Aron Oliver Fredriksen Werner

Design of a smart emotion-aware reflection system for teachers

Master's thesis in Informatics
Supervisor: Kshitij Sharma
January 2023

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Aron Oliver Fredriksen Werner

Design of a smart emotion-aware reflection system for teachers

Master's thesis in Informatics
Supervisor: Kshitij Sharma
January 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Abstract

Learning Analytics Dashboards provide the user with important information about learning progress. However, few dashboards incorporate emotional metrics as part of their indicators. The emotional state of students is closely tied to learning outcomes. In this thesis, a learning analytics dashboard for teachers is developed to showcase the emotional state of students during classroom activities.

The thesis presents a literature review to establish the state of the art of modern learning analytics dashboards. Highlighting what information they display, and which visualization techniques they utilize. Additionally, the review looks at emotions and physiological data and their relation to learning processes. The results of the review provide the foundation for the development of a dashboard capable of iterating upon the state of the art. Next, the methodology describes how the dashboard was designed, developed, and evaluated using a design-based research method across two cycles. The resulting dashboard can show the emotional state of the class in real time as well as various progress trackers. The final evaluation revealed that the dashboard provides the teacher with many useful features, and the chosen visualization types were largely effective at conveying information.

Sammendrag

Læringsanalyse dashboard gir brukeren viktig informasjon om læringsprosesser. Til nå er det få dashboard som tar i bruk følelser som del av indikatorene sine. Følelser og sinnstilstand har mye å si for hvordan man lærer. I denne masteroppgaven er det utviklet et læringsanalyse dashboard for lærere som gir et overblikk over følelsene og sinnstilstanden til elevene under klasseromsaktiviteter.

Masteroppgaven presenterer en litteraturstudie av læringsanalyse dashboard med hensikt å vise hvordan og hvilke informasjonstyper moderne dashboard formidler. I tillegg er det sett på hvordan følelser og fysiologiske data påvirker læring. Litteraturstudien danner grunnlaget for å utvikle et nytt dashboard som bygger videre på nåværende dashboard. Metodologien beskriver deretter hvordan dashboardet ble designet, utviklet, og evaluert ved bruk av forkningsmetoden design-based research over to sykluser. Resultatet var et læringsanalyse dashboard som viser følelsene og diverse læringsinformasjon om elevene i sanntid. Den siste evalueringsfasen viste at dashboardet tilbyr læreren med mange nyttige funksjoner og at visualiseringstypene var i stor grad effektive i sin formidlingsevne.

Preface

This master's thesis is the final delivery in the Master's degree program in Informatics at the Norweigan University of Science and Technology. Parts of this thesis build and extend upon previous work that was done in the course IT3915. The following parts include the first two sections of the *Introduction* [Chapter 1], which are adapted, *Literature Review* [Chapter 2], and the preliminary design mockups from the first cycle of DBR.

The master thesis is supervised by associate professor Kshitij Sharma.

Table of Contents

List of Figures

List of Tables

List of abbreviations (or Symbols)

LAD	Learning Analytics Dashboard,
HR	Heart Rate
HRV	Heart Rate Variability
DBR	Design-Based Research
SWOT	Strengths-Weaknesses-Opportunities-Threats
IBI	Inter-beat Intervals

1 Introduction

Learning analytics dashboards (LAD) have become a popular instrument of study in the past few years. The goal of these dashboards is to faciliate student learning and improve learning outcomes. A learning analytics dashboard typically displays a number of statistics and data relevant to a specific learning context. For example, tracking course activities and results of students. The key facet of LADs is their ability to convey and display information that is meaningful and contributes to learning processes in order to maximize learning outcomes (Ramaswami et al., 2019). This can be achieved by taking advantage of humans' ability to process a lot of data when conveyed in a good visual manner (Shemwell, 2005) and providing actionable information to the user. Furthermore, a learning analytics dashboard is typically oriented towards the teacher or the student, and specifications will differ between them. To illustrate, a LAD centered toward a teacher may showcase aggregated and individual statistics concerning the whole class. While a LAD centered toward a student might track and display their individual performance. Considering this, it is important to evaluate which metrics and visualization techniques to employ based on their relevance to learning outcomes and target users.

1.1 Background and Motivation

One metric that is closely related to learning and academic achievement is student engagement (Pardo et al., 2017) (Chen, 2017). In a learning context, engagement can be defined in various ways, but it is commonly characterized as "[…] the degree of attention, curiosity, passion, and interest that students show throughout their involvement in learning environments" (Shu, 2018). Engagement is widely characterized as a multidimensional construct. Categorically, engagement can be analyzed across these dimensions: cognitive engagement, which refers to the investment and deliberate efforts to acquire challenging skills; behavioral engagement, which examines active participation and involvement in learning activities; and emotional engagement, which includes both positive and negative responses to the teacher, fellow students, and the school environment (Shu, 2018). Although this description is suitable, it should be mentioned that engagement has been described in multiple ways across different studies.

The literature review in (Werner, 2022) as part of the course IT3915 highlights that LADs of the last seven years rarely include engagement as part of their chosen metrics. And when it is included, it is usually calculated using a primitive approach, for instance, measuring activity level through log data. There are better and more advanced approaches that can be used to capture engagement. A more sophisticated approach involves examining the heart rate (HR) and facial expressions of the learner. Emotions can be captured by doing facial expression analysis, and sensors can capture the heart rate (Mohammad Nehal et al., 2021) (Dissanayake et al., 2019). By examining the HR and the emotions of the students, it is possible to provide the teacher with an overview of the emotional state of the class. The aim is that this will provide the teacher with actionable information that can improve learning outcomes for the students.

In addition to metric selection, it is also important to consider visualization techniques. It has been argued that LADs should be designed in a way that the user can perceive what is important at a glance (Farahmand et al., 2020). The idea is that effective visualization techniques can help the user understand the metrics better and contribute to making the information actionable. The review in (Werner, 2022) as part of the course IT3915 found that most dashboards utilize traditional visualization types. However, some teacher-centered LADs employed other approaches like a "class report card," where a face representing a student was drawn in a card format in order to visualize emotions (Gupta, 2021) and was received positively by the teacher. By using effective visualization methods, the teacher can incorporate the information more effectively in their teaching. Based on the motivations above I reached the following goals and research questions in the following section.

1.2 Goals and Research Questions

Goal *Develop a real-time learning analytics dashboard using sophisticated methods of engagement capturing and emotion detection that gives the teacher an overview of the emotional state of the class at a glance.*

Research question 1 *To what extent does the real-time engagement and emotional overview provide the teacher with useful information?*

Research question 2 *How effectively do the chosen visualization types provide the teacher with an overview of the emotional state of the class at a glance?*

1.3 Research Method

The research in this thesis is conducted by performing two cycles of Design-Based Research (DBR). DBR is a popular research methodology when studying problems in educational contexts (Anderson & Shattuck, 2012). In an educational setting, DBR starts by identifying and analyzing a problem related to learning. Once a problem has been adequately examined, a solution is designed, called an intervention. The intervention is then materialized through a development phase. The solution should be grounded in existing knowledge and conventions regarding the instrument. Once constructed, the intervention will be subject to evaluation. The resulting knowledge can then be used to redesign the intervention. This cycle of development is a key characteristic of DBR, and multiple cycles may be completed until the intervention is satisfactory.

In my case, the problem to be intervened was identified by conducting a literature review on LADs (Werner, 2022) as part of the course IT3915. From the results of the literature review, I created low-fidelity designs of a proposed dashboard that would ameliorate the issue. These low-fidelity designs would then undergo an evaluation by conducting interviews with relevant stakeholders, in this case, teachers and researchers, using a SWOT (strengths, weaknesses, opportunities, and threats) analysis. Once the interviews were conducted, the designs were compared, and one was chosen for DBR cycle two. In the second cycle, the dashboard was developed and evaluated by performing a second set of interviews with teachers and researchers.

2 Related Work

This chapter provides an overview of the literature review and previous work. It outlines the methodology, inquiry questions, and most important findings of the review and previous work. It starts by reviewing the most important findings of the last systematic literature review on LADs, followed by another literature review that examines LADs in the subsequent time frame to capture trends over time.

2.1 Review of (Schwendimann et al., 2017)

A systematic literature review was done on LADs in 2016 by Schwendimann, Rodríguez-Triana, et al. In this review, they analyzed 246 papers that were published in the time frame 2010-2015. In their review, they gave a descriptive analysis of the use contexts of LADs and their target users. Furthermore, they looked at what types of LADs had been developed, taking special interest in their purpose, data sources, indicators, and visualization types. An indicator in this context is analogous to a metric referenced in the introduction, for example, test scores or number of completed assignments. In addition to the descriptive analysis, they gave a qualitative examination of their maturity in terms of evaluation methods.

In their descriptive analysis, they found that the main target users of LADs were teachers and students. Teachers were among the target users of most of the dashboards, 75 percent, while students were the target user of 51 percent of dashboards. Although not absent, other users, such as researchers and administrators, constituted a tiny fraction of the analyzed dashboards. Regarding the purpose of the dashboards, they found that most of the LADs were used for self-monitoring or for monitoring others. Of interest are their findings regarding data sources in the dashboards. They reported that 85 percent of dashboards used log data as their data source. Physical sensor data were only used in 7 percent of dashboards.

They identified many different indicators (Over 200) and created groupings that aimed to explain what question the indicator sought to answer; for example, Learner-related indicators showed information specific to the learner, age, prior courses, etc. However, they did not provide any distribution for the groups, which could have given information on how prevalent different types of indicators are in LADs. Moreover, they found that the most common types of visualizations were bar charts, line graphs, tables, and pie charts. Furthermore, a striking number of the dashboards, 58 percent, did not have any evaluation method at all, and only 29 percent were evaluated by including the stakeholders and gathering data from real sessions.

2.2 Literature Review

A literature review was conducted to capture the state of the art of LADs since the last systematic literature review in 2015, as outlined in the previous section. Furthermore, it sought to investigate the relationship between emotions, physiological data, and learning processes. The literature review was conducted in the Autumn of 2022. The inquiry questions were as follows:

1. What information do modern learning analytics dashboards display?
    i. Which emotions do they utilize?
    ii. How do they utilize heart rate and facial expressions?
2. How are emotions and physiological data related to learning processes?

The main extension of my review is to look at the development of LADs since the last systematic review. This provides insight into how dashboards have evolved over time while also looking more specifically at how LADs utilize emotions and sophisticated ways of capturing the engagement of the learner.

2.2.1 Methodology

The first step of the review was to choose which databases to collect articles from. I chose five of the most well-known databases in Technology Enhanced Learning:

1. Wiley
2. IEEE Xplore
3. SpringerLink
4. ACM digital library
5. Science Direct

Since the review includes two different objects of study, I used different search queries for each object. For the first part of the review concerning LADs, the same query as in (Schwendimann et al., 2017) was used:

1. dashboard AND (“learning analytics” OR “educational data mining” OR “educational datamining”)

This query specifies the main target of interest, dashboards, as well as the main fields where dashboards have been utilized. One limitation of this query is that documents that do not use the term “Dashboard” for their visualization system will not be included.

For the second part of the review regarding emotions and heart rate, I used the following queries respectively:

1. Engagement AND (”emotions” OR ”heart rate”) AND (”Education” or ”Learning”)

2. Engagement” AND (”emotions” OR ”heart rate” OR ”Skin conductance”) AND (”Education” or ”Learning”)

It is important to note that the databases vary in available search scopes; for instance, you cannot restrict the search to the abstract in all databases. Because of this, the search was performed without restricting the scope, I.e., the query was searched on the whole document. Since the goal of this literature review is to capture the state of the art of dashboards, the search was restricted to the last seven years of research, 2016 -2022. As mentioned earlier, this time frame was chosen because the last systematic literature review included articles up to 2015. The same date restriction was used for the second part regarding emotions and heart rate, to capture trends over the same timeline. The search was conducted on the seventh of October, 2022. After performing the search, there was an initial process of filtering by reading through the abstract, introduction, and conclusion. If I needed more information to gauge whether the study was out of scope or otherwise unwanted, I would also look through the discussion part. For the dashboard part, I was specifically looking for either descriptions of a dashboard or its indicators and figures. However, if the paper was outside the scope of the specific learning contexts in the query, it was not included. For the second part, the same process of reading the abstract, introduction, and conclusion was done, but this time I was specifically looking at papers where they measured emotions or heart rate in an educational learning context. After this filtering process, 22 papers about dashboards were included in the review. For the second part, seven papers about emotions and eight papers about heart rate were included.

2.2.2 Results

This section gives an overview of the main findings from the review. It is ordered by the inquiry questions.

What information do they display?

To answer inquiry question 1,” What information do modern learning analytic dashboards display?” I identified the main groupings of the different indicators utilized and how they were visualized. There was a plethora of different indicators used in the dashboards; in order to aggregate the results, I grouped them into distinct categories. These are adapted from (Schwendimann et al., 2017) and are summarized in Table 2.1:

Indicator group game	Question the indicators aim to answer	Examples of indicators in this group	Paper where the indicators were found
Learner-related	Who are the learners?	Competences, age, name, prior courses	(Gupta, 2021) (Molenaar & Knoop-van Campen, 2019) (Yaëlle & Thomas, 2018) (Naif Radi et al., 2019) (Ramos-Soto et al., 2017) (Herodotou et al., 2019) (Ruipérez-Valiente et al., 2021) (Ouatiq et al., 2022) (Juan et al., 2021)
Action-related	What do they do while learning?	Number of page visits, time spent on tasks, login time	(Gupta, 2021) (Molenaar & Knoop-van Campen, 2019)

			(Yaëlle & Thomas, 2018) (Naif Radi et al., 2019) (Ramos-Soto et al., 2017) (Herodotou et al., 2019) (Ruipérez-Valiente et al., 2021) (Ouatiq et al., 2022) (Juan et al., 2021) (Cechinel et al., 2021) (Ramaswami et al., 2019) (Pozdniakov et al., 2021) (Ong & Chua, 2022) (Farahmand et al., 2020) (Andrew et al., 2022) (Jeongyun et al., 2021) (Chen et al., 2022) (Rohloff et al., 2019) (Owatari et al., 2020) (Lai et al., 2022) (Dickler et al., 2021)
Content-related	What is the content involved in their learning?	Topics covered in a subject or assignment	(Andrew et al., 2022) (Jeongyun et al., 2021) (Ouatiq et al., 2022) (Ruipérez-Valiente et al., 2021) (Rohloff et al., 2019)
Result-Related	What is the result of their learning?	Average grade, predicted grade, relative ranking within the class	(Cechinel et al., 2021) (Ramaswami et al., 2019)

			(Ong & Chua, 2022) (Farahmand et al., 2020) (Andrew et al., 2022) (Ruipérez-Valiente et al., 2021) (Ouatiq et al., 2022) (Herodotou et al., 2019) (Ramos-Soto et al., 2017) (Chen et al., 2022) (Natercia et al., 2021) (Naif Radi et al., 2019) (Yaëlle & Thomas, 2018) (Rohloff et al., 2019) (Molenaar & Knoop-van Campen, 2019) (Gupta, 2021) (Dickler et al., 2021)
Context-related	In which context does the learning take place?	Location of the learner in the classroom, geographical location	(Jeongyun et al., 2021)
Social-related	How do they interact with others during learning	Network showing communication with others in a group	(Pozdniakov et al., 2021) (Andrew et al., 2022) (Jeongyun et al., 2021) (Juan et al., 2021) (Chen et al., 2022)

Table 2.1 Overview of Different Indicator Groups

Of the 22 papers, almost all of them (21 papers, 95%) contained an action-related indicator, progress trackers were especially prevalent. A vast majority of the papers (17, 77%) contained a result-related indicator; in contrast to previous studies, it seems more common to give a predicted grade of the learner in addition to actual results. Under half of the papers (nine papers, 41%) had learner-related indicators. Equally many papers (five papers, 23%) used social-related indicators as content-related indicators (five papers, 23%). Finally, one paper (4.5%) used a context-related indicator. These are summarized in Figure 2.1.

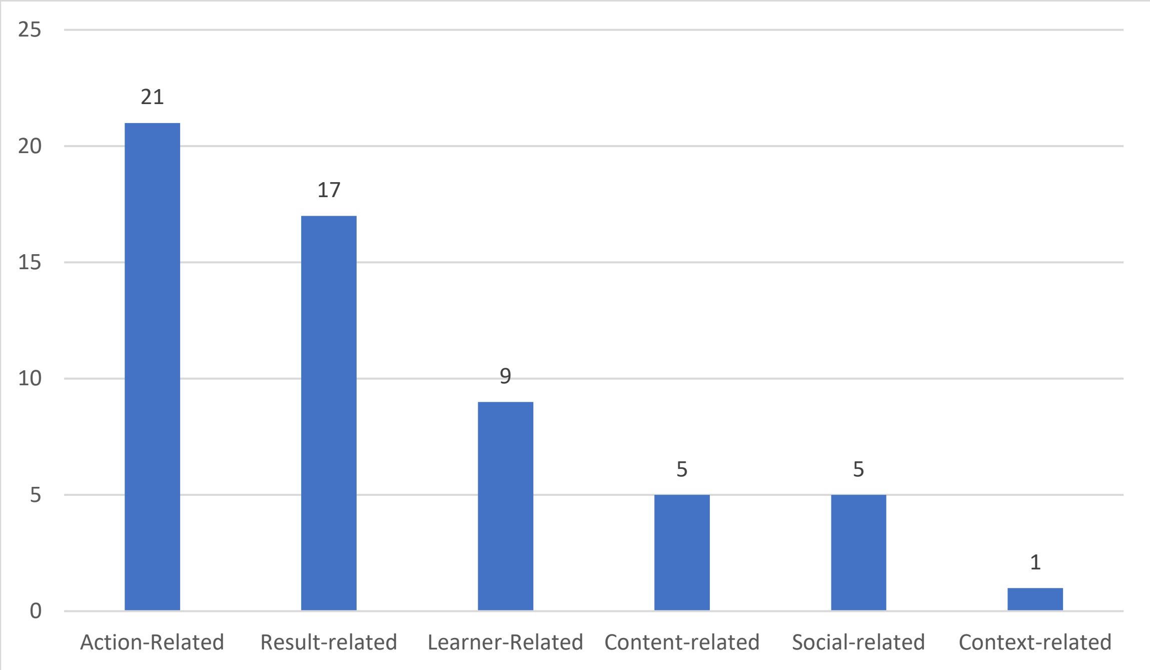

Figure 2.1 Indicators Found in Papers

There were many different types of visualizations present in the papers, the most recurring ones being bar charts (13 papers, 59%), followed by line graphs (11 papers, 50%), tables (nine papers, 41%), and text (eight papers, 36%). These and the less frequent visualizations are summarized in Figure 2.2.

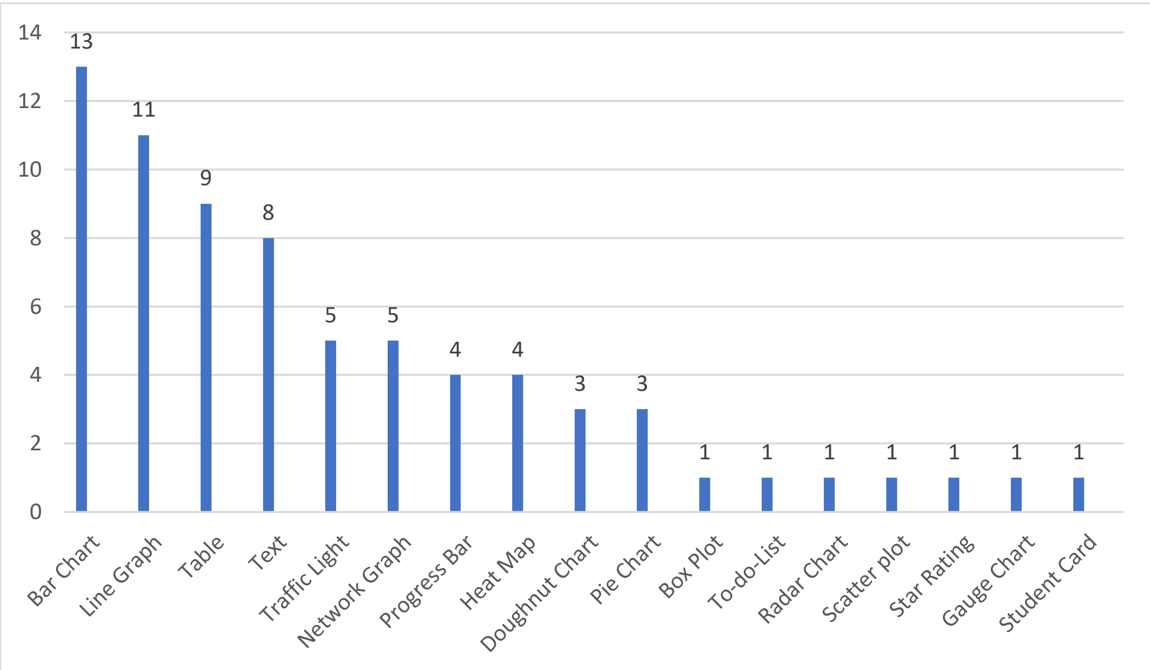

Figure 2.2 Visualizations in Papers

To answer the sub-questions of inquiry question 1, ”Which emotions do LADs utilize and how do they utilize heart rate and facial expressions) I give a descriptive summary of how many dashboards which in some way used emotions to either give some information to the user or used emotions to calculate some intermediate value (e.g., predicted grade). And which papers explicitly used heart rate and facial expressions.

Only one (4.5%) paper captured the facial expressions of students and calculated their emotions (anger, happiness, bored, emotional engagement), this was then showcased in the form of ”student cards” which included a drawn face of the student where the expression matched the calculated emotions. Three papers (14%) used log data as a form of engagement measure in a prediction algorithm to identify students at risk. The rest of the papers either used log data to showcase engagement implicitly, i.e., showing interest through activity levels, number of forum posts, number of clicks, or none at all. None of the papers used heart rate.

2.2.3 How Are Emotions and Physicological Data Related to Learning Processes

To answer how emotions and physiological data are related to learning processes, I looked at how the papers used emotions and heart rate to give a measurement that could be used to describe the learning of the subject. Below is a descriptive summary of how the different papers used heart rate and emotions.

2.2.3.1 Heart Rate

One paper used HR in addition to student self-reports to calculate the cognitive engagement of the users in a learning context (Li, Bao, et al., 2021). They argue that when the HR changes, the cognitive engagement of the learner is stronger. Another study (Dissanayake et al., 2019) used heart rate variability (HRV) to calculate the cognitive load of the learner; cognitive load refers to the amount of working memory in use. Two papers used HR in a machine learning context; the first used HR in addition to other physiological data to calculate the drowsiness of learners and to predict their wakeful state (Kawamura et al., 2021). While the other used HRV to classify which online learning topics the students found hard or easy (Pham & Wang, 2018). To increase engagement in online classes, one paper used HR as a substitute for in-person social cues to communicate how engaged you were to other learners (Han et al., 2022). Moreover, HR and HRV were used to capture the disengagement of learners who were looking at tutorial videos in a massive open online course (MOOC) (Xiao & Wang, 2016). Another paper used HR to capture moments when learners felt frustration and anxiety to increase engagement while learning (Chen et al., 2017). Finally, researchers in (Di Lascio et al., 2018) used physiological data, including HRV, to differentiate between engaged and non-engaged students by calculating their emotional engagement.

2.2.3.2 Emotions

In (Subramainan et al., 2016), the researchers argue that emotions are prevalent in educational settings and that they have a large impact on engagement and academic performance. Researchers in (Li, Yue, et al., 2021) propose a framework to capture students’ emotions by using a convolutional neural network (CNN) which is fed facial expression data. Another study found a link between positive emotions and higher academic engagement and, in turn, higher academic performance (Alfredo et al., 2021). Furthermore, the paper in (Mohammad Nehal et al., 2021) captured the emotional engagement of students by analyzing facial expressions from lecture videos, they argue facial expressions are the most potent way of capturing emotional engagement. Researchers in (Xiao et al., 2021) suggest teachers should pay particularly close attention to the emotional engagement of students in online learning, as this setting is prone to induce tiredness and disinterest in students. In a game-based learning context, researchers used data from facial expressions as features in a machine-learning algorithm to calculate emotions (Manuel et al., 2019). They also argue higher emotional engagement leads to increased motivation and learning. Finally, (Sannyuya et al., 2022) looked at the interplay of emotional and cognitive engagement as a way to predict learning achievement. They used text classification to capture the emotional and cognitive states of the learners. They found that emotional and cognitive engagement have an interplay that impacts learning achievement; positive and confused emotions contribute more to high-level cognitive engagement, while low-level cognitive engagement may lead to bad emotional engagement and learning achievement.

2.2.4 Summary

In this literature review, I have reviewed modern LADs from the last seven years, looking at their most prolific information indicators and visualizations. I also investigated how they utilize emotions and specifically how they use physiological data such as the heart rate and facial expressions of the learners. Action-related indicators, along with result indicators, are the most prevalent conveyed information. While traditional bar charts, line graphs, and tables are the most frequent visualizations. Almost none of the LADs utilize emotions explicitly, and you can only gauge student engagement implicitly from activity trackers, and even then, you can only measure one axis of engagement. Thus, it is fair to conclude that modern LADs, by and large, do not utilize emotions in any meaningful capacity.

Over the same period, I reviewed how emotions and physiological data are related to learning processes. I found that heart rate and heart rate variability were used in many different learning contexts, such as online learning and classroom environments (Han et al., 2022) (Li, Bao, et al., 2021). Heart rate was most often used as a way to measure the engagement of students, but also, in some cases, to give an indication of their cognitive load (Dissanayake et al., 2019). Emotions were also used in various learning contexts to quantify the emotional engagement of students. This was most often measured by using facial expressions, but also self-reports and machine-learning techniques were used (Li, Yue, et al., 2021) (Alfredo et al., 2021) (Sannyuya et al., 2022). From this, it is natural to conclude that heart rate and emotions are good measurements to calculate student engagement.

This review provides the knowledge foundation for the development of the dashboard in this thesis and highlights a gap in the current research regarding the utilization of emotions and HR in current dashboards. It also outlines features and information that are typically included in modern LADs.

3 Methodology

This chapter outlines the research methodology employed in this thesis. It describes the DBR process I took and each cycle of DBR in detail.

3.1 Design-based Research Strategy

Since the goal of my thesis was to develop a learning analytics dashboard, a design-based research approach was logical given the following rationales: Developing software systems lends itself especially well to an iterative approach, and this fits well with the iterative nature of DBR (Barab & Squire, 2004). Following this, it was possible to utilize the feedback I received from stakeholders between cycles to improve the system. Furthermore, DBR is well-suited for research in educational contexts since it stresses collaboration with educational stakeholders, which affected both design and functionality of the dashboard, since I could tailor it to meet their specific expectations. Moreover, DBR focuses on realizing practical solutions to problems and using these solutions in real-world settings, which is a central aim of what learning analytics dashboards aim to do. The reasons outlined above are largely in line with the characteristics of DBR as outlined in (Reeves, 2006). Ultimately though, DBR is not intended as a replacement for other methodologies (Wang & Hannafin, 2005), but in my case allowed for a flexible approach to answering my research questions, making use of different combinations of methods for development, data collection, and analysis.

3.2 Cycle 1

The first cycle started with a literature review on LADs, as outlined in section 2.2. The steps taken in cycle 1 are summarized in Figure 3.1. From the results of the literature review, I made *four* low-fidelity designs of the dashboard. This means I decided on a preliminary choice of metrics and indicators, as well as the visualization types, based on my findings in the review. Since the literature review showed that LADs use a vast number of visualization types, I made each design utilize a different set of metrics/indicators and visualization choices. I also included some other metrics that have been related to learning processes in addition to engagement and facial emotions. To provide a clearer picture of the students' emotional state. These additional metrics will be explained in the sections addressing the relevant designs.

This approach of focusing on different sets of features for each design was done to showcase multiple different solutions to the teachers and experts during the interviews, in order to maximize the chances of positive evaluations. The low-fidelity designs were made using Figma (Figma), which is a design tool that is used to create wireframes.

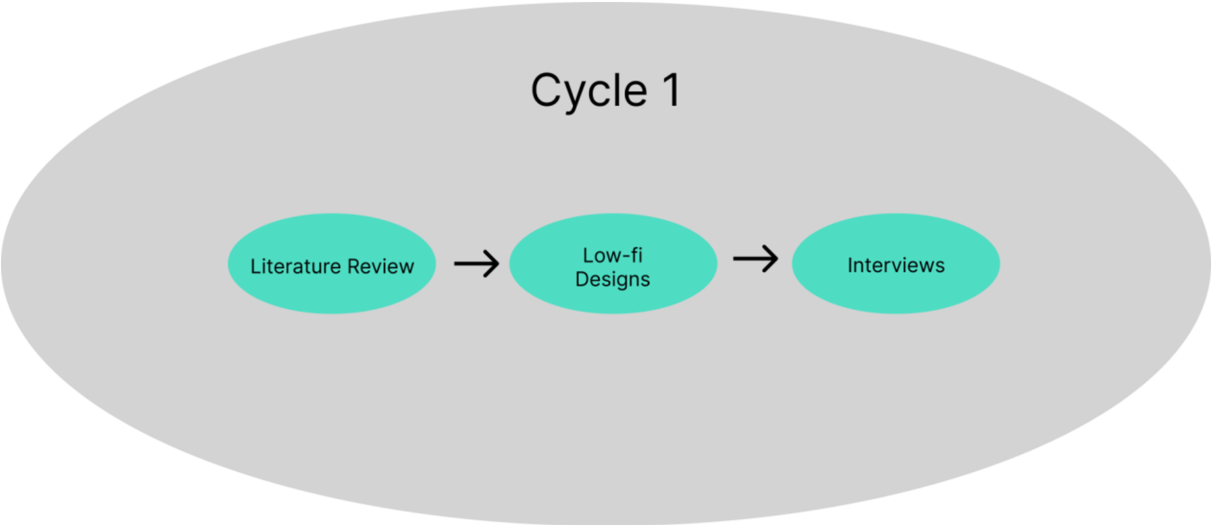

Figure 3.1 First Cycle of DBR

3.2.1 Preliminary Designs

The first design takes inspiration from the pilot's dashboard in an airplane. The idea is that the barometers will give the teacher an indication of the emotional state of the class at a quick glance. It displays five emotional metrics: Engagement, Happiness, Anger, Entertainment, and Stress. In addition to engagement and facial emotions described in the literature review in section 2.2.3, this design includes entertainment and stress metrics. Entertainment is a measure of how "fun" something is and has been correlated with HR (Yannakakis et al., 2008). Stress has also been correlated with HR and has been used in educational contexts (Sharma et al., 2022).

The barometers are color-coded according to a classification of relative increase compared to a baseline. The colors indicate whether there is an increase or decrease compared to the baseline. Green would mean that there is a relative increase compared to the baseline, while red would indicate a relative decrease. The graph also displays the metrics over time as well as by topic. In the lower-left corner, there is a box containing various descriptives about the current topic and the time spent on this topic, and the last topic. The barometer boxes are clickable, highlighted by the black outline, so the teacher can choose which metrics they want to see the graph of. They can also click on the boxes on the graph to see the metric over time or by topic.

Figure 3.2 Mock 1

The second design is more detailed. It uses the same emotional metrics as the first design, but now all the graphs are always visible. The graphs show the individual metrics over time and include a color icon in the corner indicating the state at that point in time. It also contains bar charts that show different descriptives about the current topic. The first bar chart shows the time spent on each topic. The second bar chart shows the average score on each topic. The third bar chart shows the number of problems completed on each topic. The final bar chart shows the progress in percentage on each topic.

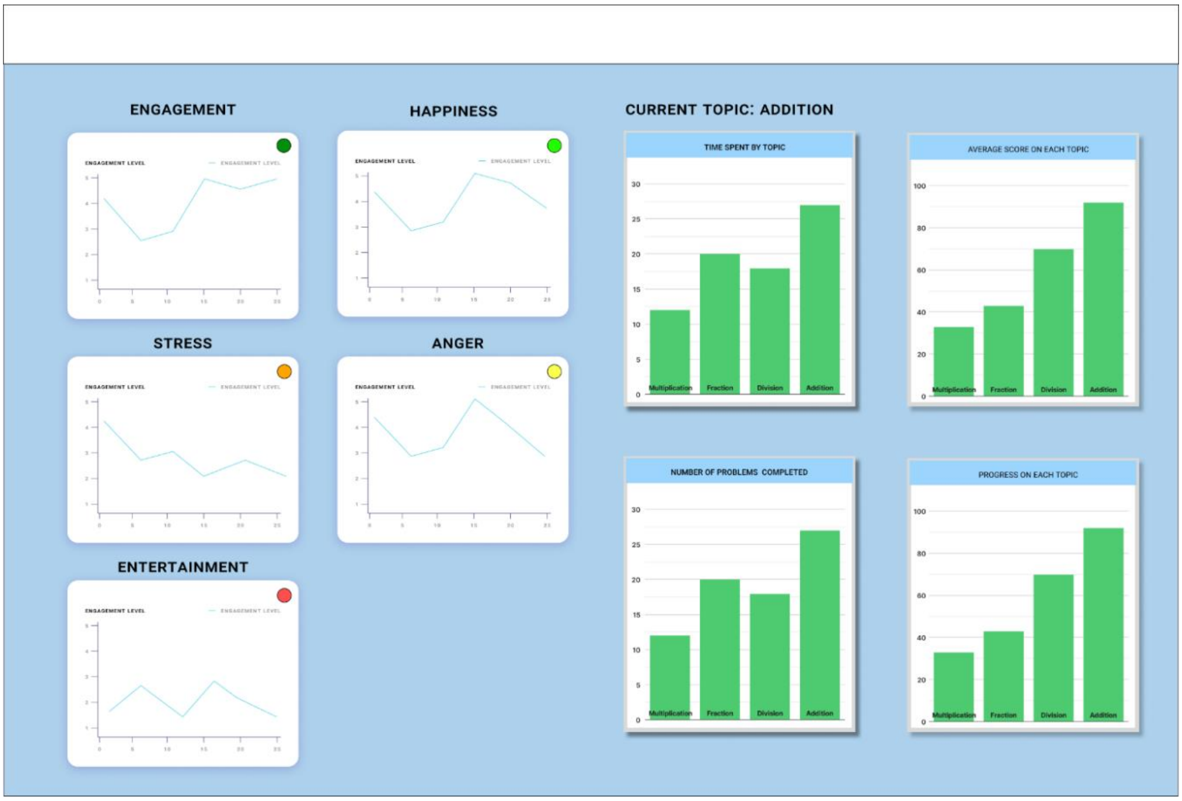

Figure 3.3 Mock 2

The third design has several different visualizations than the previous two. Firstly, the metrics happiness and anger are now aggregated in the emoji. Secondly, an alerts box is added in the top right corner that shows alerts on specific events; Happiness is low, anger is high, with a corresponding timestamp. It also contains a “Today’s Performance” box containing descriptives about the current topic. On the left is a pie chart showing which topics have the highest average time spent, as an indication of difficulty. Lastly, only engagement is now shown on the graph, and it also has a color icon on the corner to indicate a positive or negative state.

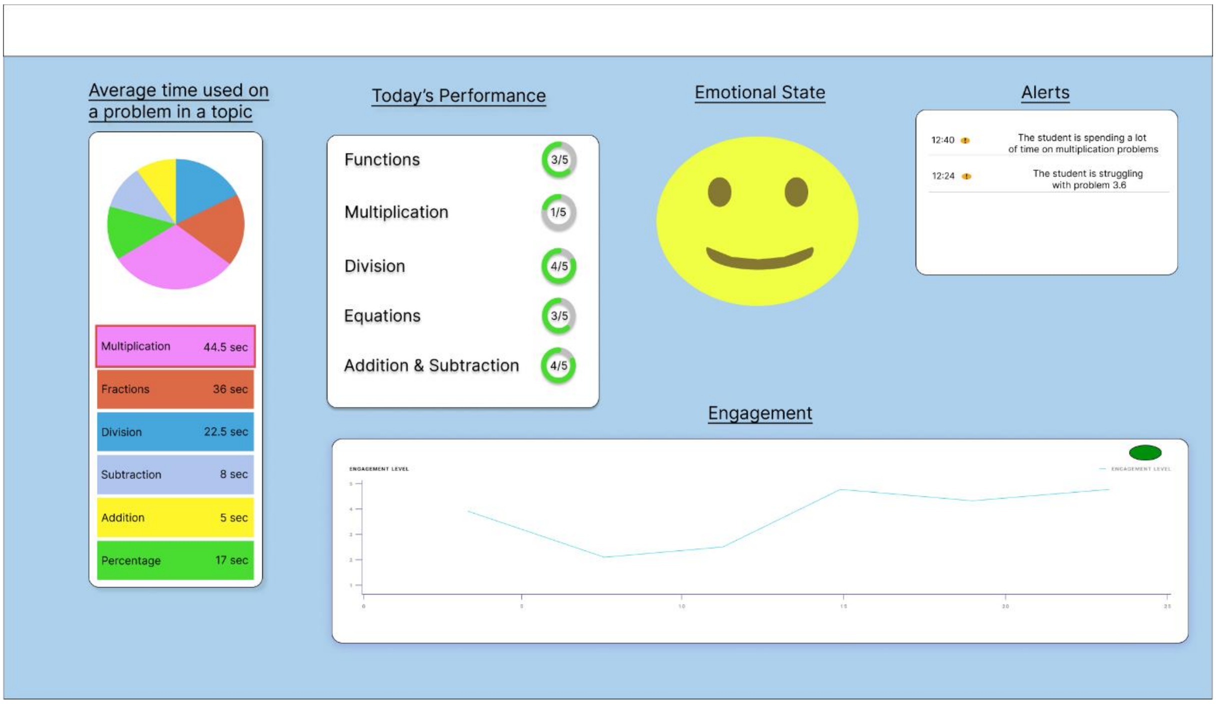

Figure 3.4 Mock 3

The fourth and final mock uses a blend of the different indicators and metrics from the previous mocks. The emotional metrics are now indicated by a larger color circle to indicate a positive or negative state. They are also now clickable, like in mock 1, to show the corresponding metric over time. In addition, an exercise overview is now shown, indicating which exercises have been completed. Red indicating not all students have completed that problem, green indicating everyone has completed it, and grey indicating no attempts. It is also possible to select the topic for the exercise view, multiplication, division, and so forth.

Figure 3.5 Mock 4

3.2.2 Selection of Teachers and Experts

After the low-fidelity designs were completed, a set of interviews with a number of teachers and experts were conducted. Following the guidelines of DBR, the interviewees were chosen based on their background as a teacher or their domain knowledge in educational research and LADs. The thought was that feedback from people with these skills and backgrounds would provide meaningful contributions to the development of the dashboard. In total, one teacher and two experts were interviewed, all were female. The teacher worked in an elementary school and had some knowledge of LADs beforehand. The experts had backgrounds in the development of learning analytics tools and conducting research in the field of learning analytics. They also had experience working with educators to implement different learning analytics tools.

3.2.3 Interview Process & SWOT Analysis

During the interviews, I and the expert/teacher completed a SWOT analysis of all the proposed designs. A SWOT analysis is a strategic planning tool that facilitates systematic analysis of strengths, weaknesses, opportunities, and threats (Weihrich, 1982), typically organized in a matrix, see Figure 3.6. For clarity, I will give a short description of the four dimensions: The dimensions are grouped into two categories; strengths and opportunities are factors that help achieve an objective. While weaknesses and threats are harmful to the objective. In addition to this, strengths and weaknesses are considered internal to the piece being analyzed. While opportunities and threats are externally dependent. For example, an opportunity regarding a learning analytics dashboard might be a recent trend that teachers prefer line charts over bar charts. It would then be an opportunity to capitalize on this trend by implementing line charts in the dashboard. It should be noted that the opportunity dimension was mostly interpreted by the participants as aspects of the design that should be added or removed and not external trends they had identified. There was also a bit of overlap regarding weaknesses and threats. Where one participant interpreted ambiguous colors as a weakness, while another thought of it as a threat.

We completed the SWOT analysis systematically, going through one design at a time, one dimension at a time. During the interview, I would first give a short description of the design, explain the measurements and functionality, and answer any potential questions before starting the analysis. During the analysis, I would take a guiding approach, which means I would ask the interviewee to give their thoughts about a particular dimension but would only intervene when they didn’t have any more points or to answer a question. After all the interviews were done, all the SWOT matrices were combined for further analysis. The maximum length of the interviews was 51.56 minutes.

Mock 1				
Strengths				
Weaknesses				
Opportunities				
Threats				

Figure 3.6 SWOT Matrix for a Single Mock

3.2.4 Combined Analysis

Before moving on to the next cycle, one of the four designs had to be chosen for continuation. This was done by using the combined SWOT matrix of all the interviews to easily compare the four designs. First, an elimination procedure was conducted, in which the two designs with the most threats were removed. This was done because threats are more detrimental than weaknesses. With two designs left, the design chosen for continuation was done by doing a qualitative weighting of some of the factors based on personal judgment and evaluation, some of which were based on what was interpreted as being more important factors for the participants during the interviews. This was done in collaboration with my supervisor, where we discussed and gave arguments for the weighting of the specific factors. This approach was chosen because the two designs left had the same number of threats. The final design was selected for continuation, and this design served as the template for the development of the dashboard in cycle 2.

3.3 Cycle 2

The second cycle of DBR consisted of two phases, as shown in Figure 3.7. The first phase was to develop the dashboard while improving the threats and weaknesses identified in cycle 1. After the dashboard was developed, an evaluation phase was done, which was conducted to assess the effectiveness of the dashboard. This section outlines each phase and discusses the approaches taken. I will use "dashboard" and "system" interchangeably in the following sections and chapters; the same applies to "client" and "student" and finally, "teacher-client" and "frontend".

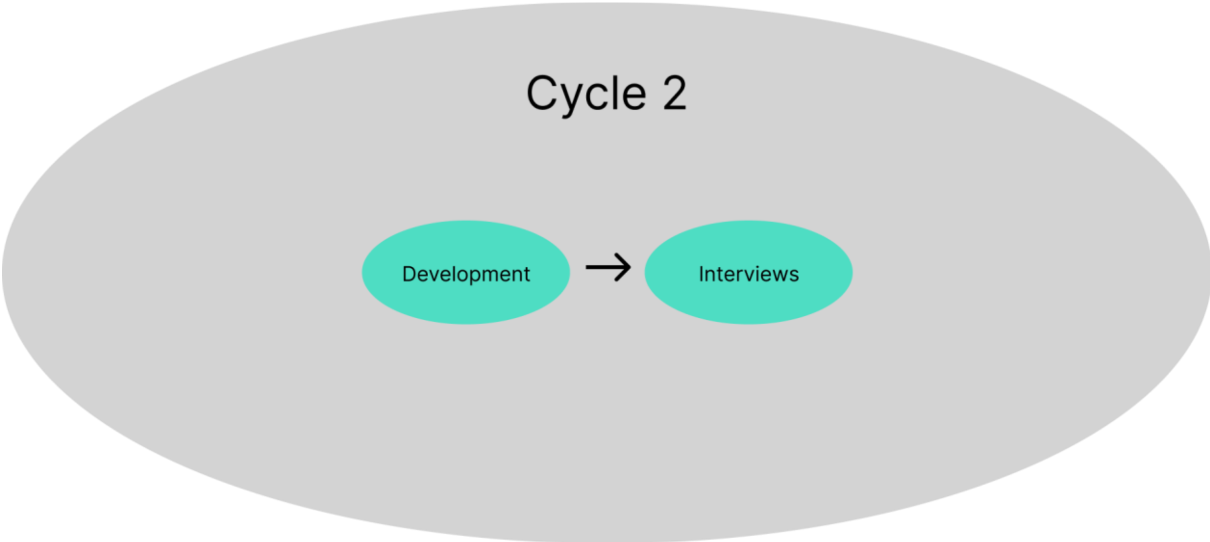

Figure 3.7 Second Cycle of DBR

3.3.1 Development

This section describes the development process and what approach I took when developing the system. It explains the overall strategy as well as some specific implementation techniques. Since some of these aspects blur the line between results and method, a special focus will be put on the tools and libraries in this section, as well as specific implementation details. The resulting architecture and functionality will be presented in the next chapter.

The development process followed an agile approach, as described in (Highsmith & Cockburn, 2001). Notably, my development process prioritized working software and stakeholder collaboration, aligning with the principles of DBR. The former was achieved by focusing on system testing, the latter by modifying the system based on the SWOT analysis conducted in the first cycle.

The system was split into three distinct parts: the frontend dashboard visualization for the teacher, the server managing student connections and processing of incoming data from the wristband sensors, and client scripts installed on student laptops to capture and transmit emotional data from webcam feeds. This approach was taken because of the nature of the devices and to make the system decoupled. The goal was to make it easier to test and allow the server and dashboard components to run on separate machines. Following the DBR guidelines of developing functioning products, I made prototypes of each part early on to allow for end-to-end testing throughout the whole development process. The intended data flow of the system is illustrated in Figure 3.8 below.

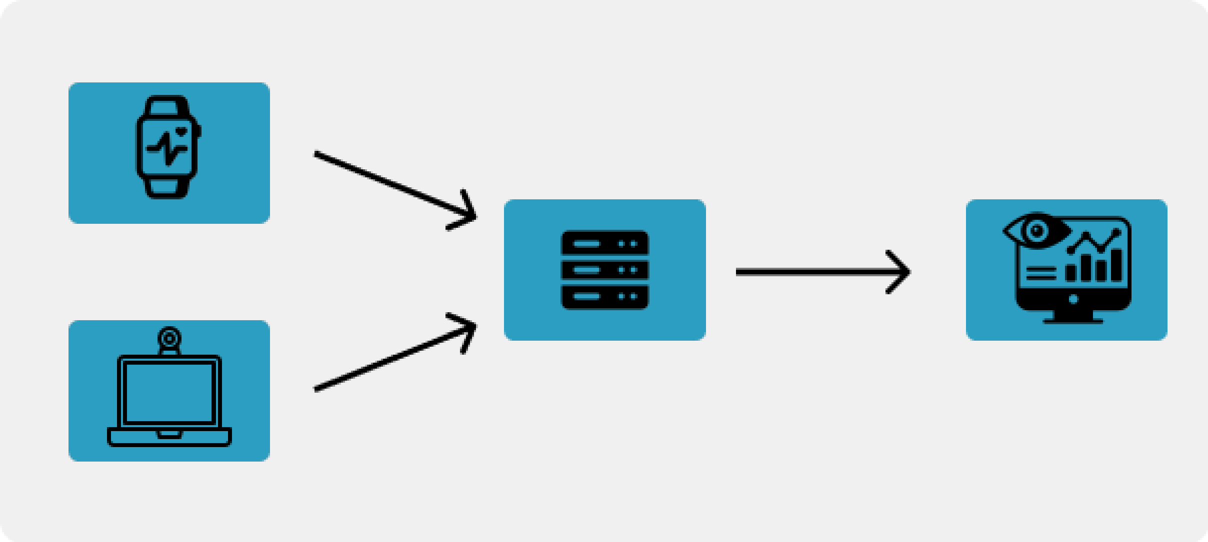

Figure 3.8 Intended Dataflow in the System

3.3.1.1 Frontend

To build the frontend part of the system, I chose React, which is a JavaScript library (React), along with TypeScript, which is a statically typed superset of JavaScript. TypeScript was chosen to ensure code reliability and maintainability as it allows for the identification of potential bugs by the nature of being statically typed. Choosing React as the framework on the frontend had multiple benefits. First of all, React is very well supported with multiple different libraries for visualizations, such as D3.js, which would allow the use of existing components instead of creating my own from scratch. Secondly, React provides good testing support with its testing library. Finally, to mitigate the time investment needed to learn new technologies, it made sense to choose a framework that I had experience with from previous work. This meant I could rapidly begin developing the dashboard without the need for considerable time investment into learning new frameworks.

The frontend maintains a socket connection to the server through which it receives aggregated emotional data of all the students. Using a socket connection is preferable when dealing with a persistent connection receiving real time data since it provides full-duplex communication. When new data arrives, the corresponding graphing components are updated. The emotional data displayed on the frontend represent an aggregated ratio between each student’s baseline for that emotion and the latest computed value. The ratio, representing relative changes in emotional state over time, offers teachers a more useful measure than absolute values, which was seen as preferable, as this eliminates the need for thorough understanding of the underlying calculations.

3.3.1.2 Server

The server is written in Python and follows a multiple-client-server architecture; it maintains socket connections to each student laptop and the frontend, as well as managing the APIs that receive data from the wristbands. To handle multiple connections simultaneously, I used the threading and multiprocessing libraries in Python. This was necessary in order to effectively manage the number of clients expected from this system. To prevent conflicts and errors when multiple threads and processes access shared data, a process-safe queue from the multiprocessing library is used. Incoming data is added to the queue, and another thread continuously consumes data from the queue and stores it in a specific "window" for each type of measurement received (e.g., happiness, stress). In this context, a window refers to a temporary storage space for data points before they are processed. Each window scales dynamically according to the number of clients connected, effectively making it a time-based window. Once enough data points have been collected for a given measurement, the data is aggregated and sent to the frontend.

There were some special considerations that had to be made when creating the server. Firstly, considering that some clients might be overrepresented in a given window, a weighted average of the data points are calculated by lowering the importance of overrepresented students before the aggregation. For each client, the weight is inversely proportional to the number of data points they contribute to the window. This weighting strategy ensures that the data received from any one student does not disproportionately influence the aggregate measurement. Furthermore, since the server handles real-time connections to a potentially large number of students, it needs to manage disconnections well. The way I solved this problem was by keeping a dictionary of connected clients. Once a client disconnected or timed out, it was removed from the dictionary. This allows the server to continue processing data even if unexpected disconnections occur.

The server also spawns processes that handle the incoming data from the wristbands. I used Empatica E4 wristbands to capture the HR (1Hz) and inter-beat intervals (IBI) (time between individual heart rates). The wristband processes connect to the streaming server provided by Empatica. The processes then receive HR values and compute the relevant measurements before putting them into the aforementioned data queue. A simple smoothing function is applied to the raw data to remove unwanted spikes in the time series. This is a simple running average on a window of 20-30 samples, depending on the measurement, with 50% of the samples overlapping between two consecutive windows. In addition, normalization of the data is done to reduce the effects of the subjective and contextual interferences that might influence the data, such as time of day, gender, age, sleep, and so forth. The data is normalized by using the first 30 samples as a baseline. The measurements put in the data queue is the ratio between the measurement computed from the latest window and the baseline.

3.3.1.3 Client Scripts

Client scripts, also written in Python, are installed on the student laptops to perform facial recognition and emotion detection. Once the script has connected to the server, it activates the webcam and begins transmitting the captured emotions. I will elaborate on the two libraries used to achieve this process.

Firstly, the webcam feed is generated by using OpenCV (OpenCV), which is a library commonly used for video and image operations. I set a cap of 10 frames per second for the webcam feed to alleviate computational load and to minimize the differences in hardware specifications on the student laptops. Secondly, each frame in the webcam feed is then processed using Py-Feat (Py-Feat). Py-Feat performs emotion detection using machine learning techniques. In my implementation, I used the pre-trained model "resmasknet" which performs emotion detection by using a residual masking network. The model is explained in (Pham et al., 2021). The output gives a probability score in the range 0 and 1, this score is indicative of the intensity of a given emotion.

The same smoothing and normalization that is done on the physiological data are performed on the facial expression data. Once the baseline is calculated, the script sends the ratios of the subsequent windows of detected emotions and the baseline to the server for aggregation.

Since the system will be used in a classroom environment, there might be interferences in the webcam feed. Other faces might be detected, such as instructors or other students. To handle this, the script will only perform emotion detection on the biggest face box in the frame. Under normal circumstances, this should correspond to the student using the laptop.

3.3.2 Evaluation

For the evaluation of the second cycle, a simple interview was developed where the participants were asked for their reactions to 16 realistic situations observed in the facial and wristband data from 12 students. The interview was structured in two strands. The first strand was about the participants' reactions to the sixteen scenarios in terms of their actions. Specifically, the participants were asked, "how would you support the students if they display the following expressions and stress levels?". The second strand was about the perceived usefulness and related challenges while using the dashboard.

3.3.2.1 Participants and Procedure

There were 12 teachers/researchers who participated in the structured interviews. There were four females and eight males. Among the participants (Three working in Sweden and Nine working in Norway), three participants were researchers who did not have teaching duties at the time of the interview; six participants had dual roles (both teaching and research), and three participants were teachers. The maximum length of the interviews was 28.93 minutes.

At the beginning of the interviews, the participants were presented with the dashboard and the functionalities of the dashboard. They were also informed about the five measurements from the dashboard. Then the participants were presented with the 16 scenarios, one after the other, and they were asked about their reaction to each of the situations. The interviewer took notes of the participants' direct responses to the questions.

3.3.2.2 Scenarios from the Dashboard

The 16 scenarios used in the interviews are given in Table 3.1. They represent realistic situations that were observed in the data of 12 students where the different measurements are in the high or low threshold.

	Engagement	**Happiness**	**Anger**	**Emotional Regulation**	**Stress**
1	High	High	Low	Low	High
2	High	High	Low	Low	Low
3	High	Low	High	Low	High
4	High	Low	High	Low	Low
5	High	Low	Low	High	High
6	High	Low	Low	High	Low
7	High	Low	Low	Low	High
8	High	Low	Low	Low	Low
9	Low	High	Low	Low	High
10	Low	High	Low	Low	Low
11	Low	Low	High	Low	High
12	Low	Low	High	Low	Low
13	Low	Low	Low	High	High
14	Low	Low	Low	High	Low
15	Low	Low	Low	Low	High
16	Low	Low	Low	Low	Low

Table 3.1 Dashboard Scenarios

4 Results

This chapter discusses the results from the two cycles of DBR. It consists of two parts. The first considers the results from the first cycle, which shows the results from the interview evaluations, as well as the combined SWOT matrices for each design. While the second discusses the results from the second cycle of DBR, which involves the dashboard system and final evaluation.

4.1 Results from Cycle 1

This Section outlines the results from cycle one. It shows the results from the interviews by showcasing excerpts relating to the specific SWOT analysis for each mock. Finally, the combined SWOT matrix is shown as the summary of the evaluations.

4.1.1 Interviews and SWOT

This section will present the result from the three interviews with the teacher and experts as described in 3.2.3. It contains excerpts from the interviews highlighting some of the main strengths, weaknesses, opportunities, and threats of each mock. The statements are organized in the same order as the designs in the previous section. For each mock, some of the key strengths, weaknesses, opportunities, and threats are highlighted with interview quotes. The combined SWOT matrix for each mock containing all the participant answers is also provided. The participants are numbered from 1 to 3. For convenience, there is a reference link to the relevant mock at the start of each corresponding section.

4.1.1.1 Mock 1

For Mock 1, all three participants highlighted the visuals of the barometers as a strength, emphasizing its simplicity, color scheme, and way of conveying information.

> **Participant 1**
>
> So, I think that the strength is the visual thing, the fact that they can immediately understand [The emotional state] if all the parameters go more to the green or to the red.

> **Participant 2**
>
> The strength here is that they can see with this bar [barometer] how much this [emotional state] changes. If it goes to green, it means they are doing quite well, and if it's close to red, it means that something is not working very well.

> **Participant 3**
>
> It's nice to see that it is quite well visualized. I like the color selection for these meters. It's very natural to understand it that way.

Participant 1 also highlighted the graph and that the teacher could choose which graph to display as a strength.

> **Participant 1**

I like the fact that you can also click [the barometer], and you can immediately see the actual real-time data.

There were different reactions regarding the information box. One participant liked the information box, highlighting it as a strength:

Participant 3

[…] And how much time they spent on the task, that sounds clear. I guess these are the strengths.

While another thought of it as a weakness:

Participant 1

This box, I think that it is not really appealing to me. Maybe you could add some icons to make it easier to understand. Because it's just text, and that's it. So maybe [it might be] a bit confusing to look at in real time. Maybe icons can help.

Regarding opportunities and threats, one participant suggested using an emoji under the barometers to improve the ambiguity of the colors. The same participant highlighted the color ambiguity of the barometers as a threat. It was also suggested that the teacher should be able to choose which metrics to include.

Participant 3

Adding emoticons at the edges of red and green. On the green, add a happy smiley and a sad smiley on the red. That can be a threat, that it can be a bit difficult to interpret or understand. I was looking at stress, and it is orange, what does that mean? No stress, high stress?

Another participant suggested that some of the metrics might be confusing as they are closely related and underlined this as a threat.

Participant 2

The threat here is that [...] I think that engagement and entertainment are very close topics, maybe it will be confusing.

Mock 1			
Strengths			
Simple and easy to understand	Graph is nice	Likes that you can click on the meters to show graph	Good choice of metrics Good color choice for barometers

Likes the meters	Barometers give immediate visual feedback	Likes the information in the box	Likes the graph
			Likes the layout

Weaknesses

How does "entertainment" relate to emotions	Doesn’t like the information box	The graph is too small, with weak colors. Should use more contrasting colors	Metrics might not be obvious to the teacher
Unnecessary information in the box.	The box is not visually appealing. Add icons to the box instead of text		The graph is small
		Prefers graph over the barometers	The graph is hard to read
Small fonts			

Opportunities

Clean up the layout	Let the teacher decide to include some of the trackers	Improve the box or remove
		Let the teacher select what information to include
Include an alert if a meter hits red	Make the box optional to make the graph bigger	Use emoji under a meter to resolve ambiguity

Threats

Related metrics might be confusing	**Misinterpreting how anger and stress metrics work, ambiguous colors**

Table 4.1 Combined SWOT Matrix for Mock 1

The main strength of mock 1 was how easy and intuitive it was to understand. Participants expressed how much they liked the immediate visual feedback provided by the barometers. They also liked the selection of metrics included in this mock as well as the other features, such as the graph and how they could select which metric to display. Some participants liked the layout, while others suggested it could be improved. The participants also suggested letting the teacher choose which metrics to include, adding alerts for critical measurement states, and adding an emoji to resolve the ambiguity of metrics. The main weakness was the information box, which participants said was both unappealing visually and containing irrelevant information. Furthermore, participants expressed displeasure with the various fonts and the size of the graph. The main threats were how related metrics might be confusing, specifically the entertainment and engagement metric. The other threat suggested was how the ambiguous colors of the stress and anger metrics might confuse the teacher, e.g., stress is in the green; is this good or bad?

4.1.1.2 Mock 2

For Mock 2, there were few strengths. However, one participant highlighted the bar charts containing useful information about the current activities of the students. And that you could see all the graphs of all the metrics at the same time.

> **Participant 1**
>
> […] This part [the bar charts] is really useful. The strength is that the teacher can see an overview of what is happening in this current activity. Another strength could be that I can see all the graphs of the parameters and how they change. So I can compare how they are going.

In contrast, there were a lot of weaknesses identified, some of the more prominent ones being hard-to-read graphs, non-contrasting and ambiguous color schemes, and unnecessary information in the bar charts. The rainbow effect of the colors was also thought of as a threat by participant 1. Lastly, all the participants expressed a preference for the barometers in mock 1.

> **Participant 3**
>
> The information shown on the right side. I would say it’s totally unnecessary. I’m not interested in the time spent by the students on different tasks. If they are doing a task at a point in time, I would like to understand if they are engaged or stressed and then, based on that, adjust the lecture accordingly.
>
> **Participant 2**

> I think as it is now, it is difficult to read. It's difficult to read the font. It needs to be bigger and bolder. Also, this green [color] maybe it's a preference, but I'm not a big fan of it. One suggestion is that you can have different colors for each bar [chart]

Participant 1

> The little circle will update […] You would obtain a rainbow effect when they change a lot. I am worried that this is not clear and would not be useful for a teacher to see these colors that change a lot. I think the bar [barometer] is more useful.

One participant suggested clustering the information to make the information less overwhelming.

Participant 2

> […] Under engagement [the graph], you have the time spent on topic and how engaged they were. The average score and how engaged they were, so you have more information clustered.

The above was also thought of as a threat by the same participant, that having all the information on each side might be overwhelming. Continuing with threats, another participant brought forth that the color from the icons might imply some meaning for the bar charts since they would sometimes have the same color, leading to confusion. Lastly, one participant also stressed that a main threat would be that it would take too long to analyze the information in all the graphs.

Participant 3

> The only threat with this one is you might spend too much time trying to see and analyze what is going on.

<table>
<tr><th colspan="5">Mock 2</th></tr>
<tr><th colspan="5">Strengths</th></tr>
<tr><td>Likes the bar chart representation</td><td>Can see all the graphs of each parameter at the same time</td><td>Provides good and relevant information</td><td>Relevant information regarding activities</td><td>Likes the bar chart overview of current activity</td></tr>
<tr><th colspan="5">Weaknesses</th></tr>
<tr><td rowspan="3">Difficult to read graphs & icons</td><td rowspan="4">Colors are less useful than meters</td><td>Bad color on the bar chart</td><td>Ambiguous colors for the icons</td><td>Bad colors</td></tr>
<tr><td rowspan="3">Hard to link emotional state and today's performance</td><td>Bar charts are not necessary during class. More important after</td><td>The icon could not be useful if it updates fast</td></tr>
<tr><td rowspan="2">Information not clear enough for a quick glimpse</td><td rowspan="2">Hard to read the graphs</td></tr>
<tr><td>Bad color schemes</td></tr>
<tr><th colspan="5">Opportunities</th></tr>
<tr><td rowspan="2">Group the activities with the emotions. Layout</td><td rowspan="2">Cluster information</td><td rowspan="2">Give the ability to zoom in on the graph</td><td colspan="2">Add contrasting colors</td></tr>
<tr><td colspan="2">Use meters from mock 1</td></tr>
</table>

Have different colors for the bar chart and icons	Change the color of the bar chart. To avoid the implication of color from icons. Neutral color	Separate the emotions and current topic on each screen	See student performance compared to the class average
Threats			
Color from icons might imply the same for the bar chart. Confusion	The color circle might be confusing	Overwhelming	Too hard to analyze and will distract the teacher for too long

Table 4.2 Combined SWOT Matrix for Mock 2

Mock 2 got conflicting feedback regarding its strengths and weaknesses. Some participants liked the bar chart representation of the current activities, while others expressed how the bar charts are too hard to parse during class. One clear strength was the ability to see all the graphs of all the measurements at the same time. The same issues regarding font, layout, and color selection as in mock 1 were highlighted as weaknesses. Additionally, the color icons could be problematic during fast updates, giving a "rainbow" effect. Participants suggested improvements could be made by using the barometers from the first mock, allowing for individual student statistics, and moving one view to another page to clean up the layout. They also suggested improving the colors and general layout. Regarding threats, they expressed that the meaning of the colors of the icons on the graphs could have implications for the meaning of the colors used in the bar charts. Moreover, the colors used in the icons could be confusing, indicating the same issue as with mock 1 for measurements such as anger and stress. Finally, participants said that the information could be too overwhelming and that it would take too much time to analyze, causing the teacher to become distracted.

4.1.1.3 Mock 3

The main strengths of Mock 3 were the textual alerts and the summary box with descriptives of today's performance. One participant also highlighted the emoji as a good and easy way of conveying the emotional state. There were also some differences in opinion regarding some of the information. For instance, one participant liked the fact that you could see the time spent, while this was seen as irrelevant by another.

Participant 1

I like the alerts a lot. I like the fact that they can see what happened that [exact] minute, and they can relate to it. Also, the performance, the fact they can see how many things the students did. […] I think that the average time, it's really meaningful too.

Participant 2

I like the alert, this is an improvement over the previous ones, it shows the state in a bigger projection. I like the pie chart that has these different colors and shows how the different topics compare in time usage. I also like the today's performance. That's a strength that they [the teacher] can see if they have completed the tasks and so on. It's a nice representation of the cycle [doughnut chart].

Participant 3

The strength, I would say, is the way the smiley captures the current emotional state, I kind of liked that the most in the whole mock. Just by looking at it, I can see how they are feeling.

For the weaknesses, the participants emphasized bad layout, color scheme, and that the "time spent" information was not needed, and finally, that there was too much information overall.

Participant 1

I can see a weakness is how everything is positioned. It's a bit unordered, I don't know where to look right now. Maybe it would be good if you let us choose what [metrics] to look at.

Participant 3

I think it is way too much information. I find it just way too much. As I said in the previous one, you need to spend quite a lot of time to see what's going on.

For opportunities, the participants suggested letting the teacher choose which metrics to see. Merging the pie chart and the "today's performance" box and removing some of the "bloat" like time spent.

Participant 1

[…] Also, I think that the today's performance and the average time use can be merged in a way because they are related. So I can compare them better.

Participant 3

How could it be improved? Just by actually removing some of the information you are offering.

The two main threats were that the teacher might spend too much time analyzing all the information, and that the alerts might be too distracting.

Participant 2

The threat can be that they [the teacher] get too distracted from the alerts, but I guess that's something we want to happen.

Mock 3				
Strengths				
Today's performance	Likes textual alerts	Likes the pie chart and time spent on information	Textual alerts are very nice	Likes the visual representation (Doughnut cycle)
Likes today's performance overview. Easy to see the progress	Emoji captures emotional state very well and is simple to understand			
Weaknesses				
Color scheme	Doesn't like the layout. Hard to know where to look	Bad font	The pie chart lacks units (explanation)	Too much information
Time spent not needed				
Opportunities				
Should be able to choose which metrics to see	Divide the screen by group to improve readability	Remove some of the indicators to reduce bloat	Merge the pie chart and today's performance in one box	See emotional state over time to link today's performance
Threats				

Alerts might distract	**It might take too much time to analyze during class**	**Too much information, might overwhelm the teacher**		

Table 4.3 Combined SWOT Matrix for Mock 3

The main strengths of mock 3 were the added textual alerts, today's performance section and the different visualizations of the emoji and doughnut cycle. Many of the same weaknesses seen in the previous mocks were also present: bad font, too much information, bad color scheme, and layout. One participant liked the time tracker, while another found it mostly irrelevant. Participants suggested that the teacher should be able to choose which metrics to include or remove some metrics to reduce bloat. They also suggested cleaning up the layout by grouping the screen into two sections. The threats were identified as the alerts' ability to distract too much, as well as the overall time it would take to analyze all the information during class. Finally, the issue of overwhelming the teacher was brought forth as a threat.

4.1.1.4 Mock 4

Starting with the strengths of Mock 4, the participants emphasized a clear layout, approval of the exercise overview, and a good selection of information being displayed. They also liked the way the emotional metrics were displayed and the ability to choose which graph to display.

> **Participant 1**
>
> The strength is that I like the way in which the exercises are ordered, it's presented clearly. And I like the fact that a teacher can see what is already done and what they are going to do.
>
> **Participant 3**
>
> I like the upper left corner part. That figure where you can click on the metric, and it shows you the progress over time. Very Clear. […] Exercise overview, it's nice to see that it didn't go well with task 1.2.

For weaknesses, some of the same points were reiterated from the previous mocks. The time tracker was still seen as unnecessary by one of the participants. Furthermore, some of the participants did not like the color scheme, highlighting that the colors for the exercise overview were not intuitive.

> **Participant 3**
>
> If you are really keen on showing this time tracker, I would say add some more information to it, because in this form, it is not informative enough that they spent five seconds on addition.

For opportunities, participants suggested adding the alert from the previous mock. They also wanted the ability to click on an exercise to see more detailed information. Some of the threats mentioned included misinterpretation of colors in the emotion icons, as well as a concern that the exercise overview might have an implication for classroom positions.

Mock 4				
Strengths				
Can choose which metric display graph of	Likes the emotional state "box". Good grouping of information	The information displayed is well-chosen; shows the correct information	Likes the time tracker	Likes the layout; clear
Likes that the teacher can see the exercise progress				
Weaknesses				
Color scheme	Colors changing in real time might confuse & overwhelm	Time tracker unnecessary	Bad colors	Too much information at once
Not very clear what the color means in the exercise overview	Graph lacks unit	The barometer from mock 1 is better than the icons		
Opportunities				
Should be able to click on an exercise to see more information	Include the task score, not just the completion status	Include an alert when metrics are critical		
Threats				

Exercise overview might imply classroom position	**Misinterpretation of colors in the emotion icons**			

Table 4.4 Combined SWOT Matrix for Mock 4

The main strength of mock 4 was the "Emotional State" box grouping the graph and all the measurements. Participants also highlighted the ability to see the exercise progress as a plus and that the overall layout was clean. Weaknesses were identified as the color icons of the measurements which could be hard to read during rapid updates, causing a "rainbow" effect. Participants also said that the barometers from mock 1 were preferable and that the time tracker was unnecessary. Improvements included letting the teacher click on an exercise to see more information, adding an alert feature when metrics are critical, and having the task score in addition to completion status. There were two threats highlighted. One was that the exercise overview might imply the classroom position of students, while the other was the issue of the colors in the icons that could be both hard to interpret and hard to analyze during rapid updates.

To summarize, the first design had the most strengths, particularly how intuitive the participants found the barometers. On the other hand, the second and third designs presented the most weaknesses and unique threats. Concerns about information overload were consistent across all designs, with participants expressing worry that teachers might spend excessive time analyzing information during class. To address this, some participants proposed solutions such as removing information or enabling teachers to select what features to include. Additionally, all designs were criticized for the ambiguity of colors and related metrics. This issue was particularly apparent for the second and fourth designs, where the use of color circles could potentially result in a confusing "rainbow effect" if emotional states fluctuated rapidly. Table 4.5, below, contains a summary of the number of items per dimension for each mock.

As described in section 3.2.4, an elimination process was completed to determine which design to progress to cycle 2. Mocks 2 and 3 were discarded due to their higher number of threats compared to mocks 1 and 4. Of the remaining designs, mock 1 was chosen for continuation. The main factor in this choice was the strong approval of the barometers, in addition, the threat of the "rainbow" effect associated with the color circles in mock 4 was considered difficult to ameliorate without an outright overhaul.

	Mock 1	**Mock 2**	**Mock 3**	**Mock 4**
Strengths	9	10	7	2
Weaknesses	5	11	9	4
Opportunities	7	6	5	3
Threats	6	8	3	2

Table 4.5 Summary of SWOT

4.2 Results from Cycle 2

This section starts by showing the finalized design of the completed dashboard along with the main functionality. Then the architecture will be shown according to the 4+1 model, and results from the evaluation phase will be presented.

4.2.1 Final Design and Functionality

The final design is shown in Figure 4.1. A number of improvements have been made to the preliminary design of mock 1. First, the information box in the lower-left corner has been improved in view of the weaknesses highlighted by the interviews. It now contains icons to provide better readability as well as a more visually pleasing design. The graph has been made bigger and now shows the baseline as a red dashed line; this was done to improve the readability, as was stressed by the participants. Furthermore, it now only shows the clicked metric over time, and the option to select time and topic was removed to reduce information bloat and emphasize the minimalist style. The graph shows a time slice consisting of the data of the measurement from the last five minutes. An icon was added to the header to show the number of connected students. Finally, based on the feedback, there is also an alert added where there will be a popup in the upper right corner whenever one of the barometers hits red. These alerts will remain for 10 seconds or whenever the teacher clicks on them.

Figure 4.1 Final Design

One of the main concerns about the preliminary design was the potential confusion about engagement and entertainment, as well as how anger and stress would work. To address the former, entertainment was replaced with another metric, emotional regulation. Which is a measure of much control individuals exert over their emotions and how they express their emotions, which can be an indication of how well one can focus on a particular task (Lerner et al., 2015). Emotional regulation is closely related to the heart rate variability of the subject (Visted et al., 2017). To address the latter, each barometer now contains a tooltip explaining the metric and what the color gradings mean. These tooltips are opened when the user clicks on the orange question mark in the corner of the barometer, as shown in Figure 4.2. Both the graph and the barometers show the relative increase for the whole class. 0.5 on the barometer equates to a 50 percent increase. The scales are larger for happiness and anger since they are more prone to relative change, e.g., a small smirk might have a large impact on the relative change compared to the baseline.

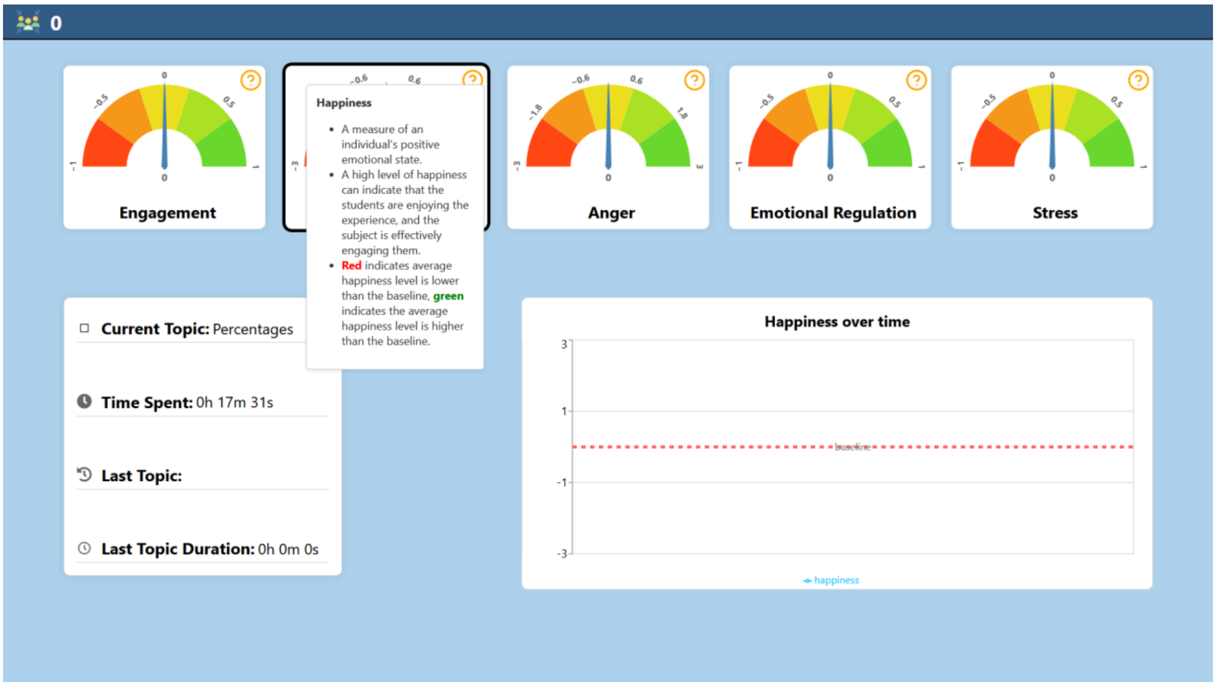

Figure 4.2 Tooltip

4.2.1.1 Measurements

This section explains how all the emotional measurements are computed, as they constitute the core functionality of the system. It starts with the three measurements computed using the HR captured with the wristband (engagement, stress, emotional regulation), then it describes the measurements computed using the webcam data.

Physiological Stress

Physiological stress is computed as the heart rate’s increasing slope. It has been shown that the more positive the slope of the heart rate is, the higher the stress is (Taelman et al., 2009)

Engagement

Engagement is measured by computing the various features computed from the heart rate, which are correlated with engagement as outlined in 2.2.3.1. The specific features used are average heart rate, max, and min heart rate, the difference between max and min heart rate, and variance of the heart rate.

Emotional Regulation

Emotional regulation is measured by calculating the heart rate variability. The HRV is calculated by computing the mean root square of successive differences (RMSSD) between normal heartbeats using the IBI data from the Empatica wristband.

Happiness & Anger

As described in 3.3.1.3, Py-feat is used to perform emotion detection on the faces of the students. The specific model employed is the “resmasknet” which performs emotion detection by using a residual masking network, as explained in (Pham et al., 2021).

4.2.2 Architecture

This section outlines the architecture of the system according to the 4+1 model in (Sommerville, 2015). It presents the Logical view through different state diagrams, as well as the Process and Physical views from the model; finally, a use-case diagram is shown. Although a part of the 4+1 model, the development view is not included; the reason is that the aspects examined in this view, such as programming languages, libraries, and tools, are already integrated in the methodology. Adding this view would not add additional insights.

4.2.2.1 Logical View

In this section, I showcase the different state diagrams for each part of the system; The frontend, the server, the Empatica processes, and finally, the client scripts on the student laptops. The state diagrams have a double-sided box to indicate the starting state. In the boxes, various events are listed, and the arrows from one box to the next indicate a state transition. The arrows are labeled with the event id to indicate which event led to the transition. It is important to note that these state diagrams serve as a high-level overview of the system, but given its concurrent nature, some parts might be in multiple states at once.

The frontend has a straightforward state, as shown in Figure 4.3. Once initialized, it will continuously attempt to connect to the server. Once connected, it will be ready to receive data. Every time data has been received, it updates the relevant visualization elements.

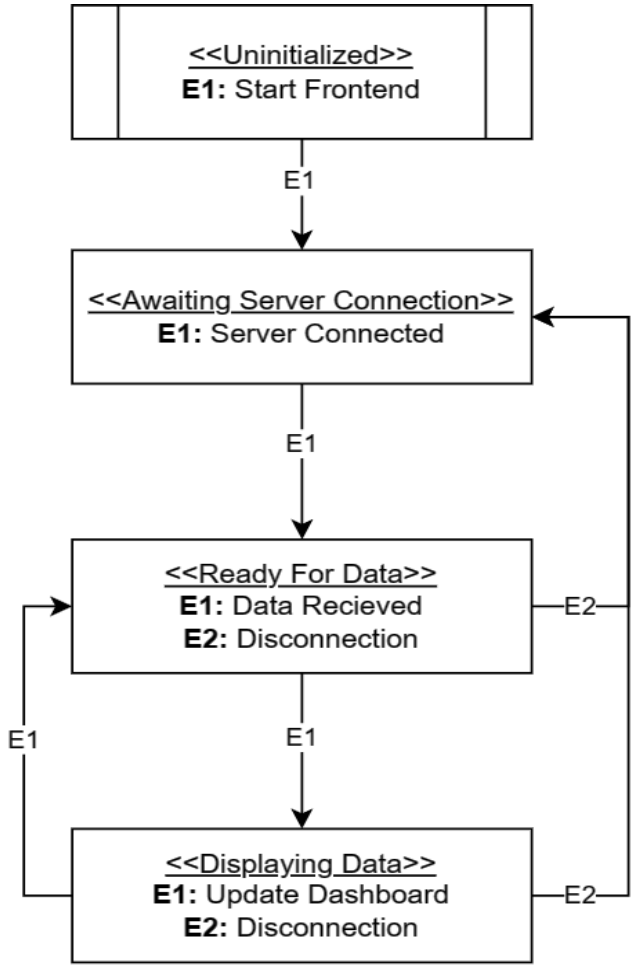

Figure 4.3 Frontend States

A sequential view of the server is shown in Figure 4.4. As mentioned previously, it is possible, and likely, that the server is in multiple states at once. For instance, the server will always listen for new connection requests.

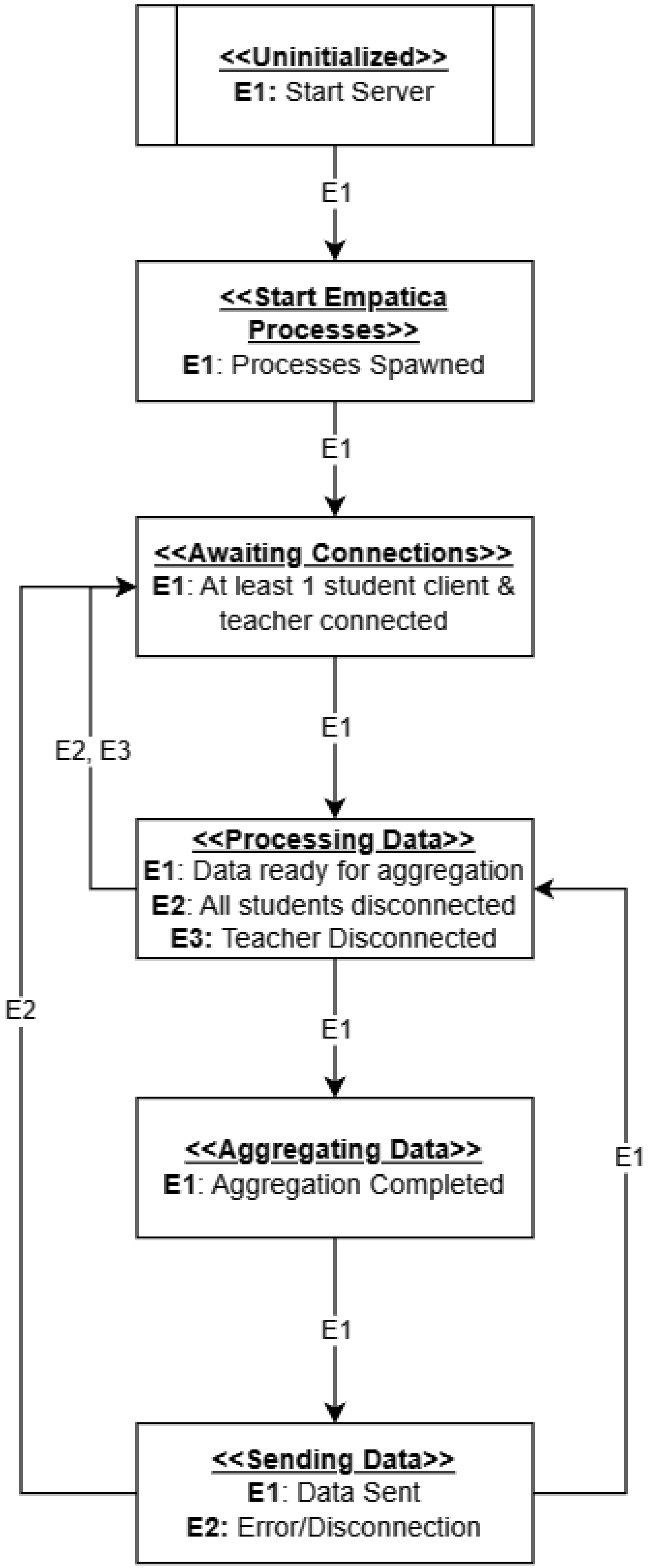

Figure 4.4 Server States

The Empatica wristband processes, initiated by the server, begin by establishing a connection to the streaming server provided by Empatica. After successfully connecting, they start receiving physiological data from the wristbands. The first N data points received are used to calculate the baseline for the associated measurements, after the baseline is computed, the ratio of the following measurements and the baseline is put into the data queue on the server. A diagram is shown in Figure 4.5.

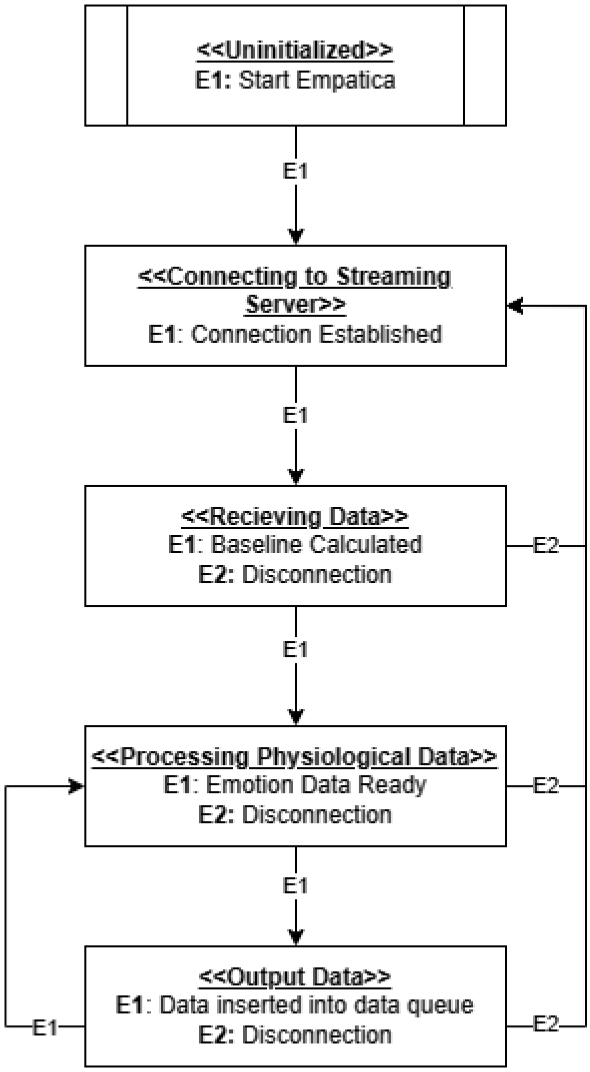

Figure 4.5 Empatica States

The student client scripts on the laptops have the states shown in Figure 4.6. After starting the script, it will try to connect to the server. Once connected, the same cycle as the wristbands will ensue. After calculating the baseline using a preliminary set of data points, the ratio of the subsequent measurements and the baseline will be sent to the server. If a disconnection occurs, the script will try to reconnect.

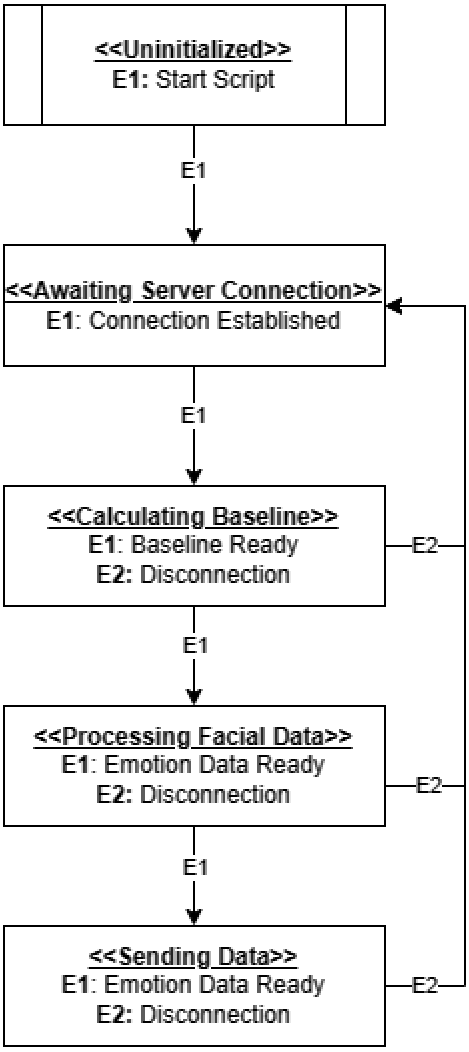

Figure 4.6 Student Laptop Script States

4.2.2.2 Process View

According to Sommerville, the process view shows how the processes of the system interact at runtime (Sommerville, 2015). The arrows represent a relationship between two objects or processes, and the direction shows the way the data flows through the system. Figure 4.7 below can be seen as a more detailed extension of Figure 3.8. seen in the methodology chapter. The student laptops are represented as individual processes $l_1 - l_n$, that receive data from their respective webcams, indicating the start of data flow. These processes are responsible for preprocessing and computing their respective measurements from the webcam feed. They send their computed measurements to the server via a socket connection. The wristbands send raw data to the server.

In the server, there is a process s_1 that is responsible for listening for incoming student laptop connections. If a student is trying to connect, a new thread c is spawned to handle all further data and requests from this student laptop client. These threads are responsible for putting the incoming data in the data queue, indicated by the "put" arrow.

The wristbands are handled by individual processes $w_1 - w_n$, these processes receive data from the wristband, compute their respective measurements and put it in the data queue.

Finally, process s_2 is responsible for handling the teacher connection (frontend). It will initially wait for incoming connection request by the frontend and then start processing data from the data queue, indicated by the "take" arrow. It aggregates the data and sends it to the frontend process through the websocket connection.

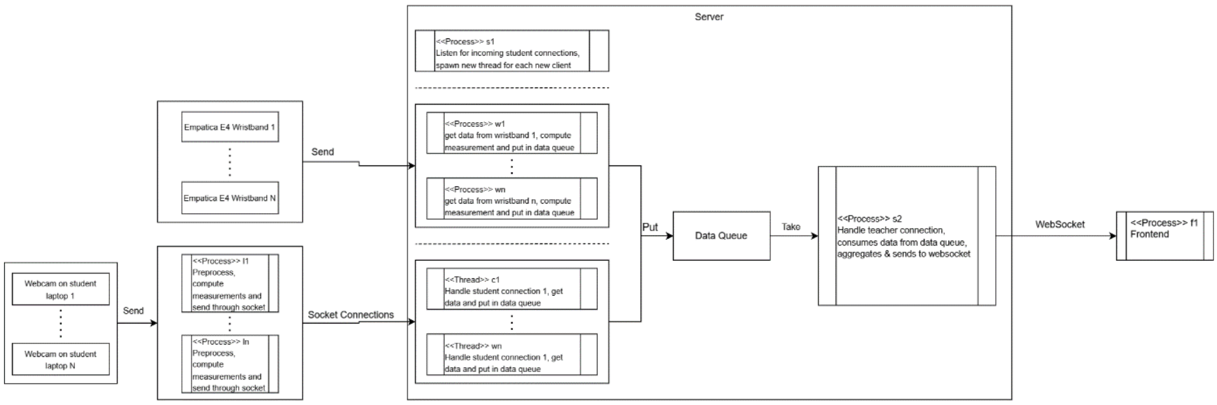

Figure 4.7 Process View

4.2.2.3 Physical View

The physical view in Figure 4.8 shows how the different software components are distributed across different machines or processors. As shown, the server and the frontend can be deployed on different machines. Additionally, an arbitrary amount of student laptops and wristbands can connect to the server. Here the API component refers to the process that is handling the wristband.

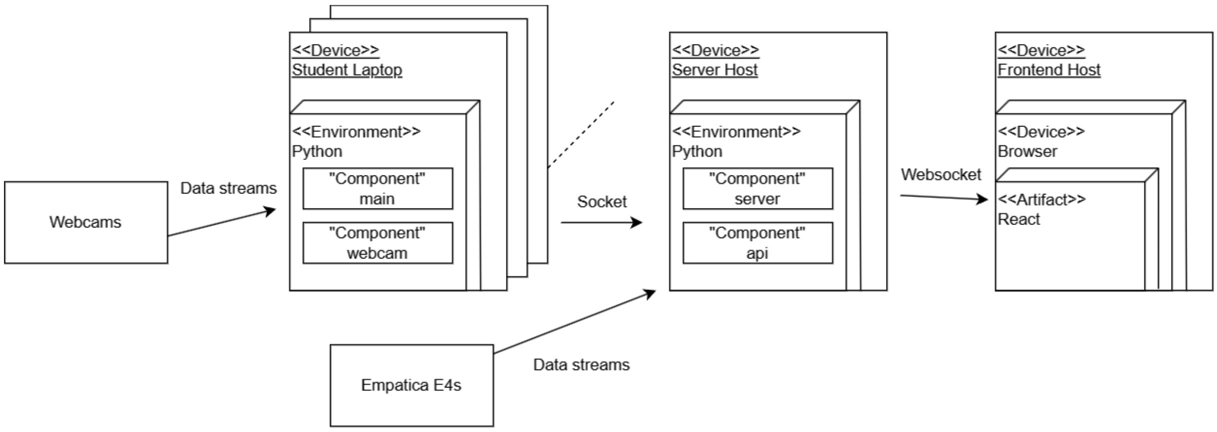

Figure 4.8 Physical View

4.2.2.4 Use Case

In Figure 4.9, I provide a use case diagram to highlight the different use cases between the actors in the system (student & teacher). The teacher is the main interactor with the system and can receive live emotional updates and notifications, view the aggregate emotional data, and look at various descriptive statistics. The student has limited interaction with the system and will only connect/disconnect to the system and transmit emotional data.

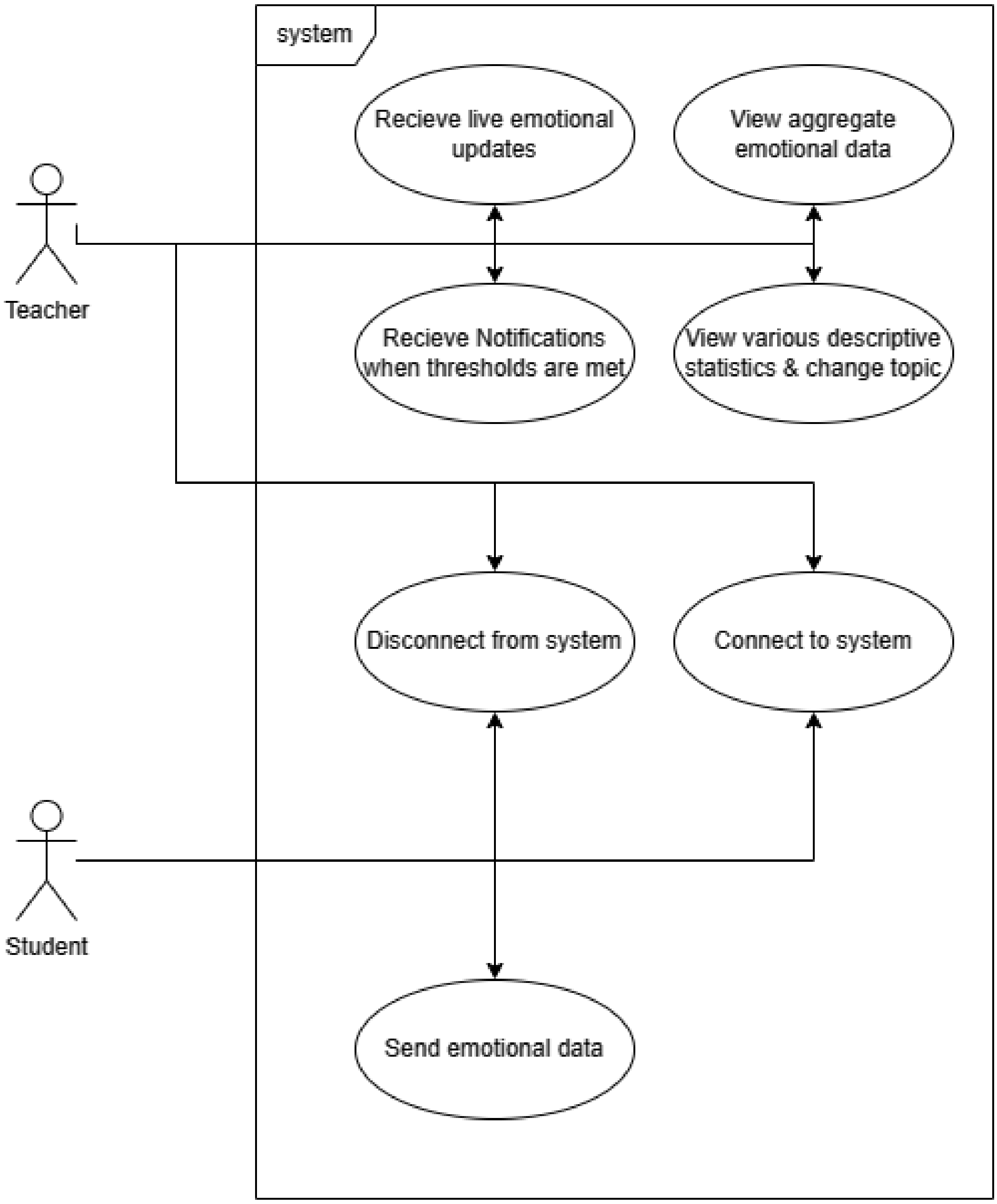

Figure 4.9 Use Case Diagram

4.3 Interview Results

From the analysis of the interviews from the participants, six themes emerged: responses for the emotions/expressions (low engagement, low happiness, high anger), responses to low emotional regulation, responses to high stress, the "promotable" behaviors, usefulness and potential uses of the dashboards, and challenges while using such dashboards.

Response to emotions/expressions: Most of the participants focused on the three aspects of the dashboard when it displayed low engagement, high anger, and low happiness from the students. This was the case whenever such a situation came up in the interviews. The teachers offered several insights about how they would address these three issues. To address the disengagement, the participants offered to assess the cause of the distraction, disengagement, and possible confusion-led disengagement and then provide solutions accordingly. Some participants mentioned providing individualized instruction to those students who clearly seemed to be distracted and disengaged. Some others suggested inquiring further and engaging the students in a dialogue about what is causing the confusion, while some others suggested that they would make the lesson plans to have more active involvement from the students. Some participants also mentioned using positive reinforcement as a way to "*tackle*" both low engagement and low happiness in the classroom. They suggested to recognize and praise classroom participation and peer help-giving behavior to "*have higher morale*" in the classroom. Regarding high anger in the classroom, most of the teachers suggested that classroom situations can be changed within a specific class period or when a teacher thinks an incident may create problems that can increase the anger in the classroom. For example, one of the participants said that, "*on bad mornings when I see that children might lose it, I was extra careful so that there are fewer problems".* Another participant mentioned, "*I modified the lesson plan [in real-time] to have the students doing activities that I thought were easier to manage*."

Response to emotional regulation: many participants expressed difficulty in having a clear understanding of this construct. The participants inquired further to clarify the meaning of emotional regulation and suggested to investigate it further in terms of which emotions are more and better regulated by the students before being able to provide feedback on it.

Response to high stress: collectively, the participants responded with the following reactions to situations involving high stress. 1) To understand the reason for the stress, for example, it could be because they are facing an impasse with the learning material, or the learning material is too difficult for their current knowledge. 2) The teacher can foster a classroom atmosphere where students feel comfortable expressing their emotions and seeking help. This can be achieved by encouraging open dialogue, promoting a non-judgmental atmosphere, and establishing a trusting relationship with the students. 3) The teacher will act with compassion and will approach the stressed students with empathy and understanding. They may engage in one-on-one conversations to inquire about the children's issues in the classroom and with the learning activity/material, allowing them to share their concerns and emotions. The teacher would actively listen, validating their discomfort, and assure them that they are not alone in their struggles. 4) The students can utilize different approaches to address their stress. Teachers can provide extra information about available material, such as solved examples in the books, provide extra hints, or ask another student for support.

They may also offer guidance on stress management techniques in the classroom, such as focusing on the easier activities, or time management strategies. 5) Recognizing the impact of short-term stress on students' ability to perform, a teacher would also consider adjusting academic expectations temporarily. This can include extending the current activity and the support provided for the student, providing additional support for the forthcoming activities, or breaking down complex tasks into smaller, more manageable steps.

The "promotable" behaviors: All the participants agreed that there were certain behavioral patterns that could be promoted in the classroom. These patterns included high engagement, high happiness, low anger, and low stress. The teachers, almost unanimously, suggested that they would keep special notes of the situation and try to emulate similar situations in the classroom. As one of the participants mentioned "*If my students are happy and stress-free, I am happy and stress-free. So, I better support [the students] in a way that keeps the overall happiness green [high, as shown in the dashboard] …*".

Usefulness of the dashboard: The participants responded with the following points about the usefulness of the dashboard 1) Close monitoring, the teachers mentioned that the dashboard allows them to closely monitor students' progress and states and provide timely feedback. They can quickly identify students who may be stressed, disengaged, sad, or angry, track their completion of tasks, and intervene as and when needed. One of the teachers suggested as a future functionality that feedback can be provided through the dashboard itself, enabling efficient communication between teachers and students.2) Data-driven insights, provide access to real-time data and analytics about student performance, progress, and engagement. Teachers mentioned about using this information to identify areas where students may be struggling, track their progress and difficulties, and make informed instructional decisions to address individual needs (by investigating further with individual students). 3) Post-class reflections and communication with colleagues, the teachers mentioned that when mapped to their own activities within the classroom, it can be a very beneficial tool to be used in the post-class reflection and also to communicate the challenges and opportunities with their colleagues. The teachers also mentioned that the dashboard could also be used to inform about what content to adjust so that the classroom can function in a smoother manner. 4) Classroom management, teachers suggested that they can organize and distribute assignments, track the overall completion status, and maintain an overview of the class's stress and engagement. It can also facilitate communication with students, allowing teachers to make activity-related announcements and updates.

Challenges while using the dashboard: Collectively, the participants raised the following issues with the dashboard and other dashboards in general. 1) To ensure the accuracy of the data representation, the participants cited the emotional regulation visualization and suggested that the dashboards should not have inaccurate or unreliable information, or information that is difficult to comprehend in a short amount of time. This can negatively affect decision-making and hinder the ability of teachers to support their students effectively. 2) Teachers mentioned that they would need appropriate training and technical support to effectively use the teacher dashboard. Lack of proper training or limited technical assistance can make it difficult for them to navigate the dashboard's features, interpret the data, and utilize the tools to enhance their teaching practices. 3) information overload, many teachers mentioned that even though the dashboards are rich in information and utility, it can cause them to be overloaded with the information that they are processing at the same time. They are mostly busy in the classroom with teaching-related activities, and adding a new technology that requires constant monitoring would add to their regular efforts. 4) Introducing a new technology like a teacher dashboard requires a well-planned implementation strategy. Lack of clear communication, resistance to change, or insufficient training and support can hinder adoption among teachers. Ensuring buy-in from all stakeholders and addressing any concerns or skepticism is essential for successful implementation. 5) Teacher dashboards store and process large amounts of sensitive student data. Ensuring the privacy and security of this data is crucial to maintain trust within the education community, according to the participants who had dual roles (i.e., both teaching and research). Safeguarding against data breaches, unauthorized access, and maintaining compliance with data protection laws can be a significant challenge.

5 Discussion

This chapter aims to discuss and reflect on the findings of evaluations of the dashboard and answer the research questions presented in section 1.3. It also discusses the knowledge gained from developing such a system and considers specific guidelines when developing an emotion-aware system.

5.1 Useful Information and Visualization

The first research question was: *To what extent does the real-time engagement and emotional overview provide the teacher with useful information?*

While the second research question was: *How effectively do the chosen visualization types provide the teacher an overview of the emotional state of the class at a glance?*

To answer these questions, I investigate the central themes found in the final evaluation of cycle 2 and the SWOT result of cycle 1.

The overview of emotions and stress seems to promote both personal feedback and feedback focusing on social dynamics. The participants proposed engaging distressed students in personal dialogue to discuss how they're feeling and to emphasize collaboration with other students to alleviate their distress. This is in contrast to dashboards, where the main focus is providing information about task progress, as found in (Molenaar & Knoop-van Campen, 2019). Here, teachers mainly gave feedback relating to progress and the task at hand after consulting a dashboard that did not contain indicators related to emotions. Furthermore, having an understanding of the emotional state of students can help teachers provide metacognitive feedback, as displayed by the participant's suggestion to teach students stress-management techniques and time-management strategies. Metacognitive feedback refers to the types of feedback that relate to how students control and influence their own learning (Molenaar & Knoop-van Campen, 2019). Previous research indicates that personal, social, and metacognitive feedback from teachers is rare (van den Bergh et al., 2013); thus, providing emotions and stress information could boost this type of feedback. Adding on, these types of interpersonal engagements between teacher and student can strengthen trust; this is especially relevant for primary school students who might not be able to express their frustrations and struggles adequately.

Additionally, it was outlined that the information conveyed by the dashboard would facilitate post-class reflection and help the teacher make informed decisions about class content. In this regard, the dashboard can provide the teacher with the means to improve their own teaching strategies and pedagogical approach. Additionally, it was stated that the dashboard provides documentation that could be used as a reference when discussing teaching approaches with colleagues. The implication here is that a teacher who may have previously had an intuition about how students react or tolerate certain topics would not feel comfortable discussing these issues with authority. Having the data from the dashboard could be helpful as leverage when discussing or partaking in decision-making processes. To build on this, future iterations of the dashboard could provide a summary of the whole session after completion. This could positively influence the post-class reflection process.

Of special interest was the dashboard's ability to provide the means to identify "promotable" behaviors by helping the participants identify patterns of behavior that they would seek to recreate, which was the case when the barometers showed low engagement, low happiness, and high anger. I have not identified this type of action response in other studies. It is an interesting observation that could have implications for the designs of future teacher dashboards. Further emphasis on highlighting such situations could be explored. For instance, parameters could be set in the dashboard to automatically inform the teacher when each of the relevant measurements is within these desired ranges or record it for future reference.

It is clear that the visualization types (barometers, graph) used in the dashboard were, by and large, effective at conveying the emotional state of the class at a glance. To understand this, I will consider the relevant feedback from the participants from the second set of interviews. It was outlined that the close monitoring of students provided by the dashboard would allow the teacher to provide timely feedback. In this sense, it was perceived that the visualizations employed in the dashboard let the teacher intervene at the right time. This would not have been possible if the visualizations were too hard to analyze or otherwise unintuitive. Furthermore, the visualizations effectively provide the teacher with the means to tackle the different emotions that students experience. This can be seen from the responses regarding the different types of feedback they would give when a specific metric is in green or red, respectively.

Furthermore, the SWOT analysis of mock 1 and the feedback from the interviews of the first cycle indicate a strong preference for the barometer visualization in contrast to the bar charts, color circles, and emojis present in the other designs. All the participants highlighted the readability and how intuitive they were, and this is a testament to its effectiveness. This type of visualization is not common in other dashboards, as established in the literature review in section 2.2. Further, this visualization was inspired by the dashboards typically seen in airplanes, which include gauges that let the pilot see exactly what meters are "okay" by seeing if the needles are in the correct spot. There might be other types of dashboards outside of educational settings that could inspire other types of visualization techniques. Moreover, considering how the barometers were especially emphasized and preferred over all the other types of visualizations, it should warrant future dashboards to explore the option of including them. However, the color coding would need to be addressed for certain measurements, such as anger and stress. The color green conveys additional information about something being "good" in addition to the intended meaning of a relative increase. A possible solution to this is to use one single color that scales in intensity from low to high in order to avoid the extra information associated with red and green.

One issue that was raised that negatively affects the usefulness of the dashboard was the hard-to-grasp emotional regulation construct. This is detrimental because if the teacher does not have a clear understanding of the measurement, it will be hard for the teacher to make any actionable response to its values. Additionally, it raises concerns relating to the documentation aspect mentioned earlier; if the teacher cannot explain the measurement clearly, it will be hard to argue for the legitimacy of the dashboard when using it in this context. A reason why this measurement was hard to grasp could be that its name is not a colloquial term – in contrast to all the other measurements which are widely used outside of research domains. This ties into another issue with training requirements and data literacy, which is a well-known obstacle with LADs, and has been discussed in previous research (Lee-Cultura et al., 2023; Ouatiq et al., 2022). Possible fixes for this could include a further emphasis on measurement explanations or removing this measurement in favor of measurements that are commonplace outside of research contexts.

Another facet that was raised was the issue of resistance to change and data privacy. As with the training requirement, these problems have been recognized in previous works (Schwendimann et al., 2017) and constitute factors that obstruct the usefulness of the dashboard. The factors that involve cultural dynamics might be hard to address, such as resistance to change. Alleviating this might involve spreading knowledge and awareness of the positive factors that such dashboards bring and focusing on the ease of using the system in practice. Specifically for my system, this would include a comprehensive guide to setting up the student laptops and running the dashboard on the teacher’s laptop.

The topics of data literacy, resistance to change, and information trust are issues that are experienced in different fields, especially concerning the adoption of machine learning techniques in the medical field. The issue here is that the algorithms are not transparent and appear as a black box where you can explain the input and output, but not the function of how you transform one to the other. Specifically for my dashboard, most of the measurements do not rely on machine learning to compute the values. This is beneficial in the sense that the function for computing them can be completely transparent and explained to teachers and administrators who might have decisional powers for the adoption of the dashboard.

Finally, information overload was raised as a challenge when using the dashboard. This was brought up in the evaluation phases. This was a problem that was anticipated, and inspired the minimalist style of the dashboard, however, it is a hard problem to address. Furthermore, it does not appear to be a unique challenge to my dashboard, as it was also identified as an obstacle for the dashboard in (Lee-Cultura et al., 2023). Making the dashboard more customizable by allowing the teacher to select or remove certain features should be explored in future iterations. However, it is also possible that further interaction and familiarity with the dashboard can help alleviate it.

To summarize, for the first research question, the findings in this thesis indicate that having real-time access to students' emotional states and engagement levels is very useful for teachers. As discussed, the information conveyed by the dashboard promotes a more personal and empathetic approach to the teacher's feedback, specifically motivating the teacher to engage in personal dialogue, and stressing collaboration between students. Moreover, the information encourages metacognitive feedback that helps students improve their own learning processes, such as stress and anger management techniques. Furthermore, the dashboard is useful for post-class reflection and decision-making concerning teaching strategies. In addition to providing the teacher with the means to improve their own teaching, it also provides them with documentation to be used while discussing teaching practices and approaches confidently with colleagues. Lastly, the dashboard helps in identifying promotable behaviors among students, which is novel compared to traditional dashboards. However, factors such as unintuitive metrics, resistance to change, information overload, and data privacy concerns are all obstacles that might hamper the usefulness of the dashboard.

Regarding the second research question, the findings point out that the visualization choices in my dashboard, particularly the barometers, were highly effective in conveying the emotional state of the class. The participants found the barometers to be highly intuitive and that they facilitated timely feedback. Moreover, the evaluations indicate that there is a preference for the barometers over traditional methods such as bar charts and line graphs as seen in the first evaluation. Following this, it is important to choose visualizations that are easy to analyze and provide sufficient information. It might also encourage more exploration into other fields for inspiration when choosing what visualizations to use. Although the visualizations I chose seem effective, some issues were brought up that should be addressed to improve the dashboard further. This includes strategies to mitigate information overload and ensure that complex constructs like emotional regulation are communicated clearly.

5.2 Guidelines and Key Considerations

The goal of the thesis was to develop a real-time learning analytics dashboard. In this part of the discussion, I will elaborate on some of the key factors that contributed to the successful development of the dashboard.

Firstly, having a preliminary design phase where you receive feedback from relevant stakeholders is crucial. The effect of having this phase allows for early identification of potential problems and issues with your designs and core functionality. Four preliminary mock-ups were made for this dashboard; however, it could have been beneficial to have had more. The reason for this is that there were multiple indicators found in the literature review that were not represented in the designs. Having a greater selection of indicators across additional mock-ups could have an impact on the final design of the system. Furthermore, it could also be worth exploring having more diverse sets of features for each design, with the idea that this could avoid getting the same feedback on multiple designs.

When developing any large software system, there is bound to be a learning curve when using new technologies. In the context of my project, having some prior experience with related emotional technologies was beneficial, however, considerable time was spent to garner sufficient domain knowledge. For instance, I chose Py-Feat to perform facial emotion detection, however, an arguably more popular tool that is commonly used for this is OpenFace (Baltrusaitis et al., 2018). Even so, OpenFace being written in C++ and not having native Python bindings would mean additional overhead to implement in my system. This overhead would mean additional development time, and considering time-constrained development is a major part of every software development, I weighed the ease of use of Py-Feat as a determining factor. Informed decisions are crucial in such scenarios, where aspects such as ease of use, reliability, and performance must be weighed carefully to match the project requirements.

The choice of emotion recognition technology had significant implications for the system design. Specifically, the use of both wearable and facial detection devices meant that the system needed to handle two different input streams. To handle this appropriately, decoupling the system into distinct components for each device made it possible to reduce the data streams into the same data, the measurements, which were represented as ratios. This allowed a central component, the server, to handle the data the same way. Additionally, reducing the system into distinct components and keeping them decoupled made it simple to test each part in isolation.

Finally, there is the issue of data integrity and accuracy. When computing measurements that rely on physiological data, one must deal with the subjective biases of the user. As mentioned in section 3.3.1.2, these include the time of day, gender, physical health, and sleep. In my system, I minimize the effect of these biases by normalizing the data using a preliminary amount of data points. Furthermore, the data that is captured from the wristband sensor might contain noise. A simple, yet effective approach to deal with this issue is to use a running average of the last N data points, this will smooth out the data and minimize the effects of unwanted spikes caused by noisy data points.

5.3 Limitations and Future Work

To address the limitations of this study, I would like to start with the first interview process and SWOT analysis. The feedback from the interviews suffers from a gender bias. This results in a sample which is not fully representative of the teacher population. A more diverse sample of participants could have affected the overall feedback for the design mock-ups. As such, the resulting design of the dashboard might cater more to the preferences of this gender. In a future study, there should be a focus on gathering a more diverse group of participants for the evaluations. Additionally, the participants found certain aspects of the SWOT analysis, particularly the opportunity dimension, challenging to comprehend during the interview process. To mitigate this in the future, one could try providing explanations beforehand and allocating additional time to thoroughly explain the dimensions with additional examples.

The main limitation of this study is that the dashboard was not tested in classroom conditions. This means that factors like student hardware, network latency and server responsiveness during a real session were not tested. This is a natural focus for a future study.

In terms of future iterations of the dashboard, there is considerable room to add more features. One feature that could be added is an end of session summary. This could be a pop-up screen showing crucial moments in the lecture where a certain metric was in a critical state. This could aid the teacher's ability to reflect on what they did during that specific moment and what to improve on for future lessons. In a similar vein, a feature could be added to maintain a history of past sessions, this could be a valuable resource for the teacher to track trends in students' emotions over time. However, one must keep in mind the feedback about information overload and bloat, as such, any potential features must be evaluated thoroughly before implementation.

The findings of this thesis provide a good foundation for future iterations of the dashboard while also providing interesting insights for teacher dashboards in general. However, I acknowledge that this is a single study, and some of the broader implications must be understood in this context.

6 Conclusion

The goal of the thesis has been fulfilled; a real-time learning analytics dashboard that gives an overview of emotions has been developed using a DBR strategy. From the evaluation of the second cycle, the dashboard effectively provides the teacher with an overview of the emotional state of the class, and the chosen visualizations effectively facilitate this. The evaluations show that my dashboard provides a lot of useful features for the teacher that are not present in traditional LADs. A novel identification was how the dashboard allowed the teacher to identify "promotable behaviors", which could have implications for teacher dashboard designs in general. Issues that plagued the dashboard were both specific to my particular dashboard and more established obstacles from previous work. Specific to my dashboard was the construct of emotional regulation, which was unintuitive and must be addressed in future iterations. Well-known problems that are found in other dashboards such as information overload, data privacy, and data literacy, were also raised and need to be addressed.

Bibliography

Alfredo, R.-M., Mirko, A., Paula, R.-Z., & Eric, O. (2021). Positive emotions, engagement, and objective academic performance: A weekly diary study. *Learning and Individual Differences*, *92*, 102087. https://doi.org/10.1016/j.lindif.2021.102087

Anderson, T., & Shattuck, J. (2012). Design-Based Research:A Decade of Progress in Education Research? *Educational Researcher*, *41*(1), 16-25. https://doi.org/10.3102/0013189x11428813

Andrew, Z., Vitomir, K., Georg, G., Srećko, J., Gabrielle, J., David, G., & Abelardo, P. (2022). Team interactions with learning analytics dashboards. *Computers & Education*, *185*, 104514. https://doi.org/10.1016/j.compedu.2022.104514

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018, 15-19 May 2018). OpenFace 2.0: Facial Behavior Analysis Toolkit. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018),

Barab, S., & Squire, K. (2004). Design-Based Research: Putting a Stake in the Ground. *Journal of the Learning Sciences*, *13*(1), 1-14. https://doi.org/10.1207/s15327809jls1301_1

Cechinel, C., De Freitas Dos Santos, M., Barrozo, C. a. S., Jesiel Emerim, Vila, E. d., Ramos, V., Primo, T., Munoz, R., & Queiroga, E. M. (2021). A Learning Analytics Dashboard for Moodle: Implementing Machine Learning Techniques to Early Detect Students at Risk of Failure. 2021 XVI Latin American Conference on Learning Technologies (LACLO),

Chen, I. S. (2017). Computer self-efficacy, learning performance, and the mediating role of learning engagement. *Computers in Human Behavior*, *72*, 362-370. https://doi.org/10.1016/j.chb.2017.02.059

Chen, J., Zhu, B., Balter, O., Xu, J., Zou, W., Hedman, A., Chen, R., & Sang, M. (2017). *FishBuddy: Promoting Student Engagement in Self-Paced Learning through Wearable Sensing*

Chen, S., Ouyang, F., & Jiao, P. (2022). Promoting student engagement in online collaborative writing through a student-facing social learning analytics tool. *Journal of Computer Assisted Learning*, *38*(1), 192-208. https://doi.org/10.1111/jcal.12604

Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive Assessment of Students' Emotional Engagement during Lectures Using Electrodermal Activity Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(3). https://doi.org/10.1145/3264913

Dickler, R., Adair, A., Gobert, J., Hussain-Abidi, H., Olsen, J., O'Brien, M., & Pedro, M. S. (2021). *Examining the Use of a Teacher Alerting Dashboard During Remote Learning* Artificial Intelligence in Education,

Dissanayake, V., Elvitigala, D. S., Zhang, H., Weerasinghe, C., & Nanayakkara, S. (2019). *CompRate: Power Efficient Heart Rate and Heart Rate Variability Monitoring on Smart Wearables* 25th ACM Symposium on Virtual Reality Software and Technology , articleno = 16 , numpages = 8, https://doi.org/10.1145/3359996.3364239

Farahmand, A., Dewan, M. A. A., & Lin, F. (2020). Student-Facing Educational Dashboard Design for Online Learners. 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech),

Figma. "Figma, the collaborative interface design tool". https://www.figma.com/

Gupta, M., Lee, Rebelsky, Allesio, Murray, Woolf, Whitehill, Arroyo. (2021). *Affective Teacher Tools: Affective Class Report Card and Dashboard* Artificial Intelligence in Education, Han, J., Yang, C.-L., Chernyshov, G., Fu, Z., Horii, R., Narumi, T., & Kunze, K. (2022). *Exploring Collective Physiology Sharing as Social Cues to Support Engagement in Online Learning* 20th International Conference on Mobile and Ubiquitous Multimedia, https://doi.org/10.1145/3490632.3497827

Herodotou, C., Hlosta, M., Boroowa, A., Rienties, B., Zdrahal, Z., & Mangafa, C. (2019). Empowering online teachers through predictive learning analytics. *British Journal of Educational Technology*, *50*(6), 3064-3079. https://doi.org/10.1111/bjet.12853

Highsmith, J., & Cockburn, A. (2001). Agile software development: the business of innovation. *Computer*, *34*(9), 120-127. https://doi.org/10.1109/2.947100

Jeongyun, H., Kwan Hoon, K., Wonjong, R., & Young Hoan, C. (2021). Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. *Computers & Education*, *163*, 104041. https://doi.org/10.1016/j.compedu.2020.104041

Juan, Z., Lingyun, H., Shan, L., Susanne, P. L., Yuxin, C., & Cindy, E. H.-S. (2021). Self-regulation and emotion matter: A case study of instructor interactions with a learning analytics dashboard. *Computers & Education*, *161*, 104061. https://doi.org/10.1016/j.compedu.2020.104061

Kawamura, R., Shirai, S., Takemura, N., Alizadeh, M., Cukurova, M., Takemura, H., & Nagahara, H. (2021). Detecting Drowsy Learners at the Wheel of e-Learning Platforms With Multimodal Learning Analytics. *IEEE Access*, *9*, 115165-115174. https://doi.org/10.1109/ACCESS.2021.3104805

Lai, H., Ameli, N., Patterson, S., Senior, A., & Lunardon, D. (2022). Development of an electronic learning progression dashboard to monitor student clinical experiences. *Journal of Dental Education*, *86*(6), 759-765. https://doi.org/10.1002/jdd.12871

Lee-Cultura, S., Sharma, K., & Giannakos, M. (2023). MultiModal Teacher Dashboards: Challenges and Opportunities of Enhancing Teacher Insights through a case study. *IEEE Transactions on Learning Technologies*, 1-19. https://doi.org/10.1109/TLT.2023.3276848

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, *66*(1), 799-823. https://doi.org/10.1146/annurev-psych-010213-115043

Li, X., Bao, H., & Li, Y. (2021, 16-20 Dec. 2021). Understanding student engagement based on multimodal data in different types of agent-based collaborative learning contexts. 2021 Tenth International Conference of Educational Innovation through Technology (EITT),

Li, X., Yue, R., Jia, W., Wang, H., & Zheng, Y. (2021, 19-21 Nov. 2021). Recognizing Students' Emotions based on Facial Expression Analysis. 2021 11th International Conference on Information Technology in Medicine and Education (ITME),

Manuel, N., Simon, G., Kristian, K., Antero, L., Stefan, H., Elise, K., Hans-Otto, K., & Korbinian, M. (2019). Increased emotional engagement in game-based learning – A machine learning approach on facial emotion detection data. *Computers & Education*, *142*, 103641. https://doi.org/10.1016/j.compedu.2019.103641

Mohammad Nehal, H., Huyen, T. T. B., Thuy, T., Ho Tran, N., Gökhan, A., & Hiroshi, U. (2021). Students' emotion extraction and visualization for engagement detection in online learning. *Procedia Computer Science*, *192*, 3423-3431. https://doi.org/10.1016/j.procs.2021.09.115

Molenaar, I., & Knoop-van Campen, C. A. N. (2019). ow Teachers Make Dashboard Information Actionable. *IEEE Transactions on Learning Technologies*, *12*(3), 347-355. https://doi.org/10.1109/TLT.2018.2851585

Naif Radi, A., Ali, D., Rabeeh Ayaz, A., Jalal, S. A., Mohammad, B., & Muhammad Ahtisham, A. (2019). An integrated framework for course adapted student learning analytics dashboard. *Computers in Human Behavior*, *92*, 679-690. https://doi.org/10.1016/j.chb.2018.03.035

Natercia, V., Pavlo, A., Denis, V., Kara, D., Anne Corinne, H.-M., & Benjamin, B. (2021). The influence of task-value scaffolding in a predictive learning analytics dashboard on learners' statistics anxiety, motivation, and performance. *Computers & Education*, *173*, 104288. https://doi.org/10.1016/j.compedu.2021.104288

Ong, S.-C., & Chua, F.-F. (2022). ChemistLab: An Educational Game with Learning Analytics Dashboard. 2022 International Conference on Advanced Learning Technologies (ICALT),

OpenCV. *OpenCV*. Retrieved 05/03/2023 from https://opencv.org/

Ouatiq, A., Riyami, B., Mansouri, K., Qbadou, M., & Aoula, E.-S. (2022). Towards the Co-Design of a Teachers' Dashboards in a Hybrid Learning Environment. 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET),

Owatari, T., Shimada, A., Minematsu, T., Hori, M., & Taniguchi, R. i. (2020, 8-11 Dec. 2020). Real-Time Learning Analytics Dashboard for Students in Online Classes. 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE),

Pardo, A., Han, F., & Ellis, R. A. (2017). Combining University Student Self-Regulated Learning Indicators and Engagement with Online Learning Events to Predict Academic Performance. *IEEE Transactions on Learning Technologies*, *10*(1), 82-92. https://doi.org/10.1109/TLT.2016.2639508

Pham, L., Vu, T. H., & Tran, T. A. (2021, 10-15 Jan. 2021). Facial Expression Recognition Using Residual Masking Network. 2020 25th International Conference on Pattern Recognition (ICPR),

Pham, P., & Wang, J. (2018). *Adaptive Review for Mobile MOOC Learning via Multimodal Physiological Signal Sensing - A Longitudinal Study* Proceedings of the 20th ACM International Conference on Multimodal Interaction, https://doi.org/10.1145/3242969.3243002

Pozdniakov, S., Martinez-Maldonado, R., Singh, S., Chen, P., Richardson, D., Bartindale, T., Olivier, P., & Gašević, D. (2021). Question-driven Learning Analytics: Designing a Teacher Dashboard for Online Breakout Rooms. 2021 International Conference on Advanced Learning Technologies (ICALT),

Py-Feat. *Py-Feat Python Facial Expression Analysis Toolbox*. Retrieved 04/03/2023 from https://py-feat.org/pages/intro.html

Ramaswami, G. S., Susnjak, T., & Mathrani, A. (2019). *Capitalizing on Learning Analytics Dashboard for Maximizing Student Outcomes*

Ramos-Soto, A., Vazquez-Barreiros, B., Bugarín, A., Gewerc, A., & Barro, S. (2017). Evaluation of a Data-To-Text System for Verbalizing a Learning Analytics Dashboard. *International Journal of Intelligent Systems*, *32*(2), 177-193. https://doi.org/10.1002/int.21835

React. *"React, a javascript library for building user interfaces"*. Retrieved 2023/08/03 from https://react.dev/

Reeves, T. (2006). Design research from a technology perspective. *Educational Design Research*, 52-66.

Rohloff, T., Sauer, D., & Meinel, C. (2019). Student Perception of a Learner Dashboard in MOOCs to Encourage Self-Regulated Learning. 2019 IEEE International Conference on Engineering, Technology and Education (TALE),

Ruipérez-Valiente, J. A., Gomez, M. J., Martínez, P. A., & Kim, Y. J. (2021). Ideating and Developing a Visualization Dashboard to Support Teachers Using Educational Games in the Classroom. *IEEE Access*, *9*, 83467-83481. https://doi.org/10.1109/ACCESS.2021.3086703

Sannyuya, L., Shiqi, L., Zhi, L., Xian, P., & Zongkai, Y. (2022). Automated detection of emotional and cognitive engagement in MOOC discussions to predict learning achievement. *Computers & Education*, *181*, 104461. https://doi.org/10.1016/j.compedu.2022.104461

Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies*, *10*, 30-41. https://doi.org/10.1109/TLT.2016.2599522

Sharma, K., Lee-Cultura, S., & Giannakos, M. (2022). Keep Calm and Do Not Carry-Forward: Toward Sensor-Data Driven AI Agent to Enhance Human Learning

[Original Research]. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.713176
Shemwell, S. (2005). Futuristic decision-making. *Executive Briefing Business Value from*.
Shu, L. (2018). Student engagement in game-based learning : a literature review. http://hdl.handle.net/2152/65951
Sommerville, I. (2015). *Software Engineering* (t. Edition, Ed.). Pearson.
Subramainan, L., Mahmoud, M. A., Ahmad, M. S., & Yusoff, M. Z. M. (2016, 23-24 Aug. 2016). Evaluating students engagement in classrooms using agent-based social simulation. 2016 2nd International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR),
Taelman, J., Vandeput, S., Spaepen, A., & Van Huffel, S. (2009, 2009//). Influence of Mental Stress on Heart Rate and Heart Rate Variability. 4th European Conference of the International Federation for Medical and Biological Engineering, Berlin, Heidelberg.
van den Bergh, L., Ros, A., & Beijaard, D. (2013). Teacher feedback during active learning: Current practices in primary schools [10.1111/j.2044-8279.2012.02073.x]. *British Journal of Educational Psychology*, *83*(2), 341-362. https://doi.org/10.1111/j.2044-8279.2012.02073.x
Visted, E., Sørensen, L., Osnes, B., Svendsen, J. L., Binder, P.-E., & Schanche, E. (2017). The Association between Self-Reported Difficulties in Emotion Regulation and Heart Rate Variability: The Salient Role of Not Accepting Negative Emotions [Original Research]. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00328
Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, *53*(4), 5-23. https://doi.org/10.1007/BF02504682
Weihrich, H. (1982). The TOWS matrix—A tool for situational analysis. *Long Range Planning*, *15*(2), 54-66. https://doi.org/10.1016/0024-6301(82)90120-0
Werner, A. (2022). Pre-Project. *IT3915 Master in Informatics, Preparatory Project NTNU*.
Xiao, M., Zhou, J., & Xu, T. (2021). *Research on the Influence of Synchronous Interactive Class on Student Online Learning Engagement* 2020 12th International Conference on Education Technology and Computers, https://doi.org/10.1145/3436756.3437028
Xiao, X., & Wang, J. (2016). *Context and Cognitive State Triggered Interventions for Mobile MOOC Learning* Proceedings of the 18th ACM International Conference on Multimodal Interaction, https://doi.org/10.1145/2993148.2993177
Yaëlle, C., & Thomas, C. (2018). Specification and evaluation of an assessment engine for educational games: Empowering educators with an assessment editor and a learning analytics dashboard. *Entertainment Computing*, *27*, 209-224. https://doi.org/10.1016/j.entcom.2018.07.003
Yannakakis, G. N., Hallam, J., & Lund, H. H. (2008). Entertainment capture through heart rate activity in physical interactive playgrounds. *User Modeling and User-Adapted Interaction*, *18*(1), 207-243. https://doi.org/10.1007/s11257-007-9036-7

Appendices

NTNU

Norwegian University of
Science and Technology