected to their course performance and degree progress*”.

## 2.4 Insights Coming From the Literature

The definitions provided in Section 2.1 are important in order to understand the state-of-the-art and the previous studies described in Section 2.3. As described in Section 2.2, the state-of-the-art is established through a literature review [Dybå and Dingsøy, 2008], and consists of several important aspects regarding LADs.

Roberts et al. [2017] and Schumacher and Ifenthaler [2018] conducted studies regarding students' opinions about LAD-features, and which features students expect to find in a LAD. Both studies resulted in similar lists with suggested features and concepts. Among those, the features that are considered the most relevant for the research to be conducted in this thesis are the “early performance detection” and the “comparison to peers”, especially valued by the students [Roberts et al., 2017]. The “early performance detection” feature was found also in Schumacher and Ifenthaler [2018]'s list, along with a feature providing support for learning activities or content based on individual performance.

Jivet et al. [2018] conducted a systematic literature review that resulted in a set of recommendations regarding the design of a LAD. Similar to what Schumacher and Ifenthaler [2018] found, some of those recommendations are based around the individual learner's performance. Jivet et al. [2018] recommend adapting the dashboard to fit and support all groups of learners, e.g., based on their performance level. It is also noted explicitly that peer comparison should be used cautiously as it can have a different impact on different learner-groups.

Sedrakyan et al. [2018a] and Papamitsiou et al. [2019] conducted studies that investigated some of the recommendations proposed by Jivet et al. [2018]. Sedrakyan et al. [2018a] looked at how to link LA with learning science. By designing dashboards with their features grounded in learning science, both the learning outcome and self-awareness of the learner could be increased. Papamitsiou et al. [2019] however, studied if and how students in different performance-level groups used the task-related visual analytics. This study revealed that the students are aware that they need support from the visual analytics, but are unsure of how to use it properly after accessing it and what action to perform next.

Based on these findings, a brainstorming session was held regarding the features to be included in the LAD that was to be designed. The brainstorming session culminated in a list of tentative features, summarized in Table 2.1.

Feature	Article
A color system that indicates if I am doing well at a glance Early alerts about performance My grades in comparison to my peers	Roberts et al. [2017]
Student grades (avg results) Problem Time Distribution (avg. time used on a problem)	Ruipérez-Valiente et al. [2017]
Line Chart to compare actual and required performance	Sedrakyan et al. [2018b]
Motivation by gamification Indicator of correct progress in the platform Time spent on different activities compared to the class avg	Ruipérez-Valiente et al. [2015]
Quiz scores over time (Promotes reflection) Compare to class functionality Game-like design, to be more relatable to students	Bodily et al. [2018]
Comparison to similarly performing peers Adaptiveness in showing stats to avoid demotivating	Jivet et al. [2018]
Adaptivity in tasks based on performance just-in-time feedback after finishing an attempt Chance to revise previous content	Schumacher and Ifenthaler [2018]

Table 2.1: List of tentative features



# Chapter 3

## Design & Implementation

This chapter gives an overview of the process of designing and developing the new SmartU interface according to the Design and Creation strategy. This is done by introducing the previous version of the SmartU system, i.e., LAERS, in Section 3.1 From LAERS to SmartU, and next describing the requirements for designing the new interface in Section 3.2 Functional Requirements. Section 3.3 Development Tools, describes the tools used in the project, followed by Section 3.4 Iterative Design Process, describing the iterative process of designing the user interface. Lastly, Section 3.5 The Current SmartU System, gives an introduction to the current SmartU system, by looking into the architecture, the mobile implementation and the challenges raised during the development process.

### 3.1 From LAERS to SmartU

SmartU is an abbreviation for *Self-assessment Measured with Analytics on Run-Time for YOU*, with its name referring to the system's own main purpose. An older version of SmartU, called LAERS [Papamitsiou and Economides, 2013], was used in previous research, i.e., there exist previous implementations of SmartU, on which this project builds upon. The basic method used in the creation of the new interface of SmartU was the Design and Creation strategy, defined in Section 1.3 Research Design & Research Methods. However, the steps in the methods were limited by the functionalities provided by the existing backend of SmartU.

### 3.2 Functional Requirements

Glinz [2007] provides multiple existing definitions of the term *Functional requirements* in order to separate it from the non-functional requirements. He splits the definitions into two categories, one with emphasis on functions and the other with

an emphasis on behavior. The functional category is about specifying “a function that a system or system component shall perform” [ISO 24765:2017(E), 2017], or in other words “what the product must do” [Robertson and Robertson, 2012, p. 10]. The behavioral category specifies how the system should behave by having specifications of inputs and outputs from the system, and how the relationship between them should be.

Non-functional requirements are a different type of requirements that focus on performance characteristics of the system, rather than its features [Glinz, 2007]. However, in the current project, as this prototype was used in a controlled environment, non-functional requirements were not considered in the design phase.

To separate the features needed for a Minimum Viable Product (MVP) from the features that would expand the MVP or the features that simply would be nice to have, three different categories were used. By using the keywords specified in RFC 2119 by Bradner [1997], a clear and precise separation is achieved. Thus, the functional requirements were split into requirements that *must*, *should* or *may* be implemented in a complete prototype. The following requirements were based on designing a user interface, drawing inspiration from the features in Table 2.1:

*Requirements that **must** be implemented:*

- User must be able to take quizzes in activities
- Activities must provide visualized statistics of the student’s progression over time
- Visualized aggregated statics on individual questions must be available to the user during a quiz
- Creation of user-accounts must result in unique users
- Logging in must enable a session which must be ended on log out
- Must be aesthetically pleasing and user-friendly on a standard desktop setup (e.g using Google Chrome on a 1920x1080 display)

*Requirements that **should** be implemented:*

- Be supported equally on mobile and desktop
- Feature multiple connected visualizations or graphs displaying statistics from previous attempts in a quiz
- A user should get an indication of their skill level for an activity and quiz

- Provide opportunity to preview ones' last attempted activities in the system
- Support multiple available activities (e.g., for different topics within a course)

*Requirements that **may** be implemented:*

- Have badges/achievements
- Login via Feide for easier user-creation
- Support multiple available courses (e.g., for different subjects)
- Feature for submitting bugs and errors on the page

### 3.3 Development Tools

To design and create a functional interface for the adaptive assessment system, it was important to choose the correct tools. This section presents the tools used in the design and implementation phases.

#### 3.3.1 Figma

Figma [2019] is a design and collaboration tool, enabling teams to work together in real-time on the same files, designing and creating wireframes for a project. This tool was especially important in order to create a functioning prototype and enabled us to efficiently progress within the design phase by allowing concurrent user access in the same file.

#### 3.3.2 Google Drive

Google [2019] is a cloud storage service, which provided a convenient way of storing, collaborating on and sharing files within the project. By using the cloud, efficiency and performance were improved due to increased accessibility and collaboration. Note that no files containing personal data were stored on Google Drive.

#### 3.3.3 SharePoint

SharePoint Microsoft [2020], which is a part of the Microsoft Office 365 package, provided safe and secure cloud storage of sensitive data and test results of the study. The data agreement between Microsoft and NTNU ensured safe accessible cloud storage, thus enabling greater efficiency, accessibility, collaboration and performance when processing the data.

### 3.3.4 Github

Github [2019] was used as source control in the project to manage source code and the development process across the developers' clients. Usage improved efficiency, redundancy and programming performance of the application.

### 3.3.5 Docker

Docker [2020] was an essential tool for the development and deployment of the project due to its convenient standardization of creating a host environment (container), ready to run, on any machine. Docker improved efficiency, performance, accessibility and the overall workflow of programming and testing the application during development.

## 3.4 Iterative Design Process

To create an application facilitating users' self-assessment and learning outcome, the application had to be user friendly and meet users' needs. With the features from Table 2.1, the requirements from Section 3.2 Functional Requirements and the existing backend limitations in mind, an iterative design process was conducted, starting from a low-fidelity paper ideation, developing into a high-fidelity interactive prototype and a final user interface.

### 3.4.1 Iteration 1

As a starting point, the initial low-fidelity paper ideation was created using only pen and paper. Those sketches were very basic compared to the final prototype, but were useful for determining how the user would interact with the interface, as they allowed for easy and rapid changes to the design. As the sketches in this phase of the design were of a low-fidelity standard, the components were only rough outlines of finished components. Thus, the components were easily changeable as none of the designers were heavily invested in the sketches.

As seen in Figure 3.1, the first iteration contained selected features from Table 2.1 and the requirements-list in Section 3.2. These features included "Announcements", "Available Quizzes", "Completed Quizzes", a "Progress Chart" and "Recommended Readings". As the design was created early in the process, all requirements of the application were not yet known. Thus, inspiration to the design was collected from the old version of SmartU, the list of features from the literature and the functional requirements present at that time.

Due to the time constraints of the project, no user-tests were performed in this iteration. Instead, the researchers had an analysis session going through the

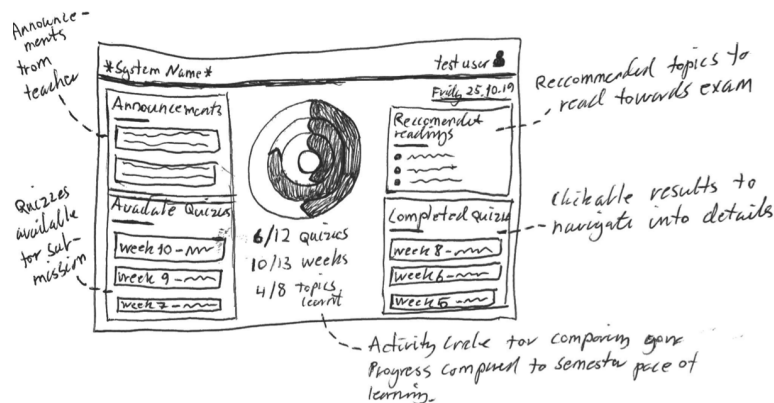
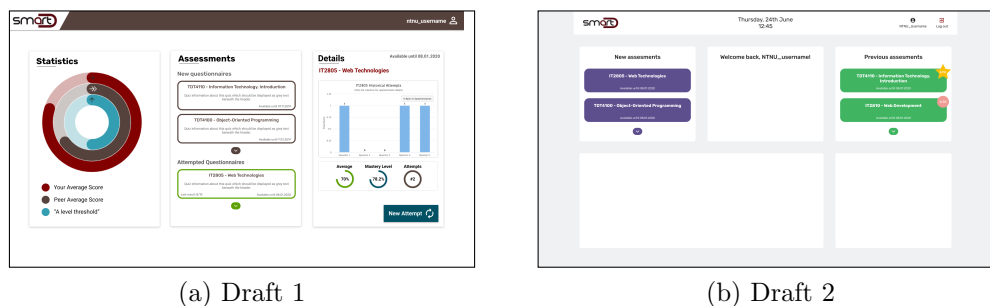


Figure 3.1: Iteration 1 - Pen &amp; Paper draft

sketches to determine what to bring to the next iteration. As the features in the sketches were evaluated, the functional requirements were also expanded with the features perceived as important.

### 3.4.2 Iteration 2

After the first iteration, it was decided to further develop some of the features from the initial draft. Two new drafts were created using Figma, focused on testing different aspects from the two paper ideation sketches. The first draft (Figure 3.2a) used the previous “Progress Chart” as a basis, while the second draft (Figure 3.2b) focused on utilizing informational features in a modular design. Using a mix of the circular progress graphs and line charts was perceived as an effective approach for comparing the learner’s statistics to their peers’.



(a) Draft 1

(b) Draft 2

Figure 3.2: Iteration 2 - High Fidelity Drafts

The user-test for this iteration was conducted by asking random students at

NTNU campus Gløshaugen to voluntarily share their opinions about the two different dashboard versions found in Figure 3.2. The participants were asked to provide a reasoning behind their choice, based on which interface was perceived as most intuitive and usable. Out of the 15 participants, 4 of them preferred draft 1 and 11 of them preferred draft 2. Pros and cons were mentioned for both interfaces, with draft 1 being criticized for showing too much incoherent information and looking very technical, but also being praised for displaying much information for the expert-users knowing how to use the system. Draft 2 was criticized for having less content displayed than draft 1 and using unnatural colors for the components. However, draft 2 also received praise for being more modern, having more “lively” colors and a more modular layout than draft 1.

After evaluating the results from the user-test, draft 2 was chosen as the basis for the next iteration, due to its modularity and the ease in changing tiles within the same view, i.e., its adaptivity. By further developing this trait, the system would be able to display different types of views for each user, depending on the user’s performance level.

### 3.4.3 Iteration 3

After acquiring more insight and knowledge of what was perceived as good in the interfaces, and based on draft 2 from iteration 2, a functioning prototype was created using Figma. New features from Table 2.1 and the requirements in Section 3.2 were also implemented, as new pages were designed to create a fully functional prototype for test-subjects to explore in a controlled environment. The activity view shown in Figure 3.3 was one of the most feature packed views in the application. This view contained many of the features mentioned in Table 2.1.

The new features implemented in design iteration 3, briefly shown in Figure 3.3, were:

- A progression chart with the ability to compare performance with peers using the quiz scores over time.
- A quick overview of the progress in the activity performance using colors to indicate the achieved mastery-level in each attempt.
- Previous attempts of the activity, showing performance quickly by color-coding the score and displaying stats from the attempts.
- Page for adaptive assessment questions.
- Statistics for a specific question in an assessment.

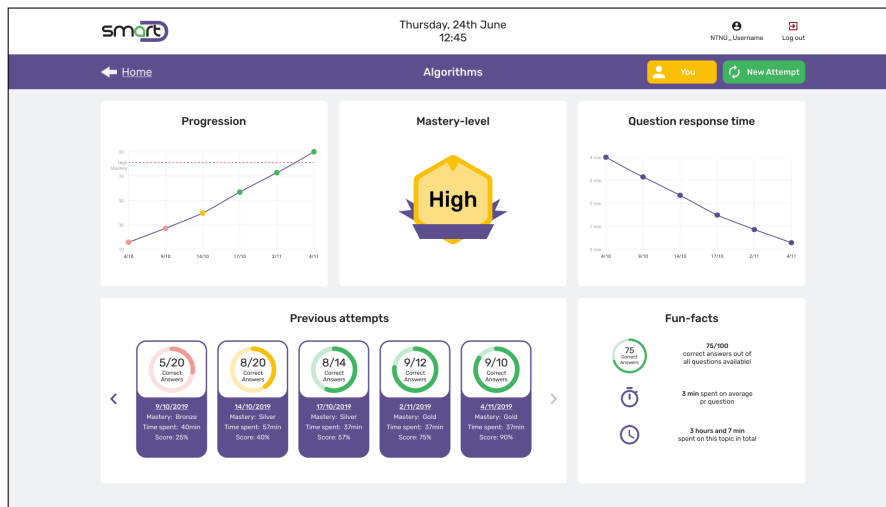


Figure 3.3: Iteration 3 - Activity View

- Icon/badge to indicate mastery-level and increase the game-based feeling of the user-interface.

After the design of the new pages and their respective features was completed, a new user-test was conducted with random students at NTNU campus Gløshaugen voluntarily participating. This user-test was conducted by providing the participants with an instance of the interactive prototype and a set of tasks. The testing in this iteration was more comprehensive and took more time per participant than in iteration 2. Volunteers were also harder to find, as the duration and complexity of the task increased. In this iteration, only three students volunteered to participate. However, the received responses provided valuable insight for further development in the next design iteration.

Responses from the user-test suggested that there was generally a good flow in the system, having easy navigation with the buttons being designed to appear clickable. The way colors were used in graphs and visualizations to indicate mastery-level/performance was praised and made the system more coherent. One participant also commented on the way the same components were re-used throughout the system: *“I can see that different components in the system are used multiple times in different views. For me this provides a feeling of consistency and makes the components easier to recognize.”* The interface also received criticism for the way peer comparison was designed. The general feedback was that it felt messy and not properly planned before being designed. There were comments on the peer graphs and visualizations being ambiguous as to whether they represented an average of the peers or one specific peer. The full set of responses from the user-test can be found in Appendix B.1.

### 3.4.4 Iteration 4

While iteration 3 introduced new features, iteration 4 focused on further developing the existing features to match how the users would like them incorporated into the interface. Based on the insight received through the feedback from the user-test conducted at the end of Iteration 3, changes were applied to meet the users' requirements. The graphs and visualizations displayed when the peer comparison mode was turned on were changed in order to be less ambiguous.

As can be seen in Figure 3.4, the lines in the comparison of progress got indicators below the graph, stating that the purple line represents your progress while the red line represents the average of the peers' progress. The chart for question response time was changed to show the student's average response time of correct, wrong and all answers when peer comparison was not activated. When peer comparison was activated it would show the student's average response time compared to the average aggregated response time of all peers. The mastery-level was also changed to show the student's mastery-level compared to the average mastery-level based on all peers when peer comparison mode was activated.

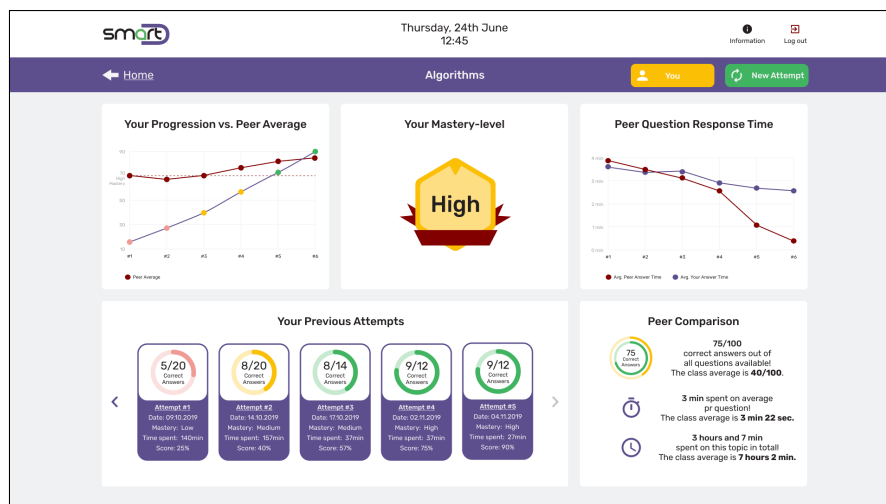


Figure 3.4: Iteration 4 - Activity View

To get accurate results from the user-test, the connections between the different views in the functional prototype were improved to be as close to a complete version as possible. Figure 3.5 illustrates the complete network of connections and interactions between the views in Figma, with each blue line representing one or several interactions connecting a component and/or views.

After applying those changes, a new user-test was conducted to verify whether the changes were an improvement. The new testing followed the same protocol as



in Iteration 3. In this iteration, the number of participating students increased to 5, as more students volunteered.

The user-test provided valuable insight with the overall feedback being positive. One participant commented on the use of colors to indicate performance, that it was easy to get a feeling of the overall performance by looking at the dashboard. However, there were still some areas that could be improved. One participant felt like the view displaying the contact information and how to use the site was crammed with text. Thus, the design was slightly altered by making the crucial information bold, enabling the users to extract the information at a glance.

There were comments on the available tools during a quiz. For instance, two of the participants did not understand the role of the timer. When the purpose of that feature was explained after the user-test, both participants stated that they would probably understand its functionality if the timer was counting incrementally. The task-related statistics that were available for each question during a quiz, also received comments. Specifically, the users commented on the visualizations displaying performance and effort, which were labeled as “cryptic”, without any textual explanation on what the visualization displayed. However, all participants managed to infer that a task was hard by looking at the provided visualized task-related statistics. Based on the feedback on the above visualizations, a textual explanation for each visualization was created, and made available by clicking on a question mark icon, placed within the box the visualization was located in. The complete results from the user-test can be found in Appendix B.1.

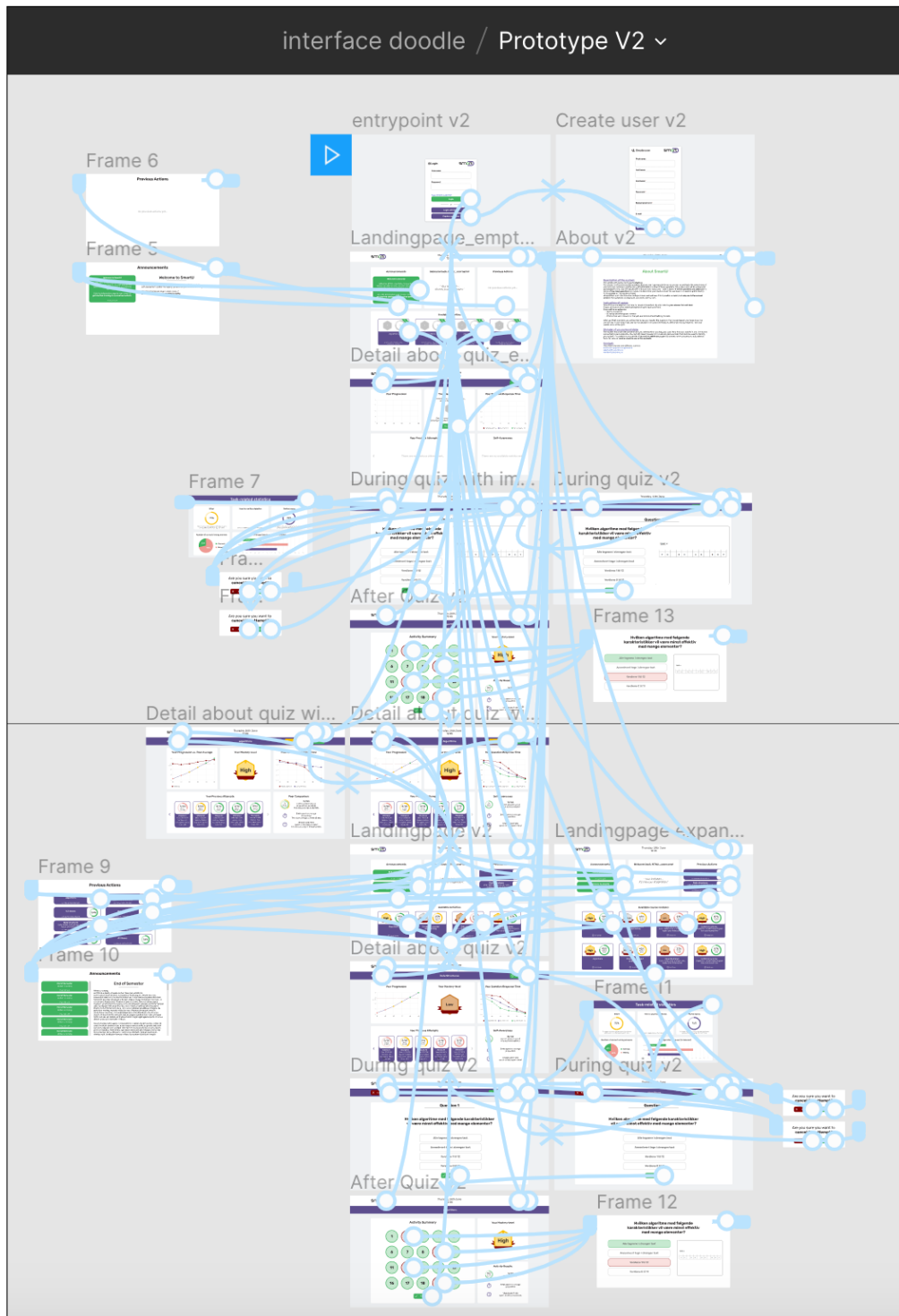


Figure 3.5: Iteration 4 - Prototype Network

### 3.4.5 Final user interface

As a result of the iterative design process, a collection of wireframes representing each view a user could access was created. The finalized user interface within Figma consisted of 19 individual frames, which functioned as guidelines for the implementation of the actual application. An overview of all wireframes can be found in Figure 3.6. Detailed images of all the wireframes can be found in Appendix B.2. In the following subsections we elaborate on the design choices and the different views in the final user interface.

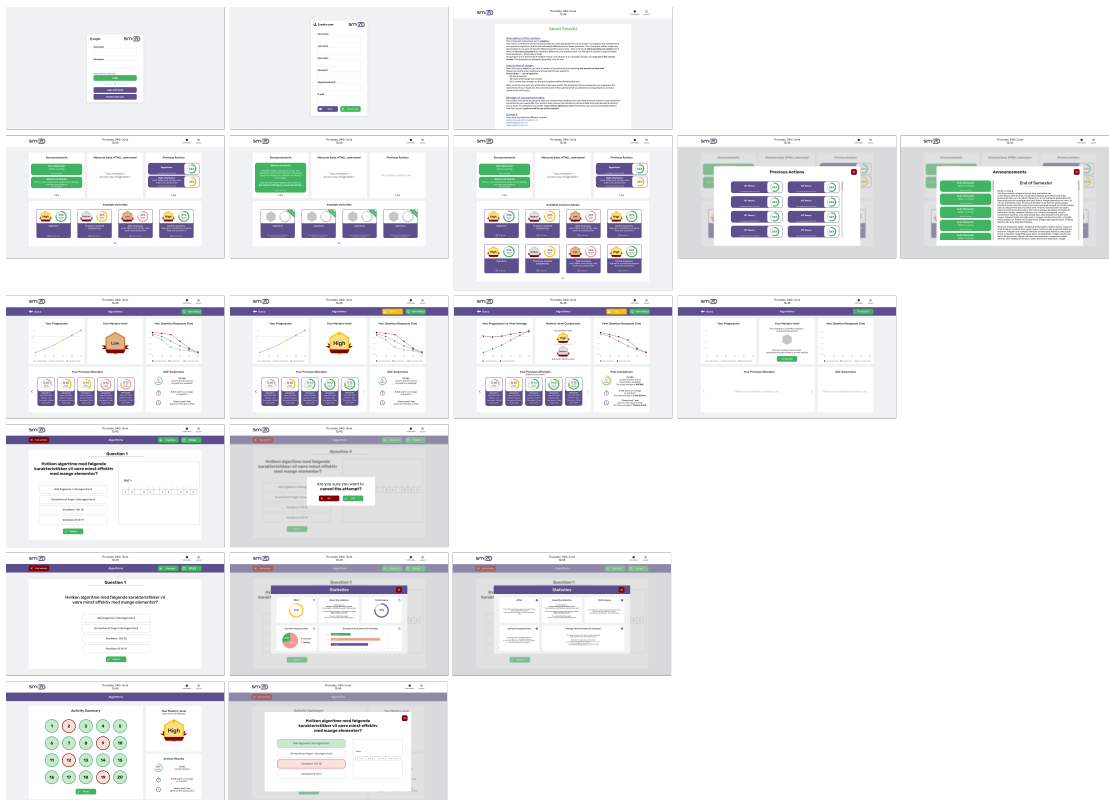


Figure 3.6: Final user interface wireframes

#### Colors and font

The user interface was designed using the color palette found in Figure 3.7 and using a font called Rubik, available for free at Google Fonts. Using this font and these six colors, it was possible to make the interface look simplistic with a “game-like” feeling, as defined in the requirements in Section 3.2, yet making it easy for the users to perceive the purpose of the elements using different colors. The following is an explanation of the colors and what they represent in the interface:

- Red (#820000) was primarily used as a contrast to the other colors in terms of visualizing ways of exiting current views or canceling other processes. It could also be used to represent wrong answers to a question. An exception is the use of a "red ribbon" around the mastery level badge.
- Yellow (#FBBF05) was used as a "medium achieving" color in terms of graphs and performance but was also used as a secondary color of buttons.
- Green (#40B461) was used to promote positive feelings and elements available for interaction, such as buttons for progressing further in a process or great performance in terms of graphs and statistics.
- Black (#231F20) was used as a less intense version of black, more in line with the other colors used in the interface.
- Purple (#5D4E8E) was the primary color of the system used throughout the application in either navigation bars or other elements needing contrasts.
- Pink (#EF9A93) was used as a secondary color to red in order to show a less intense way of promoting failed answers and poor performance in terms of graphs and statistics.

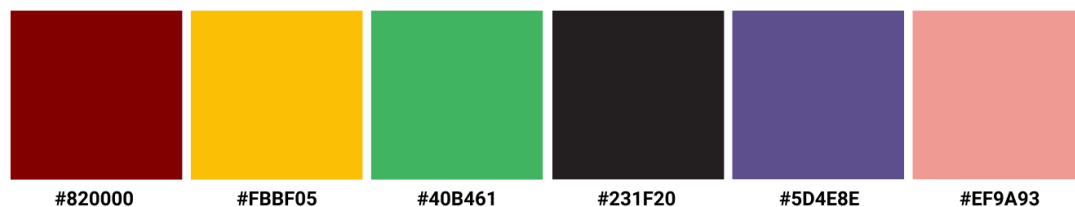


Figure 3.7: Color palette

### Login and user creation

The observations, during the iterations of user-testing, showed that the participants were split equally between creating a user and logging in with Feide and using the user-creation page in SmartU. Thus, both options remained in the final wireframes. The final wireframes for login and user-creation views can be found in Figure 3.8.

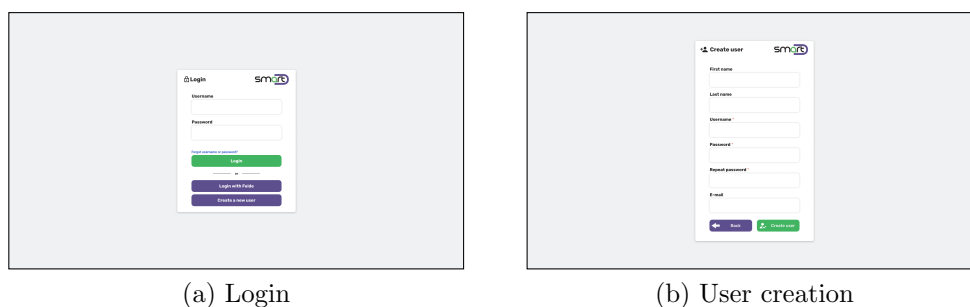


Figure 3.8: Login and user creation

### Landing page and activity-dashboard

The dashboard displayed in Figure 3.9(a) is the main dashboard and the landing page of the application after logging in. It provides access to announcements, to the learner’s last attempted quizzes and to all activities currently available to use. By selecting an activity from the dashboard in Figure 3.9(a), the learner will be directed to the corresponding activity-dashboard displayed in Figure 3.9(b).

The activity-dashboard displays information about a specific activity, including the learner’s stats based on their performance in the previous attempts in quizzes associated with that activity. To the left, the learner’s progress in performance is displayed by mapping the attempts in a two-dimensional chart where the axes represent correctness and attempt number. To the right, a chart of the development of the learner’s average response time is displayed, where the axes represent minutes to answer and attempt number. Note that this chart has three different lines, representing the time spent to provide correct answers, wrong answers and the average of all answers. The view also displays the learner’s overall mastery-level in this specific activity, aggregated stats from the learner’s attempts and all previous attempts with stats from that attempt.

The “peer”-button in the purple navigation bar allows the learner to toggle between the dashboards in Figure 3.9(b) and Figure 3.9(c). The difference between these two interfaces is that Figure 3.9(b) displays the learner’s own stats and Figure 3.9(c) displays the learner’s stats in comparison to the average of all peers.

### Quiz and task-related statistics

Figure 3.10(a) illustrates the user-interface for taking a quiz. There are two versions of this interface, with and without an image, as not all questions have an associated picture. In this interface, the learner can select an option to submit an answer on a specific question or click on the statistics button in the purple navigation-bar to access statistics for the given question.

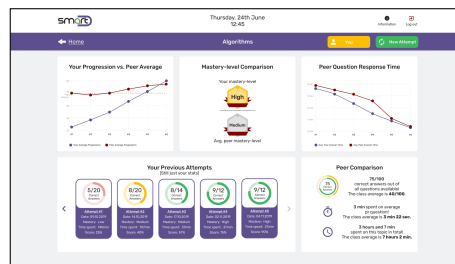
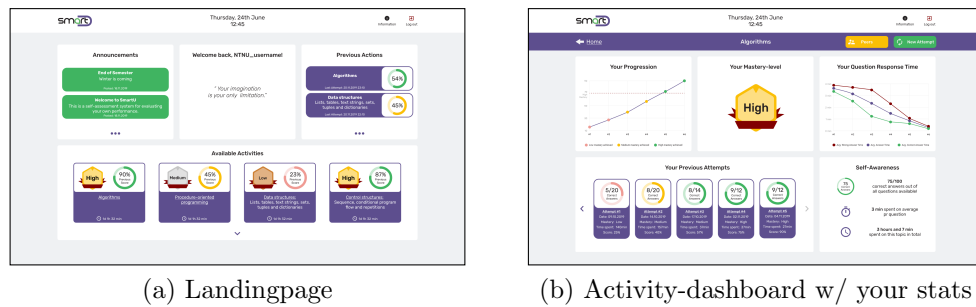
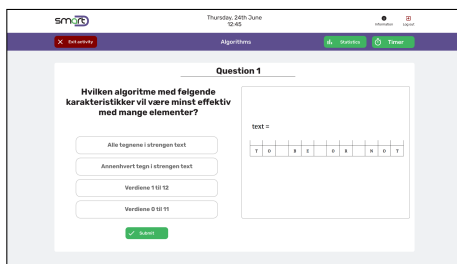


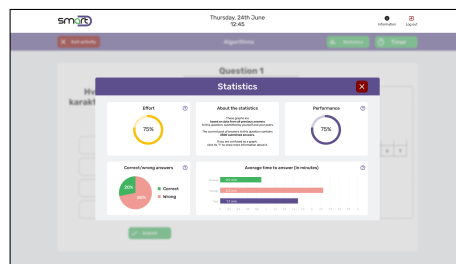
Figure 3.9: Dashboard interfaces

When accessing the task-related statistics during a quiz, the view in Figure 3.10(b) is displayed. This view provides access to 4 different visualizations based on aggregated data from all previous answers, by all users, to this specific question. The visualizations display the percentage of correct/wrong answers, the percentage of previous respondents putting in effort to answer the question, the average time to answer (correctly, wrongly and in total) and the overall performance on this question. By clicking the question-mark on the upper right corner of a visualization, a description of what the given visualization represents is provided.

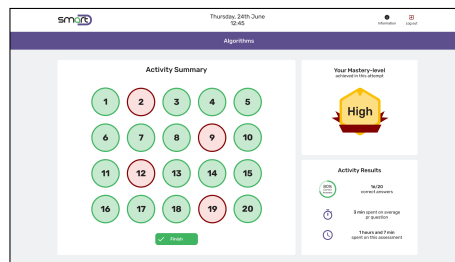
Figure 3.10(c) shows the interface the learner meets when finishing a quiz, and is a summary of their performance in the quiz. The learner gets an overview of some analytical statistics of their performance, the mastery-level they achieved in this specific attempt and an overview of the questions, with an indication of the correctness of their answer. By selecting a question, the learner can see what their answer was and what the correct answer to the question was. This immediate feedback is provided in order to trigger self-reflection and start a self-evaluation process when finishing a quiz.



(a) Interface for an active quiz



(b) Task-related statistics during an assessment



(c) Dashboard for post-quiz summary

Figure 3.10: Quiz interfaces

## 3.5 The Current SmartU System

This project utilizes parts of the backend the old SmartU system provided, with a new RESTful API, as explained in Section 3.5.3. Through the iterative design process, a new user interface was designed and implemented in the current version of SmartU, utilizing the data the backend makes available, to meet the requirements defined in Section 3.2 Functional Requirements.

### 3.5.1 Architecture

The system is a typical modern website application, based on the three-tier client-server software architecture pattern. The application consists of the user interface (Presentation Tier), a RESTful API (Application Tier) and a MySQL database (Data Tier). The project aimed at developing a new user interface for the existing SmartU system, with a strict focus on the presentation tier of the system as the backend already existed (Section 3.1 From LAERS to SmartU). However, due to limitations originating from the older backend services, some alterations were made to modernize and convert the backend templating to a modular and RESTful API.

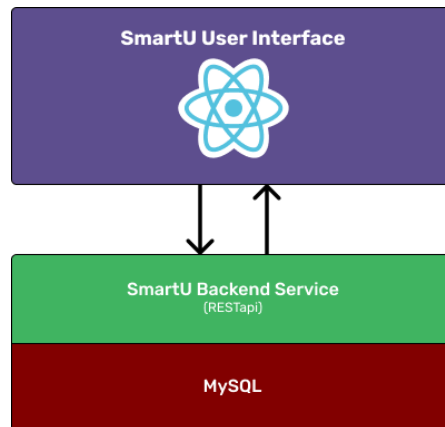


Figure 3.11: Application architecture

Although the architecture in Figure 3.11 is simple, the presentation tier has additional complexity as the system is built with a modular design. To make the user interface modular and easily maintainable, it was decided to use the well-known framework ReactJS [2019]. By introducing ReactJS to the presentation tier, it was possible to create views consisting of smaller components. These components are highly maintainable and can easily be swapped out or altered to edit the website. Additionally, it is possible to easily introduce new modules or views by adding more components to the system and connecting them to the SmartU Backend Service.



The SmartU Backend Service was initially built using the PHP template engine and this technology was sufficient for the previous versions. However, as a modern client-side interface was introduced, there was a need for replacing this technology with something more flexible for further and future expansion. Using PHP Slim, it was possible to preserve the core functionality of the existing backend, but rather than using the PHP template engine for server-side rendering, PHP Slim was used to provide data to ReactJS on the client-side through the HTTP protocol.

### 3.5.2 Mobile implementation

To ensure a seamless experience, a mobile version of the interface was implemented. The mobile version is based on traditional CSS media-queries, thus making the application accessible on all devices by adjusting to the device's screen size.

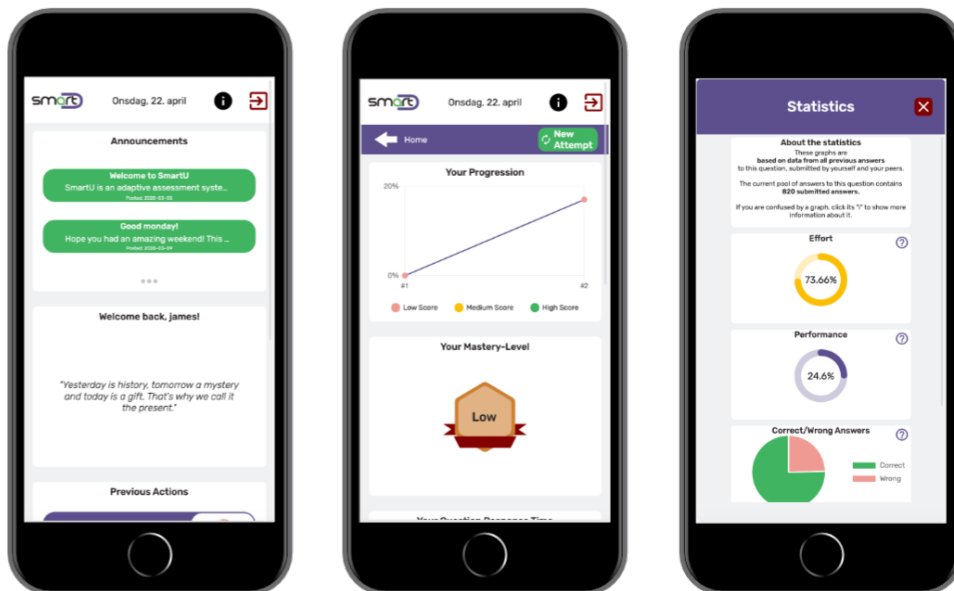


Figure 3.12: Mobile screen scaling

As seen in Figure 3.12, the application resizes to an iPhone 8 such that the user may be able to log in and use the application. All functionality is preserved in the mobile version, however, there is room for improvement. Due to time constraints and desktop being the focus of the study, the timeframe of this study did not allow for a fully user-tested and seamless experience on mobile. However, the mobile

sketches were implemented in the final version of the system, and the sketches for the mobile design can be found in Appendix B.2 for further development.

### 3.5.3 Challenges during development

#### Backend improvements

The system was initially planned to be built upon the older backend service of the previous version of SmartU, but due to limitations of the older systems using server-side templates and not RESTful API's, there was a need for changes to cater for the needs of this project. Using the same technologies (MySQL and PHP) and core algorithms, the backend was rebuilt from the bottom and up, making a modular and maintainable API for further development and future expansion.

#### Working with existing MySQL database

Implications can arise when working with and adapting to previously built systems. In this case, there was an existing MySQL database with variables to be preserved in the newer system. This was a challenge due to the nature of building a completely new frontend with extended functionality. New database fields had to be created, some left to be unused and others had to be re-purposed in order to fit the new project.

#### Feide login

As mentioned in Section 3.4.5, Feide login was an initially planned feature for the application. However, due to time constraints and the process of applying for access to an external authentication service, it was deemed more important to focus on other features during the development. However, the system was designed with these features in mind, so the supporting infrastructure required to implement such a feature should be easily implemented in the next version of the system.

# Chapter 4

## Methods

This chapter describes of how the study was conducted, according to the chosen research design and research methods explained in Section 1.3. Section 4.1 Study Design describes how the observation and user-testing were conducted, including the sample of participants, the physical test set-up and the procedure followed during testing. Section 4.2 Data Collection describes the data collection process (i.e., the usage data from SmartU, the observation-notes, the questionnaire and the interviews). Finally, Section 4.3 Data Analysis describes how the collected data was analyzed through appropriate qualitative and quantitative methods.

### 4.1 Study Design

This section describes how the experimental study was conducted, focusing on the participants of the study and the physical test-setup, and explaining the steps of the experiment itself.

#### 4.1.1 Participants

In total, there were 27 participants successfully engaging in the study. For recruiting participants, we reached out to people in common study areas on the campus, explaining the steps of the study and informing them there would be handed out gift cards for all participants that would respond to the questionnaire after finishing the user-testing of the system. Initially 40 user-tests were scheduled, but due to the outbreak of the Covid-19 virus, only 27 could be conducted prior to the lock-down of the university.

The sample of 27 participants consisted of 15 (55.56%) males, 11 (40.74%) females and 1 (3.7%) non-binary gender and included students spread across different years of study ( $M = 3.4$ ;  $SD = 1.5$ ;  $min = 1$ ;  $max = 5$ ), aged 19–27 years-old

( $M = 22.4$  ;  $SD = 2.0$ ). Most students were enrolled in a programming related study-program (77%), but some (23%) were enrolled in other programs such as chemistry or biology.

### 4.1.2 Physical test-setup

A room was booked with the sole purpose to be used as a testing-lab for the study, in order to have a controlled environment while conducting the testing. The researchers ensured that drinks and snacks were available to all participants during the testing, to facilitate the easiness of the environment. The participants used two laptops (belonging to the Department of Computer Science at NTNU) on which the test-suite had been set up beforehand (Figure 4.1). The test-suites on both laptops were reset between the testing-sessions. Due to the Covid-19 outbreak all laptops were also cleansed thoroughly with antibacterial wipes.



Figure 4.1: Test-setup without person

### 4.1.3 Procedure

As mentioned in Section 1.3, the study followed an experimental strategy, using a static group comparison design. The participants were split randomly into two groups; 12 participants were assigned to the group with access to task-related statistics (i.e., the experimental group) and 15 participants were assigned to the

groups with no access to the task-related statistics (i.e., the control group). Other than the access to the task-related statistics, access to the features of the system was the same in both groups. None of the participants had any previous experience with the SmartU system. A maximum of two participants could take the tests at the same time, with both participants belonging to the same group. This was done to ensure a controlled atmosphere and to ensure the two observers could focus on one participant each at a time. It also made the distribution of information to the participants easier. Each session lasted for a maximum of 45 minutes.

The study itself consisted of three steps to consistently present the same kind of experience and environment for all participants within the same group throughout the study. By following these steps, it was easy to keep track of the process of collecting data and consent during the testing.

### **1. Briefing**

Prior to the study, all participants were asked to provide their consent by signing a form, available in Appendix A, to satisfy the requirements for handling the personal data submitted by the participants. Next, all participants were briefly introduced to the SmartU system and its adaptiveness, using a printout of step-by-step explanations, available in Appendix A. The participants in the experimental group were given the same information as the control group, in addition to being introduced to the task-related statistics and how to use them. The printouts were available to the participants throughout the entire experiment upon being handed out. Finally, the researchers informed the participants that an observer would be observing the system, the participants' interaction with the system and environment, surroundings and atmosphere in the test-room during the testing.

### **2. User-testing and observations**

To conduct the testing, the participants were asked to create their own user-profile in the SmartU system and get familiar with the system before continuing. After getting familiar with the system, the participants were asked to imagine sitting at home, using the system to practice their skills in the syllabus for the introductory course used in the study. The reason the participants were asked to imagine sitting at home was to relieve them from any external pressure regarding performance. It was also restated that the quizzes in SmartU were self-assessments that were not connected to the grade in a course in any way.

The participants were asked to complete two or three quizzes within the available activity, depending on the time spent on each attempt. Each attempt consisted of a quiz with approximately 12 questions, with two to four possible answers, where only one was correct. Due to the adaptive nature of the quiz (Appendix A), only one answer could be submitted to a question, and the participant could not

return to a previous question or change an answer once submitted. The difficulty of the questions delivered to a participant during a quiz was determined based on the correctness of the participant's answers submitted to the previous questions, the weights of difficulty of all answered questions, as well as all remaining questions.

While the participants were using SmartU, from creating a new user to finishing the last activity and logging out of the system, the researchers acted as observers following the observation guide. The observation as a data collection method is elaborated on in Section 4.2.2.

### 3. Debriefing

The last step of the study consisted of a post-experiment questionnaire, available in Appendix A, which the participants had to fill in individually on the same computer they used for the testing, immediately after the testing was completed. The questionnaire was used for collecting quantitative data, as explained in Section 4.2.3. Upon completion, all participants received a 75 NOK gift card to be used in the campus cafeteria, as a reward for voluntarily contributing their time and data to the research.

The final step of the study consisted of a follow-up interview, which was conducted three to four days after the testing was conducted. Ten (10) participants were selected randomly for an interview, with both groups having 5 participants interviewed. Due to the outbreak of Covid-19 and the university closing the campus, the interviews had to be conducted online, where the free online conference room software *Whereby* [2020] was used and *Discord* [2020], another free online conference room software, was being used as a backup. *Open Broadcaster Software* [2020], a free and open-source software for recording and live streaming, was used to record the audio from the interviews. Although web-cameras were used during the interviews to facilitate the easiness of the environment, the video was not recorded nor stored. All interviews were conducted in Norwegian, as all interviewees were more comfortable with Norwegian than English. The interviews are described further in Section 4.2.4.

## 4.2 Data Collection

The following section describes the data collection methods used in the study, by looking into the different methods used and the origin of the constructs used in the questionnaire. Choices made to alter and customize the methods or constructs to fit the context of this study are also explained.

### 4.2.1 Usage data collected from SmartU

For the SmartU system to function as intended, a database was necessary. The SmartU database contained information about all users, all activities and their quizzes, and all attempts in quizzes performed by any user. It also contained the users' performance statistics, used to populate the graphs and visualizations throughout the system. During an attempt in a quiz, key statistics, such as how many times and for how long the task-related statics were viewed, were also added to the users' statistics in the database. As the study aims to investigate effects in performance, such statistics were vital to collect in order to conduct the study. Table 4.1 shows the data collected from the system about the participant's performance and usage, along with the corresponding data types and the range of their values.

Name	Origin of data	Range of values
Number of questions answered	The total number of questions answered	$\geq 0$
Total answer time	The time aggregated on answering the questions	$\geq 0\text{msec}$
Correct answers	The total number of questions answered <i>correctly</i>	$\geq 0$
Total correct answer time	The time aggregated on <i>correctly</i> answering the questions	$\geq 0\text{msec}$
Wrong answers	The total number of questions answered <i>wrongly</i>	$\geq 0$
Total wrong answer time	The time aggregated on <i>wrongly</i> answering the questions	$\geq 0\text{msec}$
Effort put into tasks	If participant spent more time than the effort-threshold for the question	0 - 1
Performance	Number of correct answers divided by the total number of questions answered	0 - 1
Statistics opened*	Number of times task-related statistics was opened	$\geq 0$
Statistics time*	For how long task-related statistics was opened	$\geq 0\text{msec}$

\* - Only collected from participants in the experimental group

Table 4.1: Usage data collected from SmartU

### 4.2.2 User-test & observation

As mentioned in Section 4.1.3, observations were important to monitor interactions with the system, the atmosphere and the surroundings of the testing environment. Following the structured observation guide, available in Appendix A, the observers were able to make meaningful and consistent notes about the test environment, the participants' emotional state, unexpected events, utilization of aids and external disturbances during the user-testing [Oates, 2005].

As discussed by Musante and DeWalt [2010], observations of participants can provide several advantages to the study. Thus it was important to note the above mentioned aspects of the testing environment, as they could explain unexpected results in the usage data collected from participants' interaction with SmartU. The observation guide informed the observer on how to start and end the observation, and how to behave while conducting it. It also provided a matrix with categories and examples of what should be noted during the observation, shown in Table 4.2.

<b>Category</b>	<b>Includes</b>	<b>Researchers should note</b>
Test environ- ment	Ambiance, atmosphere, alterations to environ- ment	Any differences in test- environment between tests should be noted
Subjects emo- tional state	The users state of mind	Is the user stressed or nervous? Does the state of mind change during the tests? Is the test- subject focused on the task?
Unexpected events	Failing to navigate, bugs in the system	Did anything unexpected happen during testing? Note when, why and how it happened
Help	Frequency, topic of ques- tions	What does the user ask you to help out with and how often/ many times are you asked to help?
Disturbances	People entering/leaving the room, phone notifi- cations, noises, etc	Any unwanted noises or distur- bances hindering the focus on the testing should be noted

Table 4.2: Examples of events/actions to note during observation

### 4.2.3 Questionnaire

The questionnaire featured in the debrief of the study was the main source of quantitative data regarding the participants' opinions and attitudes towards the



SmartU system. In accordance with the approved application submitted to Norwegian Centre for Research Data (NSD) regarding the handling of personal data, the questionnaire was made using Microsoft Forms.

The questionnaire was employed to collect data both for the Design & Creation strategy and for the Experiment strategy on different features of the system, by being split into sets of items ordered into categories, summarized in Table 4.3. Separating the questionnaire-items into categories or sets of items relating to the same attitude or perception, one achieves a more realistic image of the respondents' opinions, without relying too heavily on particular items [Oppenheim, 2000]. As the items in this questionnaire captured the opinions and feelings of the participants, such as their attitude or motivation of use, the items were mostly closed, attitudinal questions, with the available answers being a 5-point Likert Scale ranging from 1 to 5. One represented "Strongly Disagree", three represented "Neutral" and five represented "Strongly Agree" [Oppenheim, 2000].

Category	Acronym	Constructs
System Usability Score	SUS	Usability score
Overall evaluation of the usability of SmartU	OEUS	General Usability
Attitude towards Graphs and Visualizations	AGV	Usage Motivation Usefulness Positive and Negative emotions Intent for further use
Usability of Graphs and Visualizations	UGV	Usability of features Intuitiveness Usefulness
Graphs and visualizations in an assessment	GVA	Usability of features Usage Usefulness Intuitiveness Usefulness Motivation Intent for further use Positive and Negative emotions

Table 4.3: Overall categories in the questionnaire and their constructs

The first two categories, *System Usability Score (SUS)* and *Overall evaluation of the usability of SmartU (OEUS)* (Table 4.3) contained questions regarding the usability of the system and were related to the Design & Creation strategy, estab-

lishing whether the SmartU user-interface met the requirements of usability posed by the participants of the study. While the SUS-schema is a standardized method of measuring usability [Brooke et al., 1996], the constructs in the OEUS-category were self-created to assess whether the system met the definition of usability as described in ISO 9241-210:2019: *extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.*

The next two categories, *Attitude towards Graphs and Visualizations (AGV)* and *Usability of Graphs and Visualizations (UGV)* (Table 4.3) contained questions regarding the participants' attitudes towards the graphs and visualizations in the dashboards in the system, and their perceived usability. The questions and constructs in these categories were adapted from previous relevant studies conducted by Roca et al. [2006]; Chiu and Wang [2008]; Lee [2010] and altered to fit the context and the system used in this study.

The last category, *Graphs and visualizations in an assessment (GVA)* (Table 4.3) contained questions regarding the graphs and visualizations available in the task-related statistics during a quiz. This category covered the participants' perceived usage and usefulness, and their attitude towards such visualized statistics. The questions and constructs in this category were also adapted from the previous relevant studies conducted by Roca et al. [2006]; Chiu and Wang [2008]; Lee [2010] and altered to fit the context and the system used in this study.

#### 4.2.4 Interviews

As mentioned in the last part of Section 4.1.3, the interviews were conducted three to four days after the user-testing and observation and after answering the questionnaire. The interviews followed a semi-structured approach and an interview guide, available in Appendix A. This guide contained a list of situations requiring extra attention by the interviewer and a set of predefined questions, based on interesting overall trends found in the answers from the questionnaire. The questions were used as conversation-starters to let the interviewees tell about their experiences, feelings and thoughts and thus allowing the interviewer to probe with appropriate follow-up questions [Oates, 2005]. Examples of the predefined questions can be found in the bullet-point list below.

- Could you mention some features that you found useful in the system?
- How did you feel about the statistics provided during an assessment?
- Could you mention some features which were easy or hard to understand?
- Did you feel like the statistics helped you understand the scope of the question? How did the statistics help you?

## 4.3 Data Analysis

This section covers how the analysis of the collected data was performed. Given the mixed methods approach in collecting data, both qualitative and quantitative data analysis approaches were necessary.

### 4.3.1 Qualitative analysis

To analyze the interviews described in Section 4.2.4, the interviews had to be transcribed. The transcribed interviews can be found in Appendix A.6. While transcribing the interviews the broad and recurring themes were noted for later use in the analysis.

For the analysis itself, *NVivo*, i.e., a qualitative data analysis software, was used to code the interviews into categories or nodes by associating a quote or statement with one or more categories. The categories used for the analysis were initially based on a deductive approach, using theories based on the literature review [Oates, 2005]. However, as the coding was iterative, the notes taken when transcribing and repeatedly reading through the interviews, quickly formed new categories. Thus, the analysis also followed an inductive approach [Oates, 2005].

When clusters of similar or connected content appeared within a category, the category was split into two, more granular categories, having their original one as their parent category. Similarly, small categories were merged with other categories. Thus, a tree-like structure was established, providing a detailed separation of the content of the interviews.

In the last iteration, the data was cleansed, preparing it for presentation. In regard to working with qualitative data, it was important that the actual quotes of the interviewees were intact and not altered. Thus, when an answer to a question or a quote was ambiguous, the correct context was added, followed by “red. anm.”, an abbreviation for *redaksjonell anmerkning*, meaning *editorial/writers remarks* in English. These remarks were clearly marked with parentheses.

### 4.3.2 Quantitative analysis

The quantitative analysis of the data involved aggregating data from the answers to the questionnaire and the usage data from SmartU. The quantitative data were further processed using IBM’s Statistical Product and Service Solutions (SPSS), to extract meaningful descriptives, and perform Correlation analysis and Independent Samples T-tests, as mentioned in Section 1.3.

#### 1. Correlation

Pearson product-moment correlation coefficient (Pearson’s Correlation) was used

to find meaningful relationships between the quantitative data. The relationship was deemed significant when Sig(2-Tailed) was lower or equal to 0.05, and even more significant if lower or equal to 0.01.

## **2. Independent samples T-tests**

Independent Samples T-tests (T-tests) were used to compare the variables of the two different groups of participants in the study. In order to separate the groups, a boolean value was used as a group identifier. As explained in Section 4.1.3, only the experimental group had access to the task-related statistics, thus having the value of 1 in the t-test calculation.

# Chapter 5

## Results

This chapter presents the results of the conducted experiment, and gives an overview of the findings of the study. Section 5.1 provides insight into the participating subjects of the study, Section 5.2 outlines the descriptive statistics of the variables, Section 5.3 presents the correlations and Section 5.4 presents the Independent Samples T-tests results. Section 5.5 presents the findings from the SUS-schemas, and Section 5.6 presents the main findings from the interviews.

### 5.1 Participants & Demographics

The sample of participants in this study was briefly introduced in Section 4.1.1. As Figure 5.1(a) shows, the experiment had more male (55.56%) than female (40.74%) participants. 3.7% of the participants identified as a different gender than the traditional binary genders.

Figure 5.1(b), and Table 5.1, show the distribution of the current year of study for all participants. The majority of the participants were in their 5th year (37.04%), followed by the students currently in their 1st year and 3rd year (both 18.52%). The students in their 4th year (14.81%) and 2nd year (11.11%) had the lowest number of representatives among the participants. The age distribution of the participants ranged from 19 to 27 ( $M = 22.48$ ).

Furthermore, the participants were divided into two groups, a control group and an experimental group, as shown in Table 5.2.

	N	Minimum	Maximum	Mean	Std. Deviation
Age	27	19	27	22.4815	2.06380
Year of Study	27	1	5	3.4074	1.55066

Table 5.1: Sample of total participants

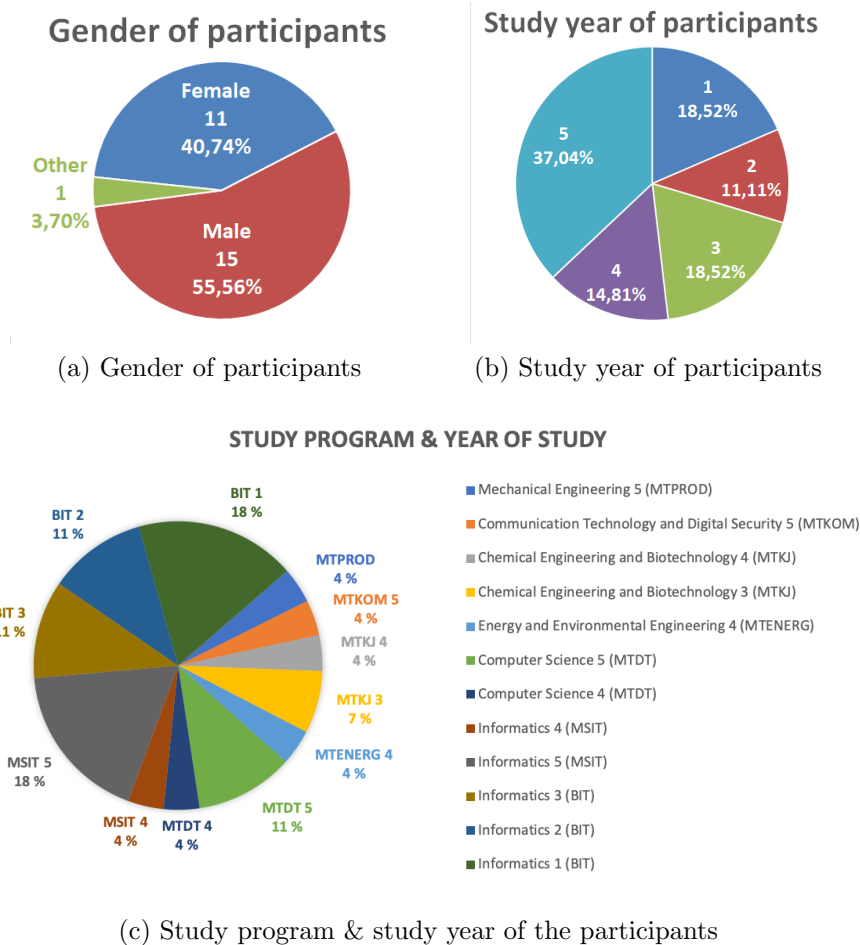


Figure 5.1: Gender and year of study of the participants

As presented in Figure 5.1(c) most participants in the study were enrolled in a programming related study-program (BIT, MIT, MSIT, MTDT, MTKOM). However, there were participants enrolled in other study-programs (MTKJ, MTPROD, MTENERG), but all students had participated in an introductory course of programming during their first year of study.

## 5.2 Descriptives

As a result of the analysis of the answers from the questionnaire in Section 4.2.3 and the data collection from SmartU, the following sections present the descriptives and mean values aggregated using IBM's SPSS.

Group 1 (control)	N	Minimum	Maximum	Mean	Std. Deviation
Age	15	19	26	21.9333	2.25093
Year of Study	15	1	5	3.0667	1.66762
Group 2 (experimental)	N	Minimum	Maximum	Mean	Std. Deviation
Age	12	20	27	23.1667	1.64225
Year of Study	12	1	5	3.8333	1.33712

Table 5.2: Sample of participants in groups

### 5.2.1 Questionnaire mean variables

Table 5.3 presents the descriptive statistics for the four constructs from the questionnaire. The results give the impression that there is a high attitude towards SmartU’s use of visualizations and graphs in general (Attitude towards Graphs and Visualizations (AGV) and Usability of Graphs and Visualizations (UGV)). However, there is a lower minimum score towards Overall evaluation of the usability of SmartU (OEUS), indicating that the overall usability is fluctuating from user to user. There is also a diverse set of opinions regarding the attitude and usefulness towards Graphs and visualizations in an assessment (GVA), considering the standard deviation. The complete set of descriptives, including descriptives of each individual question of the questionnaire, can be found in Appendix C.1.

	N	Min	Max	Mean	Std Dev
Overall evaluation of the usability of SmartU	27	2.67	4.89	4.26	.49
Attitude towards Graphs and Visualizations	27	3.29	5.00	4.39	.43
Usability of Graphs and Visualizations	27	3.20	5.00	4.33	.51
Graphs and Visualizations in an assessment	12	2.53	4.47	3.50	.55

Table 5.3: Questionnaire mean variables of categories

When looking into the descriptives of each group separately in Table 5.4, there are some key differences that should be noted. Compared to the experimental group, the control group’s OEUS Mean and Minimum value suggests that the overall usability is better than when the visualized task-related statistics are introduced. This might indicate that the visualized task-related statistics introduce a new level of complexity to the platform, increasing the required effort to understand the visualizations.

	Group 1 (control), N=15				Group 2 (experimental), N=12			
	Min	Max	Mean	Std dev	Min	Max	Mean	Std Dev
Overall evaluation of the usability of SmartU	3.56	4.89	4.33	.38	2.67	4.89	4.17	.60
Attitude towards Graphs and Visualizations	3.29	5.00	4.45	.46	3.59	4.88	4.32	.41
Usability of Graphs and Visualizations	3.20	5.00	4.37	.55	3.60	5.00	4.27	.46
Graphs and Visualizations in an assessment					2.53	4.47	3.50	.55

Table 5.4: Questionnaire mean variables between groups

### 5.2.2 SmartU aggregated descriptives

The quantitative data collection from the users' interaction with SmartU during the study, as mentioned in Section 4.2.1, made it possible to aggregate the users' performance and effort put into the quizzes. Table 5.5 shows the descriptives of the usage data for the control and experimental group.

It appears that the usage data for the two groups are similar. However, there is a difference in the effort put into answering a question. The experimental group, having visualized task-related statistics available, seems more willing to spend extra time answering a question during a quiz. On average, the users in the experimental group spent  $\sim 57$  seconds in total, during the study, looking at statistics.

Group 1 (control)	N	Minimum	Maximum	Mean	Std. Deviation
Effort	15	680.00	1000.00	915.00	80.77
Performance	15	462.00	889.00	645.87	125.77
Mastery Level	15	1	3	2.53	.743
Group 2 (experimental)	N	Minimum	Maximum	Mean	Std. Deviation
Effort	12	861.00	1000.00	955.92	43.03
Performance	12	458.00	917.00	640.08	127.44
Mastery Level	12	1	3	2.58	.669
Statistics Clicks	12	1	11	5.25	2.864
Statistics Time Spent	12	5.73	204.72	57.47	51.47

Table 5.5: Descriptive of usage data for the groups



## 5.3 Correlations

Table 5.6 gives an overview of the Pearson Correlations between the questionnaire categories and learners' effort and performance in SmartU. The correlation is deemed significant at the 0.05 and 0.01 level (2-tailed). Variables compared in the table: Overall evaluation of the usability of SmartU (OEUS), Attitude towards Graphs and Visualizations (AGV), Usability of Graphs and Visualizations (UGV), Graphs and visualizations in an assessment (GVA), Effort put into questions during quizzes (Effort) and Performance of answering correctly (Perf).

As a result of the Correlation analysis, strong positive relations were discovered between AGV and OEUS ( $r = .746$ ,  $n = 27$ ,  $p = .000$ ), UGV and OEUS ( $r = .732$ ,  $n = 27$ ,  $p = .000$ ) and UGV and AGV ( $r = .758$ ,  $n = 27$ ,  $p = .000$ ). These correlations indicate that there is a continuous level of motivation and attitude towards visualizations, graphs and usefulness throughout the whole SmartU system.

Other positive correlations were detected between Effort and AGV ( $r = .446$ ,  $n = 27$ ,  $p = .020$ ), and between Effort and UGV ( $r = .503$ ,  $n = 27$ ,  $p = .007$ ). These correlations indicate that there is a connection between the Effort of users in quizzes and the motivation and attitude towards graphs and visualizations of the system. However, there is no statistically significant relation between Effort and OEUS, thus it might seem that the overall usefulness of the system has a lesser impact on effort than visualizations and graphs.

	Mean (std dev)		OEUS	AGV	UGV	GVA	Effort	Perf
OEUS	4.259	Pearson Corr.	1					
(N=27)	(0.485)	Sig. (2-tailed)						
AGV	4.392	Pearson Corr.	<b>.746**</b>	1				
(N=27)	(0.433)	Sig. (2-tailed)	.000					
UGV	4.326	Pearson Corr.	<b>.732**</b>	<b>.758**</b>	1			
(N=27)	(0.505)	Sig. (2-tailed)	.000	.000				
GVA	3.499	Pearson Corr.	-.133	-.161	.239	1		
(N=12)	(0.546)	Sig. (2-tailed)	.727	.617	.455			
Effort	933	Pearson Corr.	.332	<b>.446*</b>	<b>.503**</b>	.286	1	
(N=27)	(68.74)	Sig. (2-tailed)	.091	.020	.007	.368		
Perf	643	Pearson Corr.	-.072	-.131	-.074	.139	.175	1
(N=27)	(124.09)	Sig. (2-tailed)	.720	.515	.713	.666	.383	

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\* . Correlation is significant at the 0.01 level (2-tailed).

---

Overall evaluation of the usability of SmartU (OEUS)  
 Attitude towards Graphs and Visualizations (AGV)  
 Usability of Graphs and Visualizations (UGV)  
 Graphs and visualizations in an assessment (GVA)

---

Table 5.6: Correlation table for questionnaire

## 5.4 Independent Samples T-test

The Independent Samples T-test was run using a cut point of 1, where the experimental group having statistics is defined as 1 and the control group not being able to access statistics is defined as 0.

As seen in Table 5.7 and Table 5.8, there was no statistically significant difference of Sig. (2-tailed) less than 0.05 in either T-test. This could indicate that there are no significant differences between the two groups, however there is a small noteworthy mention in Effort. The T-test of Effort ( $t(25) = 1.581$ ,  $p = .127$ ) has a resemblance to Table 5.4, which indicates that the participants in the experimental group tend to spend more time and effort answering questions, thus emphasizing this behavior when having the visualized task-related statistics available.

On the other hand, the lack of statistical difference may indicate that there is no noteworthy impact on performance, effort or mastery level when introducing the visualized task-related statistics.

Equal Variances Assumed	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Performance	-0.118	25	0.907	-5.783	48.996
Effort	1.581	25	0.127	40.917	25.888
Mastery Level	0.181	25	0.857	0.05	0.276

*Significantly different when "Sig. (2-tailed)" is less than 0.05*

Table 5.7: Independent Samples T-test for analyzing the effect of learners performance using statistics

Equal Variances Assumed	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Overall evaluation of the usability of SmartU	-.882	25	.386	-.117	.189
Attitude towards Graphs and Visualizations	-.785	25	.440	-.133	.169
Usability of Graphs and Visualizations	-.538	25	.595	-.107	.198

*Significantly different when "Sig. (2-tailed)" is less than 0.05*

Table 5.8: Independent Samples T-test for the difference of OEUS, AGV and UGV with and without statistics.

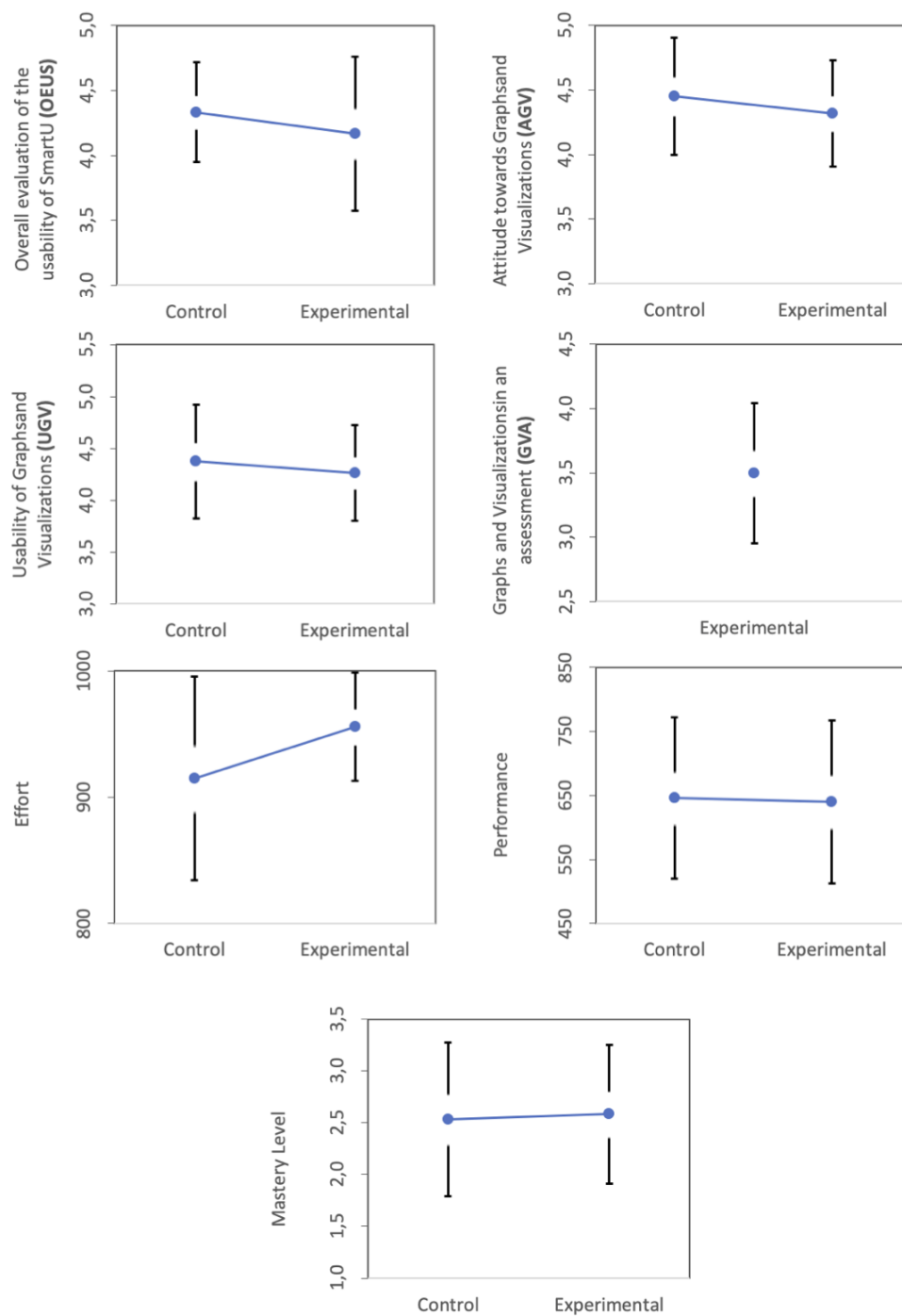


Figure 5.2: Graphical representation of Independent Samples T-tests between the control and experimental group. The dots represent the mean value while the black bars represent the standard deviation. Statistical significance is not marked due to lack of significant results.

## 5.5 System Usability Score

The SUS-schema was the first part of the questionnaire, as explained in Section 4.2.3. Table 5.9 shows the average SUS-score for each of the groups in the study, and their average score for all questions in the SUS-schema. The best score between the two groups for each question is highlighted. The total average SUS-score for the control group was 83, the total average SUS-score for the experimental group was 84,375, while the total average SUS-Score for both groups combined was 83,61. Furthermore, from all the SUS-scores, the *lowest score* was 67,5, the *highest score* was 95, while the *median score* was 85. The full list for the SUS-scores can be found in Appendix C.2.

Question	Control group (N=15)	Experimental group (N=12)
Q1 - Frequent Use	4,27	3,92
Q2 - Unnecessarily complex	2,27	2,25
Q3 - Easy to use	4,33	4,42
Q4 - Would need support	1,33	1,25
Q5 - Well integrated functions	4,27	4,42
Q6 - Too much inconsistency	1,80	1,42
Q7 - Learn it quickly	4,53	4,58
Q8 - Slow or complicated	1,20	1,50
Q9 - Felt confident	4,00	4,25
Q10 - Required training	1,60	1,42
Average SUS-score for group	83	84,375

Table 5.9: Average SUS-score of both groups for the variables in the SUS-schema

## 5.6 Results From Interviews

This section presents the findings from the qualitative analysis described in Section 4.3.1. The results presented in this section are extracted from the 10 conducted interviews. As the findings from the analysis are statements by the participants of the study expressing their opinions and perceptions, the original statements in Norwegian will be presented along with an English translation. This is necessary in order not to lose any meaningful information or nuances in the choice of words in the translation.

The tree-like hierarchy created as a result of categorizing the statements of the participants during the analysis, contained seven categories (hereby referred

to as nodes) at the top layer. Four of the seven nodes (Features, Learning, Motivation and Remarks) contained other nodes and no direct content in the form of statements. The remaining three nodes (Clarifications, Colors and Gamification) contained statements and no other nodes.

By looking away from the nodes containing clarification statements and remarks, the rest of the nodes naturally created four interesting groups of nodes for the research, aligned with the RQs defined in Section 1.2.

### 5.6.1 Participants insights towards motivation of use

The first noticeable group emerged within the parent node containing nodes with statements regarding motivation. The use of the term “motivation” in these nodes, refers to the participants feeling more or less motivated to use the system again for educational purposes and whether there was an increase in the motivation to study in general. Thus, participants could state either being motivated or demotivated, or that there was no change in their motivation. There was also a separate node containing all references to the statistics overlapping with the other motivation nodes mentioned. The nodes and their number of references are shown in Table 5.10.

Name	Files referenced	Statements in total
Negative Motivation	7	14
Positive Motivation	10	47
No Change	5	10
Statistics	8	21

Table 5.10: Nodes making up the “Motivation”-group, with number of files referenced and number of statements.

As can be seen in Table 5.10, all 10 interviewees had statements regarding their motivation being positively impacted while using the SmartU system. Seven of the interviewees expressed a decline in their motivation. Five of the interviewees also expressed that their motivation remained unchanged for some features.

Most of the statements in both Positive Motivation and Negative Motivation revolved around the topic of comparing results, either against themselves or against their peers. One statement in particular, in Positive Motivation, highlighted this:

Norwegian:

*“... når du får informasjon om hvordan andre gjør det så blir det veldig naturlig at det blir en slags konkurranse greie da. Som gjør at du fort kan sette deg mål som: “Jeg skal gjøre det så så bra i forhold til alle andre eller i forhold til deg selv”. Så du kan konkurrere både mot deg selv og mot andre, og det hjelper på å motivere, istedenfor at du ... ja ... bare gjør det på egenhånd for å si det sånn.”*

English:

*“... when you receive information about how others are doing, it becomes very natural that there will be some kind of competition present. Which allows you to quickly set goals like: “I will be this good compared to everyone else or in comparison to yourself”. So you can compete both against yourself and against others, and that helps my motivation, instead of ... yeah ... just doing everything on your own.”*

The element of competition and comparison to others, being the topic most frequently talked about in the interviews, was further reflected on with a participant stating:

Norwegian:

*“Det som motiverer meg er jo igjen, det at jeg kan se meg selv i forbindelse med nivået jeg ligger på og spesielt at du kan se resten av de andre som bruker programmet og deres nivå.”*

English:

*“What motivates me is, again, that i can see myself and my level of knowledge, and especially that you can see the rest of the users of the system and their level.”*

However, the comparison with one’s own previous performance was also emphasized with one participant stating that:

Norwegian:

*“... hvis jeg skulle brukt dette så ville det gitt meg mer motivasjon om jeg ser en stadig growth (på grafen som viser progresjon, red. anm.), eller hvis jeg ser det kommer en decline så kunne jeg kanskje tenkt sånn «ah, shit nå har jeg faktisk ikke fokusert lenge nok, eller ikke fokusert nok» om det går over en lenger periode. Så det blir lettere å være selvkritisk og analytisk når man ser på sitt eget arbeid ...”*

English:

*“... if I were to use this it would give me more motivation if I see a steady growth (in the graph illustrating performance, red. anm.), or if I see a decline I might think «ah man, I haven’t focused for long enough, or not focused enough» if it goes on for a longer period. So it becomes easier to be self-critical and analytical when you look at your own work ...”*

And another participant stating:

Norwegian:

*“Den er vel ganske uendra sånn sett fordi at jeg føler jeg har ganske høy motivasjon selv i ulike fag så jeg trenger ikke nødvendigvis å se andre sin statistikk for å drive å på måte å konkurrere med dem. Da har jeg mer interessert i å konkurrere mot mine egne resultater.”*

English:

*“It’s more or less unchanged as I feel I have quite high motivation myself in different courses, so I don’t necessarily need to see the statistics of others to compete with them. I am more interested to compete against my own results.”*

Despite all participants having some positive feedback towards the system, there were also statements regarding their motivation being negatively influenced. The possibility of comparing one’s results to one’s previous performances and to the peers was a key example of this, proving to be a double-edged sword as it also received feedback of negatively influencing the motivation. One of the participants summarized this in this statement:

Norwegian:

*“Jeg hadde jo kanskje tatt det litt personlig om jeg hadde ligget veldig lavt og jeg fikk se at andre ligger veldig høyt, og stoffet er lett for dem. Så om dette da var frivillig og ikke hadde noe å si på karakteren så hadde jeg nok kanskje bare lest selv og tenkt at det går sikkert bra.”*

English:

*“I might have taken it a little personally if I had been placed very low and I saw that others were placed very high and that the curriculum was easy for them. So if this was voluntary and had no impact on the grade, I might have just studied by myself and thought it probably was sufficient.”*

Other participants shared this opinion, as another participant added to this by stating:

Norwegian:

*“Hvis man ligger langt bak sånn mot slutten av semesteret, så er det jo sånn «hvorfor gidde, det er jo ikke lang tid igjen». Hadde sikkert blitt demotivert om jeg hadde ligget så langt bak.”*

English:

*“If you are far behind towards the end of the semester, you’d think «why bother, there is not enough time». I would probably be demotivated if I was that far behind.”*

### 5.6.2 Perceptions towards intuitiveness of features

The second noticeable group emerged when combining the parent node containing statements about the colors in the system and the parent node related to the features of SmartU. This group revolved around the intuitiveness of the features



in the system. As seen in Table 5.11 no node had all participants present, however, all participants had at least one statement regarding whether features were easy or hard to understand.

Name	Files referenced	Statements in total
Colors	5	10
Features*	0	0
Sub-nodes of Features		
Hard to understand	8	18
Easy to understand	9	24

\* - No files are referenced as the node only contains other nodes

Table 5.11: Nodes making up the "Intuitiveness"-group, with number of files referenced and number of statements.

The statements on the intuitiveness of the features were spread among most of the features present in the system. As can be seen in Table 5.11, the average participant had several statements regarding the intuitiveness, which enhances the importance of addressing the commented aspects of the features mentioned.

One of the features being labeled as hard to understand was the visualized task-related statistics. This feature received the following feedback from two of the participants sharing the same views, when asked if there were any features that were particularly hard to understand. The first participant stated:

Norwegian:

*"Når du var inne på et spørsmål så var det det at «Okey her står det statistikk». Det jeg forventet at det skulle stå var jo hvordan du ligger an så langt i løpet av testen, men så var det jo heller statistikk på dette enkelte spørsmålet. Og da var det litt merkelig (vanskelig, red. anm.) å finne ut av hva dette her betydde for det var ikke sånn super tydelig, og det føltes ikke ut som noe man typisk bruker til vanlig heller."*

English:

*"When you were answering a question it was like «Okay it says statistics here». What I expected it to be was how you were doing so far in the quiz, but it turned out to be statistics for the overall performance on that specific question. And then it was a bit hard to figure out what it meant because it was not clear, and did not feel like something you would typically use."*

Similarly the second participant stated:

Norwegian:

*"Også var det dette med analyse delen som kanskje ikke kom så godt frem første gang, før man måtte lese det lille spørsmålstegnet med informasjonen."*

English:

*“Then it was this part about the analysis bit that was not clear the first time around, you had to read the little question-mark with information.”*

The graphs in the dashboard for a specific activity was another feature being labeled as hard to understand. When asked to elaborate on features being hard to interpret on a glance, one of the participants stated:

Norwegian:

*“Det eneste jeg var litt forvirret om er den der «Your progression», greit nok at den går oppover, meg jeg var litt usikker på hva den mente? Sånn hva var det jeg gikk opp i? Er det bare rette svar, er det kunnskap?”*

English:

*“The only thing confusing me was the «Your progression»-feature, fair enough there is an ascent, but I was a bit unsure what it represented? What did I improve at? Did it represent correct answers, is it knowledge?”*

When being asked the same question, another participant also found the graphs hard to understand by stating:

Norwegian:

*“De grafene var kanskje ikke så lette å tyde etter den første gangen når man ikke hadde så mange resultater. Man ser liksom en graf, men så er det litt vanskelig å tyde ut av den, siden man ikke har noe annet å sammenligne med. Men det vil jo egentlig være ganske intuitivt hva de grafene sier. Jeg husker ikke helt om dere hadde noe forklarende tekst eller overskrifter til de grafene jeg?”*

English:

*“The graphs were hard to interpret after the first attempt at a quiz, when you did not have many results. You can see a graph, but then it’s hard to interpret it as you do not have anything to compare you result to. However, in the end it would be quite intuitive what the graphs says. I do not remember if you had any explanatory text or title to the graphs?”*

The graphs in the activity dashboard were hard to understand, and their interconnectivity was also labeled hard to understand. When one of the participants was asked why the connection between the graphs was hard to understand, it stated:

Norwegian:

*“... det at x-aksen (på grafene i dashbordet, red. anm.) bare hadde #1, #2, #3. Jeg vet ikke, det hadde vært litt kjekt om på måte det var ekstra tekst da, som i tillegg sa dette var antall forsøk nr. da.”*

English:

*“... the fact that the x-axis (on the graphs in the dashboard, red. anm.) only had #1, #2, #3. I don’t know, but it would have been nice if there was some extra text, stating that this indicated the attempt number.”*

One of the other participants was asked directly if the participant understood that there was a connection between the two graphs in the dashboard, and the reply was similar to the previous statement:

Norwegian:

*“Jeg la ikke merke til de numrene under. Så nei, jeg gjorde ikke det.”*

English:

*“I did not notice the numbers below the graphs. So no, I did not understand that.”*

However, not all features in the system were hard to understand. There were also statements with feedback stating that most of the features were perceived as intuitive at first glance. When asked if there was anything that stood out as particularly easy to use and if the participants could say something about the use of colors, the participant responded:

Norwegian:

*“... de mastery levelene ser man ganske tydelig at gull er bra, medium sølv og low gjetter jeg på sånn bronse-aktig. Generelt så er lilla litt sånn nøytral farge synes jeg. Siden bakgrunnen er nøytral synes jeg at det hjelper på å fremheve de andre effektene sånn som f.eks. «new attempt» som lyser grønt. «Klar, ferdig, gå» her skal du begynne. Også nederst nede i rødt, gult og grønt, gir det også mening at grønt et bra, eller over gjennomsnittet eller hva det er. Gul må du jobbe litt mer og rød, her var det litt lite. Dette var også ganske intuitivt.”*

English:

*“... in the mastery levels you can clearly see that gold is good, silver is medium and I’m guessing bronze is low. In general, I think purple is kind of a neutral color. Since the background is neutral, I think it helps to highlight the other effects such as “new attempt” that glows green. «Ready, set, go», this is where you begin. Also, at the bottom, with the red, yellow and green, it also makes sense that green is good, or above average or something like that. When it is yellow you must work a little more and when getting red, its bad. This was also quite intuitive.”*

When asked if there were any features or aspects with the system being particularly easy to understand, two other participants focused more on the user interface as a whole, with the first stating:

Norwegian:

*“... det hadde en ganske nice user interface med visualiseringer som var enkle å forstå ved å bare se på de i to sekunder, for så bare å gå videre.”*

English:

*“... it was a pretty nice user interface with visualizations that was easy to understand just by looking at them for two seconds, and then you could just move on.”*

And the second participant similarly stating:

Norwegian:

*“Det er ganske simpelt design da, så det er mye som bare er naturlig å forstå hvordan det funker.”*

English:

*“The design is fairly simple, so there is a lot you just naturally understand how works.”*

### 5.6.3 Features perceived as extra useful

The third noticeable group emerged within the node named Extra useful, located within the parent node related to the features of SmartU. This group revolved around the usefulness of the features in the system. As displayed in Table 5.12, seven features in the system were explicitly mentioned when participants were asked to name one or two features in the system they perceived as extra useful.

Name	Files referenced	Statements in total
Activity Dashboard	3	4
Progression graph	3	3
Peer-comparison	6	9
Quiz adaptivity	1	1
Response Time	2	3
Quiz summary	4	6
Mastery	4	4

Table 5.12: Nodes making up the ”Usefulness”-group, with number of files referenced and number of statements.

The most mentioned feature in Table 5.12 was the peer-comparison allowing one to compare one’s results in an activity to the average of the other users of the system doing the same activity. Two of the participants who stated that the peer-comparison was useful, were also asked to explain why that feature was useful. The first participant stated that:

Norwegian:

*“... det er jo på en måte veldig interessant data, det setter jo på en måte meg opp mot de andre, og det er jo mye motivasjon også i det da.”*

English:

*“... it kind of is very interesting data, in a way it compares me to the others, and there is a lot of motivation in that.”*

The second participant answered in similar terms, stating:

Norwegian:

*“... jeg brukte i alle fall responstid og så at jeg brukte sånn ganske mye kortere tid enn andre og da burde jeg kanskje tenke: «Hmm, her bruker du så kort tid, kanskje jeg bare kan tenke litt mer på det og at jeg kanskje faktisk har feil» så det var lurt å ha.”*

English:

*“At least I used the response time and saw that I spent less time than the others and then I should probably have thought: «Hmm, here you spend a short amount of time, maybe I can just think a little more about it because I might be wrong» so it was nice to have.”*

The second most frequently mentioned feature perceived as useful was the quiz-summary, displaying one’s score and aggregated stats after submitting the final answer in an attempt at a quiz. It also enabled one to see which questions had been answered correctly and wrongly, and which answer was correct if the response was wrong. When asked to explain what was making the quiz-summary so useful, a participant replied:

Norwegian:

*“... når man var ferdig med en test, fikk man opp sånn hva som var rett og galt. Dette tror jeg er veldig lurt, i og med at man ikke får forbedret seg om man ikke vet dette ... Det også at man kunne gå inn på en gammel test å sjekke «hvilket spørsmål var dette her». Det husker jeg ganske godt.”*

English:

*“... when you finished a test, you got to see what was right and wrong. I think this is very smart, in that you do not get to improve if you do not know this ... That and that you could go on an old test to check «what question was this». I remember that quite well.”*

Another participant answered similarly to the same question stating:

Norwegian:

*“Og kanskje gjennomgangen av hva som var riktig og hva oms var galt. Så man får en god ... Da fikk jeg en sånn «dette gjorde jeg det galt på» og «dette er det jeg må jobbe mer med».”*

English:

*“And maybe the review of what was right and what was wrong. So you get a good ... Then I got a feeling like «this I did wrong» and «this is what I have to work on more».”*

The third most mentioned feature, when asked to name extra useful features in the SmartU system, was the feature assigning a mastery-level. When an attempt at a quiz was finished the user would be assigned a mastery-level on that specific attempt, and this would contribute to the overall mastery-level of the activity.

When asking two of the participants what made the mastery-level useful, the first participant replied:

Norwegian:

*“Det må vel være det at man får en på en måte level, for dere hadde jo mastery level, at du kunne ha høy prestasjon og lavere. Det var vel ganske nice å se hvordan du ligger an og hvor flink du egentlig er i pensum da.”*

English:

*“It must be that in a way you get a level, because you had the mastery level, in that you could have high performance and lower. It was nice to see how well you were performing and how good you really are in the syllabus.”*

The second participant added to this, by stating:

Norwegian:

*“Du kan se ut ifra mastery-level, ikke bare om du hadde rett, men også om du hadde rett på vanskelige spørsmål.”*

English:

*“You can see from the mastery level, not only if you were right, but also if you were right on difficult questions.”*

#### 5.6.4 Opinions on Learning-tools and gamification

The last noticeable group emerged when combining the node containing statements regarding gamification and the parent node focusing on learning-tools and digitized learning (Table 5.13).

Name	Files referenced	Statements in total
Gamification	2	3
Learning*	0	0
Sub-nodes of Learning		
Digitized learning	4	4
Tool for learning	7	12

\* - No files are referenced as the node only contains other nodes

Table 5.13: Nodes making up the ”Learning-tools and gamification”-group, with number of files referenced and number of statements.

When asking the participants if SmartU was an interesting system, and what made it so, some of the replies from the participants revolved around the aspects of digitized learning. One of the participants stated that:

Norwegian:

*“... det å få samla det på en litt bedre plass istedenfor å bare måtte hive sammen en quiz litt selv med kopierte spørsmål fra eksamenene. Fikk en liten oversikt med noen grafer å gå ut ifra med litt mer kontroll enn når man bare må gå gjennom å være fin på det selv hele tiden, da er det litt mer konkret og forholde deg til.”*

English:

*“... getting things together in a slightly better place, instead of just having to get together a quiz with copied questions from the previous exams. Initially I got a little overview with some graphs, with a little more control than when you just have to go through yourself all the time, making it a little more concrete and relatable.”*

Two of the participants also found the system interesting because it had game-like features. The first of the participants stated that:

Norwegian:

*“Altså jeg synes det var interessant fordi det minner litt mer om et spill kan du si, som er en ganske annen approach enn det systemet jeg har holdt på med tanke på ITGK da.”*

English:

*“I think it was interesting because it sort of resembles a game, which is a very different approach than the system I have been using in relation to ITGK.”*

The second participant was a little more specific pointing at the mastery-level feature, stating that:

Norwegian:

*“... det lille «ranking-systemet» dere har på midten, hvor det sier «High» eller «Medium» eller at du får en liten inndeling da, og det at du får en liten medalje, det synes jeg skaper en litt mer kul ... Ja ...”*

English:

*“... the little «ranking-system» in the middle of the page, where it says «High» or «Medium» or creating a separation of sorts, and the fact that you get a little medal, in my eyes, creates a cooler ... yeah ...”*

When asking the participants whether they would use the system again, and if so, why they would use the system, the response was mostly focused around the way the SmartU system could be used as a tool for enhanced learning. One of the participants stated that:

Norwegian:

*“Nei altså jeg er veldig fan av å bruke sånne ting istedenfor å bare bruke en bok og sitte og lese for eksempel. Så er det jo litt gøy. Det er jo en litt annen måte å lære på, det å bare få servert spørsmål og svar, i stedet for å bare sitte og lese. Det er en litt mer effektiv måte å gjøre eksamensett på, kanskje.”*

English:

*“I am a big fan of using things like this instead of for instance just sitting in reading a book. So this makes it a little fun. After all, it is a slightly different way of learning, to just be served questions and answers, rather than just sitting and reading. It’s a slightly more effective way to do exam sets, perhaps.”*

A second participant took a broader and more summarizing approach in its answer, stating that:

Norwegian:

*“Jeg ville definitivt brukt systemet, dersom det var mulig, rett og slett i de fleste fag jeg kunne brukt det til. Det ga veldig sånn, eeh... Veldig konkret rett/galt og progresjon og... Du kunne se på aspekter du ikke kunne tenkt deg frem til vanlig, som responstid. Du kan se ut ifra mastery-level, ikke bare om du hadde rett, men også om du hadde rett på vanskelige spørsmål. Eller om du bare har en superficial forståelse, med at du har rett på masse spørsmål, men det har vært lette spørsmål slik at du lurer deg selv til å tro at du kan mer enn det du egentlig kan. Og det hadde en ganske nice user interface med visualiseringer som var enkle å forstå ved å bare se på de i to sekunder, for så bare å gå videre.”*

English:

*“I would definitely use the system, if possible, in most subjects I could use it for really. It gave a, uhm ... Very concrete right / wrong and progression and... You could look at aspects you couldn’t aggregate yourself, like response time. You can interpret from the mastery level, not only if you were right, but also if you were right on difficult questions. Or if you just have a superficial understanding, in that you are right on a lot of questions, but there have been easy questions, so you are fooling yourself into thinking you know more than you really do. And it had a pretty nice user interface with visualizations that was easy to understand by just looking at them for two seconds, and then just moving on.”*



# Chapter 6

## Discussion

The findings presented in Chapter 5 provide interesting results in regard to the main hypothesis of this thesis. The hypothesis was that visualized statistics can boost learner’s motivation and thus improve their performance in self-assessments and enhance their learning outcomes. Thus, the goal of the study was to investigate how the participant students perceived the different visualizations, in terms their usefulness, how the visualizations affected their attitude and their performance. As the hypothesis was broken into three research questions, each research question will be discussed along with relevant findings in the following sections.

### 6.1 Differences in Performance

The first research question assessed the differences in performance and was articulated as “*What is the effect of different visualizations on learners’ performance in the context of an adaptive assessment system?*”.

In SmartU, there are two different ways to measure a learner’s performance. One way is to look directly at the performance on an attempt in a quiz, which is the sheer number of correct and wrong answers. However, this measure is not taking the differences in the question-difficulty into account, introduced by the adaptive nature of the quiz, as described in Appendix A. This way of measuring performance is thus used to evaluate how well a student performs on a specific attempt, which does not necessarily indicate how well the student knows the syllabus. The second approach to measure performance is to look at the mastery-level assigned to a learner in an activity upon finishing an attempt at a quiz. This measure of performance takes the question-difficulty into account, is represented on a 3-point scale (High, Medium or Low), and is used to indicate a learner’s mastery of the syllabus.

Similarly to previous studies conducted by Jivet et al. [2018] and Papamitsiou

et al. [2019], the learners' performance could be used to separate the participants into performance-groups. However, as the results in Table 5.5 and Table 5.7 shows, along with the lack of a significant correlation between the GVA and performance (Table 5.6), there is no evidence that the visualized statistics provided during a question had any statistically significant effect on performance.

Despite that there was no statistically significant difference in performance, it is worth noting that the experimental group had a marginally higher average mastery-level than the control group (Figure 5.2), indicating that the overall mastery of the course syllabus used as test-content was higher for the experimental group than for the control group. However, the number of correct answers was marginally higher for the control group than for the experimental group. This could be explained by the adaptivity of the self-assessment: the students in the experimental group had a higher mastery-level, and thus, harder questions were assigned to the students in experimental group than to the students in the control group. It is also worth noting that the control group had an average mastery-level of 2.53 and the experimental group had an average of 2.58 (Table 5.5). With the highest possible mastery-level being 3.00, both groups scored very high, making it hard to improve.

Although the results from the quantitative analysis lacked statistically significance in relation to the visualizations affecting the performance of the participants, the qualitative analysis provided some interesting findings in this regard. As presented in Section 2.3.4, Sedrakyan et al. [2018a] claim that learning processes can be positively influenced by dashboard feedback if based on underlying mechanics of learning processes. Looking at Table 5.12, the features perceived as the most useful for the participants were the *peer-comparison*, *quiz-summary*, *mastery-level* and the *activity dashboard* itself. These features are exclusively providing information about performance, either in a specific attempt or over time. The feedback provided in these features is in line with the two examples mentioned in Sedrakyan et al. [2018a], and could contribute to improving the learning outcomes, and thus affect the performance of the participants. Looking at the mean of OEUS, AGV and UGV (Table 5.3), and the way the variables significantly correlate with each other and with effort (Table 5.6), some of the statements from the interviews regarding the usefulness become even more interesting. In Section 5.6.3, the statements provided by the participants regarding the quiz-summary feature indicate that the immediate feedback received upon finishing a quiz triggered self-reflection regarding their own performance. Seen in combination with the statements regarding the mastery-level and peer-comparison, one could argue that the participants both used and understood the visualizations and found them useful. Thus, the visualizations could contribute to an increased learning outcome via the self-reflection upon finishing a quiz and the increased self-awareness of own

performance provided by the mastery-level and graphs in the activity dashboard.

## 6.2 Usefulness of Visualizations

The second research question revolved around the usefulness of the visualizations, as follows: “*What is the learners’ perceived usefulness of visualizations in an adaptive interface?*”.

Looking at the list of features for the system proposed in literature (Table 2.1), all features considered and implemented in the current version of SmartU were based on the associated previous research. The features mentioned in Table 2.1 are mainly features designed to improve learners’ awareness of their performance, either in a specific attempt or over time. Among those features are the three most useful features in SmartU as perceived by the participants, based on the conducted interviews. Table 5.12 shows the number of references made to the different mentioned features when the participants were asked to name one or two particularly useful features in the SmartU system. Peer-comparison, Quiz-summary and Mastery-level are the features having the highest frequency in this list, as mentioned previously in Section 6.1.

However, the visualized task-related statistics available during an attempt at a quiz were not mentioned once when the participants were asked for the most useful features. This finding makes sense, as the mean-value of GVA is low compared to the other variables in Table 5.3. As Table 5.4 shows, the questions about GVA were available only to the experimental group, as this is what separated the groups, as explained in Section 4.1.3. Despite not statistically significantly different, looking at Figure 5.2, it is apparent that the effort spent on tasks is higher for the experimental group than for the control group. However, there was no correlation between GVA and high effort, and the increase in effort did not seem to affect the performance. This can be explained by the fact that effort was measured by logging if the participant spent more time answering the question than the effort-threshold for the question. Thus, by adding the visualized task-related statistics to the system, the participants in the experimental group would naturally spend more time on average answering a question as they were using the extra feature. This means that the participant had to see the visualized statistics and interpret the graphs and visualizations before answering the question, thus spending more time overall.

The results from the interviews combined with the results from the statistical analysis indicate that, in the version of the system used by the control group, the features containing visualizations and graphs based on statistical data from the participants’ performance, were well received and thought of as useful features. The findings about these features being perceived as useful also align with the

findings from previous research (Table 2.1). Roberts et al. [2017] suggested that participants found that early alerts about performance, peer-comparison of grades and a color-system that indicates the performance at a glance were features perceived as useful. The present study reports the same findings for the alerts about performance and peer-comparison and the colors used in the different features. Furthermore, the present study also confirms the recommendations in Jivet et al. [2018] regarding the peer-comparison and adaptiveness, as discussed in Section 6.3. Ruipérez-Valiente et al. [2017] and Ruipérez-Valiente et al. [2015] introduced LAD extensions to existing learning platforms with features meant to promote self-reflection. The features implemented in SmartU based on these articles were explicitly perceived as useful in the interviews conducted in this study. On the other hand, the visualized task-related statistics were not perceived as useful based on the conducted interviews, despite containing several of the aspects from Table 2.1. The answers from the questionnaire indicated the visualized task-related statistics were not perceived as useful, as they only increased the effort being put into tasks without a corresponding increase in the mastery-level or performance.

### 6.3 Factors Affecting Attitude and Motivation

The third and last research question revolved around the attitude and motivation towards the visualizations as follows: *“What are the differences in learners’ motivations/attitudes towards systems that provide visualizations”*.

Although not significantly different, it is perceivable from the results of the Independent Samples T-tests in Figure 5.2 and Table 5.8 that the mean of AGV, concerning the participants’ attitudes towards the graphs and visualizations, was lower for the experimental group than for the control group. OEUS and UGV were also lower in the experimental group, and as mentioned previously in Section 6.2, the mean of GVA was much lower than the mean of the other questionnaire variables. This means that the participants in the experimental group expressed less satisfaction with the system than the control group, and that they found the visualized task-related statistics less satisfying than the other parts of the system. However, the differences in the scores from the questionnaire were not statistically significant as shown in Table 5.8. Furthermore, the scores for both groups, except for GVA, were also in the high end of the 5-point Likert scale (Table 5.4). On the other hand, the results from the SUS-schema show that the experimental group had a higher SUS-score than the control group, and that 8 out of the 10 questions had a higher score in the experimental group. This finding indicates that the experimental group found the system more usable than the control group, which contradicts some of the findings presented from the questionnaire. However, with such a low and homogeneous sample-size, the results can easily be affected by

personality types and outliers in the responses [Lewis, 2018].

According to Bangor et al. [2008], products with at least a SUS-score of 70 are passable, the better products are scoring between the high 70's and high 80's, while truly great products score higher than 90. From the results in Section 5.5, the average participant found the usability of the system high, as similarly indicated by OEUS. Bangor et al. [2008] also states that, despite having insufficient data to link individual success metrics to a high SUS-score, anecdotal evidence in their research suggests that a good task-performance could be correlated with a high SUS-score. No such correlation was found in this research, but as stated earlier, both the average performance/mastery-levels and the average SUS-scores were high. However, in line with the anecdotal evidence of Bangor et al. [2008], a trend that was discovered was that most of the SUS-scores from 75 and below were given by participants having a performance below the average performance (experimental N=3 and control N=5). Thus, it could be assumed that the attitude, motivation and perceived usability of these participants were affected by a sub-average performance in terms of the percentage of correct answers. However, the sample size of participants lacks the statistical power to prove this.

From the interviews it was clear that the participants had more statements regarding potential positive influences in their motivation than negative (Table 5.10). Most of these statements were concerning participants comparing results, either comparing against their own previous results or against their peers. The significant correlations between AGV and UGV and effort (Table 5.6), seen in relation to the findings from the interviews, could indicate that the attitude towards the graphs and visualizations and their perceived usefulness affected the effort.

As mentioned in Section 6.2, the peer comparison feature was deemed the most useful feature in the system. This feature was also particularly interesting regarding the attitude and motivation of the participants, and it can be connected to the correlation between AGV and UGV and effort. Specifically, in the results from the interviews, there were statements claiming that when receiving information about the aggregated average score of the other participants and one's own previous performances, a competition naturally emerges. The creation of goals like *being better than the average participant* or *beating your own high score* is a natural reaction to this. Thus, the results from the interviews could indicate that the graphs and visualizations displaying one's score and the peer comparison feature were contributing to the positive influence in motivation, which again led to the high level of effort on tasks.

However, Jivet et al. [2018] stated that peer comparison should be used cautiously, as different types of learners can perceive it differently. Their research found a connection between academic performance and the perception of social comparisons. The study conducted in this thesis did not include interviews of

all participants, and there were no explicit questions in the questionnaire asking about motivation regarding the peer comparison. Thus, there is no basis to make the same assumptions based on statistics, as there is insufficient data to compare mastery-level or performance against the perceptions of the peer comparison feature. However, based on the interviews alone, one can find similarities to the findings in Jivet et al. [2018], as in the present study, the participants stated that their motivation would change based on their results in the peer comparison. Some participants stated that being ahead of the average curve in the peer comparison would influence their motivation positively, as it was perceived as a “confidence boost”. Others, however, claimed that their motivation would be higher if positioned slightly behind the average curve, as it was a way of further improving their own skills more than the others. What recurred in most interviews was that most participants would feel demotivated if positioned too far behind the average curve. This was especially the case in specific contexts, like when being positioned far behind the average curve when practicing towards an exam of a course close to the examination date.

## 6.4 Research Limitations

Although the different parts of the thesis mostly went according to plan, there were steps of the research potentially not being planned carefully enough. The global outbreak of the Covid-19 pandemic also reached Norway, which forced the research to diverge from the original plans. These aspects will be covered in the following sections.

### 6.4.1 Covid-19 and its implications

While conducting the user-tests and observations as described in Section 4.1.3, the outbreak of Covid-19 reached Norway and ultimately forced the University to shut down the campus. This had an impact on various aspects of the study conducted.

#### **Final sample-size of participants**

The most notable implication of the University abruptly shutting down the campus was the fact that students and staff were not allowed to stay on campus. This happened during one of the days of the testing and implied that the remainder of the tests had to be canceled. Furthermore, the days before the testing started, some of the participants canceled their scheduled testing due to fear of the high risk of infection Covid-19 has. Although these participants were willing to conduct the tests later when Covid-19 did no longer pose a threat, the time-frame of the

study did not allow for this to happen. In total, Covid-19 had a big impact on the sample size of the participants, as the total number of participants went from 40, as originally planned, to 27. The low sample-size means that the statistical power behind the statistical analysis is low. The results from the Pearson correlations and the Independent Samples T-tests, should therefore be carefully considered against the sample-size as the correlations might not reflect a true effect. There is also a possibility that other true effects within the results could not be discovered.

### **Physical attendance needed for testing**

Another implication of Covid-19 was the strict demands for disinfection and general hygiene, as the testing was conducted in a controlled environment, requiring physical attendance. As previously stated in Section 4.1.2, all laptops and other equipment used during testing were thoroughly cleansed with antibacterial wipes between every conducted test. Furthermore, the table and chairs used by the participants were also cleaned between tests. This was a time-consuming routine and caused delays in the testing as the day progressed, due to a tight schedule, not prepared with disinfection of equipment in mind.

As the number of people infected by Covid-19 increased rapidly in Norway during the days that the experimental procedure was taking place, the uncertainty about what was going to happen was also a factor with a side-effect on the testing. Although not measured, the mood and atmosphere during the testing were pressed, most likely due to the virus. From the observation notes, all participants seemed concentrated and were quiet. However, during small-talk after completing the testing, some of them expressed concerns for being infected at the campus.

### **Interviews conducted over video**

As described in part 3 of Section 4.1.3, semi-structured interviews were conducted with selected participants. However, with a closed campus it was not possible to conduct these interviews face-to-face. Thus, Whereby and Discord were used to conduct the interviews. As mentioned, this allowed the interviewers to see the interview object, and recording of the conversation was made easy. However, some aspects to a face-to-face conversation, such as the ability to observe the body language and the fine-tuned facial mimics, were lost. Furthermore, as all participants had to stay in their homes while being interviewed, the setting and ambiance of the interview were changed. The mindset and concentration of the participants were thus most likely influenced by Covid-19 and the implications it had for the participants on a personal level.

### Working from home

In addition to the interviews being conducted remotely via video, the rest of the work with the data collection and analysis also had to be performed with the researchers being at separate locations. This caused extra challenges due to communication and collaboration being more difficult and limited exclusively over web-based services. This was especially relevant for the analysis of the data, where the researchers had to perform the coding of interviews and the statistical analysis while being apart. Thus, brainstorming and sharing opinions and explanations were more difficult, leading to the process being more time consuming and tedious than originally planned.

### 6.4.2 Diversity in participants and test activities

In retrospect, the selection of the course used as the activity for the evaluation of the system was sub-optimal. The course, i.e., TDT4100 – related to programming in Python, is a course in which computer science students had a clear advantage regarding performance in the activity, despite that the other study-programs also include this course in the curriculum for the first semester. This could have led to the less experienced participants being influenced by their attitudes towards the course and their general knowledge in Python, instead of focusing on the system and its features. Because the graphs, visualizations and mastery level are set and adjusted based on the participants' performance, having less knowledge in Python could potentially cause wrong answers, which in turn could be demoralizing and ultimately influence their understanding of the features of the system.

Looking at the descriptive statistics in Section 5.1, it is also apparent that, based on the study-program, the sample of participants was not very diverse, as 77% of the participants were enrolled in a study program related to programming.

### 6.4.3 Inexperienced researchers

The researchers of the thesis are not experienced ones, and the planning and conduct of the research were affected by this. In retrospect, it is clear that the creation and conduct of the questionnaire and the interviews in particular could have benefited from the researchers having more experience. The constructs used in the questionnaire were not as consistent as desired and the questions could have been articulated better to avoid unintended bias when being answered. In addition to this, one could argue that an additional questionnaire should have been given to the participants before the testing, to establish their attitudes, motivations and opinions before using the system. After transcribing and decoding the interviews, it became evident that the conduct of the interviews could have been executed



better. Some questions were articulated poorly by the interviewers and ended up as closed-ended questions, being answerable by yes or no, instead of being open questions and requiring the interviewees to provide a rationale of their thoughts or opinions.



# Chapter 7

## Conclusions & Future Work

The research hypothesis proposed that “visualized statistics can boost learners’ motivation and thus improve their performance in self-assessments and enhance their learning outcomes.” This hypothesis was broken down into three research questions, investigating the effect the different visualizations have on the learners’ performance, the perceived usefulness of the visualizations and the differences in learners’ motivation towards systems providing such visualizations. In order to address these questions a triangulation of research strategies was used. First, an artifact, i.e., SmartU, was designed and developed; next, SmartU was tested in a controlled environment, during an experimental study with 27 participants.

The results from this study, as discussed in Chapter 6, can be separated into two categories, the *strengths* and the *weaknesses* of the SmartU system.

### Strengths

Firstly, the average SUS-score for SmartU indicates that SmartU had good overall perceived usability. The graphs and visualizations used throughout the system had an expressed impact on the self-awareness of the participants, triggering self-reflection upon finishing an activity, potentially resulting in an increased learning outcome. Most features in SmartU providing graphs and visualizations were perceived as useful by the participants, with one exception, further discussed in the weaknesses. Furthermore, there were indications that particularly the graphs and visualizations used to compare results to peers affected the participants’ motivation and attitude. When provided in an appropriate situation and context, to a specific group of performers, these statistics triggered competitive behavior, increasing the effort put into tasks.

## Weaknesses

One of the features providing graphs and visualizations, the visualized task-related statistics, was not perceived as useful by the participants. Furthermore, despite a high average performance from participants in the study, there was no clear (explicit) indication that the graphs and visualizations improved the performance. However, what became apparent from the interviews was that comparison to peers could contribute negatively to the motivation of participants if provided in unfavorable situations or to an unsuitable group of performers.

## Future Work

Within the strengths and weaknesses of the study with SmartU lies opportunities for further research and improvement of the system. As seen in Section 6.4, the study also has limitations, some caused by the researches, others by external forces. The sum of this makes some interesting points for further research and development that would build on and iterate further on the current version of SmartU, while providing meaningful and interesting future work directions:

- Conduct more experiments with a bigger and more diverse sample of participants, utilizing other courses than the one used in this study, and making sure that the study-program does not have ascendancy over other study-programs, as the tendency was in this study.
- Investigate the effect of gamification elements in the SmartU system, i.e., through user-created challenges, awards or badges, as it was identified as a factor contributing to motivation from the interviews.
- Test the mobile and desktop versions and compare the effect the type of device has on accessibility, attitude and motivation to further use, perceived usefulness and effect on performance.
- Investigate further why the visualized task-related statistics were not perceived as useful, what kind of statistics should be used, and how such features could be implemented in a way that is intuitive to the users.

In addition, the usability of SmartU could be further improved with future development based on the results of this study and better accommodate the initial requirements:

- Continue design iterations and development for the mobile version of the system, to create a seamless experience regardless of device.

- Enhance the clarity and clarify the purpose of available tools and features during the quizzes, e.g., the visualized task-related statistics, to easier be perceived at a glance.
- Enhance the clarity and clarify the purpose of the visualizations in the activity-dashboards, e.g., through information boxes, providing information on what the visualization represents and the origin of the data used in the visualization.
- Include several possibilities of user authentication (i.e. Feide) for ease of use.



# Bibliography

- Bangor, A., Kortum, P. T., and Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction*, 24(6):574–594.
- Bodily, R., Ikahihifo, T. K., Mackley, B., and Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education*, 30(3):572–598.
- Bodily, R. and Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, 10(4):405–418.
- Bradner, S. O. (1997). Key words for use in RFCs to Indicate Requirement Levels. RFC 2119.
- Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.
- Brown, M. (2012). Learning analytics: moving from concept to practice. *EDUCAUSE Learning Initiative*, 7.
- Chiu, C.-M. and Wang, E. T. (2008). Understanding web-based learning continuance intention: The role of subjective task value. *Information & Management*, 45(3):194–201.
- Discord (2020). Discord - chat for community and friends. <https://discord.com/>, Accessed on 2020-05-20.
- Docker (2020). Package software into standardized units for development, shipment and deployment. <https://www.docker.com/resources/what-container/>, Accessed on 2020-04-22.
- Dybå, T. and Dingsøy, T. (2008). Empirical studies of agile software development: A systematic review. *Information and software technology*, 50(9-10):833–859.

- Figma (2019). Figma, the collaborative interface design tool. <https://www.figma.com/>, Accessed on 2019-12-03.
- Github (2019). Github, the worlds leading software development platform. <https://github.com/>, Accessed on 2019-12-03.
- Glinz, M. (2007). On non-functional requirements. In *15th IEEE International Requirements Engineering Conference (RE 2007)*, pages 21–26. IEEE.
- Google (2019). Google drive, a safe place for all your files. <https://www.google.com/drive/>, Accessed on 2019-12-03.
- Illeris, K. (2018). An overview of the history of learning theory. *European Journal of Education*, 53(1):86–101.
- ISO 24765:2017(E) (2017). *International Standard - Systems and Software Engineering–Vocabulary*. IEEE Std. IEEE.
- Jivet, I., Scheffel, M., Specht, M., and Drachsler, H. (2018). License to evaluate: preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pages 31–40. ACM.
- Johnson, R. B., Onwuegbuzie, A. J., and Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2):112–133.
- Klein, C., Lester, J., Nguyen, T., Justen, A., Rangwala, H., and Johri, A. (2019). Student sensemaking of learning analytics dashboard interventions in higher education. *Journal of Educational Technology Systems*, 48(1):130–154.
- Laerd Statistics (2020a). Independent t-test using spss statistics. <https://statistics.laerd.com/spss-tutorials/independent-t-test-using-spss-statistics.php>, Accessed 20.04.2020.
- Laerd Statistics (2020b). Pearson’s product-moment correlation using spss statistics. <https://statistics.laerd.com/spss-tutorials/pearsons-product-moment-correlation-using-spss-statistics.php>, Accessed 20.04.2020.
- Lee, M.-C. (2010). Explaining and predicting users’ continuance intention toward e-learning: An extension of the expectation–confirmation model. *Computers & Education*, 54(2):506–516.
- Lewis, J. R. (2018). The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, 34(7):577–590.



- Microsoft (2020). Sharepoint, team collaboration software tools. <https://www.microsoft.com/en-us/microsoft-365/sharepoint/collaboration>, Accessed on 2020-04-22.
- Musante, K. and DeWalt, B. R. (2010). *Participant observation: A guide for fieldworkers*. Rowman Altamira.
- National Council on Measurement in Education (NCME) (2017). Glossary. [https://web.archive.org/web/20170722194028/http://www.ncme.org/ncme/NCME/Resource\\_Center/Glossary/NCME/Resource\\_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorA](https://web.archive.org/web/20170722194028/http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061#anchorA).
- Oates, B. J. (2005). *Researching information systems and computing*. Sage.
- Open Broadcaster Software (2020). Open broadcaster software. <https://obsproject.com/>, Accessed on 2020-05-20.
- Oppenheim, A. N. (2000). *Questionnaire design, interviewing and attitude measurement*. Bloomsbury Publishing.
- Papamitsiou, Z., Economides, A. A., and Giannakos, M. N. (2019). Fostering learners' performance with on-demand metacognitive feedback. In *European Conference on Technology Enhanced Learning*, pages 423–435. Springer.
- Papamitsiou, Z. K. and Economides, A. A. (2013). Towards the alignment of computer-based assessment outcome with learning goals: the laers architecture. In *2013 IEEE Conference on e-Learning, e-Management and e-Services*, pages 13–17. IEEE.
- Prosser, M. and Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education*. McGraw-Hill Education (UK).
- ReactJS (2019). React, a javascript library for building user interfaces. <https://reactjs.org/>, Accessed on 2019-12-03.
- Roberts, L. D., Howell, J. A., and Seaman, K. (2017). Give me a customizable dashboard: Personalized learning analytics dashboards in higher education. *Technology, Knowledge and Learning*, 22(3):317–333.
- Robertson, S. and Robertson, J. (2012). *Mastering the requirements process: Getting requirements right*. Addison-wesley.
- Roca, J. C., Chiu, C.-M., and Martínez, F. J. (2006). Understanding e-learning continuance intention: An extension of the technology acceptance model. *International Journal of human-computer studies*, 64(8):683–696.

- Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., and Kloos, C. D. (2015). Alas-ka: A learning analytics extension for better understanding the learning process in the khan academy platform. *Computers in Human Behavior*, 47:139–148.
- Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Pijera Díaz, H. J., Santofimia Ruiz, J., and Delgado Kloos, C. (2017). Evaluation of a learning analytics application for open edx platform. *Computer Science & Information Systems*, 14(1):51–73.
- Schumacher, C. and Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78:397–407.
- Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., and Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1):30–41.
- Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., and Kirschner, P. A. (2018a). Linking learning behavior analytics and learning science concepts: designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*.
- Sedrakyan, G., Mannens, E., and Verbert, K. (2018b). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Visual Languages and Computing*, 50.
- Siemens, G. and Gasevic, D. (2012). Guest editorial-learning and knowledge analytics. *Journal of Educational Technology & Society*, 15(3):1–2.
- Vaishnavi, V., Kuechler, W., and Petter, S. (2004). Design science research in information systems. *January*, 20.
- Vieira, C., Parsons, P., and Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, 122:119–135.
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., and Mislevy, R. J. (2000). *Computerized adaptive testing: A primer*. Routledge.
- Whereby (2020). Video meetings, video conferencing and screen sharing. <https://whereby.com/>, Accessed on 2020-05-20.
- Yoo, Y., Lee, H., Jo, I.-H., and Park, Y. (2015). Educational dashboards for smart learning: Review of case studies. In *Emerging issues in smart learning*, pages 145–155. Springer.

# Appendices



# Appendix A

## Study

## A.1 Form of Consent

### **Request for participation in research project**

#### **SmartU**

##### **"User-testing of learning analytics dashboard"**

###### **Background and Purpose**

SmartU is a research project which aim is to investigate the effects visualizations have on learners/students when applied to an adaptive interface in a course at university. In relation with the research project, a learning analytics dashboard (LAD) have been developed as a mean to test the visualizations in a real life context. The user-test will contain hands-on testing of the LAD at NTNU Campus Gløshaugen in Trondheim, Norway, where the students will get familiar with the dashboard and test themselves in the available self-assessments within the system.

The participants of our project will be students at NTNU Campus Gløshaugen in Trondheim, Norway, volunteering to participate in the user-testing.

The responsible for the project will be an Associate Professor at the department of Computer and Information Science (IDI) at NTNU, Trondheim, Norway. (see general information section).

###### **What does participation in the project imply?**

For the purpose of the research project, data will be collected using observations, web-based questionnaires. Questions that will be used for the questionnaire, will concern participants' attitudes (e.g. sensemaking, satisfaction, easiness, difficult/easy/challenging parts of the activity) toward the use of visualizations in the LAD. Data regarding participants' name, email, gender, age, year of study, line of study will be collected using a questionnaire on mobile phones/tablet or computer belonging to NTNU.

The duration of the user-testing will be approximately 30 minutes, consisting of the mentioned hands-on user-testing of the system and a follow-up questionnaire prior to the user-testing.

Participants can request to see the questionnaire and ask for any additional information regarding any other data collection instrument before giving consent.

**What will happen to the information about you?**

All personal data will be treated confidentially. Only the project group (see general information section below) will have access to the personal data. The list of names of the participating students will be stored in NTNU Sharepoint according to the data processing agreement between NTNU and Microsoft. Only the researchers and data controller will have access to this site.

We state that the participants will not be recognizable in the publication. The project is scheduled for completion by June 2020, then all data will be anonymized.

**Voluntary participation**

It is voluntary to participate in the project, and you can at any time choose to withdraw your consent without stating any reason. If you decide to withdraw, all your personal data will be made anonymous.

**Participants' rights**

Participants have the right to request access to/deletion/correction/limitation of personal data, the right to data portability, and the right to send a complaint to the Data Protection Officer at NTNU or The Norwegian Data Protection Authority about the processing of personal data.

**General information-project group:**

The leader of the project is Michail Giannakos, Associate Professor at Department of Computer and Information Science at NTNU, e-mail: [michailg@ntnu.no](mailto:michailg@ntnu.no), address: Sem Sælands vei 9, IT-bygget \* 103, phone number: +47 73593469.

If you would like to participate or have any questions concerning the project, please contact:

Martin Lunde, e-mail: [martlun@stud.ntnu.no](mailto:martlun@stud.ntnu.no), mobile number: +47 46743379  
Master student at the Department of Computer and Information Science at NTNU  
and/or

Jakob Westermoen, e-mail: [jakobwe@stud.ntnu.no](mailto:jakobwe@stud.ntnu.no), mobile number: +47 98869122  
Master student at the Department of Computer and Information Science at NTNU.

Data Protection Officer (Personvernombud) at NTNU (Thomas Helgesen,  
[thomas.helgesen@ntnu.no](mailto:thomas.helgesen@ntnu.no))

The study has been notified to the 'NSD – The Norwegian Centre for Research Data AS ([personverntjenester@nsd.no](mailto:personverntjenester@nsd.no), 55 58 21 17) has assessed that the processing of personal data in this project is in accordance with data protection legislation.

**Consent for participation in the study**

I have received information about the project and I am willing to give my consent for my participation.

Participant's name: \_\_\_\_\_

\_\_\_\_\_  
(Signed by participant, date)



## A.2 Introduction To Adaptiveness and Task Related Analytics

### Adaptive self-assessments:

Self-assessment involves processes that allow students to evaluate their own progress and identify their own skill gaps and where their knowledge is weak. To further personalize the self-assessment in smartU, the activities are **adaptive**: they are designed to best fit the individual's expertise based on the previously provided responses, and to successfully detect and determine the individual's mastery level, using a minimum number of appropriate tasks.

Specifically, the activities involve multiple-choice questions with 2-4 possible answers per question - where only one is the correct. The questions are delivered to you one-by-one (one at a time). Each time you answer a question, smartU considers the correctness of your response and the difficulty of the question and decides which is the next most appropriate question for your mastery level. The goal is to detect your mastery level using the least number of needed questions. Due to the adaptivity, each time you take a self-assessment (i.e., in each *attempt*), the number of questions that will be assigned to you varies. The adaptivity works as follows:

- There is a finite number of questions in smartU's question-bank.
- Each question belongs to one of three categories of difficulty, with a different weight on each.
- The first question is randomly assigned to you.
- When you answer a question, the correctness of your response, the weights of difficulty of the question, and the weights of difficulty of all non-answered questions, are used to decide the next question you will receive. Please, note that (in each question) you **cannot change your answer** after submitting it, and you **cannot return to a previous question** that you have already answered: **you can only move to the next question**.
- When you submit an answer to a question, you will not be asked the same question again in future attempts. Note that this only affects the number of questions available to you, and not to the other users.
- The minimum number of questions that will be assigned to you in an attempt is 10, and the maximum is 20.

Due to the adaptive nature of the self-assessment and the finite number of questions in the question-bank, if you are a high-performer ("experts"), you are likely to run out of questions soon! You already master the content!

### Adaptiveness of the interface:

Based on your performance in the activities, the dashboards will change according to your mastery-level. This means that users with a high mastery-level potentially will have access to features that users with a low or medium mastery-level cannot access. This also applies to the different activities you initiate: before starting the activities, the interface is the same for all students. This can also work the other way around, depending on whether you initiate a new activity. Note that for research purposes, access to some features can be deliberately given or subtracted.

### Task-related analytics (“Statistics”):

The task-related analytics are analytics about the specific task (question) that are available along with the question via the “Statistics”-button to the right in the purple navigation-bar.

You can use the task-related analytics every time you want to infer the real requirements of a question in terms of its difficulty, the time needed to answer it, and the probability to give a correct answer, based on the question’s “history” data from peers having been collected and analyzed. This will help you self-regulate your efforts on answering the questions, better understand the true needs of the questions, and score well to the questions, accordingly.

The “Statistics” feature provides access to 4 visualized statistics based on the answers previously provided to this specific question, by all other users that have answered this question. The available statistics are Effort, Performance, Correct/Wrong answers and Time to Answer. The question mark (?) next to a visualization shows a description/explanation of the specific visualization. An explanation is also given here:

**Effort** - the average peers putting effort into this question (in percentage) (e.g., 3 peers put effort in / 4 total peers having solved the question = 75%). Effort is calculated based on one’s time-spent on answering the question, i.e., the time allocated trying to give a correct answer. Effort is an indicator of the “real” question difficulty: one can infer the difficulty of the question from the peers’ effort on it. Higher effort indicates a more difficult question.

**Performance** - the average performance (in percentage) on this question based on previous answers (e.g., 300 correct/400 total = 75% performance). Performance is an indicator of the “real” question difficulty: based on how many peers have responded to this question correctly, one can infer the difficulty. Higher performance indicates a less difficult question.

**Correct/wrong answers** - This visualization shows the differentiation of correct and wrong answers in percent. It is based on all previous answers to this question.

**Average time to answer** - This visualization shows you how many minutes have been used in order to solve this task previously. There are three columns in the graph:

- Green for the average time to provide a correct answer
- Red for the average time to provide a wrong answer
- Purple for the average time to an answer in total

The average time to answer guides you to regulate your time: if it takes significantly longer/less than the average values, it’s a sign that this question might need your attention.

## A.3 SmartU Post-Experiment Questionnaire




Smartu

\* Obligatorisk

### Personalia

Personalia is needed in order to contact you for further help, and in order to categorize test-results

1

Name \* 

Skriv inn svaret

2

Email address \*

Skriv inn svaret

3

Age \*

Tallet må være mellom 18 ~ 45

4

Gender \*

- Female
- Male
- Prefer not to say
- Annet

5

Line of study (E.g BIT, MIT, MSIT, MTD, ...) \*

Skriv inn svaret

6

Year of study \*

Tallet må være mellom 1 ~ 7

7

Username used in the SmartU-system \*

Skriv inn svaret

Neste


Smartu

\* Obligatorisk

Usability of SmartU

The System Usability Scale (SUS) is a reliable tool for measuring the usability of a user-interface. It consists of a 10 item questionnaire with five response options for respondents; from Strongly agree to Strongly disagree.

8

System Usability Scale \* 

	Strongly disagree		Neutral		Strongly agree
I think that I would like to use SmartU frequently.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found SmartU unnecessarily complex.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I thought SmartU was easy to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think that I would need the support of a technical person to be able to use SmartU.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found the various functions in SmartU were well integrated.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I thought there was too much inconsistency in SmartU.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would imagine that most people would learn to use SmartU very quickly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Strongly disagree		Neutral		Strongly agree
I found SmartU very slow or complicated to use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt very confident using SmartU.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I needed to learn a lot of things before I could get going with SmartU.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**9**  
Overall evaluation of the usability of SmartU \*

	Strongly disagree		Neutral		Strongly agree
I think that the navigation in SmartU was nearly effortless	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think that when I needed help to learn how to use SmartU, the system provided me with sufficient information.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think that a user who has never seen SmartU before can learn how to accomplish basic tasks fast.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think that experienced users should be able to complete tasks using SmartU quickly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think that it is not frequent that users make errors while using SmartU.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Strongly disagree		Neutral		Strongly agree
I think that the interaction with SmartU is clear and understandable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using SmartU makes me happy to accomplish my self-assessment tasks.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using SmartU gives me enjoyment for my continuous learning.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using SmartU leads me to exploration in my continuous self-assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[Tilbake](#) [Neste](#)

Utlever aldri passordet ditt. [Rapporter misbruk](#)

Dette innholdet er opprettet av eieren av skjemaet. Data du sender inn, sendes til skjemaeieren.

Lvert av Microsoft Forms | [Personvern og informasjonskapsler](#) | [Bruksvilkår](#)

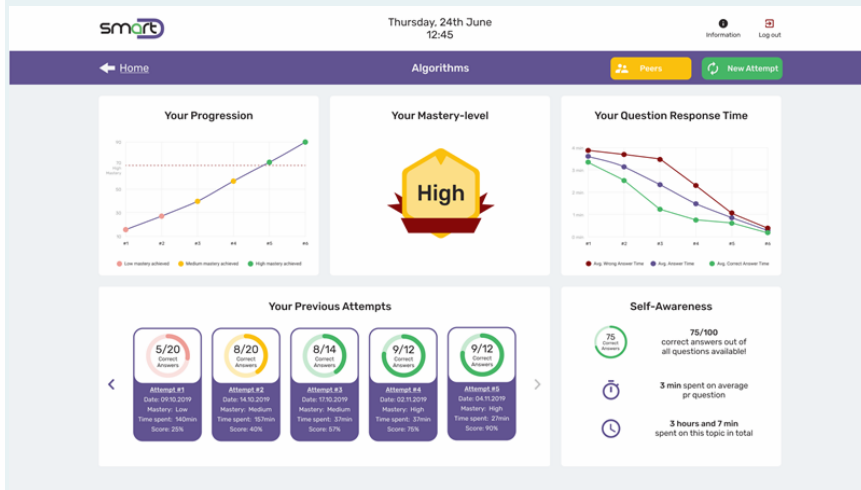
Smartu

\* Obligatorisk

Graphs and visualizations used in the system

This section consists of questions regarding the usage of graphs and visualizations in the graphical user interface of the SmartU system and NOT the available statistics during an assessment.

The answers will be on a scale from 1 to 5, press the option that is the closest match to your experience.



10

Usage of the available visualizations \*

Strongly disagree                      Neutral                      Strongly agree                      Not relevant

It was easy to understand that the graphs reflected my skill-level







	Strongly disagree		Neutral		Strongly agree	Not relevant
By utilizing the visualizations continuously I gain better understanding of my own progress	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The design and colors in the visualizations made me feel happy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The design and colors in the visualizations made me feel confident and comfortable while using SmartU.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would like to continue to use SmartU to increase my skills	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

13

Other comments regarding the use of visualizations outside an assessment

Skriv inn svaret

Tilbake

Neste

Utlever aldri passordet ditt. [Rapporter misbruk](#)

Dette innholdet er opprettet av eieren av skjemaet. Data du sender inn, sendes til skjemaieren.

## Smartu


\* Obligatorisk

## Graphs and visualizations used in the system

This sections consists of questions regarding the usage of graphs and visualizations in the graphical user interface of the SmartU system and NOT the available statistics during an assessment.

The answers will be on a scale from 1 to 5, press the option that is the closest match to your experience.

14

Available activities in homepage \* 

**Available Activities**


Strongly  
disagree

Neutral

Strongly  
agreeNot  
relevant

It was easy to understand that a card represents an available activity in a course (e.g ITGK)

It was easy to understand that the "previous score"-graph represented the latest score achieved in the specific activity

	Strongly disagree		Neutral		Strongly agree	Not relevant
It was easy to understand that the mastery-level within an activity card represented the mastery-level of that activity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt motivated to achieve a high mastery-level in all available activities when using the system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found it useful to be able to quickly gain insight of my performance in a topic by looking at a card	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found the cards useful in terms of presenting important data of a topic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

15

Your progression and question response time \*



	Strongly disagree		Neutral		Strongly agree	Not relevant
I believe that the provided visualizations made information easy to extract	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found it useful to be able to see the progression of the time I spent answering questions in an assessment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was able to understand the connection between the different lines in the response time chart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found it useful to be able to find statistics from a the same attempt in both graphs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[Tilbake](#) [Neste](#)

Utlever aldri passordet ditt. [Rapporter misbruk](#)

Dette innholdet er opprettet av eieren av skjemaet. Data du sender inn, sendes til skjemaieieren.  
Levert av Microsoft Forms | [Personvern og informasjonskapsler](#) | [Brukstilvår](#)

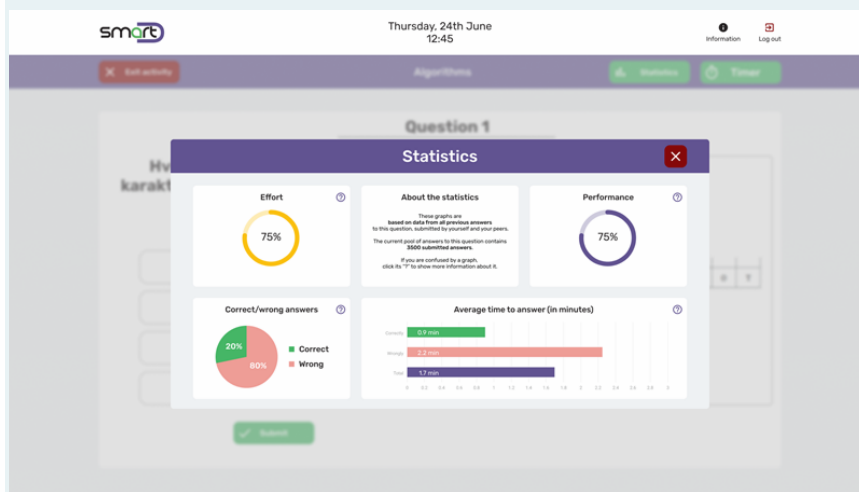
Smartu

\* Obligatorisk

Graphs and visualizations used in an assessment

This section consists of questions regarding the usage of graphs and visualizations in used in an assessment. This covers both the images used beside the answer-options and the graphs/visualizations under the "statistics".

The answers will be on a scale from 1 to 5, press the option that is the closest match to your experience.



16

Usage of the available visualizations \*

Strongly disagree

Neutral

Strongly agree

Not relevant

It was easy to understand that the visualizations in the statistics represented aggregated data based on all user's previous answers to a specific question







	Strongly disagree		Neutral		Strongly agree	Not relevant
Using the statistics during an assessment provided enjoyment and motivation for future self-awareness reflection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>19</b>						
Other comments on the available statistics during an assessment						
<input type="text" value="Skriv inn svaret"/>						
<a href="#">Tilbake</a>		<a href="#">Neste</a>				
Utlever aldri passordet ditt. <a href="#">Rapporter misbruk</a>						
Dette innholdet er opprettet av eieren av skjemaet. Data du sender inn, sendes til skjemaeieren. Levert av Microsoft Forms   <a href="#">Personvern og informasjonskapsler</a>   <a href="#">Brukstilvår</a>						

Inndeling 6 ...

Thank you for your participation!

20

Is it okay if we contact you for a short interview (10-15 mins) about your experience with SmartU? \*

Yes

No

+ Legg til ny

## A.4 Observation Guide

### Observation Guide

**Introduction:** Before the experimental study and observation begins, the observer should introduce itself and the role it will take during the experimental study. A connection should be established between the observer and the test-subject, creating a relaxed and comfortable setting for the study to be conducted in. The perfect environment for the study to be conducted in is if the study could be conducted in an environment-setting familiar to the test-subjects so they would interact with the system as they normally would.

**Interaction with test-subject:** In the role of an observer, interaction with the test-subject should be kept to a minimum, except for when the test-subject asks for help. It is important that when providing help, the help does not take too long, or become very frequent.

**Capturing contextual information:** Capture the ambience, atmosphere and environment in the room during the test. Also try to capture the test-subjects' emotional state, are they stressed/nervous/relaxed? Does the emotional state change during the test?

The aim of this point is to develop a set of notes that describe the scene as much as possible.

**Capture crucial information:** When observing, you should note what the test-subject is doing while testing. You do not have to take notes of every single action or mouse-click, but your notes should summarize the events that happened and unexpected actions the user made.

**Examples of events/actions to note:**

<b>Category</b>	<b>Includes</b>	<b>Researchers should note</b>
Test environment	Ambience, atmosphere, alterations to environment	Any differences in test-environment between tests should be noted.
Subjects emotional state	The users state of mind	Is the user stressed or nervous? Does the state of mind change during the tests? Is the test-subject focused on the task?
Unexpected events	Failing to navigate, bugs in the system	Did anything unexpected happen during testing? Note when, why and how it happened
Help	Frequency, topic of questions	What does the user ask you to help out with and how often/ many times are you asked to help?
Disturbances	People entering/leaving the room, phone notifications, noises, etc	Any unwanted noises or disturbances hindering the focus on the testing should be noted

**End of testing and observation:** It is important to thank the test-subjects for their time and to provide them with the interviewer's contact details. Depending on circumstances, it may also be worth letting respondents know how they can obtain the project reports because this provides them with a sense of ownership of the material that they have shared.

A list of the specific behaviours being observed. These can include verbal and physical behaviours which indicate the outcome. Examples of what a mild and a strong version of the behaviours look like. A place to log how often the behaviour is observed, with space to make qualitative notes describing how the behaviour was exhibited. No more than five things to observe at group level. Indication of when the observation should be performed. The duration of the observation.

## A.5 Interview Guide

### Semistructured Interview Guide

**Capturing answers:** Recording of answers will be done through taking notes and audio recording. This procedure allows the interviewer to highlight key points to probe further on relevant topics, while making sure no information is lost due to the transcripts of the recordings.

**Develop a rapport with the respondent:** Obtaining meaningful information from respondents will be easier if they are comfortable opening up to the interviewer. This can be done by asking nonprobing questions related to their hobbies, their spare time and so on.

**Ask questions that lead detailed answers:** It is important that you phrase questions in a way that gets respondents to provide detailed answers, rather than simple “Yes” or “No” answers.

**Examples of questions:**

- Did you find the SmartU system interesting? If you did, how so?
- How did you feel about the homepage and presented visualizations?
- Do you think that the system would improve your motivation of studying? If you do, why would it do that?
- Could you mention some features that you found useful in the system?
- How did you feel about the statistics provided during an assessment?
- Could you mention some features which were easy or hard to understand?
- Did you feel like the statistics helped you understand the scope of the question? How did the statistics help you?
- Would you use the system again? If so, what encourages you to do so?
- What do you think of the mastery element to visualize your knowledge?

It is good to have a set of questions on hand, but the interviewer must also be prepared to expand on, or probe, the predetermined questions as the need arises. This is the essence of qualitative interviews.

**End the interview:** Deciding when to end an interview may depend on a number of factors. E.g. interviewers may feel that they have exhausted their questions, and that they are no longer getting new information or if the respondent seems tired or has other commitments to attend to. It is good practice for interviewers to summarize the key points that they feel the respondent has provided, because this gives the respondent a final chance to expand or clarify any points. Finally, it is important to thank the respondent for their time and to provide them with the interviewer's contact details. Depending on circumstances, it may also be worth letting respondents know how they can obtain the project reports because this provides them with a sense of ownership of the material that they have shared.



## **A.6 Interviews**

The following section presents all the interviews conducted during the study in the native language Norwegian.

## A.6.1 Interview 1

### Intervju nr. 1 – (statistikk)

---

H: Syns du SmartU var et interessant system? Hvis du synes det, hvordan var det interessant?

S: Nei det var jo interessant synes jeg. Det kan det jo være nyttig for ITGK-studenter for å gå over og lære pensum, og det er vel det som er det beste med systemet. Og så at det kanskje er litt mere lavere terskel for å sette seg inn i et datasystem enn å drive å sitte og lese spørsmål og gjøre oppgaver i ITGK boka liksom.

H: Og du fikk med deg når vi bruker-testa systemet at det er ikke nødvendigvis er bare ITGK vi kan ha i systemet, du kan ha alle mulige fag?

S: ja, det fikk jeg med meg!

H: Ja, bra. Kunne du gitt konkrete eksempler på noen features du synes var ekstra nyttige som ble gitt i systemet?

S: Godt spørsmål. Det må vel være det at man får en på en måte level, for dere hadde jo mastery level, at du kunne ha høy prestasjonen og lavere. Det var vel ganske nice å se hvordan du ligger an og hvor flink du egentlig er i pensum da. Det var vel det mest nyttige. Ellers så er det også litt interessant å se på og statistikkene når du kommer ut etter testen og se åssen du har gjort i forhold til alle andre.

H: Synes du statistikken du snakker om hjalp på motivasjonen din med tanke på studiet?

S: Nei det ville jeg nok ikke si.

H: Har du noen hadde noen grunn til at det ikke påvirker motivasjonen eller var motivasjonen bare heilt uendra?

S: Den er vel ganske uendra sånn sett fordi at jeg føler jeg har ganske høy motivasjon selv i ulike fag så jeg trenger ikke nødvendigvis å se andre sin statistikk for å drive å på måte å konkurrere med dem. Da har jeg mer interessert i å konkurrere mot mine egne resultater.

H: Hva syns du om statistikken som blei tilgjengelige under et spørsmål, altså den statistikk-knappen?

S: Jeg benytta meg vel ikke så altfor mye av den. Jeg har følte egentlig ikke at den ga meg så veldig mye. Den kunne kanskje gjort det, om man virkelig var i en sånn eksamens fase da, så kunne det kanskje vært litt mere interessant å se hvordan et sånt spørsmål ligger an i forhold til hva som kanskje kan komme på eksamen og sånt noe, og hvor relevant det vil være. Men når man bare tar det litt sånn underveis, før en sånn setting, så er ikke det noe jeg bryr meg så veldig mye om.

H: Du snakket litt om statistikk i stad, etter en quiz, men det var også statistikk på dashbordet foran aktiviteter, der du så for eksempel masterien din og tidligere forsøk og sånt. Hvordan syns du det visuelle og på en måte grafer og sånt fungerte der?

S: Nei jeg synes også fungert ganske greit. Det var vel kanskje x-aksen som tok litt tid å på en måte tolke seg litt fram til, for det står jo bare nr én, nr to og nr tre. Så det var litt sånn derre confusing da så det tok litt tid på måte å tolke de grafene. Men de var jo nyttige når du først klart å forstå akkurat hva de prøvde å si da.

H: Skjønte du da at det var sammenheng mellom forsøk i den ene og den 2.?  
Hvis du har lyst så kan jeg jo vise deg grafene nå, hvis du ønsker?

S: Nei det går fint, jeg husker ganske greit og jeg synes det var en sammenheng sånn sett.

H: Ja, det var bare vanskelig å skjønne hva på måte grafen representerte i seg selv da?

S: Mer det at x-aksen bare hadde nr én, nr to og nr tre. Jeg vet ikke, det hadde vært litt kjekt om på måte det var et ekstra navn da, som i tillegg sa på dette var antall forsøk nr da. Husker ikke helt hva slags type navngivning aksene heter. Men det er litt sånn.

H: Ja, jeg skjønner. Men følte du at, ja for reference da, så gir jo grafen deg din progresjon og en spørsmåls respons tid. Syns du det var nyttige grafer? Fikk du noe ut av det liksom, eller var det bare en graf som var der?

S: Både og. Jeg er ikke så stor på benytte meg og sånne grafer selv, men selvsagt er det jo nyttig å se på en måte mønstre da, om du har brukt kortere tid på en quiz da, hvor du har flere feil og se et sånt mønster og se om det er et mønster på det da. Det er jo veldig interessant egentlig.

H: Benytta du deg av såkalt peer comparison, at du sammenligner deg sjøl mot gjennomsnittet av de andre?

S: Ja, det gjorde jeg. Det var egentlig veldig greit. Kjekt å bare se sånn nysgjerrighets messig.

H: Følte du noe ekstra motivasjon, når du så hvordan de andre lå an kontra hvordan du lå an?

S: Ikke noe mer enn han at man kanskje blir glad for at man selv eventuelt ligger bedre da enn resten. Jeg kunne vel kanskje blitt litt mer motivert og, om jeg så at jeg lå mye lavere enn alle de andre.

H: Så hvis du lå rett bak gjennomsnittet så hadde du lagt inn en ekstra innsats for å komme på området over kurven?

S: Njaaeeei, da skulle jeg vært ganske langt bak før det hadde motivert meg skikkelig. Om jeg hadde vært ekstremt dårlig i forhold til gjennomsnittet så tror jeg det kanskje hadde påvirket meg, men om jeg ligger veldig nærme gjennomsnittet så ville jeg egentlig ikke bry meg, fordi

at jeg vil nok vite selv, selv i en situasjon hvor dette systemet er i produksjon, om jeg på en måte har slurva når jeg har gjort spørsmålene eller ja sånne ting.

H: Så det hadde ikke påvirka deg hvis du lå cirka på gjennomsnitt, men om du hadde lagt langt bak, da du lagt inn et ekstra gir? ok kult.

## A.6.2 Interview 2

### Intervju nr. 2

---

H: Synes du SmartU var et interessant system? Hvis du synes det, hvorfor var det interessant? Eventuelt hvorfor var det ikke interessant?

S: Ja, jeg synes konseptet er ganske genialt. Det er jo sånne systemer som sikkert kommer til å bli tatt godt i bruk i nærheten av eksamensperioder og sånt, da det gjerne er da man ønsker å teste kunnskapen sin og sånt. Jeg føler at den løsningen jeg prøvde fremstiller veldig de kravene jeg ville hatt til et sånt system som jeg ville brukt selv. Syns det var enkelt og intuitivt å bruke.

H: Har du noe konkrete eksempler på features du synes var ekstra nyttige i systemet?

S: hmm... Features som f.eks det at det står hvilket nivå man er på. Det er ganske kult tt det var en slags medalje og hvilket nivå man har. Det var også kult å kunne sjekke den grafen for å se hvordan man gjør det nå, sammenlignet med tidligere ganger man har gjort testene. Det var ganske nice.

H: Du testet på torsdag? Gjorde du ikke? Eller testa du på onsdag?

S: Onsdag tror jeg.

H: Når du var ferdig med en quiz, så kom du til en slags oversiktsside. Synes du den var nyttig?

S: Ja

H: Her fikk du jo tilgang på riktig svar til hva du svarte feil på feks. på det spesifikke forsøket.

S: Ja den ja! Det var veldig nyttig. Det eneste jeg tror jeg tenkte på, var at når man trykker på en man har feil på, så kommer det opp alternativene med en rød og en grønn... Nei bare glem det, hahah... Nå skulle jeg si at det var vanskelig å skjønne hvilken man hadde svart riktig på, men den man har tatt feil på er jo sikkert den røde. Bare glem det. Jeg synes det funket bra!

H: Var det noen features du synes var lette eller vanskelige å forstå? Sånn ikke for å rippe opp i det du nettopp sa, men om du husker noe som var litt ekstra lett å skjønne eller noe du måtte spørre om å få forklart?

S: Nei, de grafene var kanskje ikke så lette å tyde etter den første gangen når man ikke hadde så mange resultater. Man ser liksom en graf, men så er det litt vanskelig å tyde ut av den, siden man ikke har noe annet å sammenligne med. Men det vil jo egentlig være ganske intuitivt hva de grafene sier. Jeg husker ikke helt om dere hadde noe forklarende tekst eller overskrifter til de grafene jeg?

H: Tror det var noen overskrift, men ikke noe forklaring med tekst annet enn det skrevet som var først.

S: Ja, men hvis det var overskrifter som sier at det er en graf over tidligere resultater, så vil det jo være ganske intuitivt.

H: Var det noe som var ekstra lett å skjønne da? Noe som du f.eks husker ekstra godt?

S: Generelt sett var det veldig lett å skjønne hvordan selve testen skulle gjennomføres. Sånn hvordan man skulle svare på spørsmål og bevege seg videre. Når resultatene ble displayed og hvordan resultatene hang sammen.

H: Brukte du den sammenlignings-funksjonen?

S: Hæ?

H: På forsiden på dashboardet? Hva skal man si? Der man kunne sammenligne seg med andre medstudenter, med en sånn gul knapp.

S: JA! Det gjorde jeg. Jeg så at jeg brukte mye kortere tid enn andre!

H: \*Viser en screenshot av dashboardet med knappen\*

S: Ja, jeg brukte den.

H: Ble du motivert når du så hvordan du lå an i forhold til andre folk?

S: Ja.

H: Hvordan ble du motivert? Lå du rett foran? Lå du rett bak? Hva var det som gjorde at du ble motivert når du så det?

S: Jeg husker ikke helt hvordan jeg lå, men jeg vet at jeg har ganske mye konkurranseinstinkt, og man vil jo ikke gjøre det dårligere enn andre som også tar det. Så det vil jo kanskje gjøre at man bruker den mer, eller leser mer før man tar testen for å gjøre det bedre på testen, for man vil jo gjøre det bra om man ser at andre gjør det bra.

H: Hvis du lå rett foran, hadde du blitt mer eller mindre motivert enn om du lå rett bak gjennomsnittet?

S: Jeg ble egentlig ganske motivert av å gjøre det bedre enn folk.

H: Så du ville lagt en ekstra innsats dersom du lå bak for å komme foran?

S: Ja, og hvis jeg lå foran ville jeg også lagt en god innsats for å fortsette å ligge foran, siden jeg blir motivert av å ligge foran.

H: Hvis du da ligger langt bak, ville du blitt dobbelt så motivert siden det er dobbelt så langt opp? Eller hadde du blitt demotivert fordi det er for mye arbeid?

S: Jeg tror jeg hadde blitt litt demotivert, men jeg tror ikke det hadde hindret meg fra å bruke appen videre. Tror jeg bare hadde blitt misfornøyd og irritert, men tror jeg hadde brukt testen liksom.

H: Vi kan gå over til noe litt annet da. Synes du det var intuitivt, eller skjønte du at det var en sammenheng mellom grafen til venstre og grafen til høyre? Din progresjon og din responstid?

S: Nei, det tror jeg ikke at jeg tenkte over egentlig.

H: Synes du at mastery level, på en måte, tilstrekkelig visualiserte din kunnskap? Nå innenfor ITGK, men det kunne jo vært hva som helst da. Synes du mastery levelen din representerte din kunnskap?

S: Jeg husker ikke helt hvilken mastery level jeg fikk, men jeg tror jeg følte det da. Hva stod det? Var det et tall? Jeg husker ikke helt hvordan det så ut.

H: \*Viser et screenshot av hvordan mastery level badge ser ut\*

S: Ja, den skjønte jeg ja. Og jeg følte/skjønte at den representerte det nivået jeg var på da.

H: Har du noe andre synspunkter eller liknende? Noe på hjertet?

S: Jeg likte fargene, de var fine! Hahah

### A.6.3 Interview 3

## Intervju nr. 3 - (statistikk)

---

H: Vi begynner med et av spørsmålene vi har lagd da ... Det er jo om du syns SmartU var et interessant system, og hvorfor var det interessant, og eventuelt hvorfor var det ikke interessant?

S: Okey ... Jeg syns det var interessant for deg gir deg en ganske god oversikt over hvordan du ligger an relativt til fremgang i faget, og det kan være praktisk i forhold til selvevaluering hvor du helt konkret får beskjed: her burde jeg jobbe videre på dette, og her er jeg litt svak. Og her er jeg veldig god, så da burde jeg kanskje ikke bruke fullt så mye tid der med det første, og det er jo ganske kjekt å vite. Så sånn sett er det interessant.

S: Men så er det det der med hvordan et spesifikt spørsmål fungerer med statistikken ... Altså statistikken til et enkelt spørsmål tror jeg ikke jeg brukte like mye, men det er ikke sååå interessant, det vil jeg heller finne ut av i det jeg jobber med oppgaven, hvor vanskelig den er. Det er i alle fall sånn jeg jobber da ...

H: Ikke sant ... Har du noen andre sånn, konkrete features som du synes var ekstra nyttige?

S: Hm ... Det var jo bare sånn statistikken over hvordan du ligger an og det at du kan sammenligne det med andre som var viktig.

H: Så det er det som liksom er essensen av programmet for din del?

S: Ja.

H: Var det noen features du syns var på en måte ... ekstra lette eller ekstra vanskelige å forstå, som du husker igjen som typisk; «Ah, den skjønte jeg med en gang» eller «dette her forstår jeg ikke hvordan jeg skal bruke ...» ikke nødvendigvis som at det var unyttig, men som i at du ikke skjønte hvordan du skulle benytte deg av featuren?

S: Okey, så ... I alle fall den oversiktssiden som var når du ikke var inne på noe spørsmål, den var ganske grei å forstå. Da skjønte jeg at dette var sånn samlet sett for alle spørsmål du har gjort, men så når du var inne på et spørsmål så var det det at «Okey her står det statistikk». Men det jeg forventet at det skulle stå var jo hvordan du ligger så langt i løpet av testen, men så var det jo heller statistikk på dette enkelte spørsmålet og da var det litt merkelig å finne ut av hva dette her betydde for det var ikke sånn super tydelig, og det følte ikke ut som noe man typisk bruker til vanlig heller. Så det var kanskje en litt merkelig måte å tenke på syns jeg ...?

H: Kult, det var bra feedback! Og, hm ... Du følte ikke da at denne statistikken hjalp deg å forstå «scopet» på spørsmålet, du følte kanskje da det var mer i veien, en slags forstyrrelse?

S: Eeh ... Det var jo relativt greit at det bare var skjult bak en statistikk-knapp, og jeg føler at du har ikke bruk for all den statistikken, for da må du liksom bruke tid på å tolke det samtidig som du jobber med spørsmålene og det blir litt sånn at du går vekk fra det du egentlig jobber med og tenker på noe annet. Og akkurat det syns jeg kan være litt forstyrrende. Men det kunne jo være at det gikk ann å korte det ned til å bare være «Åja så mange syns det har vært et utfordrende spørsmål» eller noe tilsvarende, så du ikke trenger tolke så mye annet ut fra det.

H: Okei, så problemet du hadde da med den statistikken var at det kanskje var for komplekst til at du bare kunne se og så bare forstå alt, men at det var for mye som skjedde da?



S: Ja!

H: Har du noen eksempler da, altså du nevner jo dette med at du bare skulle se at «så mange synes det er vanskelig» og that's it. Har du noe andre eksempel på hva som eventuelt kunne vært der da?

S: Sånn istedenfor det som stod der?

H: Ja, typisk hvis du hadde trykket på den knappen. Hvis du på en måte, i en perfekt verden, så hadde du fått opp nøyaktig det du tenkte skulle være under den knappen, hva hadde vært der da?

S: Ømm ... Hm ... Det er jo litt sånn ... Det jeg trodde først når jeg trykket på den knappen var jo det at står hvordan du ligger an så langt. Og da tenker jeg det også kunne stått en setning eller to om «Hei, du gjør det så langt så [] mye bedre enn de andre gjør og bare litt om hvor vanskelig dette spørsmålet er»

H: Ikke sant ... Kjempe feedback det. Ja ... Brukte du den funksjonen som ble kalt «Peer-comparison» når du var på dashbordet for en aktivitet, slik at du kan sammenligne deg med de andre som har tatt samme quiz?

S: Eeh ... Var det det på hver enkelt quiz?

H: Nei, det var ... Skal vi se ... Jeg skruer på «delt skjerm her» (Viser skjermbilde fra Figma) ... Nå kan du se! Den gule knappen her.

S: Ja

H: Brukte du den?

S: Nei, det husker jeg ikke noe særlig. Du kan jo prøve å trykke på den, hva skjer da?

H: Da kommer du til det vinduet her (Peer-comparison mode).

S: Aaah, okay ...

H: Da får du liksom sammenlignet deg selv opp mot det andre har gjort.

S: Ja! Det så jeg jo litte gran på.

H: Ja, da får du jo litt sammenligning, altså din average (score) mot de andres samlede average (score).

S: Ja. Hva innebærer progresjon?

H: Progresjon er jo da dine forsøk mappet i en graf, så du ser kurven på din progresjon mot den gjennomsnittlige kurven til de andre.

S: Okey okey, ja, for jeg brukte i alle fall responstid og så at jeg brukte sånn ganske mye kortere tid enn andre og da burde jeg kanskje tenke: «Hmm, her bruker du så kort tid, kanskje jeg bare kan tenke litt mer på det og at jeg kanskje faktisk har feil» så det var lurt å ha.

H: Følte du at du ble motivert når du brukte den statistikken og så hvordan du la an i forhold til de andre?

S: Ja ... Litte gran. Du har jo lyst til å gruse motstanden skikkelig, ikke sant? Da kan det jo være litt motiverende. Kanskje en litt hard måte å si det på, men ja ...

H: Nei nei, jeg forstår den altså! Men si at du ligger rett BAK gjennomsnittet ... Hadde du blitt mer motivert da, til å komme over (gjennomsnittet) og så gruse de, eller er du mer motivert om du ligger rett over gjennomsnittet for så å øke forsranget?

S: Jeg blir motivert begge gangene syns jeg. Og da hjelper det å si «jeg skal virkelig pushe på, jeg skal bli så mye bedre enn gjennomsnittet», men så tenker jeg at om du er litt svakere elev, så kan det jo være ikke fullt så motiverende å få beskjed om at du ligger så så langt bak.

H: Hva da om du ligger helt sist? Sånn typisk LANGT bak, hadde du blitt motivert da, eller hadde gått over i demotivasjon? Om du skjønner spørsmålet?

S: Ja, jeg skjønner spørsmålet ... Jeg har på en måte ... Jeg er sjeldent i den situasjonen, men jeg synes iallfall det ville vært motiverende for akkurat meg, men jeg vet ikke med andre.

H: Ja, så du er egentlig motivert uansett da, så lenge du ser hvordan de andre ligger an?

S: Ja.

H: Ja ... Vi gir jo en slags «master-level», syns du det var intuitivt hva denne her «masteryen» skulle representere?

S: Ja, sånn nogenlundene.

H: Okey, hva representerte «masteryen» for deg? Sånn, hva tenker du at den representerer?

S: Hvor godt du har svart på de siste spørsmålene og hvor stor andel av de som er korrekte, og jeg tror ikke tidsbruken spilte noen rolle der ...?

H: Nei, det tror jeg du har helt rett i, haha. Hvis jeg sier at den representerer din kunnskap i aktiviteten ... Så den masteryen som er i dashbordet skal da representere, kall det din «overall knowledge», syns du den klarte å få frem det?

S: Ja, den representerer jo litte gran, men det kommer jo litt an på hvor høyt nivå du sikter til selv, om du har veldig høye krav for det innebærer at du skal forstå akkurat det og det innenfor det emnet så kan det hende at det ikke er en god nok indikator for det, sånn i seg selv, men da er man ganske streng med seg selv da. Men generelt sett syntes jeg at den representerer det ganske greit.

## A.6.4 Interview 4

### Intervju nr. 4 – (statistikk)

---

H: Syntes du SmartU var et interessant system? Og hvis ja, hvorfor det? Og Hvis nei, hvorfor ikke?

S: Ja, altså jeg lurer på, har dere noen mock-ups eller noe sånt så jeg kunne tittet på?

H: Ja! Jeg kan dele skjermen min i alle fall, så kan du se.

S: Ja, Nice!

S: Når jeg, på en måte, ble introdusert til det, var min naturlige reaksjon å sammenligne med like applikasjoner med omtrent samme formål som det dere hadde. Og når jeg da ble presentert med informasjonen var det på en måte en helt ny vinkling ved å sammenligne seg med andre. Det er jo mange andre applikasjoner hvor du får vite hva din karakter er og hva snitt karakteren er, men det her var mye mer inn i hvert enkelt spørsmål. Sånn umiddelbart ... Sånn en generell tanke er at jeg har ikke helt lært meg å bruke den informasjonen til noe nyttig ... Så, det er kanskje greit at noen sliter med et spørsmål og at det er få som klarer å få det riktig, men det gir ikke meg så veldig mye. Men det gir meg kanskje litt sympati da, hvis jeg syns at spørsmålet er veldig vanskelig, så føler jeg meg litt bedre av å se at det er andre som bruker lang tid på spørsmålet, men umiddelbart så ville jeg kanskje tenkt at ... Eeh ... At når jeg sitter og løser de oppgavene, så hadde jeg nok tenkt at av all den dataen som er samlet inn på denne applikasjonen, så er det jo mange som sikkert har kjørt gjennom mange ganger. Og etter hvert som du blir god på tema gjennom et semester, så gjør du oppgavene bare kjappere og kjappere. Kanskje du har regnet på bits før, slik at du vet hva den bitstrengen er representert som, og da bare trykker du med gang. Det er liksom ... hvis jeg blir presentert for en oppgave noen vanligvis bruker 1 min på, så syns jeg på en måte at ... Ja, jeg følte kanskje litt at disse tallene kan bli misvisende da. Ehh ... sånn at hvis det for eksempel er noen funksjoner som sier at første gangen noen ble presentert for spørsmålet så brukte de så lang tid. Da altså at man kunne sett på hvor mye man hadde vært borti det eller vært eksponert for de og de spørsmålene, da får man litt mere data og skjønner kanskje litt mer. Men jeg er nok mer opptatt av å finne ut hva svaret på spørsmålet er enn hvor mange andre som har fått riktig på det. Det er morsomt sånn sett, når jeg gikk tilbake til spørsmålet og så hva som var fasit, da hadde det vært nyttig å få den recapen, men akkurat jeg klarte ikke å finne verdien i hvor mange som sliter. Men det er jo deilig da, når jeg sitter og undrer, og så trykker jeg og så ser jeg «Ah jeg er ikke den eneste», og det er fint. Og så var det når jeg satt og løste spørsmålene første gangen, så visste jeg ikke hvor mange spørsmål som var igjen, det var kanskje det at jeg ikke skjønnte interfacet eller noe sånn. Men ofte er det motiverende å se at «nå har jeg gjennomført så mange, og det er så mange igjen». Men når det på en måte ikke kommer tydelig frem ... Nå husker jeg ikke helt hvordan mock-upen ser ut da, om det er innlysende etter andre gang jeg går igjennom, men i alle fall slik jeg opplevde da, så syns jeg at det var litt sånn. Ofte så er det greit å vite når man kan løpe på do eller ta seg en kaffe, spesielt når man blir tatt tiden på, på hvert spørsmål, og det hele tiden innsamles data hele tiden når jeg er på quizen.

H: Om jeg kommer med litt tilleggsinfo da. Fordi ... Ja, systemet er jo adaptivt, så ... Nå var det veldig mange som endte opp med 12 spørsmål da ... Men slik det skal være er jo egentlig at du ikke helt vet hvor mange spørsmål du får, så det går liksom ikke helt an å si at «nå er det slutt på quizen» fordi om du gjør det veldig bra, eller veldig dårlig. Si, teoretisk sett så trenger systemet bare 5 spørsmål for å avgjøre hvor du ligger, men det kan trenge opp til 20, slik at det er litt vanskelig å si når quizen faktisk er slutt. I tillegg til det er det jo også slik at systemet bare bruker et spørsmål en gang. Så du vil aldri få samme spørsmål to ganger, så derfor vil alltid vite at det spørsmålet du er på nå vil du og

de andre bare svare på en gang. Så du for eksempel vil aldri treffe spørsmålet «hva blir denne bit-rekka i desimaler» mer enn en gang. Så, ja. Det er i alle fall slik vi vil det skal funke, haha.

S: Da stod det garantert ... Jeg husker i alle fall nå når du sier at, spesielt vanskelighetsgrad og det med spørsmål stod jo på det arket (introduksjonsarket til adaptivitet). Men ofte hvis jeg skulle hoppet ut i en sånn applikasjon, så ville jeg nok ikke ha satt meg ned og lest et a4 ark, men heller testet det, kjørt noen iterasjoner og da hadde gjort meg disse bemerkningene. Og det virker jo som det absolutt er gjennomtenkt, men kanskje prøve å få frem det da?

H: Ja, kan vi da konkludere med at systemet gir litt lite informasjon for førstegangsbrukere?

S: Hmm ... Altså da må man definere hva ... For jeg syns at det å gjennomføre det er veldig intuitivt. Du trykker på den «aktiviteten» du skal, du blir presentert for spørsmål. Men det er med en gang du begynner å få sånne som begynner å tenke sånn ... «Ja, hvor mange har svart på dette spørsmålet og hvor lang tid brukte de?», at da hvis man har fått presisert det her; at første gang dette spørsmålet ble presentert for brukerne, eller formulert en annen lekker setning der. Bare sånn at når det er såpas basert på statistikk, så burde det også fremkomme ganske tydelig sånn at man tenker mer sånn «ÅJA», i stedet for at det er sånn gjennomsnittlig tid. For jeg vet i alle fall hvordan lignende applikasjoner brukes som IntoIT og Memorizer, at der er det på en måte når jeg ser hva som er snittet på en klasse der, så er det på en måte satt en standard inne i hodet mitt da, særlig at alle får A første gangen de går gjennom det her. Men jeg skjønner jo også at det snittet går veldig opp når eksamen nærmer seg også da. Så gjerne kanskje presiser akkurat det (at snittet deres er basert på at ting bare kan gjøres en gang), men igjen så er jo det heller ikke noe som er kritisk for applikasjonen, jeg syns jo bare det at å få tilgang til spørsmål som tester deg på emnet er det som er kjernen, og det som er viktigst. Å ha et godt spekter av spørsmål som man kan få testet ferdighetene sine og få lære ut ifra de.

H: Ikke sant ... Benyttet du deg av sammenlignings-funksjonen på dashbordet, så du kunne sammenligne deg med hvordan de andre i klassen hadde gjort det. Så fikk du opp din mot de andre sin graf?

S: ja ... Jeg trykket på den og jeg så på den. Og det er jo på en måte veldig interessant data, det setter jo på en måte meg opp mot de andre, og det er jo mye motivasjon også i det da. Hvis jeg for eksempel hadde vært i en klasse og så skjønner jeg at «dæven jeg ligger jo langt bak snittet her» så måtte jeg jo ta meg sammen ... Eeh ... Men i den settingen jeg skulle simulere: «jeg setter meg ned hjemme for å jobbe med det», så ville jeg igjen tenkt litt mer som Memorizer da, at når du ser snittet er en A og du fikk en C, så tenker ikke jeg sånn «Hmm ...». Men det er veldig interessant å se hvis jeg sliter med en oppgave, for da er det jo veldig deilig å tenke sånn «jaa ...». Og hvis jeg da skulle brukt det her mer aktivt, så ville jeg jo kanskje sett på enkelt-spørsmålene og tenkt: hvor mange har svart riktig på det her, og da tenkt at dette er mer en «no-brainer» og svart med en gang, så jeg slipper å bruke tid. Og om jeg ser at mange svarer feil ville jeg tenkt at dette er noe lureri og ville da kikket litt ekstra. Men jeg syns jo at det var interessant å kunne følge en progresjon og kunne se hvor ...

**\*internett falt ut\*** ja ...

H: Hvis du ser på bildet jeg delte nå (sammenligning med peers). Du ser da at du har den røde linja som da er en average av hva klassen har gjort, og den lilla er din performance. Så da ser vi her at ca. midt på grafen så ligger du litt under gjennomsnittet. Ville du vært motivert hvis du lå her, altså på den nest siste (like under gjennomsnittet) eller her (rett over gjennomsnittet)? Altså, gir det deg mer motivasjon om du ligger rett over gjennomsnittet, eller gir det deg mer motivasjon om du ligger rett under for å komme over gjennomsnittet?

S: Hvis det er snakk om hvor motivert jeg blir, så vill jeg nok heller ha sett på hvor mye riktig jeg får egentlig.

H: Okay, så hvordan de andre gjør det, er ikke så viktig for din motivasjon?

S: Neeei ... Men igjen så har jo jeg ikke vært presentert overfor et verktøy ved tidligere eksamensperioder, så hvis jeg skal kunne se en verdi i det, så må jeg kanskje ... Eller igjen da, nå fikk jeg jo vite at det var gjennomtenkt at statistikken på et spørsmål og sånt er jo satt på bakgrunn at ... Det skal gi mening, men når jeg blir presentert for sånne ting, så tenker jeg med en gang sånn «Naah, her har jeg sett så veldig mye, ja ...» Det er så mye støy da, i tallene, at jeg kanskje automatisk ser litt bort fra det. Men hvis jeg hadde sett for meg at dette her var min første eksamensperiode og jeg setter med ned med dette verktøyet og jeg har blanke ark, så ville jeg jo absolutt hatt interesse av å se hvordan jeg var i forhold til andre. Men hvorvidt jeg er mer ... Altså jeg syns kanskje at ... Det høres kanskje teit ut ... Men det lille «ranking-systemet» dere har på midten, hvor det sier «High» eller «Medium» eller at du får en liten inndeling da, og det at du får en liten medalje, det syns jeg skaper en litt mer kul ... Ja ...

H: Mhm. Så du er mer opptatt av den «gamification» delen da på en måte. At du får en medalje, enn at du får «cold, hard facts» og grafer hvor du kan sammenligne?

S: Ja ... Jeg syns i alle fall, slik det er i dag, at de tå sammenligne meg med alle andre er jo også på en måte ... Jeg syns jo kanskje at det ville vært mer deprimerende i fag som Algdatt for eksempel.

H: Ja, det tror jeg vi to andre også kan være enig i, haha.

S: Haha, ja. Og da er det jo på en måte ja ... Jeg ville jo kanskje tenkt litt mer sånn, hvis jeg hadde gjort det mye bedre enn gjennomsnittet så hadde jeg sett mye på det og tenkt sånn «ah, det her er jo dødsfett». Men hvis jeg dagen før eksamen hadde satt meg ned der og så hadde jeg sett at jeg lå under, da hadde jeg tenkt «ah, shit».

H: Ikke sant.

S: Og så hadde det kanskje vært litt interessant å se på hva de tallene tar bakgrunn i på en måte. For hvis jeg da ser at average tar utgangspunkt i hvor mange riktige du får hver gang, eller om den tar utgangspunkt i hvor mange ganger du har utført spørsmålet og alle da pleier å få 9 riktige på hver eneste. Altså at hvis man kunne fått litt mer bakgrunn, så jeg kunne sette meg ned og lese, og tenke «Åja, det er på bakgrunn av det her!» så ville jeg nok tenkt at ... Men det er mye mulig at det faktisk fremgikk i dette dokumentet deres, men umiddelbart så tenkte jeg ikke at det var sånn.

H: Men så bra, da tenker jeg kanskje at vi burde runde av der!

## A.6.5 Interview 5

### Intervju nr. 5

---

H: Synes du SmartU var et interessant system? Hvis ja, hvorfor synes du det var interessant, og hvis nei, hvorfor ikke?

S: Ja, jeg synes det var interessant i den forstand at det var egentlig sånn man forbereder seg til mange eksamener uansett. Det å få samla det på en litt bedre plass istedenfor å bare måtte hive sammen en quiz litt selv med kopierte spørsmål fra eksamenene. Fikk en liten oversikt med noe grafer å gå ut ifra med litt mer kontroll enn når man bare må gå gjennom å være fin på det selv hele tiden, da er det litt mer konkret og forholde deg til.

H: Synes du de grafene som var tilgjengelig var lette å forstå? Synes du de passa inn i dashboardet og skjønte du hva de presenterte?

S: Ja, synes det var ganske greit. Har ikke noe spesielt å si på det.

H: Synes du at det var intuitivt sammenheng mellom grafene? At de representerte ulike sider av samme datasett?

S: Litt usikker, for jeg husker ikke alt.

H: \*Viser screenshot av dashboard for reference\* - Var det intuitivt at de representerer samme data? Bare på litt forskjellige måter?

S: Ja, ja, ja sånn ja. Det til høyre der ja.

H: Ja, og det at der er intuitivt at det er sammenheng mellom disse to grafene, eller om du ser på det som to helt isolerte grafer som ikke har noe med hverandre å gjøre?

S: Når en liksom leser at tiden går ned og progresjonen går opp, så synes jeg at det er intuitivt ja.

H: For du forstår at prikk nummer 1 er sammen forsøk?

S: Ja

H: Så bra. Synes du at disse grafene bidrar til å øke din motivasjon? Hvis du skulle brukt dette systemet, hadde du blitt motivert av å se at progresjonen din gikk opp og responstiden din gikk ned?

S: Absolutt, det blir litt konkurranse med seg selv, og litt mindre ork istedenfor å bare måtte trykke gjennom å få mye av det samme opp igjen. Få frem litt tydeligere hva som mangler og hva man ikke trenger å bruke mer tid på.

H: Brukte du sammenlignings-funksjonen? En sånn gul knapp som det stod peers på. Hvor du kunne sammenligne deg selv mot gjennomsnittet av de andre i klassen?

S: Ja

H: Ble du da motivert når du så hvordan du lå an i forhold til de andre? Hadde det noe påvirkning på motivasjonen din?

S: Jeg lå vel litt jevnt over egentlig, så kanskje ikke noe spesielt på motivasjonen min. Det var jo kanskje litt morsomt å se at progresjonen min lå greit an.

H: Så du får ikke noe indre motivasjon for å jobbe på for å holde deg over? Det var mer sånn «Nå ligger jeg over, det var gøy å se»?

S: Ja, hvertfall den settingen jeg var i, men det er klart at før eksamen så hadde det kanskje vært en annen greie selvfølgelig. Når motivasjonen er å få bedre karakterer enn andre, så er det klart at jeg kan se for meg at jeg blir motivert.

H: Hvis du hadde lagt rett under, hadde du følt mer på motivasjonen da?

S: Det tror jeg nok helt klart ja!

H: «Nå skal jeg kjøre på, skal ikke ligge under streken»

S: Ja, det tror jeg hadde vært mer motiverende ja.

H: Si en enda mer ekstrem situasjon da, om du hadde ligget lang bak. Hadde det vært dobbelt så motiverende fordi det var dobbelt så langt opp? Eller hadde det blitt overveldende og grenset mer mot demotiverende.

S: Det kommer kanskje litt an på hvilken tid det er, hvis det f.eks er nær eksamen så tror jeg det sikkert hadde vært veldig kjekt å se hvor langt bak man ligger, men midt i semesteret tror jeg det hadde vært en litt mer «wake up call» og innse at man kanskje må jobbe litt mer.

H: Ja, du brukte kanskje den statistikk siden, altså den oppsummeringssiden på slutten av quizen, som viste hvor bra du gjorde det på dette forsøket, og du fikk tilgang til rett/feil svar og sånt. Kan du fortelle litt hva du synes om den siden?

S: Ja, den likte jeg veldig godt egentlig. Litt oppsummert og kunne trekke frem det man gjorde feil og fikk den tilbakemeldinga. Litt sånn oversikt.

H: Syns du det på en måte... Ble du mer motivert til å lære når du fikk en slags fasit på at det var veldig få som fikk riktig. De aller fleste hadde hvertfall 1 forsøk hvor de fikk en feil. Ble du på en måte motivert til å få 12/12 når du så at du fikk en rød prikk og fikk se hva som var feil?

S: Jeg synes jeg ble litt motivert og ville ha alle 12.

H: Følte du at du fikk noe ut av statistikken som var tilgjengelig angående tid og sånt? Så stod det liksom sånn, «Du hadde så og så rask responstid».

S: Ja, jeg synes det forteller en god del om hvor god kontroll man har på en måte. Hvis man er altfor rask, så kunne det bare vært en vill gjetning, men jeg tror man må prøve å sikte seg inn på en sweetspot hvor man har kontroll.

H: Ikke sant, men ikke bruke unødvendig tid kanskje?

S: Ja.

H: Så er det jo sånn at både for et forsøk og på selve dashboardet, så viser vi jo såkalt mastery level. Hva tenker du at mastery level skulle representere?

S: Nåværende ferdighetsnivå.

H: Hva legger du i ferdighetsnivå?

S: Tenkte bare hvor godt man kan stoffet.

H: Er du enig i at det er en forskjell i hvor godt du gjorde det på en quiz og hvor godt du kan stoffet?

S: Ja, det er klart.

H: Hvilken av de 2 vil du si at mastery level representerer?

S: Hvertfall i begynnelsen sier den hvor godt man gjorde det på en quiz, men det er klart at når man har gjort mange quizzes, så vil det representere hvor godt man kan stoffet.

H: Er det noen features du synes var ekstra nyttige som du husker som «en feature som var nyttig»? Som du benytta deg ekstra mye av da.

S: Jeg mener å huske at det var sånn at de feilene jeg gjorde skulle liksom dukke opp. Så det likte jeg veldig godt, at man får litt tilbakemelding på det man ikke kan så godt. Og det som man har svart rett på 2-3 ganger allerede blir lagt litt bort.

H: Du får spørsmålet én gang liksom, og om det er riktig så har du svart riktig på en måte.

S: Ja, det er ikke noe man trenger å svare flere ganger på, når man kan det.

H: Hvis vi skal stille spørsmålet litt annerledes. Er det noe features du husker som er spesielt lette å skjønne hvordan man skal bruke? Eller motsatt? Noe som var helt uforståelig hvordan man skulle bruke, ikke unyttig men vanskelig å forstå konseptet rundt da?

S: Hmm... Nå må jeg tenke litt...



H: Det er lov å svare nei om du ikke kommer på noe.

S: Jeg kommer ikke på noe akkurat nå.

H: Si du skulle brukt systemet på nytt. Hva er det som motiverer deg til å gjøre det?

S: Egentlig sånn helhetlig å ha oversikten på det, og ha alt samla på en plass. Får statistikk. Ja, egentlig bare jevnt over at det er der og ikke trenger å forholde seg til så mye annet når det kommer til å lære meg den type spørsmål.

H: Så er det jo sånn at vi går en interaksjonsmaster, så vi er jo også litt interessert i dine tanker om fargebruk og sånt. Hva synes du om uttrykket siden gir? Føler du at det er et system som passer for skole? Syns du det er for «corporate»? Føler du deg tilfreds i dette systemet?

S: Jeg synes det funker godt som en sånn skole-hjelpemiddel. Det var intuitivt design og sånt. Farger og ja.

H: Fikk du inntrykket av at forskjellige farger representerte forskjellige ting? Og isåfall, har du lyst til å fortelle litt om det? \*viser screenshot av dashboard for reference\*

S: Ja, jeg synes egentlig det var ganske greit. Mye grønt. Disse oppsummeringsknappene på slutten hadde grønt når det var rett og ja.

H: Hva er det gul, grønt og rødt representerer?

S: Rødt -> lavt, gult -> middels og grønt -> høyt

H: Hva er det du tror lilla representerer?

S: Tenker ikke på lilla sånn fryktelig mye egentlig. Er bare en gjennomgående farge på en måte.

H: Men du klarte å skille forskjellen på hva du hadde trykt på og ikke hadde trykt på av svar under quizen?

S: Ja, det kan jeg ikke huske at ikke gjorde iallefall.

H: Ville du brukt systemet igjen hvis du hadde sjansen?

S: Ja.

## A.6.6 Interview 6

### Intervju nr. 6

---

S: kan jeg bare spørre om en ting før vi starter? Jeg husker dere hadde bilder av selve nettsiden på spørreundersøkelsen? Kunne jeg ha fått de å se på slik at svarene mine ikke kun går på hukommelse?

H: Ja! Jeg kan dele skjermen min, så du kan se på bildene fra Figma.

S: Supert! Det så kjempegreitt ut.

H: Såå ... Ja. Syns du SmartU var et interessant system? Hvis ja, hvorfor var det interessant og hvis nei, hvorfor var det uinteressant?

S: Jeg syns det var veldig interessant. Det har et aspekt av læring, istedenfor at du bare tar en svær bulk så du for eksempel bruker andre nettsider, eller andre muligheter med en stor «load» av masse spørsmål, så skulle dette prøve å tilrettelegge for hva du trenger å vite og den «reaffirmer» litt av det du allerede kan, men den skal prøve å veilede deg til det du ikke kan og prøve å lære deg det i stedet. Heller enn at du bare går gjennom 500 forskjellige spørsmål om og om igjen.

H: Hva syns du om selve, vi kan kalle det dashbordet for en aktivitet, der du ble ... Ser du musepekeren på skjermen min?

S: Jepp, det gjør jeg.

H: Type, dette her er jo da typisk et dashbord, og dette har kan være noen andres dashbord for samme aktiviteten. Hva syns du om dette, og visualiseringen som blir presenter her?

S: Visualiseringen er veldig greie. De er store de er veldig lette å se på første øyekast. Det eneste jeg var litt forvirret om er den der «Your progression», greit nok at den går oppover, meg jeg var litt usikker på hva den mente? Sånn hva var det jeg gikk opp i? Er det bare rette svar, er det kunnskap? Men på den andre siden var det responstid, da går det i sekunder og så kan du se basert på hva slags type spørsmål du hadde rett og galt og se hvordan det var. Og så er det jo de andre aspektene ... er veldig store, sånn som mastery-level og dine tidligere attempts og self-awareness var alle veldig intuitive, mens «Your progression» var litt sånn forvirrende, utenom at den gikk oppover.

H: Syns du at det du får se på dashbordet er ... gir deg økt motivasjon til å studere? Hvis du skjønner spørsmålet, altså får du noe økt motivasjon av å få statistikk på denne måten og visuelt se fremgangen din, eller måtte gud forby, tilbakegangen din?

S: Ehh ... Jeg er en veldig data-orientert person, så jeg syns det er veldig interessant å kunne se hva som er gjort bra og hva som er gjort dårlig så jeg kan se ... på en måte hvordan den dataen går da, bakover i tid fra tidligere attempts. Så for meg hvis jeg skulle brukt dette så ville det gitt meg mer motivasjon om jeg ser en stadig growth, eller hvis jeg ser det kommer en decline så kunne jeg kanskje tenkt sånn «ah, shit nå har jeg faktisk ikke fokusert lenge nok, eller ikke fokusert nok» om det går over en lenger periode. Så det blir lettere å være selvkritisk og analytisk når man ser på sitt eget arbeid, enn bare «okei, nå jobba jeg i dag, for et par dager siden og noen dager før det igjen» hvordan ligger jeg egentlig an, mens her står det mer konkret.

H: Har du noen features du syns var ekstra nyttige? Altså som du husker som at dette brukte jeg, og disse, en eller to featurene er essensen av programmet for meg? Disse husker jeg ekstra godt?

S: Hmm ...

H: Hva er det første du kommer på liksom?

S: Jeg synes det var veldig interessant med responstid, for jeg har liksom aldri tenkt over hvor lang tid jeg bruker på å svare på forskjellige svar. Så det var veldig stilig å kunne se de forskjellige måtene med om det var riktig eller galt, og jeg kan ikke huske hva den i midten var, men ... Forskjellen på feil og riktig på hvor mye tid du brukte på et spørsmål? Jeg hadde aldri tenkt sånn på, men det gir mening at du bruker mye lenger tid på et spørsmål som du ikke vet, og dermed også får feil på. Og spørsmål du vet, så svarer du med en gang fordi du vet svaret, og jo fortere du kan få tilbake svar synes jeg var veldig interessant og veldig bra.

H: Skjønte du at det var sammenheng mellom de to grafene, altså at begge deler tar utgangspunkt i et attemp?

S: Ja, da sånn punktene var jo en attemp hver. Så, jo lenger til høyre du gikk jo nyere var attempen. Tilsvarende var lenger til venstre lenger tilbake i tid?

H: Syns du det var nyttig da å se progresjonen din opp mot responstiden din?

S: Ja! Jeg følte det var kurant å kunne vite forskjell på at du kunne se at du både får riktig, men også raskere riktig fordi at ... du holder på.

H: Ja. Brukte du den ... Ja, hvis du ser på skjermen nå. Den gule knappen her, som da representere sammenligning med de andre i klassen?

S: Jeg tror hvis jeg trykket på den, så kom det ikke noe, eller så trykket jeg ikke på den i det hele tatt. Men jeg følte jeg trykket på alt som var på nettsiden ...

H: Okay ... Men hvis du ser på det bildet her, så er det et eksempel på hvordan siden kunne sett ut.

S: Riktig.

H: Blir du motivert av å se dine stats opp mot andres stats?

S: Ømh ... Det kommer an på, for det er jo hvordan du ligger an i forhold til de andre hvor motivasjonen kommer eller forsvinner, og det er jo veldig fra person til person. Jeg tror at generelt hadde jeg vært motivert til å jobbe meg opp uansett om de andre var bedre eller dårligere enn meg, men så hadde jeg bare jobbet uansett. Så det hadde ikke vært så veldig viktig for meg, men kanskje det hadde vært greit å vite sånn generelt sett hvordan man ligger an i løpet, hvis det ikke bare er til eksamensøving da, men mer generelt til å øve igjennom året. Da kan du se «Okey jeg faller ikke langt bak» eller «sånn her ligger resten av folkene an». Opp mot eksamensøving generelt så tror jeg ikke jeg hadde hatt så veldig nytte av det, for da har du lyst til å bare bli bedre på stoffet.

H: Riktig ... Så da kan vi konkludere med at det ikke har så mye å si hvordan de andre gjør det? Du er mer opptatt av å bare fokusere på deg selv og ... Du er mer fokusert på å få 12/12 enn at grafen din skal være høyere enn gjennomsnittet?

S: Ja, jeg vil si det.

H: Så er det jo sånn at det er ei side som kommer opp på slutten av en quiz. Kan ikke du fortelle litt om hvordan du brukte den, og hva du fikk ut av det?

S: Første gang jeg brukte den ble jeg litt forvirret siden jeg trodde jeg kunne se spørsmålene. Men det ble ikke gitt noe indikasjon på at jeg kunne trykke på tallene (kulene som viser om du har rett

eller galt), mens på andre forsøk bare prøvde jeg å trykke, for jeg tenkte «det må jo være noe der», og da kunne jeg jo gå veldig konkret tilbake til spørsmålene jeg fikk feil med en gang og se hva jeg fikk feil på. Jeg prøvde å tenke på at om det var et spørsmål jeg fikk to ganger på rad, eller to attempts på rad. Om jeg var usikker på hva jeg hadde svart tidligere, om jeg for eksempel hadde svart litt fort, så kunne jeg tenkt at noen spørsmål du var usikker på så kunne du svart, men også huket av en boks for «dette er jeg usikker på». Da kunne du sett spørsmålet du hadde svart feil eller riktig på, men det var et spørsmål du hadde «gjettet» på, så kunne du sett at det var det som var riktig svar så du liksom «reaffirmer» i stedet for at du blir litt usikker på om du fikk rett eller galt på det.

H: Ikke sant, så du kunne markert det til oppsummeringen, det var DETTE spørsmålet du var usikker på?

S: Ja for eksempel. Eller at det da hadde blitt en blå/grønn eller blå/rød da basert på at du gjetta på det spørsmålet, enn at du trodde du hadde rett.

H: Ikke sant, kul feedback! Hmm ... Ja. Jeg har egentlig gått gjennom veldig mye av det vi lurte på ... Men, vi kan jo ta et siste spørsmål. Egentlig ganske lett sånn sett. Ville du brukt systemet igjen? Og hvis ja, hva er det som motiverer deg til å bruke det?

S: Eeh ... Jeg ville definitivt brukt systemet, dersom det var mulig, rett og slett i de fleste fag jeg kunne brukt det til. Det ga veldig sånn, eeh ... Veldig konkret rett/galt og progresjon og ... Du kunne se på aspekter du ikke kunne tenkt deg frem til vanlig, som responstid. Du kan se ut ifra mastery-level, ikke bare om du hadde rett, men også om du hadde rett på vanskelige spørsmål. Eller om du bare har en superficial forståelse, med at du har rett på masse spørsmål, men det har vært lette spørsmål slik at du lurte deg selv til å tro at du kan mer enn det du egentlig kan. Og det hadde en ganske nice User interface med visualiseringer som var enkle å forstå ved å bare se på de i to sekunder, for så bare å gå videre.

H: Ikke sant ... Ja, da tror jeg ikke vi har flere spørsmål.

## A.6.7 Interview 7

### Intervju nr. 7

---

H: Synes du SmartU var et interessant system? Hvis ja, hva gjorde det interessant? Hvis nei, hvorfor var det uinteressant?

S: Det var jo interessant, mest fordi jeg er vant til NTNU systemene som er ganske dårlige. Og vi har ingenting som kan teste hvordan man ligger an før sånn eksamen. Det er liksom da man finner ut hvor gode man er i et fag. Det er egentlig bare hvordan jeg synes det er interessant å finne ut hvordan jeg ligger an mens semesteret går da.

H: Nå vet jeg ikke om du husker konkret hvordan systemet så ut? \*Viser screenshot av dashboard\*. Fikk du noe ut av de dashboardene her, og hva synes du om visualiseringene som er tilgjengelig?

S: Jeg likte det detail view når du skulle inn på en quiz hvor man har oversikt over statistikk og sånt. Siden det ikke er for mye der, og du skjønner egentlig hva alt betyr på en måte. Annet enn at sånn, kanskje fordi jeg ikke ser på alt med en gang, men bare på de grafene, så synes jeg at de tekstene var litt sånn små. Så når jeg først så på grafene skjønte jeg ikke helt hva som var greia før jeg så teksten... Det er ganske simpelt design da, så det er mye som bare er naturlig å forstå hvordan det funker.

H: Kan du si noe om fargebruken her?

S: Det er jo det med grønn, rød og gul. At det er bra med grønt og dårlig med rødt i guess. Ikke så mye annet enn det.

H: Er det noe som indikerer at noe er mer klikkbart enn annet?

S: Det blir jo fort om man ser på landingpage, så er det alt som er lilla. Det er typisk knapper. Eller det som har bilde også lilla under.

H: Var det noe du tenkte på når du prøvde systemet, eller var det noe du tenkte på nå som jeg legger det litt i fanget ditt.

S: Nei jeg skjønte jo på en måte at det skulle være knapper da, siden det skiller seg ut. Men jeg tror ikke at jeg tenkte det, bare følte det sånn.

H: Så det falt bare litt naturlig at det her er noe annet enn det andre da kanskje?

S: Ja

H: Hvis du måtte velge liksom sånn 1 eller 2 features som på en måte var selve kjernen av programmet for deg? Ikke den der quiz modulen da, for den er veldig obvious. Liksom noen ting som du brukte, som du følte var mest nyttig for meg?

S: Det er jo den der statistikken tenker jeg. Det er vel egentlig den største delen annet enn quizen som forteller deg noe om deg.

H: Så du tenker i dashboardet for en aktivitet?

S: Ja. Kommer ikke på noe mer, men kanskje de der mastery levelene sier litt om hvordan man ligger an. Er det ikke det appen skal brukes til? Å si noe om hvordan man ligger an og teste seg selv i fag?

H: Er det noen features som du husker som veldig lette eller veldig vanskelige å forstå hva det skulle brukes til?

S: Hmm...

H: Var det noen ting som du liksom forstod med en gang, eller var det noe du lurte på hvorfor eksisterte, for du skjønnte ikke hva du skulle bruke det til?

S: Nei jeg forstod hvordan man skulle komme seg til en quiz ganske lett da. Bare å trykke på available activities og new attempt, for det stod ganske tydelig. Det er kanskje en liten detalje da på en måte, men på landingpagen så er det de der under announcements og previous actions så er det 3 prikker. Jeg skjønnte ikke helt at de skulle åpne en ny fane med ny informasjon. Jeg trodde det var sånn en type settings ellerno. Det er en sånn mindre ting da.

H: \*Viser summary siden\* Kan du fortelle oss litt hva den siden her var, og hva du brukte den til?

S: Den sier jo litt om hvor du har feil og riktig spørsmål og gir en kjapp oversikt over hoved detaljene over quizen din da på en måte. På siden der, activity results. Jeg skjønnte ikke helt at man kunne trykke på dem for å se riktig svar. Jeg trodde bare det var sånn «her fikk du feil, men burde ha husket hva som var greia liksom»

H: Var det fordi det ikke kom noe hover-effekt når du hovret?

S: Nei, mest fordi det var det jeg tenkte med en gang jeg kom dit.

H: Det var ikke intuitivt å prøve å trykke på dem? De følte ikke den inviterte til trykk på en måte? Det var bare en grønn ring?

S: Nei

H: Og de activity results. Brukte du dette til noe? Ga det deg noe innsikt liksom?

S: Det sier jo på en måte hvor sikker jeg er på hvertfall den quizen da. Jeg brukte jo mindre tid, hvertfall på det ITGK greiene, så så jeg at jeg brukte lite tid på alle spørsmålene, så jeg kunne hvertfall mange av svarene.

H: Så var det jo også en funksjonalitet som tilbyr deg sammenligning med de andre i klassen. Som var tilgjengelig med de andre knappene her. Brukte du den noe?

S: Jeg brukte den, men skjønte ikke helt hva «peers» betydde. Jeg trodde det betydde sånn «sjefen» eller han over meg.

H: Det er ikke et helt vanlig ord kanskje, for deg å bruke?

S: Men jeg trykket på den og skjønte hva den gjorde etter at jeg trykket på den.

H: Du føler kanskje at det burde stått noe annet enn «peer»? F.eks «compare»?

S: For det var sammenligning med de andre studentene, var det ikke? Kanskje «compare to other students», men akkurat det er litt langt da. Men noe i den duren.

H: Når vi først har kommet inn på «comparison», følte du deg motivert når du ble sammenlignet med gjennomsnittet i klassen?

S: Ja! Det handler jo om å være best, så lenge jeg ligger best an.

H: Du vet jo ikke at du ligger best an da, du vet jo bare at du er bedre enn gjennomsnittet.

S: Det er sant, men jeg tenker det at man vil på en måte ligge litt over gjennomsnittet. Jeg ville følt, hvis jeg hadde ligget bak på en måte, så hadde jeg sikkert prøvd å øvd litt mer og tatt en quiz til på en måte.

H: Vil du si at du hadde vært mer motivert til å ta en quiz til om du hadde lagt rett bak gjennomsnittet enn om du hadde lagt rett foran gjennomsnittet?

S: Mhm, Ja.

H: Men hva er forskjellen i tankegangen om du ligger rett bak kontra rett foran?

S: Hvis jeg ligger foran, så vet jeg på en måte at jeg har litt mer kontroll enn de fleste andre. Hvertfall over gjennomsnittet. Hvis jeg ligger bak... man vil jo alltid ligge på samme nivå som resten av klassen. Det er ikke gøy å være han på sånn gruppeøving hvor du er han som ikke kan noe og må sitte der liksom.

H: Så hvis du lå rett bak, så hadde du vært motivert for å ta et ekstra tak for å komme rett over?

S: Ja

H: Si du hadde ligget langt bak, hadde du da vært dobbelt så motivert fordi det var dobbelt så langt opp til gjennomsnittet, eller hadde det hatt en annen effekt?

S: Jeg er litt usikker... Det spørres hvor langt inn i semesteret man er da, hvis man er i starten av semesteret ville man jo anta at man alltid ligger litt sånn bak. Og da er det jo ikke lenge igjen til å kunne lese. Men hvis man ligger langt bak sånn mot slutten av semesteret, så er

det jo sånn «hvorfor gidde, det er jo ikke lang tid igjen». Hadde sikkert blitt demotivert om jeg hadde ligget så langt bak.

H: Si at vi tar utgangspunkt i det siste du sa da, «hvorfor gidde, hva er vitsen?». Hvis vi er der at nummer 1 får deg til å tenke det. Hvilken av disse prikkene er nærme nok gjennomsnittet til at du skal tenke at «Nei, shit, det er 2 uker igjen, dette kan jeg klare.» Hvilken prikk må du opp til da?

S: Nummer 4 tror jeg. Da er det ikke så stort gap da. Da tenker jeg at det hadde gått.

H: Skjønte du at det var en slags sammenheng mellom din progresjon og din responstid på et spørsmål?

S: Nei.

H: Hvis jeg forteller deg at begge deler er utifra samme forsøk?

S: Åja, jeg ser det nå, men tenkte ikke over det når jeg så på det først. Jeg tenkte det bare var to forskjellige datasett på en måte.

H: Så du koblet liksom ikke at, si forsøk nummer 4 fikk du rett over 50% og at du brukte veldig mye kortere tid på å svare riktig og en ganske god dip i hvor lang tid du brukte på å svare feil.

S: Jeg la ikke merke til de nummerende under. Så nei, jeg gjorde ikke det.



## A.6.8 Interview 8

### Intervju nr. 8 – (statistikk)

---

H: Synes du systemet var interessant? Og isåfall, hva var det som var interessant, eller i verste fall, hva var det som var uinteressant med systemet?

S: Jeg syntes det var interessant, fordi jeg ser på en måte at det kan brukes til veldig mange ting. Om det er liksom ITGK eller om det skulle være teoritentamen til å ta bil-lappen f.eks. Så det synes jeg var kult, at det var et typisk system som kan brukes på mange multiple-choice type temaer da, og flere fag. Så det synes jeg at var kult. Også likte jeg sånn grafisk at det kom frem veldig tydelig hvordan du lå an. Og det kom frem tydelig, iogmed at det på en måte var fylt opp en sirkel med 100% eller om det var sånn som dette mastery level greiene. Så det synes jeg var gode features.

H: Hvis du tenker tilbake på hele programmet som en helhet. Hvis du har type 1 eller 2 features som merker seg ordentlig ut. Så hvis du liksom tenker tilbake igjen og tenker sånn: «Det var denne tingen her, eller disse 2/3 tingene som var programmet for min del». Hvilke features hadde det vært?

S: Sånn type spørsmålene og det at man gjorde disse testene. Men jeg likte da også at når man var ferdig med en test, fikk man opp sånn hva som var rett og galt. Dette tror jeg er veldig lurt, iogmed at man ikke får forbedret seg om man ikke vet dette. Det også at man kunne gå inn på en gammel test å sjekke «hvilket spørsmål var dette her». Det husker jeg ganske godt. Også var det dette her med å kunne sammenligne seg med andre i dette her sånn analyse delen. Det kom ikke så godt frem aller først, hva det var, for jeg måtte lese den beskrivelsen et par ganger før det gikk inn. Men da etter å ha skjønt det, så ga det ganske mening, og det kan være en ganske kul sånn ting for å ha litt konkurranse og sånt. Og det kan jo være motivasjon det å ha konkurranse.

H: Hvis vi stiller spørsmålet på en litt annen måte da. Husker du noen features hvor du liksom umiddelbart tenkte at «dette er sånn man bruker den» eller motsatt da «Dette her forstår jeg ikke i det hele tatt». Var det features som du catcha med en gang, eller features som du ikke forstod i det hele tatt?

S: Jeg syns de som var veldig lett og intuitive å forstå var hvordan man lager bruker. Også hvordan testene foregår: «begynn en ny test her». Det var veldig intuitivt i tillegg til at man fikk disse svarene etterpå med fargekodene f.eks og hvor mange man fikk riktig. Også var det dette med analyse delen som kanskje ikke kom så godt frem første gang, før man måtte lese det lille spørsmålstegnet med informasjonen.

H: Ja, for da er du inne på det inne i quizen, den statistikken som var tilgjengelig på ett spørsmål.

S: ja, stemmer! Så det var det som ikke var så intuitivt først, men det å lage bruker og gjøre testen skjønnte man veldig fort.

H: Du nevner jo det med den statistikken. Første gang du så den da, litt mer i detalj, hva følte du første gang du så den?

S: Ja, for da var det liksom ikke så mye data på den. Som jeg på en måte... Jo! For det var jo en del ulike ting. Det var blant annet en graf som ikke hadde noe data på, så da skjønte jeg ikke helt den. Men det kan jo være fordi jeg ikke hadde tatt spørsmålet selv?

H: Ah, ja, det var vel kanskje den nede til høyre var det ikke? Den største av de?

S: Ja, det var sånn en graf liksom, som skulle vise noe fremgang typ.

H: Den var vel faktisk ødelagt når du testa tror jeg.

--- Mini avbrekk hvor vi forklarte at vi fikset grafen etter hennes testing ---

S: Det var egentlig den som gjorde at jeg ikke skjønte hva den skulle gi meg, for de andre grafene forstod jeg når jeg trykket på spørsmålstegn. Så det var det som var problemet der.

H: Men de andre statistikkene på den som var tilgjengelig. Hjalp de deg på en måte til å forstå noe? Eller var det bare statistikk som var der?

S: For min del var det litt statistikk som var der, og jeg tror jeg kanskje kunne brukt det som motivasjonsfaktor, heller enn å sammenligne seg med andre og på en måte ha internkonkurranse da på en måte. Men akkurat da jeg gjorde testen, følte det mest ut som statistikk.

H: Det som på en måte var litt sånn intensjonen da vi lagde det, var at man for et overblikk over hvor vanskelig spørsmålet er. Er dette noe «jeg burde skjønt sånn» eller noe «jeg burde bruke mye tid på»? Var det en tanke som streifet deg under quizen?

S: Ja, nei, det var en veldig god intensjon, jeg tror jeg bare typisk når jeg testa den 3 ganger... Jeg tror at hvis man hadde brukt programmet litt hyppigere og at du hadde vært litt mer inn i det, så hadde det kanskje gitt deg enda mer enn en testkanin. Det er derfor det kanskje også ble litt sånn statistikk, siden jeg er litt sånn prøvekanin. For man skjønner intensjonen veldig, og det er definitivt lurt det synes jeg.

H: Men så nevner du jo dette her med sammenligning med andre. \*viser screenshot av dashboard\* Hadde du peers tilgjengelig under din prøverunde?

S: Jeg lurer på det.

H: isåfall brukte du den?

S: Jeg lurer på om jeg klikket på den, men husker ikke helt hva som kom opp.

H: \*viser screenshot av peer dashboard\* Hva får du ut av det grensesnittet her?

S: Altså, det første jeg ser som er lettest sånn visuelt, er det som er i midten. Ditt nivå på dette her i «algoritmer» er bedre enn den gjennomsnittlige. Også er det vel antall ganger på x-aksen på de grafene?

H: Ja, antall attempts.

S: Så typisk sånn at man begynner litt treigere, også går det fortere og fortere jo lenger man kommer i det. Jeg tenker hvertfall sånn hvis det går bra, så er det litt sånn motivasjonsfaktor å gjøre det bedre enn sånn gjennomsnittlig, så kunne det gjort at jeg ville gjort testen kanskje enda en gang for å høyne nivået.

H: Hvis vi tar utgangspunkt i forsøk 5 og 6 i eksempel grafene. Hva ville gjort deg mest motivert? Det å ligge rett bak gjennomsnittet? Eller det å ligge rett foran gjennomsnittet?

S: Rett over tror jeg. Fordi da viser man at man har en god progresjon. Det tror jeg. Eller kanskje liksom...

H: Om man ser vekk fra de andre forsøkene, blir du mer motivert dersom du ligger over eller under, uavhengig progresjonen tidligere liksom?

S: Ja, jeg tror det. Fordi det tilsynelatende ser ut som om det er bedre.

H: Du blir mer motivert når du allerede er bedre?

S: Ja. Eller, hva skal man si. Jeg tror at jeg vil at man skal ligge over. Det gjør meg motivert om jeg ser at den ligger over og hvis den er under, så har jeg lyst til å dytte den opp. Så da er det jo på en måte sånn sett at man er mer motivert når den ligger under, fordi man vil få den opp. Så liksom rett under, for da vil man liksom pushe den opp for å få den over gjennomsnittet. Så når den da er over så er det liksom sånn «okei, her er det bra.»

H: Hvis vi tar utgangspunkt i det forsøket her hvor du ligger rett under gjennomsnittet. Vil du liksom bli 5 ganger mer motivert om du ligger så langt under som du gjør på forsøk nummer 1, eller vil det at du ligger såpass langt bak... det øker liksom ikke motivasjonen, eller vil du bli enda mer motivert når du ser at det er langt opp til gjennomsnittet?

S: Jeg tror jeg ville blitt... jeg tror det finnes en balanse der, at det er sånn at man skal ligge litt bak, men ikke for langt bak, for da blir det demotiverende igjen. For da blir det sånn «Oh shit, jeg er så mye dårligere enn gjennomsnittet, og det er ikke noe gøy, jeg går ut av appen». Jeg tror det finnes en balanse der, helt klart. Som kanskje ligger rundt... Si forskjellen på mellom forsøk 3 og 4. Det tror jeg på en måte er hvor min balanse ville ligget. Hvis jeg hadde vært under der, så ville det vært litt kjedelig, men over så hadde jeg tenkt at jeg kunne klart dette her.

H: Hvis vi ser litt på fargen som er brukt her da. Vi går jo interaksjonsdesign, blant annet. Kan du fortelle oss litt om hva de forskjellige fargene her representerer?

S: Ja altså f.eks de mastery levelene ser man ganske tydelig at gull er bra, medium sølv og low gjetter jeg på sånn bronze aktig. Generelt så er det lilla litt sånn nøytral farge synes jeg. Siden bakgrunnen er nøytral synes jeg at det hjelper på å fremheve de andre effektene sånn som f.eks «new attempt» som lyser grønt. «Klar, ferdig, gå» her skal du begynne. Også nederst nede i rødt, gult og grønt, gir det også mening at grønt et bra, eller over gjennomsnittet eller hva det er. Gul må du jobbe litt mer og rød, her var det litt lite. Dette var også ganske intuitivt.

H: Hva synes du om fargevalget på grafene?

S: Jeg synes den lilla var bra, men den andre kunne kanskje stått ut litt mer. Hvertfall som jeg ser på skjermen nå da, så kan de kanskje bli litt like de fargene.

H: Synes du det er bedre fargevalg på den som ikke er sammenligning, men bare dine stats?

S: Ja, jeg synes det, men synes fortsatt at det røde og lille fortsatt går litt sånn sammen igjen. Kunne kanskje gjort det litt tydeligere.

H: Ville du brukt systemet igjen? Hvis du ville det, hva er på en måte motivasjonen din til å bruke systemet?

S: Jeg kunne gjerne tenkt meg å bruke systemet igjen, for jeg har en del fag som er sånn multiple choice. Det å da kunne ha en nettside som da gir deg spørsmål sånn hele tiden kunne definitivt vært en god øving for å øve seg på disse multiple choice spørsmålene.

## A.6.9 Interview 9

### Intervju nr. 9

---

H: Da kjører vi bare på med det første spørsmålet vi har forberedt som er, synes du SmartU er et interessant system? Hvis ja, hvorfor det? Og tilsvarende hvis nei, hvorfor ikke?

S: Ømh ja ... Altså jeg synes det var interessant fordi det minner litt mer om et spill kan du si, som er en ganske annen approach enn det systemet jeg har holdt på med tanke på ITGK da. Det var interessant fordi du kunne se statistikk og sammenligne med andre, i tillegg til det at når du får statistikk på deg selv kan du forbedre deg, og så får du bedre statistikk etter hvert som du forbedrer deg.

H: Ja, kult. Du snakker jo litt om statistikk, hva kan du fortelle oss om fargene som ble brukt i statistikken, hvis du husker det?

S: Ja, det var vel gult, eller ...? Rød, gul og grønn var det det systemet?

H: Ja, for eksempel ...

S: Ja, ja, ja! Jeg likte det veldig godt. Det er jo veldig kjekt at du hvis du først får gul da for eksempel, så kan du for eksempel jobbe videre for å få grønn. Så jeg synes det er ganske positivt at du kan gå på «hjemmesiden din» eller der du får statistikken din og prøve å få alt grønt da.

H: Og hva var det da de forskjellige fargene representerte da sa du?

S: Det er vel på en måte ... mestringsnivået ditt? Er det ikke det?

H: Jo, det kan man vel si. Så, hva synes du da om, vi kan kalle det dashboard-sidene, da for eksempel for en spesiell aktivitet, som for eksempel ITGK. Hva kan du fortelle om dashboardet og sånt?

S: Dashboardet var der du kunne se statistikken som sammenlignet med resten av folkene?

H: Jeg kan dele skjerm, så du kan se ... Om jeg får det til da ... Sånn! Disse sidene her ...

S: Ah, ja de ja!

\*Intervjuobjektet lener seg frem inn i web-kameraet sitt\*

H: Haha, var godt vi ikke tar video-opptak av det her!

S: HAHAHA! Okey ... Ja ... Den husker jeg, den var jo ganske grei, for da får du det overviewet av mestringsnivået ditt over de forskjellige aktivitetene. Og den i midten der, den første du ser, er det det generelle mestringsnivået? Ja, det er jo ganske greit å få se den, og prøve å få den til å være «High». Det er ganske nice.

H: Og hvis du skal fortelle oss på en måte ... Hvis du ser programmet, type denne siden her og så var det en side etter en quiz, og selve quizen ... Hvis du tenker på en eller to features som du husker ekstra godt, som du tenker «Dette var programmet for min del», «Det var dette jeg brukte, som gav meg ekstra mye verdi». Hva er det for deg? Om du tenker på en eller to features ...

S: Hmm ... En eller to ja ... Altså det at jeg kan se meg selv sammenlignet med resten av klassen eller resten av de som gjør det samme, føler jeg hjelper meg med at jeg jeg kan se hvor jeg ligger i forhold til andre. I tillegg til at du kan se progresjonen sammenlignet med alle, ikke bare den endelige scoren. Og spesielt under programmet, at ... du på en måte ... greit nok, tiden var kanskje litt

stressende, men, det er kult å få se statistikk på hvor raskt du selv svarer på spørsmål, og i tillegg om at svaret er galt eller rett hvor lang tid du brukte på det. Det var stilig og nice.

H: Nice. Hvis vi stiller det samme spørsmålet, men på en litt annen måte da ... Om du tenker om det var noen features som var veldig lette å skjønne hvordan man brukte? Var det liksom features du bare tenkte «ah, selvfølgelig det er sånn her man bruker det!»?

S: Ja, altså det er jo ikke noe som var sånn spesielt vanskelig føler jeg ... Det var rett frem, kanskje på dashbordet, det å velge aktivitet? Det var veldig obvious at her trykker du deg inn på en aktivitet, du gjør den ferdig, og så ja ...

H: Ja, så klikkflyten på en måte? Hvordan du navigerer var lett å forstå da?

S: Ja.

H: Var det da noe som var veldig vanskelig å skjønne hvordan du skulle gjøre? Altså det stikk motsatte da, at du liksom tenkte sånn «Okey, hva gjør jeg med det her? Hvordan bruker man det?».

S: Jaa ... Noe som jeg husker som kanskje var bitte litt forvirrende var etter du var ferdig med en quiz, så skjønnte jeg ikke med en gang liksom ... Du kunne jo se hva som var rett og galt på et spørsmål med hva som var grønt og rødt. Men jeg skjønnte ikke med gang at man kunne trykke på de (runde kulene). Det var liksom ikke noe jeg skjønnte før jeg gikk over med musa og det endra seg til at man kunne trykke på det. Så kanskje det var noe jeg ikke helt skjønnte?

H: Ja, ikke sant ... Ehh ... Skal vi se, jeg skal bare dele skjerm igjen her ... Her!

S: Der ja!

H: Det var jo denne siden du snakket om, som kommer etter en quiz. Hvordan brukte denne siden da?

S: Ømh ... Ja, fordi der kan du se hva du svarte, mastery-level og ja ... Det jeg så på det som var at det var en statisk side som viste resultater og så kunne du trykke på neste og gå videre. Men så viste det seg jo at du kunne gå inn å se akkurat hva du valgte og hva som var rett og se mer detaljer liksom. Det var ikke helt tydelig første gangen jeg var inne på den, at det ikke bare var en side du skulle se på og så gå videre liksom.

H: Ja, så det burde kanskje vært noe forklarende tekst da? Eller har du noe forslag? Si du skulle hatt det perfekte utgaven av denne siden, hvordan hadde du endret slik det ser ut i dag da?

S: Ja si du hadde hatt en liten setning som sa «klikk her for å se detaljer» er jo ganske vanlig å ha hvis det faktisk er sånn at du kan se mer detaljer.

H: Ja ikke sant, nice feedback. Med tanke på detaljnivå som blir vist på selve summary-dashbordet, hadde du forventet å få mer data om et spørsmål når du trykket inn på summary-siden enn det som blir vist nå? Annet enn hva som var riktig svar og det du svarte? Hadde du forventet å se statistikk på hvor lang tid du brukte på det spørsmålet for eksempel?

S: Ja, det hadde vært litt kult å få vite at «du brukte AKKURAT så mye tid på AKKURAT dette spørsmålet», men det er ikke noe jeg nødvendigvis hadde forventet nei.

H: Nei, riktig. Hmm ... Jeg må bare finne jukse-arket mitt. Vi kan jo spørre om hvordan statistikken påvirket motivasjonen din?

S: Ja. For meg syns jeg det er ganske positivt om jeg vet at jeg ligger i det øverste sjiktet. Så hvis jeg får hørt tilbake at jeg har medium nivå, eller at jeg sammenlignet med andre ligger rett på snittet, så vil jeg gjerne prøve å gjøre det enda bedre og gjerne gjøre en quiz eller en aktivitet igjen for å ligge øverst. Så det var veldig engasjerende for meg å se at jeg ikke lå der med en gang, men så kunne jeg gjøre noe med et ved å bare ta aktiviteten på nytt med en gang.

H: Føler du at det er mer motiverende jo lenger ned du ligger i forhold til andre peers? Eller hva tenker du om det?

S: Ja altså, sånn sett så er det jo motiverende om jeg vet jeg ligger godt an. Men jeg vil jo gjerne gjøre en aktivitet til jeg ligger der jeg vil ligge. Så det kan jo egentlig gå begge veier ...

H: Hvis du ser på skjermen nå, så viser vi jo sammenlignings-modusen på dashbordet for en aktivitet. Jeg vet ikke om du brukte den når du testet?

S: Jo, jeg tror det?

H: Okey, for det du sier da er at ... Om du ser hvor musepekeren er på skjermen nå, så er den over forsøk nummer 6 på grafen her ... Du sier da at du er mer motivert når du ligger rett over snittet, enn tilsvarende når du ligger rett under?

S: Ja. For da vil jeg prøve å opprettholde forspranget, sånn at hvis jeg vet at jeg først har vært der, så vil jeg ikke duppe under.

H: Vil det si at du er VELDIG motivert når du ligger over snittet og litt motivert når du ligger under snittet? Eller betyr det at du er motivert når du ligger over snittet og umotivert når du ligger under snittet?

S: Eeh ... Det er nok den første ... Jeg ville ikke blitt umotivert når jeg ligger under, men jeg vil fortsatt være litt motivert til å gjøre det bedre. Men definitivt være mest motivert om jeg ligger over.

H: Ja. Og hvor langt ned under snittet skal du ligge før du blir umotivert?

S: Tja ... Litt under går jo fint, men om det blir langt ned så vil jeg kanskje begynne å bli umotivert. Om alle andre gjør det bra og jeg gjør det dårlig så orker jeg ikke en gang å prøve.

H: Hadde du hatt lyst da til å ha tilgang til dette om du lå mellom forsøk 1 til 3 (et godt stykke bak snittet)? Er det noe du ville sett på da, eller ville du heller ønske at du kun hadde tilgang fra forsøk 3 og opp?

S: Hmm ...

H: Med tanke på motivasjon og at målet her er at du skal gjøre det best mulig.

S: Ja, det hadde kanskje vært en god ide, sånn om jeg kom inn i det litt sånn halvveis og jeg så at jeg lå langt bak så hadde ikke det vært så motiverende. Så kanskje jeg hadde gjort to – tre ganger og fått en litt mer gjennomsnittlig score så hadde det vært bedre for motivasjonen.

H: Og tilsvarende på denne som vi har kalt «mastery-level comparison» ... Hvis du får tildelt «low» og det står en stor bronsemedalje på toppen til deg, men gjennomsnittet er «high» og det står en stor gullmedalje under streken. Hva hadde du følt da med tanke på attitude og motivasjon for å lære mer i programmet?

S: Da hadde jeg sikkert blitt litt demotivert kanskje? Jeg hadde jo kanskje prøvd å jobbe mer? Men hadde jeg sett en sånn stor bronsemedalje og alle andre hadde vært «high» så hadde jeg sikkert gjort minimum nødvendig og så bare gått ut.

H: Og sett i lys av at dette her er ikke er noe obligatorisk aktivitet, men heller et tilbud som ... altså ... som kan komme. Det har jo ikke noe med skolen annet enn at du kan øve til fag for din egen del. Om du da hadde sett at du lå såpas langt bak snittet som for eksempel forsøk 1 på grafen i Figma viser, hadde du da fortsatt å bruke programmet? Eller hadde du tenkt «nei, vet du hva, dette var teit, jeg trenger ikke dette programmet» og fortsatt å lese på egenhånd?

S: Ja jeg hadde jo kanskje tatt det litt personlig om jeg hadde ligget veldig lavt og jeg fikk se at andre ligger veldig høyt, og stoffet er lett for dem. Så om dette da var frivillig og ikke hadde noe å si på karakteren så hadde jeg nok kanskje bare lest selv og tenkt at det går sikkert bra.

H: Så du på en måte, ja ... Ikke for å putte ord i munnen din, men det blir du blir såpas demotivert at du føler at du ikke trenger at programmet forteller deg at du er dårlig? Du vil heller at programmet sier ifra når du har fremgang, eller når du er god?

S: Ja.

H: Nice ... Så på en måte ... Hvis du da skulle brukt systemet ... Eller, ville du brukt systemet? Og hva er det i så fall som motiverer deg?

S: Jeg ville garantert brukt systemet om jeg hadde muligheten. Det som motiverer meg er jo igjen, det at jeg kan se meg selv i forbindelse med nivået jeg ligger på og spesielt at du kan se resten av de andre som bruker programmet og deres nivå. Det er jo ganske vanskelig å få kontroll over ellers, med mindre du kjenner alle andre som tar faget eller dere har en felles chattekanal dere snakker i. Så det er sykt motiverende for meg og se hvordan andre gjør det og så sette meg selv i dette ved at jeg får et nivå jeg og. Så jeg hadde nok brukt dette garantert.

H: Okey, da tror jeg ikke vi har flere spørsmål, så da runder vi av med mindre du har noe mer på hjertet du vil snakke om, som du husker ekstra godt eller ...?

S: Nei, det eneste for meg var bare at jeg ikke forstod det etter en aktivitet at du kunne trykke videre på det for å få detaljer. Men det var det eneste jeg hadde som var sånn ekstra.



## A.6.10 Interview 10

### Intervju nr. 10 – (statistikk)

---

H: Syns du SmartU var et interessant system? Hvis du synes det, hva var det som gjorde det interessant, og hvis ikke, hva var det som gjorde det uinteressant?

S: Hmm ... Nei, jeg syns det var interessant og gøy. Jeg har prøvd et par andre tester med spørsmål fra fag, software lignende ting tidligere. Det som var veldig interessant her var jo å få masse statistikk på hvordan andre hadde gjort testene og hvordan de gjorde det. Det gjør det jo veldig lett å evaluere sin egen ... Altså hvor bra man gjør det selv i forhold til andre. Skaffe seg litt feeling på ting som "dette her burde jeg kunne" og "det går greit om jeg ikke vet det her" om alle gjør det dårlig på noe.

H: Ikke sant ... Og når du sier når alle gjør det dårlig på noe, snakker du da om den statistikken som var tilgjengelig under et spørsmål eller tenker du mer på at de jevnt over gjør det dårlig på den statistikken på dashbordet

S: Ja, jeg tenkte på den statistikken på et spørsmål, om det er mange som gjør det dårlig på et spesifikt spørsmål.

H: Ja, for det var jo to typer statistikk, hvis man kan si det sånn. Det var jo statistikk i form av grafene på dashbordet og så var det statistikk som var tilgjengelig under et spørsmål. Hvis vi tar for oss det som var på dashbordet først, altså grafene som viser progresjon og responstid. Hvordan syns du det dashbordet var? og disse visualiseringene som blir presenter, fikk du noe ut av disse?

S: Ja ... Ehh ... Nå tok jo jeg bare to tester, så det ble jo ikke så veldig stort datagrunnlag for å danne grafer. Men jeg kan se for meg at om du bruker det mer, så hadde det vært veldig interessant å se progresjonen, hvordan det utvikler seg. Ikke bare på hvor myr rett jeg får, men også på hvor kjapt det går da.

H: Mhm, gjorde du deg noen tanker om valg av farger på selve dashbordet? Enten i grafene, eller bare generelt på dashbordet?

S: Nei ... Ikke noe spesielle tanker rundt fargevalg nei.

H: Nei, okei. Vil du si at dette dashbordet da kunne vært med og økt motivasjonen din for å ... ja, kall det studere. Ta flere quizer, bli bedre i da for eksempel ITGK?

S: Ja, selvfølgelig.

H: På hvilken måte da?

S: Ehh ... Altså når du får informasjon om hvordan andre gjør det så blir det veldig naturlig at det blir en slags konkurranse greie da. Som gjør at du fort kan sette deg mål som: "Jeg skal høre det så så bra i forhold til alle andre eller i forhold til deg selv". Så du kan konkurrere både mot deg selv og mot andre, og det hjelper på å motivere, istedenfor at du ... ja ... bare gjør det på egenhånd for å si det sånn.

H: Ikke sant, og da snakker du også om det å bruke sammenlignings-funksjonen som var tilgjengelig?

S: Ja.

H: Og bare for å recappe, så kan du der se på gjennomsnittet til de andre i klassen da.

S: Jeg ville bare sett på grafene og sammenlignet den average i forhold til deg selv da.

H: Vil du si at du ble motivert når du første gang du så denne sammenlignings greia og du så din graf plottet opp mot gjennomsnittet av de andre sin graf ... Ble du da motivert til å jobbe mer?

S: Hmm ... Ja! Eller første gang jeg gjennomførte testen, så visste jeg jo ikke at det var en graf på slutten hvor man så alle andre og min egen. Så så jeg den og da neste gang jeg kjørte gjennom ble jeg enda mer motivert. Så ble jeg mest motivert helt på slutten når jeg skulle sammenligne med kompisen min som satt ved siden av og også gjorde testen.

H: Ja, ikke sant ...

S: Og det hadde jo vært veldig kult om man hadde sammenligning mot individuelle, så man kunne invitere og sammenligne direkte mot en annen.

H: Ah, ja kul tanke! Vil du si du blir mer motivert da ... i kontekst av sammenligning og progresjon ... Vil du si du blir mer motivert om du ser du ligger rett foran gjennomsnittet eller blir du mer motivert om du ligger rett bak snittet?

S: Hmm ... Det er vanskelig å si ....

H: Da motivert som "motivert til å jobbe mer".

S: Jeg tror jeg er den typen som hadde blitt mest motivert om jeg lå litt foran, for da er det mye større ... Det er viktigere for meg å opprettholde at jeg ER best, enn å være under snittet og ta igjen alle andre. Men det er sikkert veldig subjektivt.

H: Ja, selvfølgelig. Men vi du si at du er skikkelig motivert om du ligger foran og litt motivert hvis du ligger bak? Eller vil du si du er motivert hvis du ligger foran og umotivert hvis du ligger bak? Hvis du skjønner spørsmålet ...?

S: Hmm ... Det kommer an på tidsperspektivet her ... Hvis jeg ligger under etter å ha fullført veldig mange over lang tid, så vil jeg si jeg er veldig lite motivert. Men om jeg kan se en progresjon på at jeg nærmer meg resten eller at man ser at jeg kommer til å ta igjen de andre om 4-5 tester eller dager, så vil jeg si at det hadde motivert meg mer.

H: Så hvis du da lå langt bak, så hadde du ... Da hadde du bare vært skikkelig demotivert? Da skulle du ikke ønske du kunne bruke det til å bli skikkelig god og gå forbi? Du hadde kanskje ikke tenkt i de banene da?

S: Jo, hadde jeg vært VELDIG langt bak så hadde jeg sikkert blitt motivert til å gjøre det bedre ... For da blir det jo litt sånn "nå må jeg jo faktisk stå på". Men om jeg ligger rett under, så vil jeg ikke si det hadde motivert meg til å skaffe meg den lille som hadde tatt meg over da.

H: Hadde du hatt lyst til å få beskjed om at du lå skikkelig langt bak?

S: Ja.

H: Liksom, hvis du hadde ... ja ... Si du hadde "Low" mastery-level, som da er at du ikke mestrer innholdet og du lå et hav bak snittet. Hadde du da hatt lyst til å ha tilgang til sammenligning av statistikken? Hadde du da hatt lyst til å få det i fleisen da "Du er skikkelig dårlig, og gjennomsnittet er skikkelig bra"?

S: Ja, det vil jeg si. Det tror jeg at jeg hadde hatt godt av.

H: Hmm, okei ... Hvis du tenker på hele systemet da. Kan du gi oss typisk en eller to features som er selve essensen for deg? Sånn hvis du tenker på systemet, så tenker du kanskje "ah, disse to greiene her eller denne greia her er liksom det som gjør dette systemet bra"?

S: Hmm ... Tja, det må jo bli sammenligningen med grafer. Og kanskje gjennomgangen av hva som var riktig og hva oms var galt. Så man får en god ... Da fikk jeg en sånn "dette gjorde jeg det galt på" og "dette er det jeg må jobbe mer med". Det var nok de to viktigste bitene som ga mest verdi for min del da.

H: Hvis vi stiller spørsmålet i en litt annen vinkel. Var det noen features som var helt umulig å forstå hvordan man skulle bruke?

S: Hmm ... Ikke som jeg kommer på ... Eeh ... Det er jo alltid slik at når du bruker et nytt system sp må du orientere deg og sånne ting. Jeg tror ... jeg slet med å finne ut hvordan jeg skulle starte en test om ikke observatøren hadde pekt det ut for meg. Han sa etter hvert "trykk der" og det var først da jeg så hvor knappen for å starte en ny test var.

H: Var det første gang du skulle starte en test, eller når du skulle ta et nytt forsøk?

S: Ja, det var første gang jeg skulle starte.

H: Var det da noen features i andre enden? Som du tenkte "dette var så lett å forstå at ... ja ..." at du bare tok det med en gang?

S: Hmm ... Altså det meste var jo der da. Som å svare på spørsmål og få logga inn, og alle disse tingene som er ganske standard. Men også det med å sjekke statistikken og sånt mener jeg at jeg skjønnte med en gang. Når du er inne i et spørsmål, får opp statistikk og evaluerer hva de forskjellige. Det var vel 4 eller 5 statistikk-ting da inne på hvert spm. Jeg vil si jeg skjønnte hva en og hver av de var.

H: Så bra. Kan du fortelle litt hva syns om den statistikken som var tilgjengelig på et spørsmål?

S: Hmm ... Jeg syns ikke ... For min del var det ikke så interessant hvor lang tid folk hadde brukt på å svare på spørsmålet. Det var mer interessant for min del å se hvor mange som hadde svart riktig og hvor mange som hadde svart galt. Og jeg mener å huske at jeg hadde en baktanke om at hvis det var mange som hadde svart galt, så var det litt sånn ... Da hadde jeg litt en følelse at det kunne være ett lure-spørsmål, eller et litt sånn tricky spørsmål da. Hvis det var veldig høy prosent som hadde svart galt. Men jeg vil si jeg brukte den grafen et par ganger, og kunne nok brukt den enda mer. Men jeg var innom der et par ganger for å se hvordan andre hadde gjort det på samme spørsmål.

H: Ja. Så vil du da si at den statistikken hjalp deg å skjønne "scopet" til spørsmålet, altså omfanget?

S: ...

H: Type brukte du flere av grafene? Eller så du kun på ... "Okey det er kun 25% som har rett her, og 75% som har feil, så her er det enten et vanskelig spørsmål eller så er det lureri"? Du satte ikke dette i sammenheng med de andre grafene?

S: Nei, jeg sammenlignet veldig lite informasjon fra de forskjellige grafene på et spørsmål.

H: Ja. Så statistikken på et spørsmål sa deg egentlig bare "her er det mange som har svart feil" så lukket du det?

S: Ja. Eller ... Mer, dette er noe jeg mest sannsynlig må se nøye over da. Eller ikke ... ja.

H: Husker du noe om hvordan statistikken fikk deg til å føle? Altså ble du stresset når du sjekket statistikken? Ble du lettet, glad eller trist?

S: Jeg ble nok overrasket når det var veldig mange som hadde svart feil. Ømh ... vil jeg si. Men ikke lettet eller glad eller ... Ingen sterke følelser vil jeg si.

H: Ble du da overasket fordi du selv oppfattet spørsmålet som lett? Eller ... Hvorfor akkurat overasket?

S: Hmm ... ja, det var nok fordi jeg så på det som relativt trivielt ja.

H: Okey. Kult.

S: Men det var vel sikkert også der jeg svarte feil da ...

H: Ja, ikke sant, haha. Men syns du at du fikk hjelp av fargesettingen på grafene når du skulle ... når du åpnet statistikken. Syns du at du fikk hjelp av fargene, av måten ting var ... fargelagt på?

S: Det ... kan jeg egentlig ikke huske. Men det var ingen ting som tilsa at det IKKE hjalp.

H: Okey, så helt nøytralt da?

S: ja.

H: Okey. Ville du brukt systemet igjen? Og i så fall hva er det som oppmuntrer deg til å bruke det igjen?

S: Nei altså jeg er veldig fan av å bruke sånne ting istedenfor å bare bruke en bok og sitte og lese for eksempel. Så er det jo litt gøy. Det er jo en litt annen måte å lære på, det å bare få servert spørsmål og svar, i stedet for å bare sitte og lese. Det er en litt mer effektiv måte å gjøre eksamensett på, kanskje.

H: Okey, effektiv hvordan da?

S: Nei altså ved at man slipper å bla rundt i en PDF eller i ark man har printet ut for å sjekke løsningsforslag, for det får du servert helt på slutten når du har fullført testen. Eller det blir vel mer svarene da, og ikke løsningsforslag med utregning osv. Men ja, så er det mer funksjonalitet enn om jeg bare skulle gjort det selv i og med at det er all informasjon fra flere folk som faktisk bidrar da. Det gjør det jo interessant vil jeg si.

H: Så du vil si at programmet gir deg økt innsikt i din egen kompetanse da?

S: Ja, både min egen, men også alle andre som også tar faget da. Så det er jo for så vidt ... Jeg vet ikke om jeg vil si motiverende, men fort om du sliter med noe i pensum da, så kan det være litt sånn trøstende å se at det er faktisk mange andre som sliter med de samme tingene.

H: Så har vi et spørsmål om design til slutt. Hva synes du om det overordnede designet av programmet?

S: nei, hmm ...

H: Det kan være hva som helst, form, farge, hvordan ting ser ut ... runde hjørner eller ikke.

S: jeg tror ikke jeg rakk å danne meg så mange tanker rundt det. Men jeg tror det var veldig mye farger ... eeh ... på dashbordet, i forskjellig form og farge, og sånne ting. Jeg har en liten følelse av det kanskje var litt rotete?

H: Lite oppfølgings spørsmål ... Hvordan synes du flyten var i programmet? Fra du lagde en bruker til du hadde fullført et par quizer?

S: hmm ... Den var grei. Noe jeg kanskje kunne savnet var ... å ikke bli sendt dirkete tilbake til dashbordet når du var ferdig med en quiz, men heller bare starte en ny quiz med en gang. Jeg vet ikke om det var funksjonalitet for det, men jeg havnet i alle fall rett tilbake i dashbordet?

H: Sånn at du kunne gått rett fra quiz-summary og startet på en ny quiz?

S: Ja, for eksempel.

H: Så om du hadde hatt en egen knapp i stedet for bare "finish", typisk "prøv igjen"?

S: Ja, for så vidt, i stedet for at det blir veldig punktum da, så kan man prøve å holde det gående på en måte? Jeg vet ikke helt, det er litt vanskelig å sette ord på ... Men ja.

H: Men da har vi vel dekket alle punktene, og det var veldig bra info. Da runder vi av med mindre du har noe mer på hjertet?

S: Nei jeg har ikke noe, annet enn at det var et veldig kult program. Bra jobbet!



# Appendix B

## Design, Implementation & Product

### B.1 User Tests

## B.1.1 User Test 1

### Usertest questions

1. You have a Feide account from NTNU and want to login to SmartU using this account.
2. After logging in, what do you see?  
Please explain the features.
3. There have been some announcements in the class, what are these and when where they posted?
4. Imagine you have visited the SmartU application before, what did you do the last time you visited?
5. You want to see detailed information about the "Algorithms" assessment. How do you proceed?
6. As you have attempted the assessment a couple of times before, what can you tell about your performance? Please explain what you think the different elements does.
7. What is your current mastery-level in Algorithms?
8. You would like to compare your progression to other peers in your class, how would you proceed to accomplish this?
9. Please explain how the ("new") graphs work. What do they tell you?
10. How would you proceed in order to make a new attempt to the assessment?
11. How do you answer a question?
12. How do you answer a question when there is an image shown?
13. Can you exit the quiz? If yes, how do you do that? (press no)
14. The quiz is finished, please summarize what you see in the summary-page.  
How much time did you spend and what was your average time pr. question?
15. Which questions did you answer incorrectly? How do you find out what the correct answer to one of these questions is?
16. You are sick and tired of the assessment-system, how do you log out?



	Subject 1	Subject 2	Subject 3
1	Logger inn lett	Logger inn med Feide	Mener det ikke ligner på feide sin innlogging, men logger inn likevell
2	Hvorfor er quotet her? Announcements? Hva er det til og hvorfor er det her? Ser ut som kurs-modul er koblet opp mot et fag	Ser annonseringer først, de er store og fine, lett å lese de Ser et quote, hva jeg gjode sist og jeg ser jeg har noen tilgjengelige moduler.	Ser announcements Ser det finnes en bruker som er trykbar i navbaren, hva skjer om jeg trykker på den? Ser at vedkommende ikke er så flink i noen av modulene som er tilgjengelig, ser også at modulene som ligger der er forskjellige fag/quizer
3	Skjønte lett hvordan man leste annonseringer	Skjønte lett hvordan man leste annonseringer	Skjønte lett hvordan man leste annonseringer
4	Skjønte ganske fort hvordan man finner ut hva man gjorde sist man var pålogget, noe forvirret av datoene som var skrevet feil i prototypen	Veldig lett	Veldig lett
5	Klikker på algoritmekortet med en gang for å finne ut mer om "algoritmer quizen"	skjønte det med en gang	Skjønner det med en gang
6	Previous attempts: tidligere forsøk på akkurat denne quizen Fun-facts: random stats fra quizen som er litt annerledes enn de fra grafene Progression: Kan se han er blitt linært bedre over tid ut fra grafen	Previous attempts: Jeg ser jeg har blitt bedre ut i fra mine forsøk med statistikk på Progression: denne grafen viser også at jeg har blitt bedre over tid Ser også at tiden jeg bruker pr spm går ned etterhvert som jeg blir bedre	Progression: Ser at grafen har en linær økning Fun-facts: Ser at det er et par fun-facts om performancen min til nå Previous attempts: Ser at det er tidligere forsøk i kortene på bunnen og at de er knyttet til progression grafen som ligger over
7	High	High	High mastery
8	Trykker på peer med en gang	Trykker på peer med en gang	Trykker på peer med en gang

9	Skjønner comparison mellom seg selv og peer på performance/progression, bruker litt mer tid på å forstå fun facts	Progression er ikke helt intuitiv om du ikke husker hvordan grafen så ut før du trykket på "peer"  Ser også at fun-fact endret seg	Peer comparison blir lagt til på progression og fun facts.  "Jeg synes det er litt vanskelig å skjønne hva som menes med de nye grafene, de er litt difuse... Vet liksom ikke om grafen med den røde streken er en avg av de andre eller en annen spesifikk person?"  Grønn og gul farge kan være litt vanskelig å skille mellom ved første blick om det går fort
10	good	good	good
11	good	good	good
12	good	good	good
13	good	good	good
other comments	Quotet var litt forvirrende, det ble litt blikkfang og det andre blitt støyete pga det.	Forventer å få opp feide sin login-portal når det trykkes på login med feide	Ser at ulike komponenter i systemet brukes flere ganger, det gir følelse av helhet og gjør komponentene lettere å kjenne igjen
	Skulle ønske jeg fikk noe mer ut av å ha oppnådd en high mastery enn kun en gull-medalje det stod high på	Savner noe som indikerer hva som er din graf og hva som er peer sin graf, kanskje en forklaring på hva som er hvilken farge eller tilsvarende	Hvem sin mastery-level er det jeg ser når jeg har trykket på "peer"-knappen? Er det min eller en kombinasjon av min egen og de andres?
			Skjønner ikke hvorfor antall spørsmål jeg har svart på endres på de forskjellige attempstene jeg har hatt
			Peer comparison oppfattes av meg som litt "hulter til bulter" slik det er nå, ikke helt klart hva som er hva

## B.1.2 User Test 2

### Ustertest Questions v2

1. You have never used SmartU before and need an account to login. What do you do?  
*Du har aldri brukt SmartU systemet før og trenger en brukerkonto for å logge inn. Hva gjør du da?*
2. After logging in, what do you see? Please explain the features.  
*Etter du har logget inn, hva ser du av elementer og features?*
3. There have been an announcement in the course, when was it posted? Are there more announcements?  
*Det har kommet et varsel i faget fra foreleser, når ble dette postet? Finns det flere varsler?*
4. You are a bit unsure on how to use the system. How would you proceed in order to find more information on how to use it?  
*Du er litt usikker på hva du skal gjøre videre, så hvordan ville du gått frem for å finne ut mer informasjon om dette?*
5. You want to use this system as it is intended, how do you start an activity?  
*Du vil ønske å bruke systemet til det det er lagt for, hvordan starter du en ny aktivitet?*
6. There are a couple of actions available when conducting an activity, before answering a question. What are these actions, and how do they work?  
*Det er en del funksjoner tilgjengelig under en aktivitet, men før man besvarer et spørsmål. Hva er disse funksjonene og hvordan fungerer de?*
7. You wish to complete the activity by answering the last question, what can you say about your performance in the activity?  
*Du ønsker å fullføre aktiviteten din ved å svare på det siste spørsmålet. Hva kan du fortelle om din prestasjon?*
8. Imagine that you have completed several activities now. After pressing the "finish" button, what can you tell about your overall performance and your last activities?  
*Se for deg at du har fullført flere aktiviteter over tid. Etter at du går videre med "finish" knappen, hva kan du fortelle oss om den oppsummerte prestasjonen din og dine siste aktiviteter?*
9. You would like to compare yourself to other classmates, how would you proceed?  
*Du har lyst til å sammenligne deg med andre klassekamerater, hvordan gjør du dette?*
10. What do the metrics available in the peer-comparison interface tell in terms of your performance?

*Hva forteller den tilgjengelige statistikken deg om din prestasjon i forhold til dine klassekamerater?*

11. You would like to improve on another activity where you have a low mastery level. How would you proceed in order to find and select this activity?  
*Du har lyst til å forbedre en annen aktivitet hvor du har et lavt mastery level. Hvordan går du frem for å finne og velge denne aktiviteten?*
  
12. You choose to start a new attempt at the activity, but this time you want to have a closer look at the available statistics to the question. What kind of information can you aggregate from the different data?  
*Du velger å starte et nytt forsøk på aktiviteten, men denne gangen har du lyst til å ta en nærmere kikk på hva slags statistikk som er tilgjengelig til spørsmålet. Hva slags informasjon får du fremstilt, og hva forstår du ut fra denne informasjonen?*
  
13. You realize that you don't want to finish the activity, and want to abort the attempt. How do you proceed?  
*Du innser at du ikke har lyst til å fullføre aktiviteten, og vil avslutte forsøket. Hvordan gjør du dette?*

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
1	Fyller ut form og trykker create.	Logger lett inn.	Logger inn med feide	New user -> create (ble kontoen egentlig opprettet nå?). Logger inn	Logger inn med feide
2	Det mest interessant er kortene i bunn.	Skjønner alt.	Forklarer alt bra	Beskjed fra fagansvarlig/studass, tidligere ting	ser på velkommen tilbake, announcements og activities
3	Ingen svar men finner announcement modal.	24. nov, ingen varsler, finner modulen lett.	Trykker se mer på announcement	klikker på announce av og på. "Ser ikke at det er flere".	Fant enkelt frem til 3 dotter og antar at det betyr at det er en oversikt.
4	Kikker på kortene med en gang.	Trykker "Information", lese alt, nyttig info	Trykker "Information", nyttig	Trykker på "information" og tenker "Aaah, må jeg lese alt dette?". Lettere om det ble presentert annerledes, kanskje mer punktvis.	Finner enkelt informasjon-knappen. Misforstår litt hva som går/ikke går mtp. bold skriften før punkter. "Not available" blir tolket som "available"
5	Trykker "Try activity" (Hvorfor må man trykke try og ikke start?)	Trykker Algorithms ("Har du starta nå?" "neeei") Trykker try etterhvert	Algorithms -> Try activity (går fort)	Trykker "algorithms". Her er det noe nytt siden ikke det er noe stats. (tomt dashboard). Trykker try activity.	Går riktig inn på try activity
6	Statistikk om ting <u>du</u> har svart på	Skjul/vis timer, hvorfor må jeg starte timeren? er noe statistikk for peers.	Vet ikke hva timer gjorde. Statistikk, viser sikkert mine stats	Regner med at timer viser tid de har brukt. Stats er hvordan oppgaven er løst av andre, det likte jeg.	Timer: viser tid/timer deg selv. Statistikk: Generell statistikk på spørsmålene fra alle brukere. Kan avslutte quiz.
7	Fikk mastery high, antall rette svar. "Hvor ble det av statistikken?"	Feil på 4 spørsmål, skjønner stats, mastery sier jeg gjorde det bra. Hva betyr egt. mastery? Er det basert på peer performance eller fagets læringsmål?	Riktig og galt (16/20 riktig), litt stats.	Wow high mastery, kult! Lett å skjønner hvor mye tid man har brukt, veldig bra med ikonene. Ikke så lett å skjønne at man kan klikke de røde. Hjelper kanskje litt med hover?	4 Feil og hva som var rett. Mastery level (basert på hvor mye som var feil?). Forskjellige stats over quizen. Ser enkelt hva det riktige svaret var på feil spørsmål.
8	Peers: Hvordan jeg vs andre?	Er algorithms et fag eller en aktivitet? Skjønner ellers det meste. Previous attempts virker som en quiz. Statistikk ser ut som det kommer fra et fag.	Progresjon er linka mot previous attempts og jeg ser jeg bruker mindre tid.	Usikker på hva progression er ut fra, men skjønner jeg er blitt bedre. Ser jeg bruker mindre tid. Ser etter litt at det er for en ting og ikke totalt for hele platformen.	Progresjon: Hvor bra over de forskjellige forsøkene. Forstår tidsgraf greit. Forstår attempts og self-awareness godt.
9	Trykker på peers lett	Trykker peers med en gang	Trykker på peers	Trykker peers, lett å skjønne.	Finner peers fort.
10	"At jeg er bedre enn andre"	Skjønner grafene godt på peers	Ser ut som jeg har vokst godt og tatt igjen de andre. Ser på time at jeg er litt tregere enn de andre.	Har bedre progresjon enn average. Ganske lik svartid.	Ser average på andre og seg selv. Lurer litt på om mastery level er sammenlignet med andre.
11	Finner lett frem	går "tilbake" og trykker "Data structure"	Skjønner ikke at er inne på Algoritmer. Går fort når det forstås at det er "subjects".	Home -> low	Går tilbake og finner data structures. Litt usikker på "aktivitet" formulering.
12	Skjønner ikke effort eller performance (må utbedres). Skjønner de to nederste grafene. Gir mer mening når det blir gitt mer kontekst.	"Kake diagram er om meg". (Avg-, effort- og performance-time er om peers. Skjønner ikke performance. Resonnerer at spm er vanskelig.	Hva er effort? Skjønnte ikke dette. Usikker på om correct/wrong er meg eller alle? men det er sikkert alle. Tror denne infoen sier meg at spørsmål er vanskelig.	Usikker på om correct/wrong er basert på meg eller andre. Effort: hvor mange bruker tid/skipper ikke. Performance: Hvor mange gjør det bra. Ser ut fra stats at spørsmålet er vanskelig.	Ser på statistikk. Hva er effort og performance? Kake og average er lett å forstå. Basert på stats, virker det som om det er et vanskelig spørsmål.
13	lett	Avslutter quiz lett	Avslutter lett	lett	Lett
other comments		Previous attempts: Hva om jeg vil ta alle spørsmål for å pugge, går ikke det? Er spørsmål knyttet til et kapittel? Gir litt mer mening med mer kontekst om at systemet tilhører itg og har lest mer om "Information."	Usikker på når det er statistikk for meg eller andre.	Veldig mye er veldig lett å skjønne.	Previous attempts: ser på farger for å se prestasjon kjøpt.
		Liker kapittel inndeling bedre enn en "Ove til eksamen"-quiz, men med mulighet til å ha en quiz med alle spørsmål. Det er en use-case jeg hadde brukt og ville da forventet en mastery uten adaptivens-konseptet. Trenger bare være åpent rett før eksamen. Hadde brukt systemet som et supplement til boka og et morsomt avbrekk.	La ikke så mye merke til dato i navbar	Previous attempts viser litt mye info, nice om man kan klikke inn og se mer.	Previous actions på forsiden. Nyeste aktiviteter man har gjort og % score som man fikk rett på aktiviteten totalt.
			Ganske intuitivt hele tiden!	Available activities er tydelig og lett å forstå.	Information: Blokk med tekst, men kjøpt å lese bold og det som stikker seg ut.
				Quote tar litt mye plass kanskje?	

				Tenkte på % grafen til previous actions på hvordan det ligger ann eller hvor mye av fullført av tilgjengelig materiale. Kan også være hvor bra man gjorde det på siste forsøk.	
				Ønsker mer feedback på handlinger.	

## B.2 Final User Interface

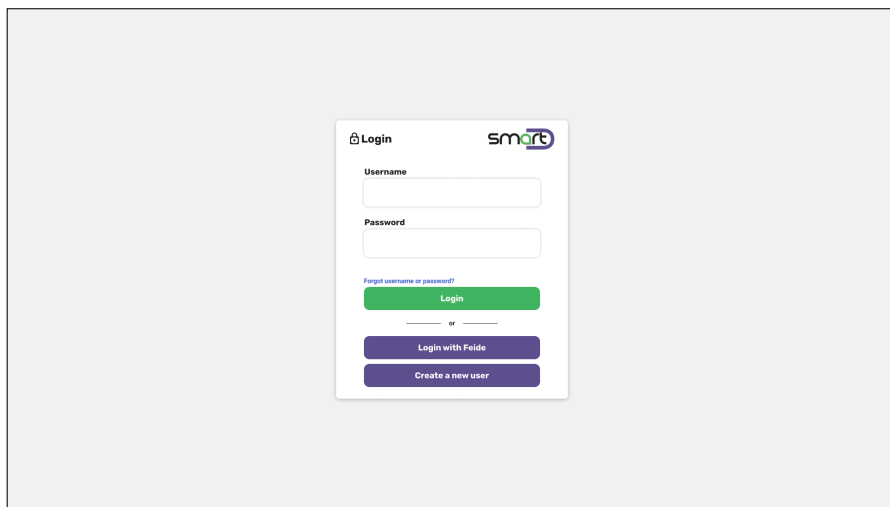


Figure B.1: UI Entrypoint

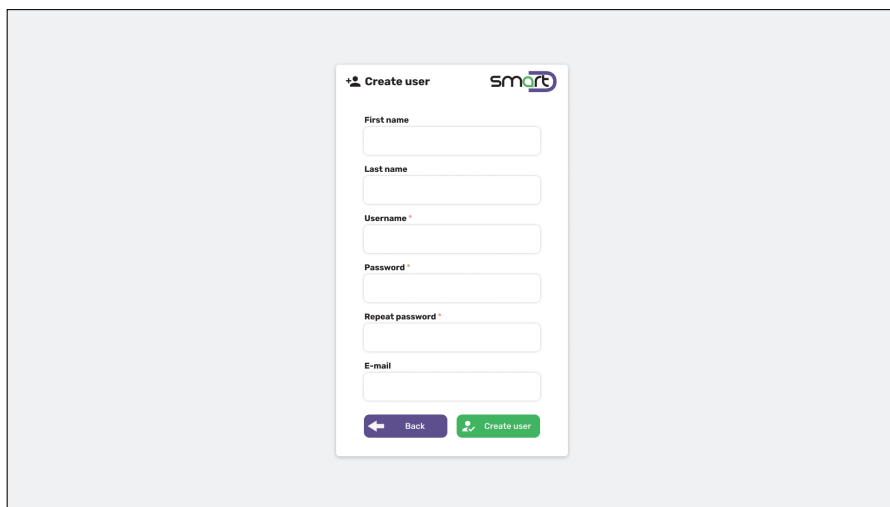


Figure B.2: UI Create User

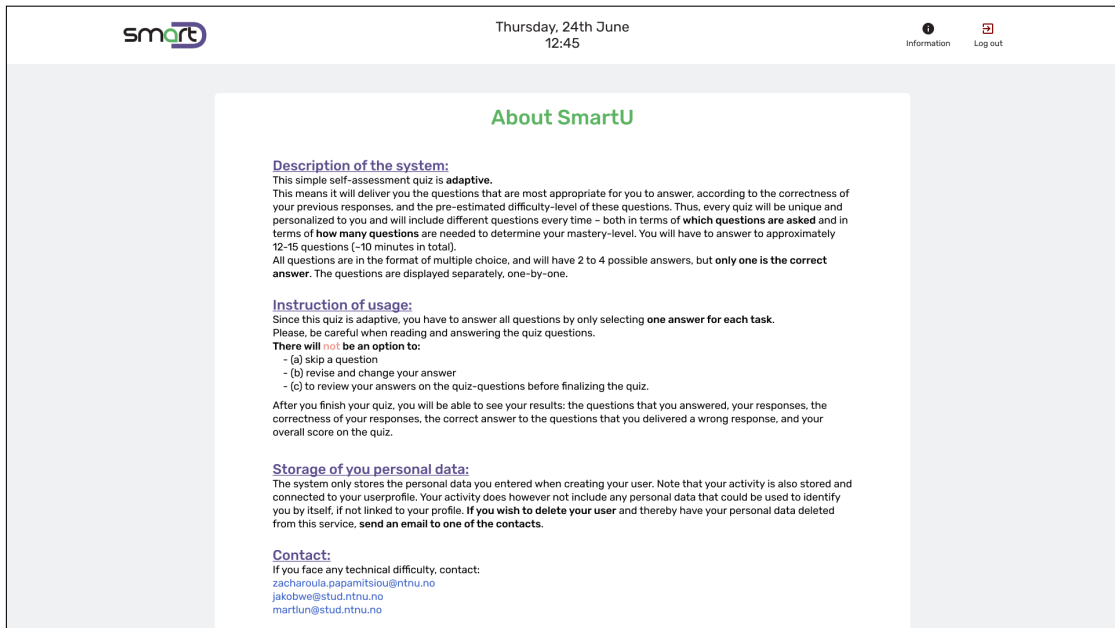


Figure B.3: UI About

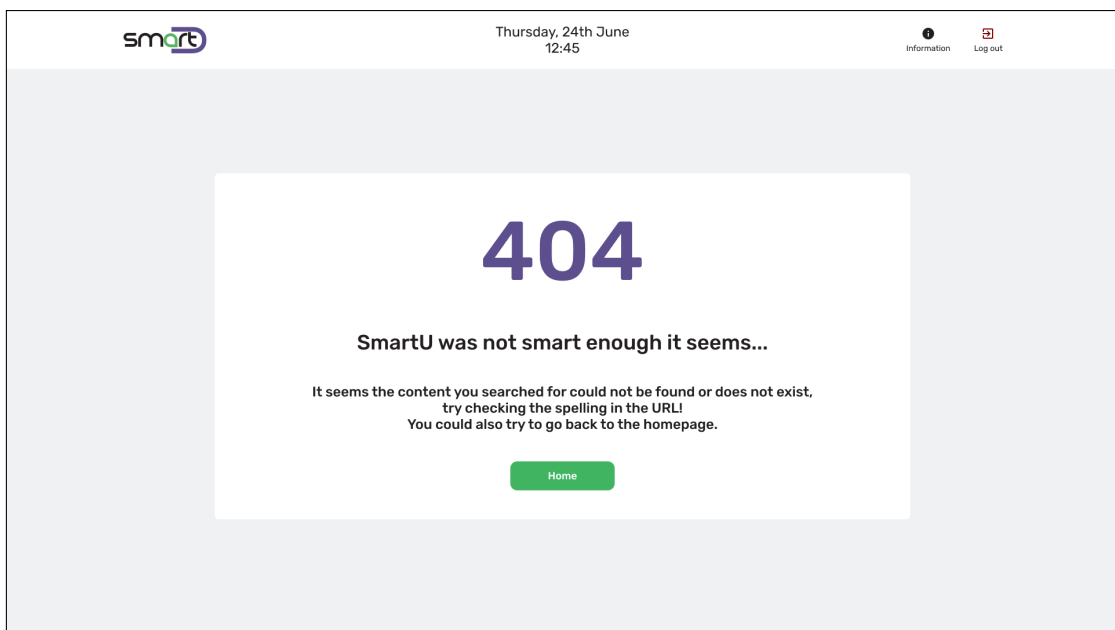


Figure B.4: UI 404 page



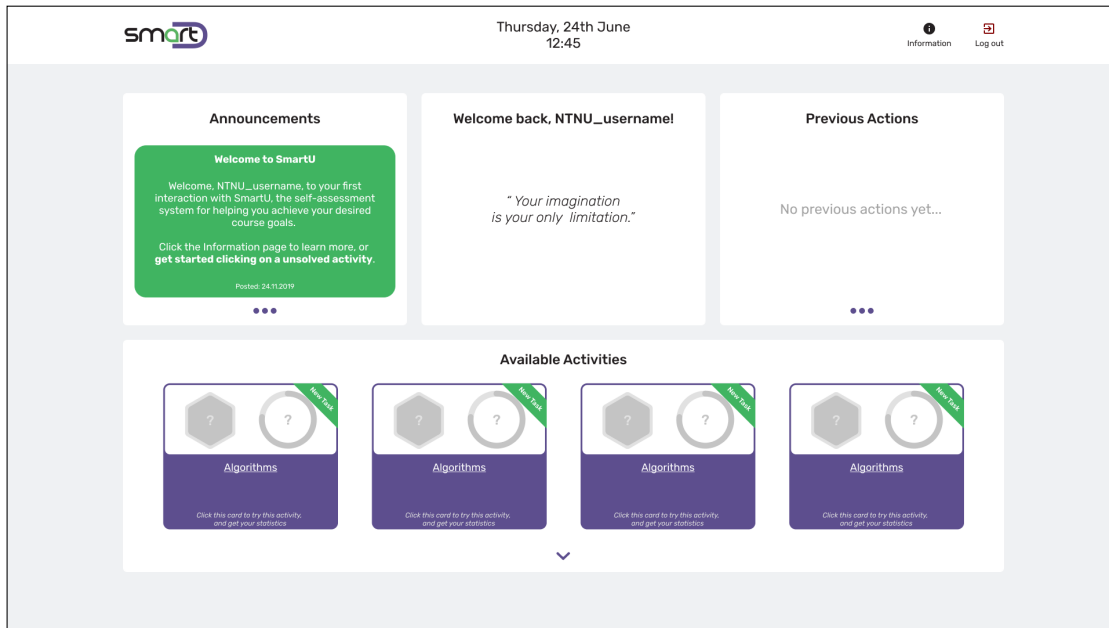


Figure B.5: UI Landingpage Empty

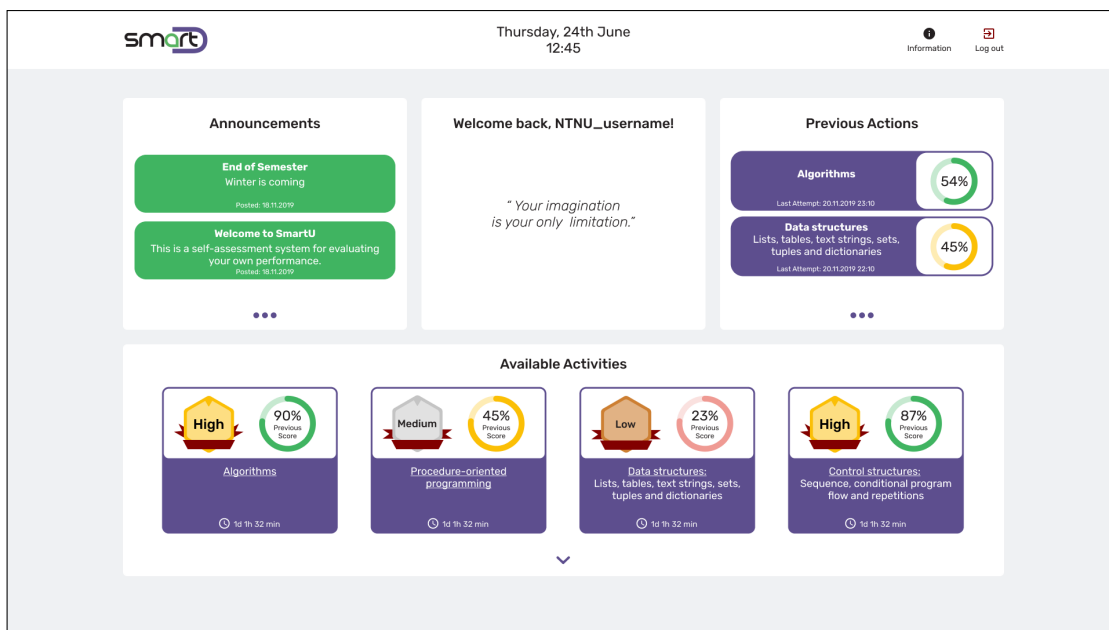


Figure B.6: UI Landingpage

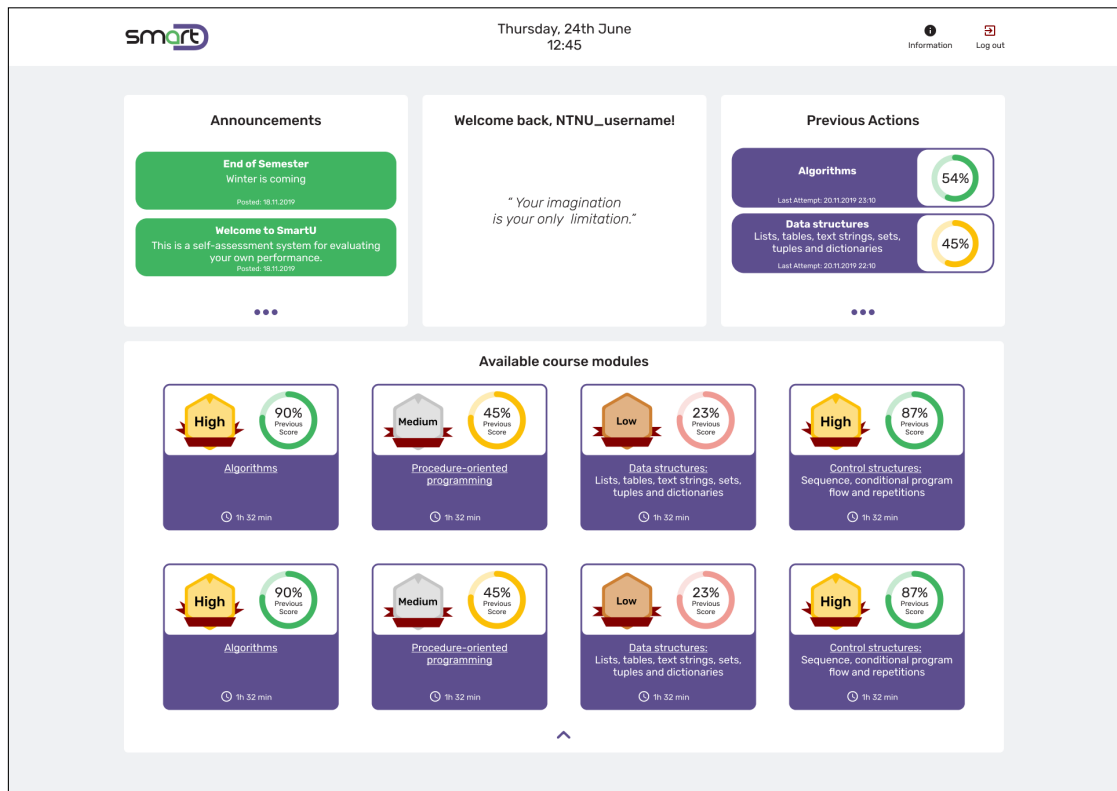


Figure B.7: UI Landingpage Expanded

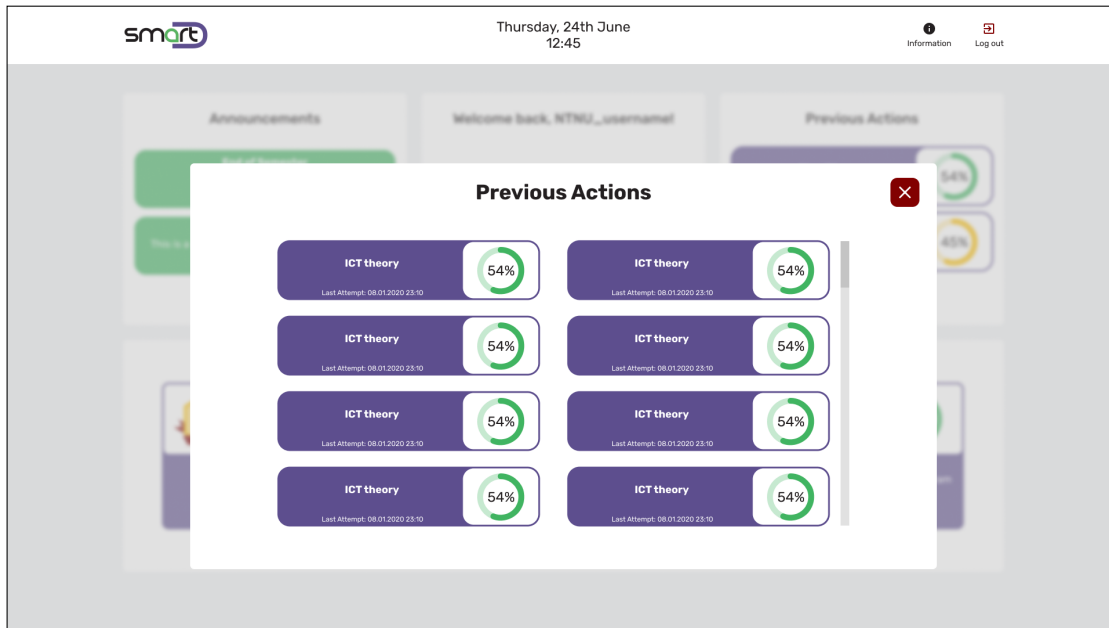


Figure B.8: UI Landingpage Modal - Previous Actions

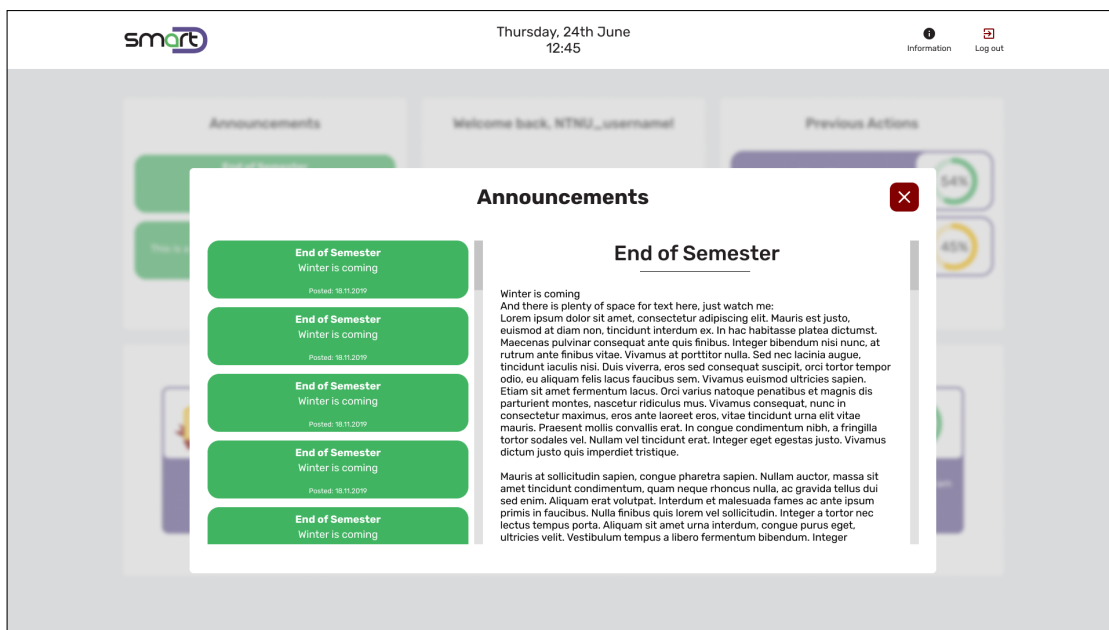


Figure B.9: UI Landingpage Modal - Announcements

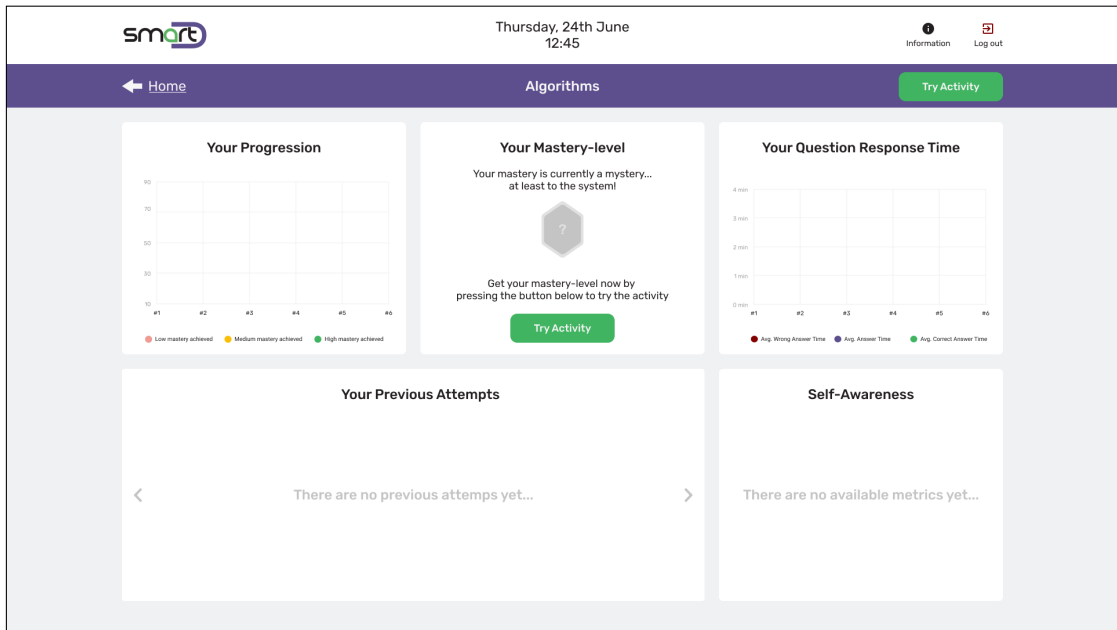


Figure B.10: UI Quiz Details Empty

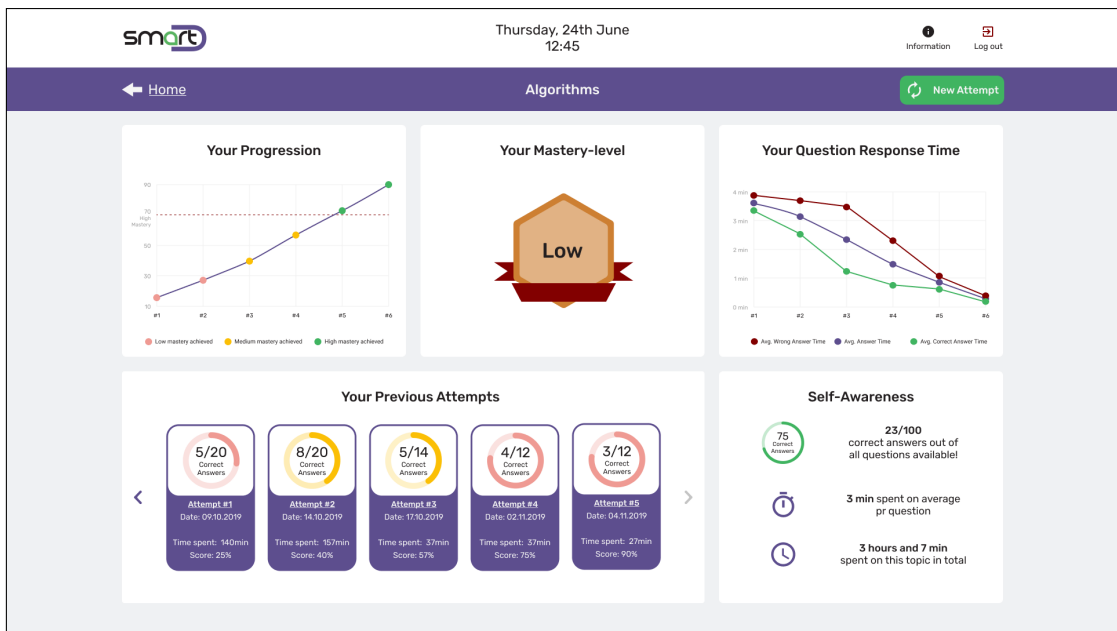


Figure B.11: UI Quiz Details

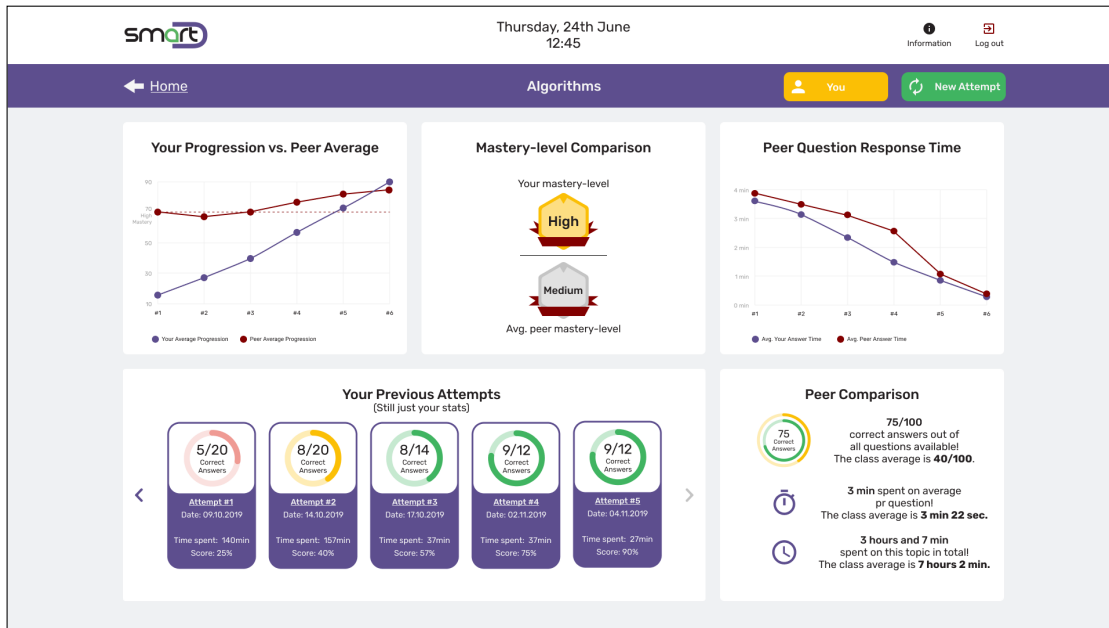


Figure B.12: UI Quiz Details with buttons

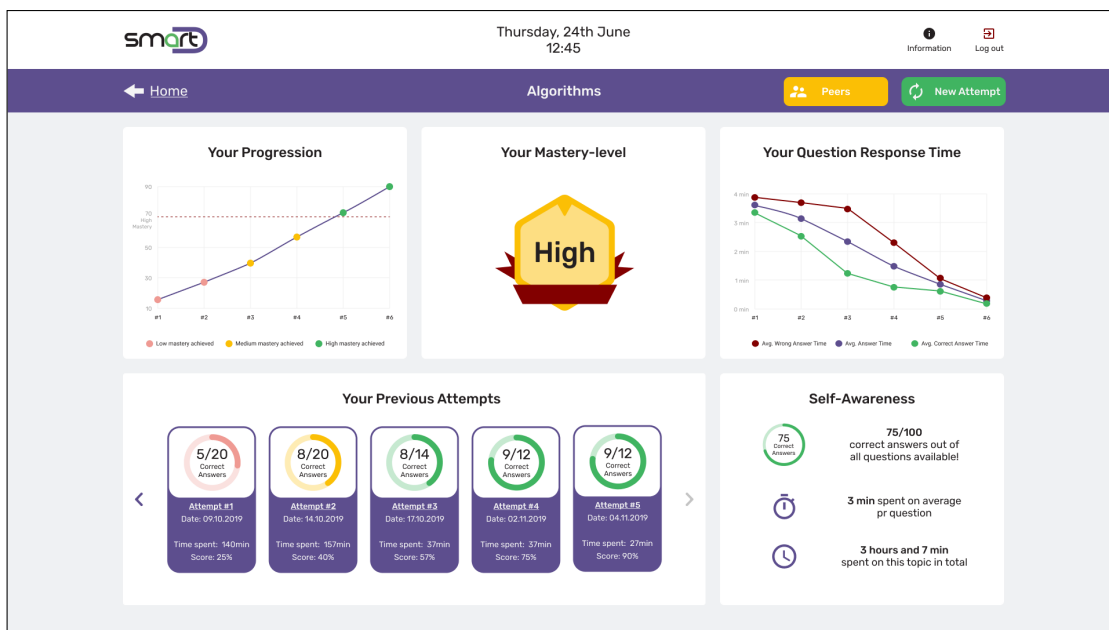


Figure B.13: UI Quiz Details with buttons 2

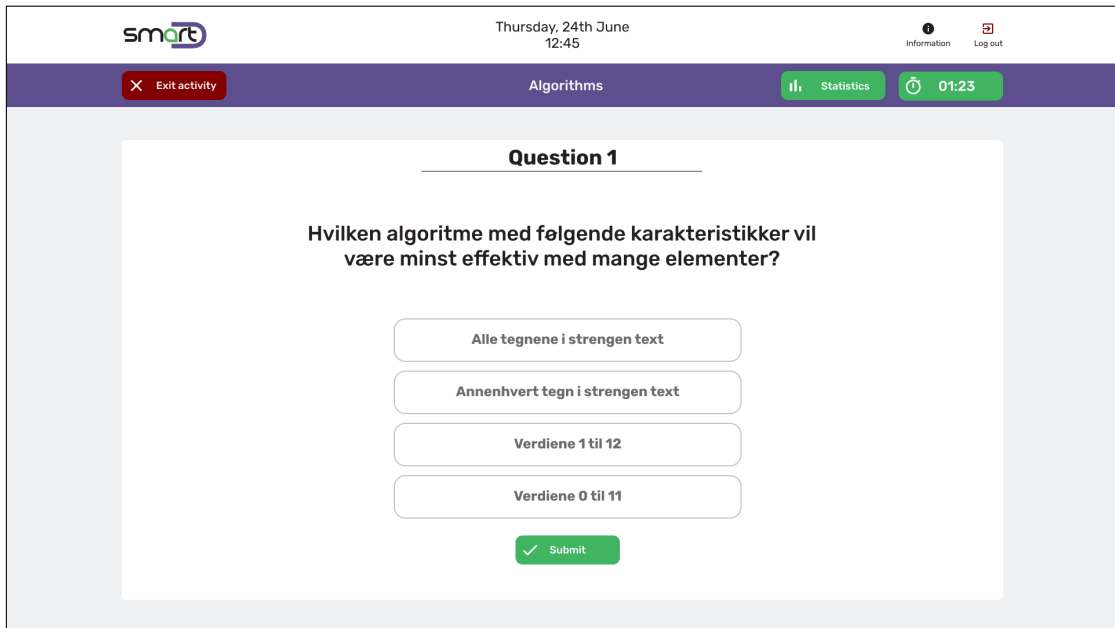


Figure B.14: UI During Quiz

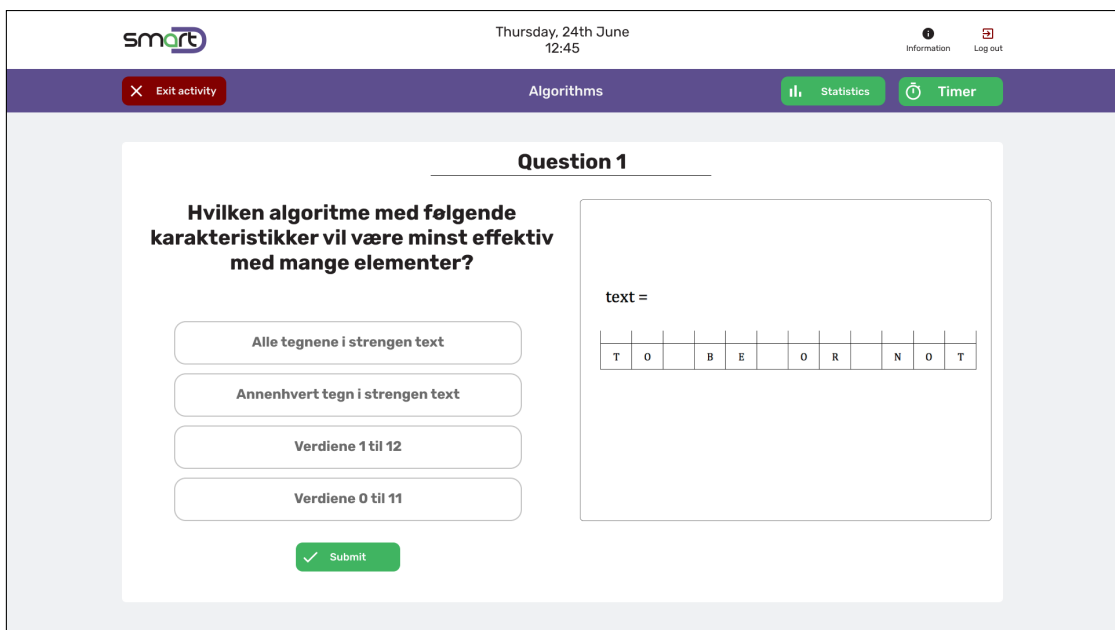


Figure B.15: UI During Quiz with img

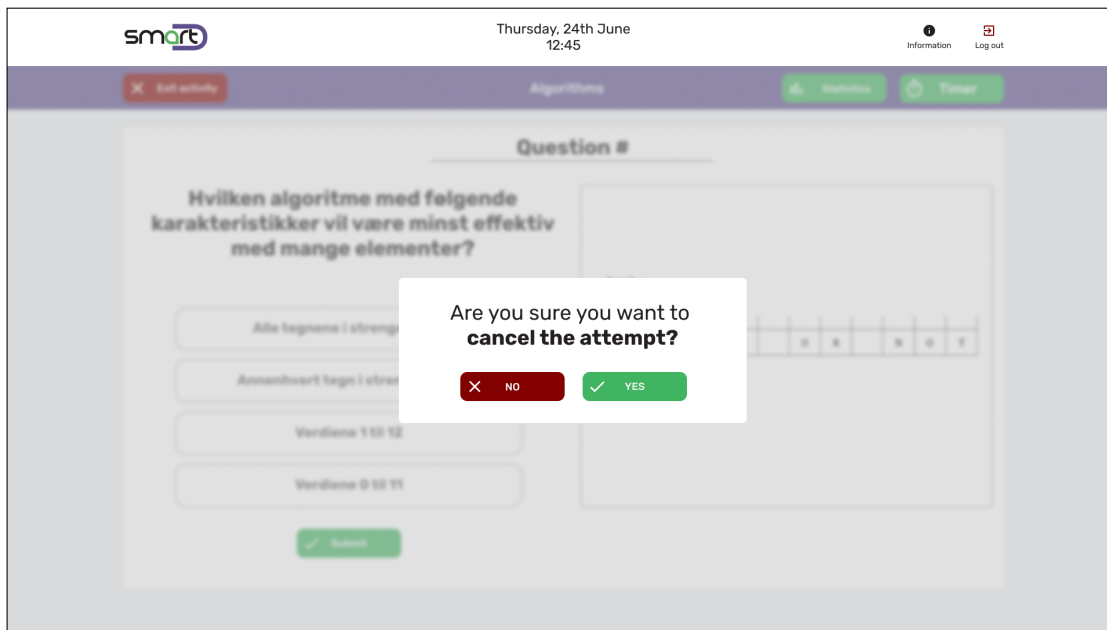


Figure B.16: UI During Quiz Modal

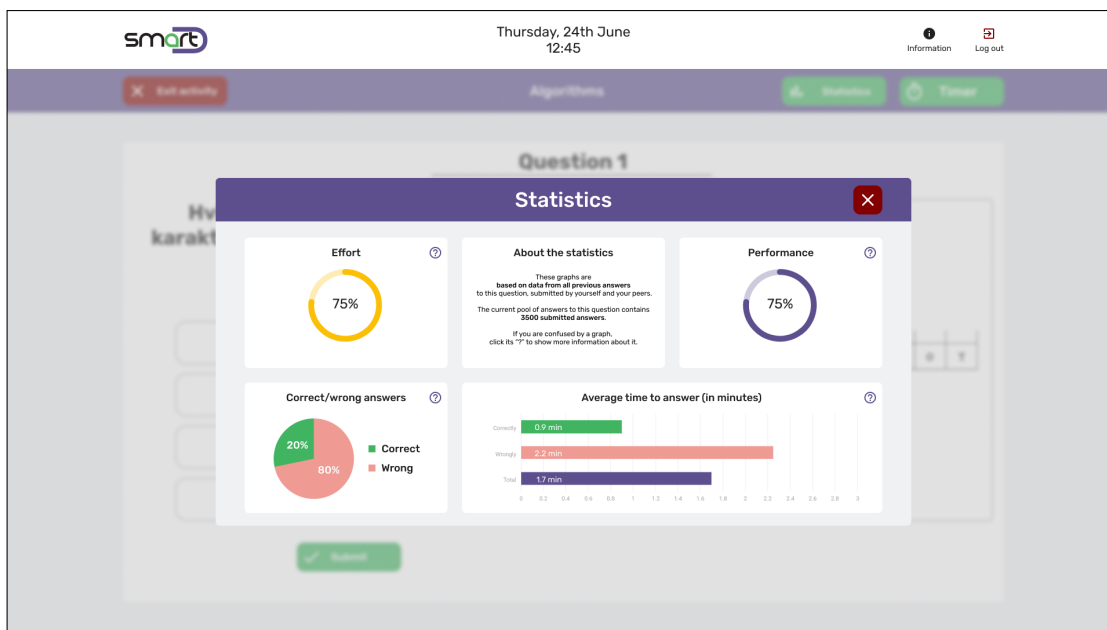


Figure B.17: UI During Quiz - statistics

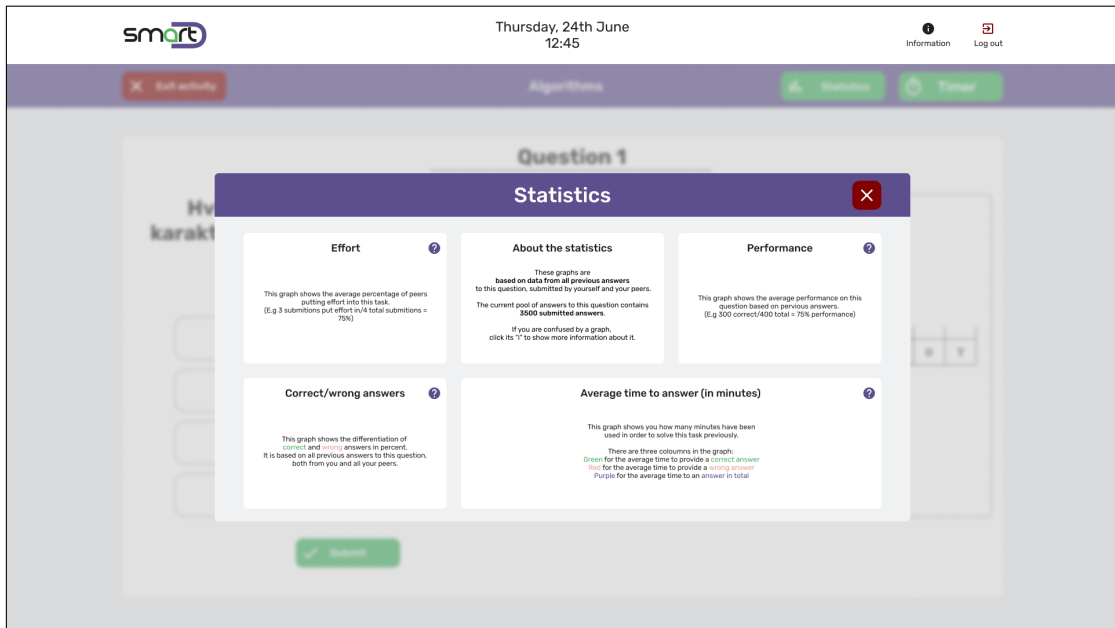


Figure B.18: UI During Quiz - statistics information

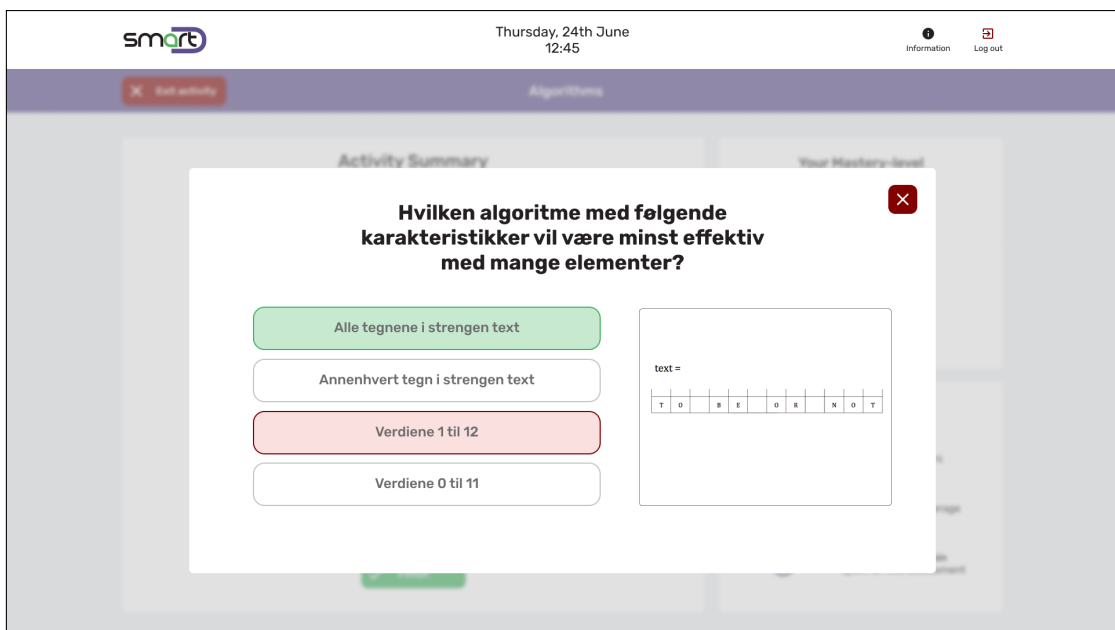


Figure B.19: UI Quiz Summary



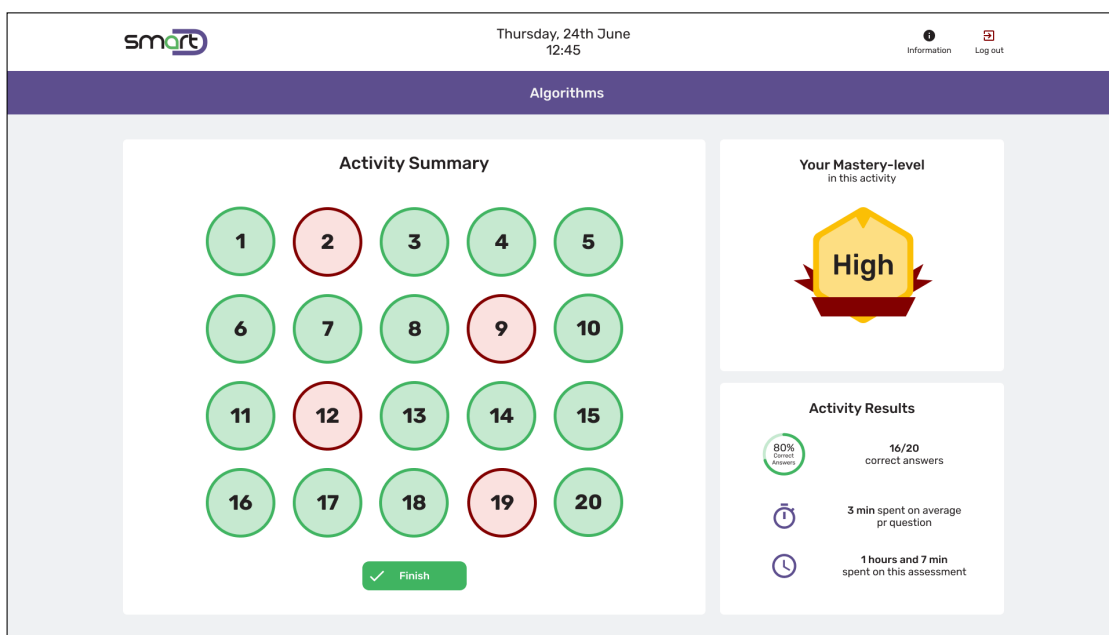


Figure B.20: UI Quiz Summary - Modal

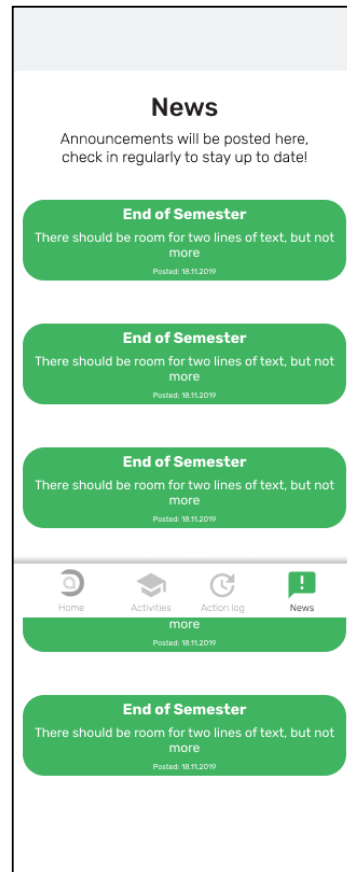
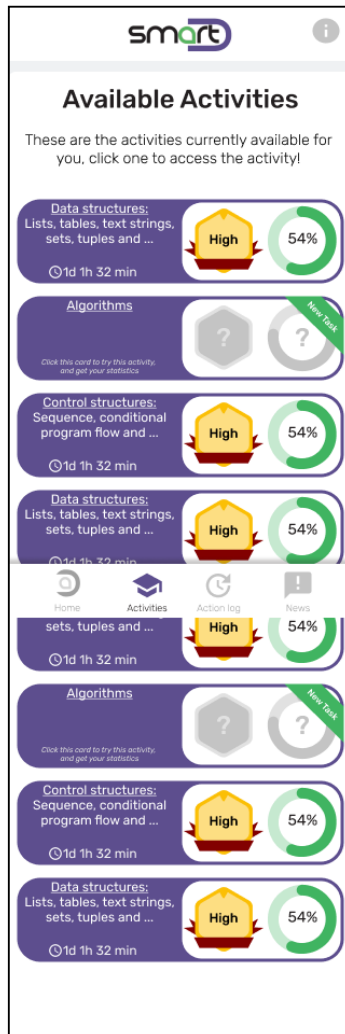
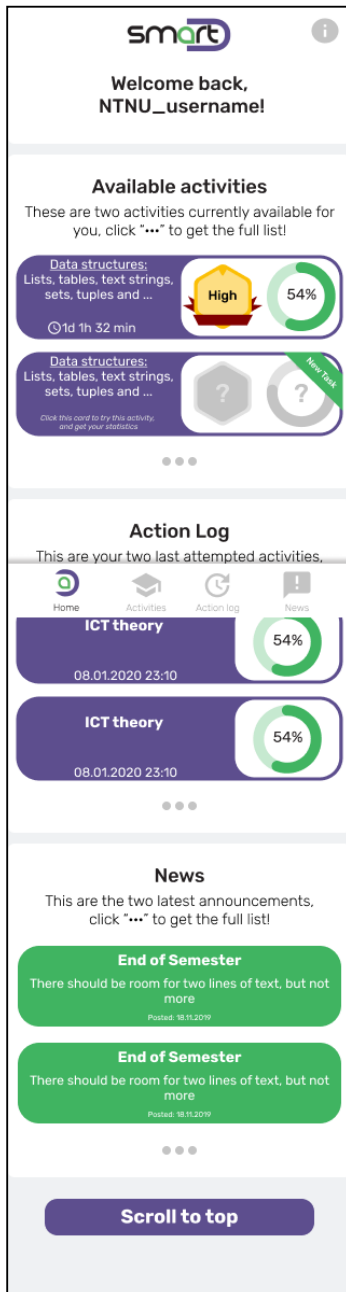
## B.3 Mobile User Interface

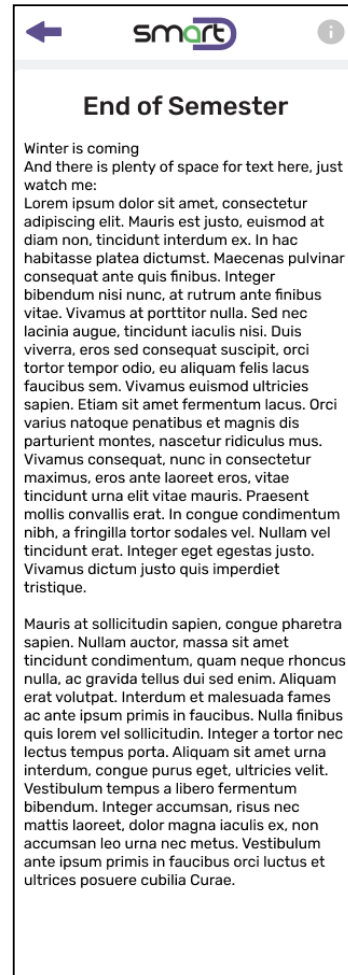
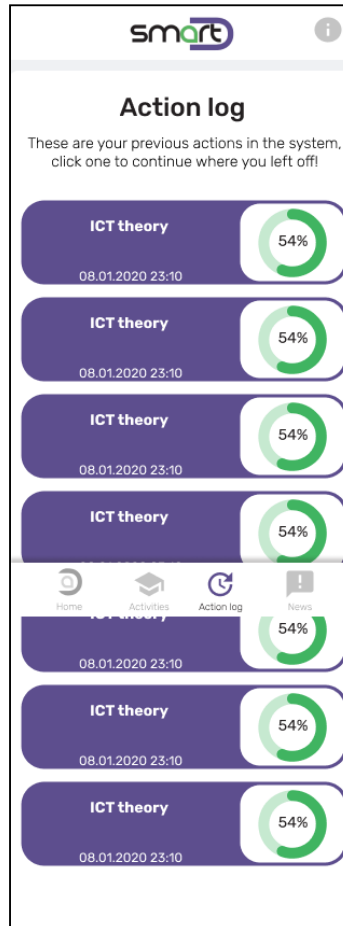
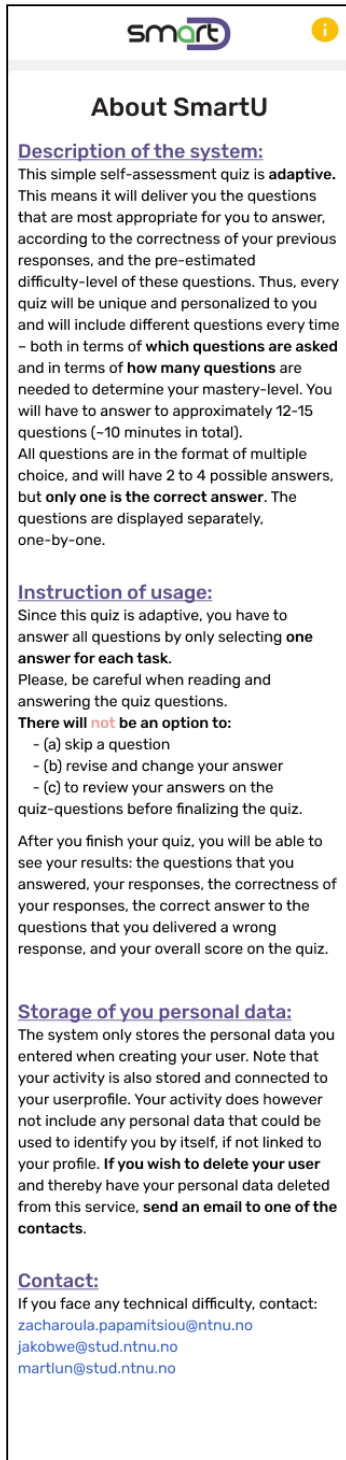
The image displays three mobile user interface screens for the SmartU application, each enclosed in a black border.

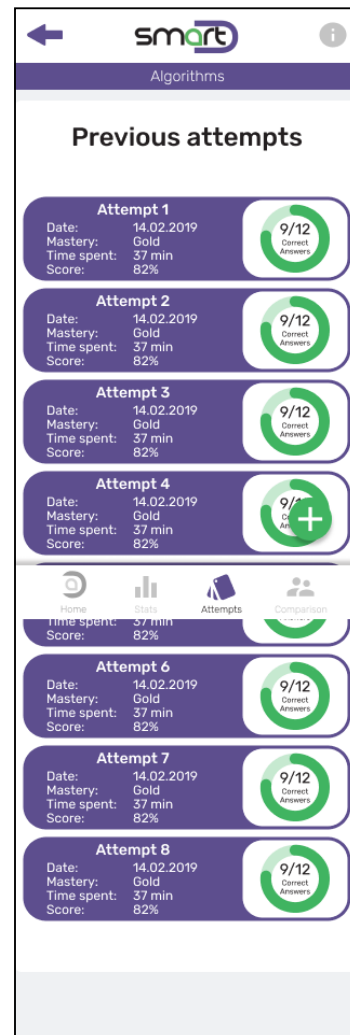
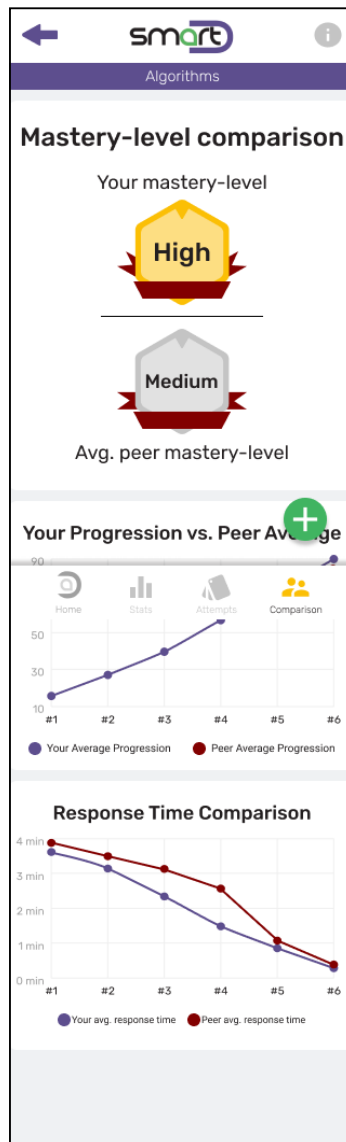
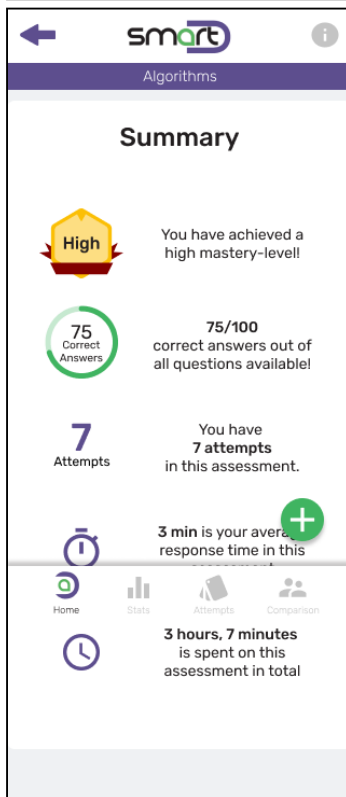
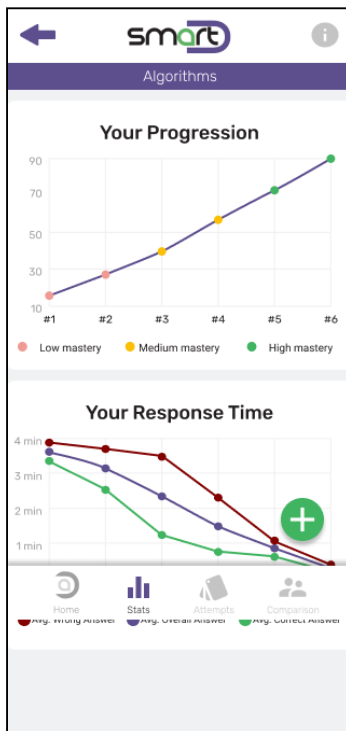
**Screen 1: Error 404**  
The screen features the SmartU logo at the top. Below it, the text reads: "Error! 404 SmartU was not smart enough it seems... It looks like the content you searched for could not be found or does not exist, try checking the spelling in the URL! You could also try to go back to the homepage." A green button labeled "Home" is positioned at the bottom.

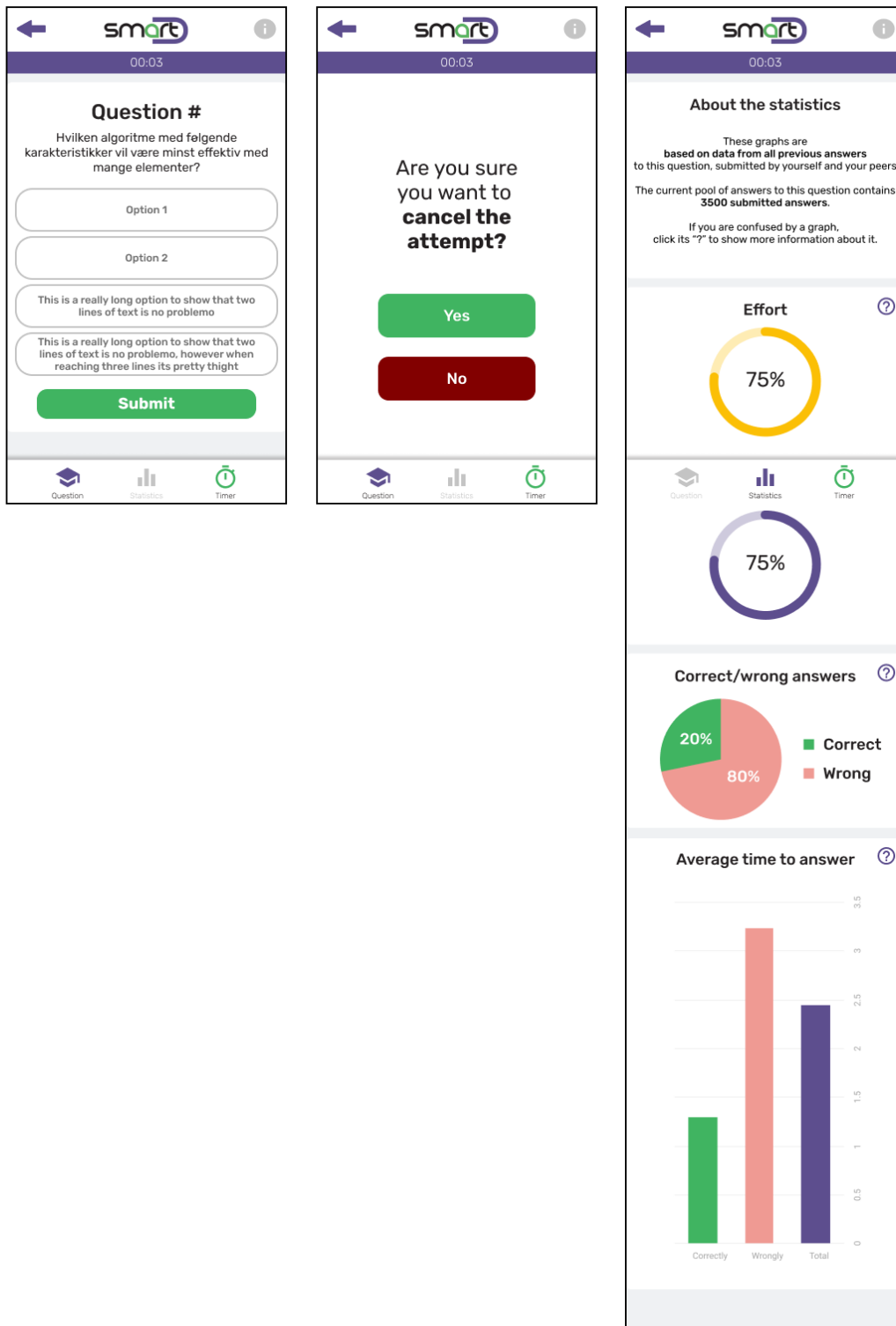
**Screen 2: Create user**  
The screen has a header with a person icon, the text "Create user", and the SmartU logo. It is divided into two sections: "System info" and "Personalia".  
The "System info" section contains three input fields: "Username \*", "Password \*", and "Repeat password \*".  
The "Personalia" section contains three input fields: "First name", "Last name", and "E-mail".  
A green button labeled "Create user" is located at the bottom.

**Screen 3: Login**  
The screen features the SmartU logo at the top. Below it, the text reads: "Login".  
It contains two input fields: "Username" and "Password".  
A link labeled "Forgot password?" is located below the password field.  
A green button labeled "Login" is positioned below the input fields.  
Below the button, the text reads: "Dont have an account? Sign up!".  
The SmartU logo is displayed at the bottom of the screen.









# Appendix C

## Results

### C.1 Questionnaire & Descriptives

**Overall evaluation of the usability of SmartU (OEUS)**

*Mean is based on Likert Scale from (Strongly Disagree) 1 to 5 (Strongly Agree)*

Questionnaire Question	Control Group			Experimental Group		
	N	Mean	Std Dev	N	Mean	Std Dev
"I think that the navigation in SmartU was nearly effortless."	15	4.13	.834	12	4.08	.515
"I think that when I needed help to learn how to use SmartU, the system provided me with sufficient information."	15	3.87	.640	12	3.42	.900
"I think that a user who has never seen SmartU before can learn how to accomplish basic tasks fast."	15	4.53	.516	12	4.50	.674
"I think that experienced users should be able to complete tasks using SmartU quickly."	15	4.93	.258	12	4.83	.389
"I think that it is not frequent that users make errors while using SmartU."	15	3.87	.915	12	4.00	.739
"I think that the interaction with SmartU is clear and understandable."	15	4.53	.640	12	4.33	.888
"Using SmartU makes me happy to accomplish my self-assessment tasks."	15	4.40	.632	12	3.92	.996
"Using SmartU gives me enjoyment for my continuous learning."	15	4.53	.640	12	4.25	.866
"Using SmartU leads me to exploration in my continuous self-assessment."	15	4.20	1.146	12	4.17	.937

Table C.1: Questionnaire Descriptives OEUS



**Attitude towards Graphs And Visualizations (AGV)***Mean is based on Likert Scale from (Strongly Disagree) 1 to 5 (Strongly Agree)*

Questionnaire Question	Control Group			Experimental Group		
	N	Mean	Std Dev	N	Mean	Std Dev
"It was easy to understand that the graphs reflected my skill-level."	15	4.73	.458	12	4.25	1.357
"I was able to make sense of the graphs and extract information regarding my skills."	15	4.33	.724	12	4.17	1.115
"I think the visualizations helped my understanding of my current knowledge level."	15	4.67	.488	12	4.25	.622
"I believe that a graphical representation is more fitting than a textual representation of the data displayed in the graphs."	15	4.73	.458	12	4.17	1.267
"I found it useful to view stats of my performance over time."	15	4.53	.743	12	4.17	.577
"I understood that the mastery-level reflects my level of knowledge as a result of previous performances."	15	4.00	1.134	12	4.50	1.168
"I found it easy to understand that the color green was used to represent a positive association."	15	4.60	.910	12	4.58	.669
"I found it easy to understand that the color yellow was used to represent a mediocre/neutral association."	15	4.40	.986	10	4.80	.422
"I found it easy to understand that the color red was used to represent a negative association."	15	4.53	1.125	12	4.75	.452

**Continues in the following table C.3**

Table C.2: Questionnaire Descriptives AGV (Part 1)

**Attitude towards Graphs And Visualizations (AGV)***Mean is based on Likert Scale from (Strongly Disagree) 1 to 5 (Strongly Agree)*

Questionnaire Question	Control Group			Experimental Group		
	N	Mean	Std Dev	N	Mean	Std Dev
"I think that the graphical representations provided all the information I needed."	15	4.33	.724	12	4.17	1.030
"I believe that the provided visualizations made information easy to extract."	15	4.47	.640	12	4.42	.669
"I was motivated to further use the system to see my progression unfold in the visualizations."	15	4.80	.561	12	4.50	.522
"I was motivated to further use the system to achieve the highest mastery-level."	15	4.53	.743	12	4.33	.492
"By utilizing the visualizations continuously I gain better understanding of my own progress."	15	4.47	.743	12	4.25	.622
"The design and colors in the visualizations made me feel happy."	15	4.07	.961	12	4.00	.953
"The design and colors in the visualizations made me feel confident and comfortable while using SmartU."	15	4.07	.961	11	4.00	.894
"I would like to continue to use SmartU to increase my skills."	15	4.40	.632	12	4.25	.452

Table C.3: Questionnaire Descriptives AGV (Part 2)

**Usability of Graphs and Visualizations (UGV)***Mean is based on Likert Scale from (Strongly Disagree) 1 to 5 (Strongly Agree)*

Questionnaire Question	Control Group			Experimental Group		
	N	Mean	Std Dev	N	Mean	Std Dev
"It was easy to understand that a card represents an available activity in a course (e.g ITGK)."	15	3.73	.799	12	4.00	1.129
"It was easy to understand that the "previous score"-graph represented the latest score achieved in the specific activity."	15	4.47	.834	12	4.50	.674
"It was easy to understand that the mastery-level within an activity card represented the mastery-level of that activity."	15	4.80	.414	12	4.58	.669
"I felt motivated to achieve a high mastery-level in all available activities when using the system."	15	4.53	.640	12	4.75	.452
"I found it useful to be able to quickly gain insight of my performance in a topic by looking at a card."	15	4.40	.828	12	4.17	.577
"I found the cards useful in terms of presenting important data of a topic."	15	4.27	.961	12	4.17	.577
"I believe that the provided visualizations made information easy to extract."	15	4.53	.640	12	4.17	.835
"I found it useful to be able to see the progression of the time I spent answering questions in an assessment."	15	4.47	1.060	12	4.00	.739
"I was able to understand the connection between the different lines in the response time chart."	15	4.13	1.125	12	4.08	.996
"I found it useful to be able to find statistics from a the same attempt in both graphs."	15	4.40	.986	12	4.25	.754

Table C.4: Questionnaire Descriptives UGV

**Graphs and visualizations in an Assessment (GVA)***Mean is based on Likert Scale from (Strongly Disagree) 1 to 5 (Strongly Agree)*

Questionnaire Question	Experimental Group		
	N	Mean	Std Dev
"It was easy to understand that the visualizations in the statistics represented aggregated data based on all user's previous answers to a specific question."	12	4.50	.674
"I frequently used the statistics during an assessment to get more information about a question."	12	2.58	.900
"I believe that a graphical representation of data is more descriptive than text in this context."	12	4.25	1.357
"I read the text in "About the statistics" and found it informative."	9	3.11	1.537
"I found it useful to be able to retrieve statistics about a question to see the previous overall performance."	12	3.58	1.311
"I was able to deduct the difficulty of a question based on the provided statistics of a question."	11	3.73	1.191
"I found it useful to be able to see how much effort others had put into a specific question."	12	3.58	1.084
"I found the statistics to be a useful tool when conducting an assessment."	12	3.42	1.084
"I felt motivated when a question had a high degree of previous correct answers."	10	2.30	.823
"I was motivated to answer a question when the effort-level of a question was high."	11	3.73	.786
"I felt motivated when I got a correct answer to a question perceived as difficult."	11	4.27	.786
"I felt more informed and confident conducting an assessment when utilizing the statistics."	11	3.45	.820
"I felt like the statistics made me more effective when answering questions."	12	2.92	1.084
"By using the statistics during an assessment I feel motivated to continue using SmartU in the future."	12	3.50	.905
"Using the statistics during an assessment provided enjoyment and motivation for future self-awareness reflection."	12	3.58	.996

Table C.5: Questionnaire Descriptives GVA

## C.2 SUS-Scores

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total Score
P1	5	1	5	1	5	1	4	1	3	2	90
P2	5	1	5	1	5	1	5	1	4	2	92.5
P3	5	5	5	5	5	1	4	1	3	1	72.5
P4	4	1	5	1	4	1	5	1	4	1	92.5
P5	3	3	4	1	4	1	4	1	3	2	75
P6	4	3	4	2	4	3	4	2	3	2	67.5
P7	4	4	4	1	4	1	4	2	4	3	72.5
P8	5	1	4	1	4	2	5	1	5	2	90
P9	4	2	5	1	5	2	4	1	5	1	90
P10	5	2	5	1	4	2	5	2	5	1	90
P11	4	1	5	1	5	4	5	1	5	1	90
P12	4	2	4	1	3	2	5	1	4	1	82.5
P13	5	3	3	1	4	4	5	1	4	2	75
P14	3	2	4	1	4	1	4	1	4	1	82.5
P15	4	3	3	1	4	1	5	1	4	2	80
P16	3	1	5	1	5	1	5	1	5	1	95
P17	4	2	4	1	4	1	5	2	4	1	85
P18	3	2	4	2	4	2	3	1	4	2	72.5
P19	4	2	5	1	5	2	5	1	4	2	87.5
P20	3	4	5	1	4	1	5	2	5	1	82.5
P21	4	2	5	1	5	1	5	1	4	1	92.5
P22	4	2	4	3	4	2	5	2	4	2	75
P23	5	4	4	1	4	2	3	2	4	1	75
P24	5	2	4	1	4	1	5	2	4	3	82.5
P25	4	4	5	1	5	1	5	1	5	1	90
P26	4	1	4	1	4	1	5	2	3	1	85
P27	4	1	4	1	5	2	4	1	5	1	90
Average											83.61

	Positive result		Neutral result		Negative result
--	-----------------	--	----------------	--	-----------------

Table C.6: Results from the SUS-test, based on standard SUS-schema

