Understanding Collaboration in Virtual Labs:

A Learning Analytics Framework

Hanna Birkeland

Supervisor: Barbara Wasson
Co-supervisor: Mohammad Khalil

Master's Thesis
Department of Information Science and Media Studies
University of Bergen

May 31, 2022

Acknowledgements

I would like to thank my supervisor Barbara Wasson and my co-supervisor Mohammad Khalil for their invaluable guidance, feedback and expertise.

Also, big thanks to all the participants that contributed to this project for your time and valuable feedback.

Thank you to my friends and family for all the support and motivation. A special thank you to my fellow master students at lesesal 644; Hedda Bakken, Lars Holth and Gina Farsirotos. Thank you for all the laughs and for keeping me motivated throughout this thesis.

Last but not least, thank you to my partner and best friend Magnus Sture for all his love and support, and for always being there after long days of writing.

Abstract

Online education is increasing and progress within technology has inspired the development of virtual laboratories, which allow students to conduct experiments online. One of the main challenges of virtual laboratory environments is facilitating collaboration similar to those existing in physical laboratory settings. This research explores how one can obtain a better understanding of collaboration in virtual labs through the use of learning analytics.

The research work of this thesis was carried out within the frame of design science research, where the main contribution is an artefact in the form of a learning analytics framework. The aim of the artefact is to provide a guiding framework for the integration of learning analytics to better understand and support learning and collaboration in virtual labs. The artefact was evaluated in two iterations using semi-structured interviews with seven experts.

It was found through the artefact development process that social network analysis, statistical analysis, natural language processing, and sentiment analysis are valuable data analysis methods for identifying patterns within collaboration in virtual labs. A proposal of a learning analytics dashboard has proved to be a valuable tool to visualise the analysis to the stakeholders in question (students and instructor). The overall reception of the framework was understandable and well-presented.

The contribution of this research provides opportunities for future work which involves putting the framework into practice. The implementation of learning analytics to support collaboration in virtual labs can make it easier for students to reflect on their own performances and thereafter improve from it, as well as supporting instructors to reflect on their teaching methods and provide assistance to students in need.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The importance of technology in education is increasing. The momentum has been even larger after the outbreak of the COVID-19 pandemic when typical in-person classroom teaching was suspended worldwide and forced the educational systems to move into virtual environments. In biosciences, where laboratories are at the core of undergraduate education, finding a feasible way of conducting laboratory work in a virtual space is challenging, yet crucial. The challenge is represented by replicating hands-on exercises and teamwork online to meet the pedagogical standards of university discourses and the desired outcome of such exercises. The progress within technology and communication networks has made it possible to develop virtual and remote laboratories that allow students to conduct experiments online and find a way around the limitations of physical laboratories (Alkhaldi et al., 2016). Collaboration and teamwork are common practice in physical laboratories (Teng et al., 2016). With the increase in teaching and learning online, the challenge of facilitating cooperative learning emerges in these online environments, and learning analytics can be used to better understand learning performances during online laboratory collaboration. Learning analytics may offer students and teachers insight into the interactions within a group. Such information benefits teachers to facilitate their teaching to each group, and students to self-reflect during collaboration.

1.1 Motivation

This research is a part of an ongoing European Erasmus+ project at the Centre for the Science of Learning Technology (SLATE) named European Network for Virtual lab Interactive Simulated Online learning 2027 (ENVISION_2027). The objective of the project is to innovate new e-learning course modules, which will provide the students with the possibility to carry out laboratory exercises online and engage actively in teamwork in digital environments.

This research is further motivated by the need for transitioning from conventional classroom teaching into digital remote teaching due to the pandemic, and the rapid development of communication technology within education. As collaboration and teamwork is common practice in physical laboratories, the need to explore how to improve the learning outcomes of collaboration within a virtual lab environment is present.

1.2 Research Problem

The aim of this research is to provide and evaluate a framework for the integration of learning analytics to better understand and facilitate learning performances and collaboration in virtual labs. The learning analytics framework intends to describe how learning analytics can better support digital learning for students of higher education with an example from the biosciences.

Based on the purpose of this research, the research questions are defined as follows:

RQ1: What is the current state of research on the use of learning analytics to understand and support collaboration in virtual labs?

RQ2: How can learning analytics support collaboration in virtual labs?

RQ3: What aspects are important when designing learning analytics implementations for collaboration in virtual labs?

1.3 Thesis outline

The thesis is organised as follows:
Chapter 2: Literature Review presents relevant theory and a literature review.
Chapter 3: Methodology describes the methods used in this thesis, which are guided by a design science research methodology.
Chapter 4: Artefact Development presents the artefact development process, including two iterations of development and evaluation.
Chapter 5: Final Artefact presents the final artefact.
Chapter 6: Discussion presents a discussion of the research and answers to the

research questions.

Chapter 7: Conclusion provides some conclusions and ideas for future work.

Chapter 2

Literature Review

This chapter presents the relevant literature related to the relevant research questions[1]. First, selected background theory is presented to provide an understanding of the concepts relevant to learning analytics, collaboration and virtual laboratories. Next, a systematic scoping review that reviews the state of the art of the current research on the use of learning analytics in virtual labs and collaboration, is presented. The focus of the systematic scoping review was to map current trends and challenges in order to decide which factors to consider when developing a learning analytics framework for collaboration in virtual labs.

2.1 Learning Analytics

The field of learning analytics is a growing area of technology-enhanced learning research. It has been defined by the Society for Learning Analytics Research (SoLAR) as *"the measurement, collection, analysis and reporting of data about learners and their context, for purposes of understanding and optimising learning and the environments in which it occurs"* (SoLAR, 2021). Learning analytics provides information about learning behaviour allowing students to reflect on their own performances, and teachers and tutors to assist based on each student's performance. As learning analytics provide clear benefits for teachers and learners, it is also beneficial for other groups. At the administration level, it can assist in the evaluation of institutional resources and educational offers, as well as being in the interest of researchers who are developing and evaluating data mining techniques for educational issues (C. Romero & Ventura, 2013).

[1] Part of this chapter has been published in the companion proceedings of the Learning Analytics & Knowledge conference 2022.

There are several factors within technology, education and politics that have motivated the emergence of the learning analytics field (Ferguson, 2012). Within technology, big data is a growing challenge. The term refers to large datasets of complex structures with difficulties of capturing, storing and analysing (Sagiroglu & Sinanc, 2013). This data provides useful information and deeper insights in business settings. Such large datasets are also generated within virtual learning environments, including vast amounts of interaction data. This raises the technical challenge of how to extract value from these big learning-related datasets. Within education, the increase in online learning raises the challenge of how to optimise the opportunities for learning online (Ferguson, 2012). As it offers benefits, it also paves way for problems. Students may experience feelings of isolation, technical issues or motivational loss, and teachers might struggle to identify these students. Within the political factor, there is a growing demand for educational institutions to measure and improve performance in learning. This raises the challenge of how to optimise learning at a national or international level.

2.1.1 Learning Analytics Life Cycle

Khalil & Ebner (2015) proposed a Learning Analytics Life Cycle which presents the processes that are involved in learning analytics. Visualised in Figure 2.1, the cycle consists of four main sections: *learning environments*, *big data*, *analytics* and *act*.

Learning Environment
The learning environment may be an online course, a learning management system or any virtual learning environment (Khalil & Ebner, 2015). These environments are where the stakeholders produce data. The different stakeholders which are engaged in learning analytics include learners, instructors, researchers and educational institutions, each with different objectives. The objective of learners is to personalise their online learning and strengthen their learning performances. For instructors, the intention is to promote real-time feedback to their students and enhance their teaching. The objective of researchers is evaluating and improving courses, and discovering new ways in which to deliver educational information for instance through visualisations. Lastly, for educational institutions, the aim is to accomplish higher educational goals by supporting decision-making processes.

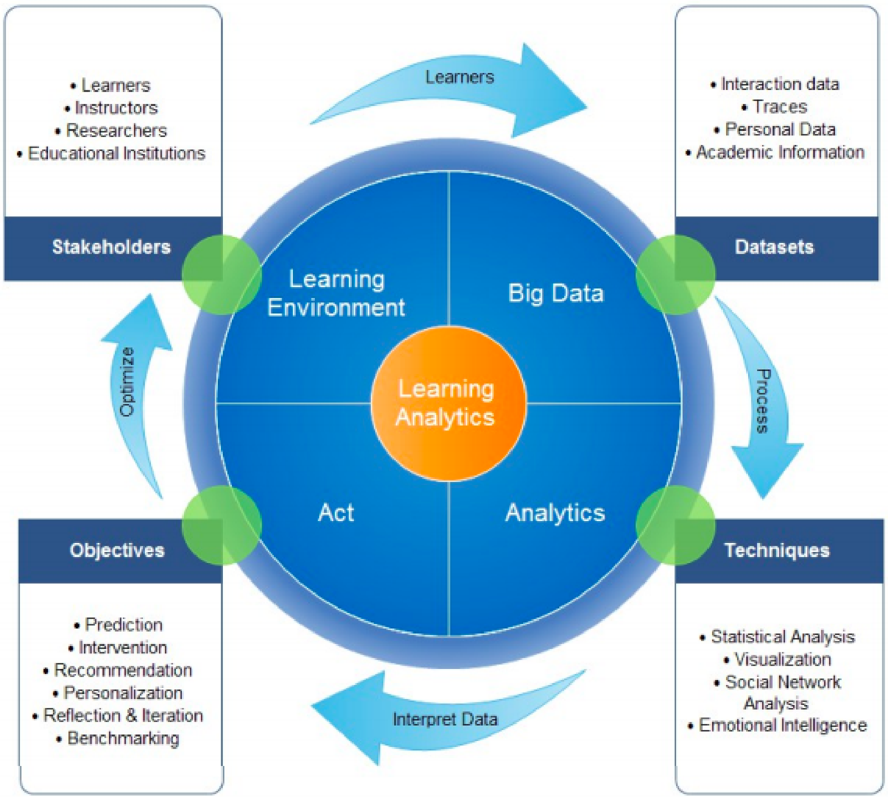


Figure 2.1: The Learning Analytics Life Cycle proposed in Khalil & Ebner (2015).

Big Data

The big data section involves the data which have been generated by learners within the learning environments. The different data processed in an educational learning environment includes, but are not restricted to interaction data, traces, personal data and academic information (Khalil & Ebner, 2015). Interaction data refers to data related to discussion forums and visualisations. Traces can be data regarding mouse clicks, number of logins, number of completed assignments, which documents have been accessed or questions asked. Personal data include name, birthdate, ID or any personal information. Lastly, academic information can be data about which courses have been attended, exams taken, grades and date of graduation.

Analytics

The data then needs to be analysed in order to retrieve meaningful information from it. This takes place in the analytics section of the cycle. The various analytical techniques used within learning analytics fall into two main categories: *quantitative* and *qualitative* analytics (Khalil & Ebner, 2015). *Quantitative* analysis deals with numbers and statistics, and consists of several subcategories for learning analytics quantitative methods. Firstly, there is statistical analysis. This type of analysis involves numerical computations and mathematical operations related to traces. Such analysis can for instance be calculation of time spent on different tasks or analysis of mouse clicks. Secondly, there are visualisations. This type of analysis enables interpretation of efforts in a learning environment where statistical information can be illustrated through charts, heat maps, evaluation models and diagrams. It is however stated that there is a lack of certainty within learning analytics as to what specifically needs to be measured in order to get a higher understanding of the learning progress. Visualisation techniques can regardless join information with meaning and help stakeholders in decision-making. Learning analytics dashboards are a type of learning analytics visualisation which are widely accepted as it offers an easy insight and provides better visibility of the analysis (Khalil & Ebner, 2015). A learning analytics dashboard can be defined as a display of analytics that reflects students' interactions, patterns of learning, performance and status through visualisations of elements such as charts and graphs (Park & Jo, 2015). Lastly, there is quantitative social network analysis. This type of analysis focuses on the relationships between different entities and allows detailed examination of networks consisting of entities and relations between them (Khalil & Ebner, 2015).

Qualitative analysis methods involve the processing of data in order to obtain more explained descriptions (Khalil & Ebner, 2015). It contains two subcategories: emotional intelligence and qualitative social network analysis. Emotional intelligence considers emotions, and can be categorised into positive and negative emotions. Emotional intelligence can help detect the levels of wellness within courses (Atif et al., 2013). Qualitative social network analysis can include analysis of interviews, surveys and observations, taking the form of virtual ethnography. Virtual ethnography is research aiming to explore social interactions taking place in a virtual environment (Given, 2008). Such interactions can take place in web-based discussion forums or chat rooms.

Act

The next stage in the cycle involves the interpretation of the analysis in order to achieve the objectives of learning analytics (Khalil & Ebner, 2015). The objectives consist of *prediction*, *intervention*, *recommendation*, *personalisation*, *reflection and iteration*, and *benchmarking*. The aim of *prediction* is the investigation of unknown numerical values such as scores and grades, in order to reveal learner activities and future performances. In doing so, appropriate *interventions* could be carried out by instructors. Appropriate interventions could help prevent dropouts and better students' success by providing assistance to those who may need it. The goal of *recommendation* in learning analytics is to provide students with recommendations based on their activities. This can be to recommend a relevant course or a book. The intention behind *personalisation* is to support learning by personalising online learning based on the students' needs and ability. *Reflection and Iteration* involves the evaluation of previous work in order to improve future learning, an iteration which can help all stakeholders in the learning analytics cycle. Lastly, there is *benchmarking* which aims to identify the best practices in order to improve performances. This also contributes to the discovery of the weak practices in learning.

2.1.2 Learning Analytics in Higher Education

In higher educational settings, the use of learning analytics are increasing. Leitner et al. (2017) present an overview of current trends of learning analytics in higher education. The results showed that Massive Open Online Course (MOOC) platforms are a popular educational environment within learning analytics research. Other key areas proved to be the enhancement of students' learning performances and behaviours, observing potential dropouts, and the use of learning analytics to perform interventions. Several limitations were discovered, which included time restraint on the existing research, the size of the datasets, and the size of the group in question. Lastly, ethical limitations concerning privacy and data ownership were discovered. The stakeholders identified seemed to be mainly researchers. The most common learning analytics methods used within the higher education domain were found to be prediction, distillation of data for human judgement, and outlier detection for discovering students at risk of dropping out (Leitner et al., 2017).

2.1.3 Learning Analytics Constraints

There are different challenges related to privacy and ownership that emerge when implementing learning analytics on educational data. Khalil & Ebner (2015) introduce an eight-dimensional figure of constraints presented in Figure 2.2, that can constrain the processes of learning analytics. Reconditions of these constraints have been updated in Prinsloo et al. (2022) and Slade et al. (2019).

Privacy
There are potential privacy issues regarding learning analytics and the analysing of student data (Khalil & Ebner, 2015). Through data analysis, personal information about students' attitudes and activities can be revealed. Also, in terms of the

objective of prediction in learning analytics, an instructor can point out students at risk, which may lead to the problem of labelling where a student is labelled either as good or bad. Anonymisation may therefore be required in order to preserve sensitive information about students.

Access

Specific authentication should ensure that only appropriate users are permitted to access specific data (Khalil & Ebner, 2015). There should be different levels of access for different learning analytics stakeholders in order to maintain students' privacy. Students should have access to view and update their own data. Teachers should also have access to students' data, not including sensitive information such as ethnicity. Decision makers should have access to data that meets the institutional perspective of preventing high dropout rates.

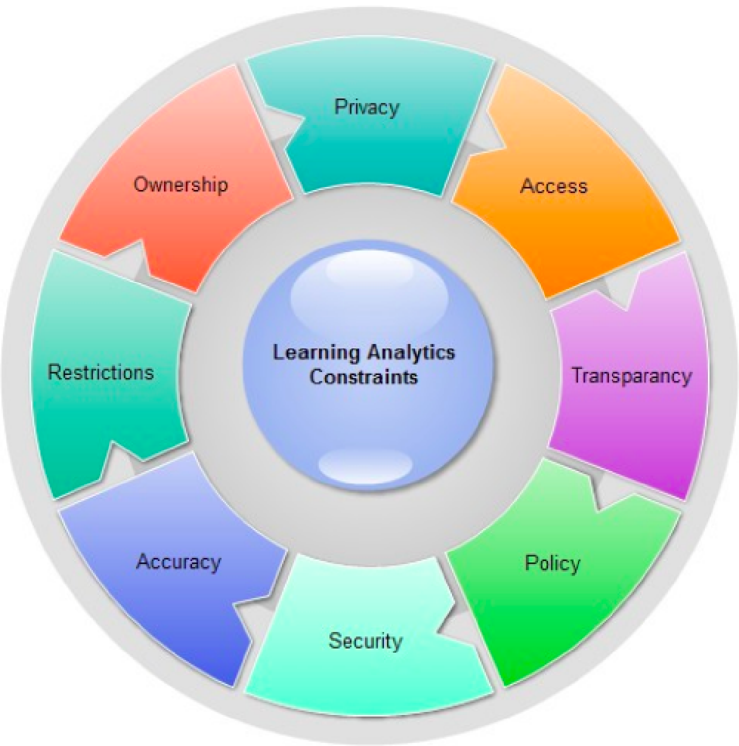


Figure 2.2: Learning Analytics Constraints from Khalil & Ebner (2015).

Transparency

The use of learning analytics should aim to be transparent (Khalil & Ebner, 2015). As a way of ensuring transparency, the institution can provide information about data collection and the usage of that data. Students may wish to know how much data is collected, how their performance is being traced, and how evaluations and interventions are handled based on this data.

Policy

Institutions adopting learning analytics must adjust their policies so that they reflect privacy and the ethical implications of the use of learning analytics (Khalil & Ebner, 2015). Khalil & Ebner (2015) presents a number of regulations that serve as an ethical learning analytics policy:

- Gathering of personal information such as date of birth, sex, address, ethnicity, occupational status, study records and qualifications.
- A description of the usage of this information, whether it is for research reasons to achieve the objectives of Learning Analytics, or for the benefit of the students, which could be the prediction of student behaviour and to provide recommendations and advice based on Learning Analytics.
- Methodology of the data gathering either by the input from the student or by other services, such as browser cookies.
- Security principles for protection of the data.
- A description of how long the student data is stored and the process of deletion.

Security

In order to maintain the safety of the students' records and analysis results, learning analytics tools should adhere to appropriate security principles (Khalil & Ebner, 2015). A widely used model that can help guide security issues is the CIA triad, which stands for Confidentiality, Integrity and Availability (Andress, 2014). Confidentiality refers to the protection of data from unauthorised access. Integrity refers to the prevention of data being altered in an unauthorised manner. Availability refers to the data being available for authorised parties to access when it is needed. In terms of learning analytics, the information of students should not be available to unauthorised parties (Khalil & Ebner, 2015). This can be achieved through encryption, which can guarantee access to only authorised personnel.

Accuracy

Learning analytics should aim to guarantee that the selection of data and its analysing generate a solid level of accuracy (Khalil & Ebner, 2015). The wrong selection of datasets could affect the accuracy and lead to learning analytics results being inaccurate. This would affect the performance of the learning analytics objectives such as predictions and interventions.

Restrictions
Legal restrictions such as copyright laws and data protection confine the benefits of employing learning analytics (Khalil & Ebner, 2015). These restrictions involve limitations for how long data can be stored, keeping the data secure from internal and external threats, the usage of data only for specific purposes and requirements for results to be as accurate as possible.

Ownership
The two main perspectives of who owns the data in the educational setting suggests the students and the institutions (Khalil & Ebner, 2015). It has been proposed by Jones et al. (2014) to unite these two perspectives so that students and the institution share the ownership, which would support the institution's needs for data and the students' needs for privacy. Institutions could utilise students' data for analysis in order to personalise the learning platforms while respecting students' rights, ensuring that student information is kept confidential.

2.2 Collaborative Learning

Collaborative learning can be defined as a *situation* in which *two or more* people *learn* or attempt to learn something *together* (Dillenbourg, 1999). This is, however, a broad definition which may be interpreted in various different ways. The element of *two or more* may include a pair, a small group of 3-6 people, a class of students or a community or society, which can range from hundreds to millions of people. The element of *learning* may include the process of problem solving, studying course materials, or learning from experience through life. The element of *together* involves different kinds of interactions, which can be in person or through the use of computers, or how the work is divided and whether the efforts are joint.

A collaborative learning situation can include various kinds of contexts and interactions. The effects of collaborative learning should therefore not be measured in general, but based on the specific interactions in which the collaborators were engaged (Dillenbourg, 1999). As the improvement of students' performances in collaborative situations is valuable for any educational institution, assessments of group performances are essential. In the context of gaining a better understanding of collaborative interactions, computer supported collaborative learning (CSCL) environments are valuable tools as they enable recordings of interactions. CSCL refers to situations where collaborative learning takes place with the assistance of computers (Stahl et al., 2006). The technology involved in the CSCL context intends to be fundamentally social, meaning the technology should encourage social interactions and collaborative learning, leading to individual learning. The unique advantage of information technology to gather and analyse interaction data can facilitate collaborative learning and offer specified guidance.

2.3 Virtual Labs

There has been an increase in interest in the development of technologies for laboratories such as virtual labs as they offer unique opportunities for students (Alam et al., 2014). Removing the barrier of time and location, it allows students to conduct experiments remotely, without having to be physically present at the laboratory. Alkhaldi et al. (2016) has classified three different categories of labs; *physical labs*, *remote labs* and *virtual labs*. *Physical labs*, also known as hands-on labs, are the traditional lab environments where students physically conduct experiments in a laboratory. In *remote labs*, experiments are conducted in a physical lab, located away from the experimenter, and the experimenter is connected to the physical lab remotely through the network. A *virtual lab*, also known as a simulated lab, is a simulation of a laboratory environment, allowing students to conduct experiments in a virtual space. The lab in question for this thesis is the virtual lab. However, remote labs are also relevant as they involve the conduction of experiments online, and will be included in the literature review. The common denominator for the terms virtual labs and remote labs will hereby be denoted as online labs.

2.3.1 Examples of Virtual Labs

There are variants of virtual labs out there. Some of these are presented here in order to provide a better understanding of what a virtual lab may involve.

Gizmos

Explore Learning Gizmos is a virtual lab which offers interactive simulations for students in the fields of maths and science, for secondary and higher education[2]. They offer a library of interactive STEM (science, technology, engineering and math) cases which allows the student to take the role of a professional trying to solve a real-world problem. Gizmos offer real-time data of students' results, allowing teachers to follow the students' progress. An example from a stoichiometry STEM case is presented in Figure 2.3, where the students are to investigate the

[2]Gizmos. https://gizmos.explorelearning.com/ (accessed 20.04.2022).

source of a legionella bacteria which has led to an outbreak of Legionnaires' disease. They are to use stoichiometry to disinfect the water supply and stop the outbreak of the disease.

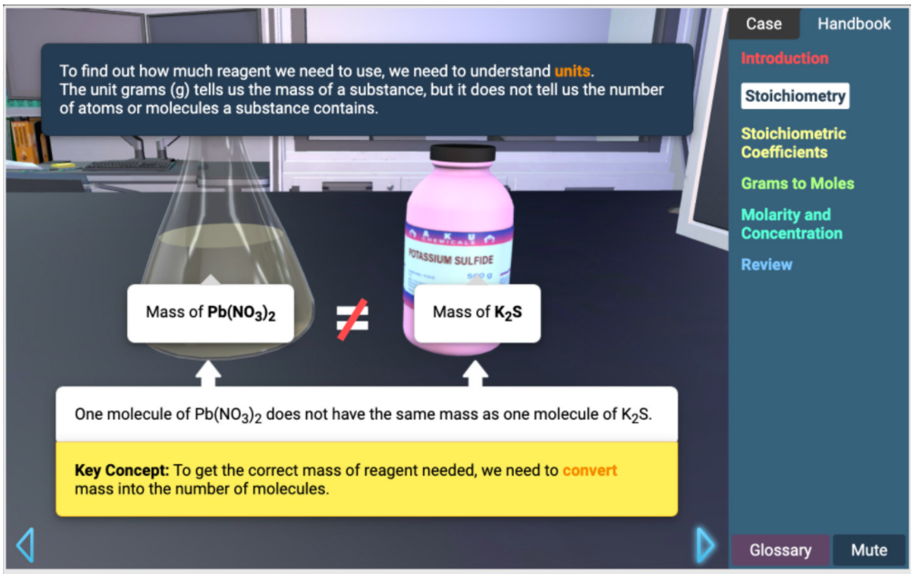


Figure 2.3: Screenshot from Gizmos - a Water Crisis: Stoichiometry STEM case.[3]

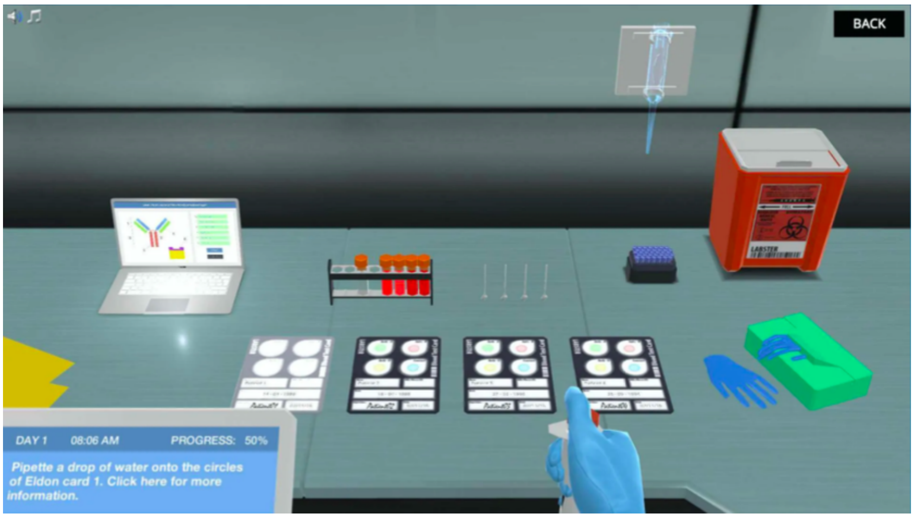


Figure 2.4: User interface of the Labster simulation "About Antibodies: Why are some blood types incompatible?"[4]

Labster

Labster is another virtual lab which offers interactive lab simulations for several sections within biosciences, including biology, chemistry, engineering, general sciences, medicine and physics[5]. Each simulation includes an AI assistant to guide

[3]Gizmos. Water Crisis: Stoichiometry [Screenshot]. https://gizmos.explorelearning.com/index.cfm?method=cResource.dspDetail&interactivecaseID=20 (accessed 20.04.2022).
[4]Labster. Antibodies: Why are some blood types incompatible? [Image]. https://www.labster.com/simulations/antibodies/ (accessed 20.04.2022).
[5]Labster. https://www.labster.com/about/ (accessed 20.04.2022).

the students, and quiz questions along the way. The teachers are provided with students' performance data. Figure 2.4 shows an example of a virtual lab session within biochemistry dealing with the concepts of antibodies. The session is called "About Antibodies: Why are some blood types incompatible?", and allows the student to examine the blood samples from a mother and her unborn child, in order to determine whether they are compatible or not.

Inq-ITS

Inq-ITS (Inquiry Intelligent Tutoring System) is a virtual science lab which offers a collection of labs for secondary school[6]. These labs include general inquiry labs where students practise individual science investigation skills; physical science labs which include topics like energy and particles; life science labs which include topics like genetics and cell health; and earth science labs which include topics like continental plate boundaries and orbital patterns. Inq-ITS uses algorithms that generate real-time student performance reports for teachers, similar to the aforementioned labs. Teachers are also provided with alerts as to which students might need assistance. This information can help teachers provide relevant feedback to students. The labs are also equipped with an AI virtual tutor which helps guide students as they work through the labs. The processes and actors involved in the Inq-ITS labs are presented in Figure 2.5.

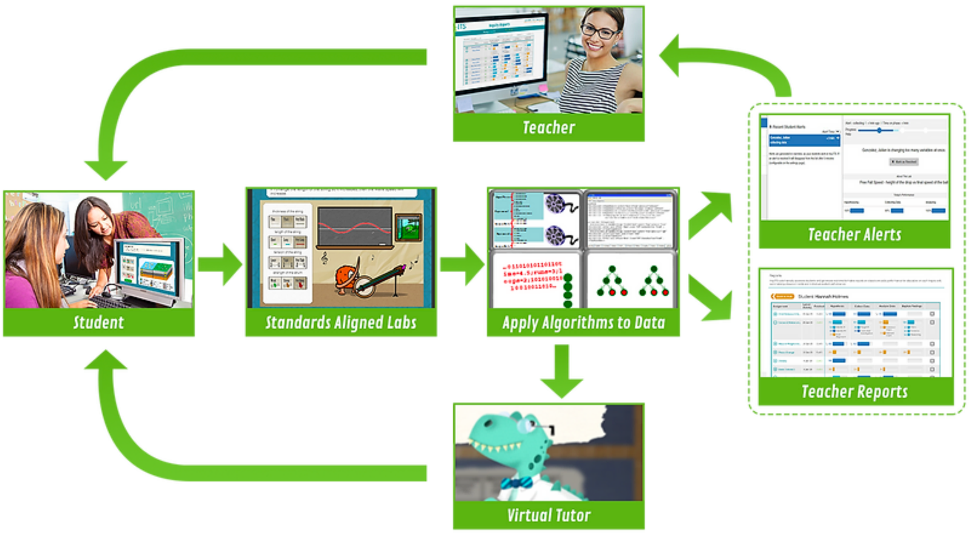


Figure 2.5: Inq-ITS in Action.[7]

[6]Inq-ITS. https://www.inqits.com/labs (accessed 20.04.2022).
[7]Inq-ITS. Inq-ITS in Action [Graph]. https://www.inqits.com/about (accessed 20.04.2022).

2.3.2 Effects of Online Labs

Various research has been conducted seeking to demonstrate what effect such laboratories have on learning outcomes (Finkelstein et al., 2005; Wiesner & Lan, 2004). Alkhaldi et al. (2016) elaborate on the findings within some of this research, observing diverse results. Some of these studies report no distinction in the performance of students between online and physical labs, whereas in others, online labs have an advantage over physical labs in terms of gaining conceptual knowledge. Alkhaldi et al. (2016) summarise that other studies show that the combination of both physical and online labs gives better results in student performance than that of solely physical labs. It is found that online labs have many benefits, although they may not replace physical labs. Integrating online labs with physical labs can lead to rich learning environments, if utilising a solid pedagogical framework that supports learning. What seems to be lacking within the area of online labs, is the research on collaboration within such environments (Alkhaldi et al., 2016). This finding supports the need and motivation for further research within this area.

2.4 Scoping Review (Birkeland et al., 2022)

A literature review has been conducted to provide a theoretical framework for this research, and to inspire the development of a learning analytics framework[8]. The review methodology used is a scoping review. Scoping reviews are valuable for providing an overview of existing literature on a given topic, giving clear insight into the volume and nature of the available literature (Peters et al., 2015). The use of this method is therefore particularly beneficial for when a certain topic has not been thoroughly reviewed, consequently making it suitable for this research, as the current research on the topic in question is limited. The value of a scoping review is the broad perspective it can provide on a specific topic, compared to a systematic review which is designed to answer a more narrow research question based on specific study settings (Peters et al., 2020).

The database used for the literature review is Web of Science, as it is a high-quality database providing access to multiple other databases. The established search string is *(learning analytics) AND (virtual lab* OR online lab* OR digital lab**

[8]This chapter has been published as a poster in the companion proceedings of the Learning Analytics & Knowledge conference 2022 as follows: Birkeland, H., Khalil, M., & Wasson, B. (2022). Learning Analytics in Collaborative Online Lab Environments: A Systematic Scoping Review. Companion Proceedings of the 12th, 92.

OR remote lab).* The search was performed on the 16th of November 2021, and resulted in 419 articles in addition to 2 other articles found through citations.

Initially, a second search was performed with the search string *(learning analytics) AND (virtual lab* OR online lab* OR digital lab* OR remote lab*) AND (collaborat* OR team*).* This search yielded less results, specifically 73, and no additional papers were identified. Thus, the results from the first search were used.

Screening of the 421 articles was carried out on the following inclusion criteria:

1. Written in English
2. Within the time period 2011 - present
3. The context is online labs or online learning environments
4. The concept is understanding collaboration between students through learning analytics
5. The participants include teachers, students, lab assistants, researchers or lab facilitators
6. Types of evidence sources include all study settings

The inclusion criteria context includes both online labs and online learning environments to capture the apparent knowledge gap identified in the existing research. The context involving other online learning environments were limited to the scope of collaboration scenarios expected to be applicable to virtual lab environments. The knowledge gap is further discussed in Chapter 6.

The process of identifying the relevant literature is represented in the PRISMA-flowchart in Figure 2.6. The query search gave a result of 419 articles to be screened, in addition to the 2 other articles found through citations. The results were filtered by papers published from 2011 to present day, and papers in English. Removing duplicates was not needed as the papers were retrieved from a single database. After filtering, the title, abstract and keywords of 411 articles were screened on the inclusion criteria, resulting in 26 articles to be further screened on the full-text reading. The final screening resulted in 11 articles to be included in the literature review. These 11 articles are presented in Table 2.1.

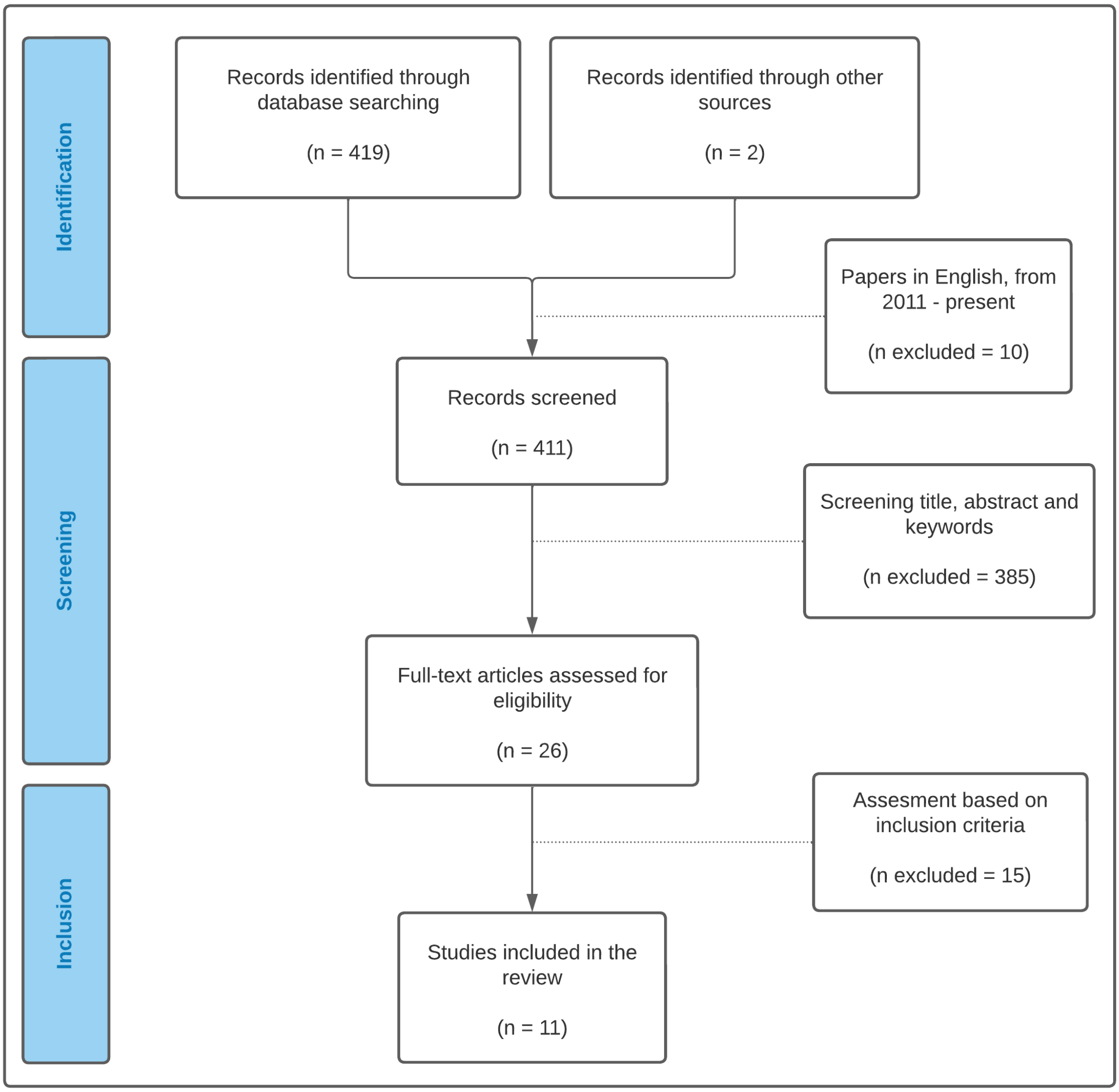


Figure 2.6: PRISMA flow diagram.

Table 2.1: List of articles used in review.

Year	Title	Citation
2021	Improve teaching with modalities and collaborative groups in an LMS: an analysis of monitoring using visualisation techniques	(Sáiz-Manzanares et al., 2021)
2021	Evaluating the efficiency of social learning networks: Perspectives for harnessing learning analytics to improve discussions	(Doleck et al., 2021)
2020	Investigating Collaboration as a Process with Theory-driven Learning Analytics	(Kent & Cukurova, 2020)
2020	Detecting the Depth and Progression of Learning in Massive Open Online Courses by Mining Discussion Data	(Pillutla et al., 2020)
2019	What information should teacher dashboards provide to help teachers interpret CSCL situations?	(van Leeuwen et al., 2019)
2016	Current and Future Developments in Remote Laboratory NetLab	(Teng et al., 2016)
2015	Design of Virtual Learning Environments: Learning Analytics and Identification of Affordances and Barriers	(Qvist et al., 2015)
2015	Automatic Assessment of Progress Using Remote Laboratories	(S. Romero et al., 2015)
2015	Using Learning Analytics to Visualise Computer Science Teamwork	(Tarmazdi et al., 2015)
2015	Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory	(Xing et al., 2015)
2014	Leveraging Non-explicit Social Communities for Learning Analytics in Mobile Remote Laboratories	(Orduña et al., 2014)

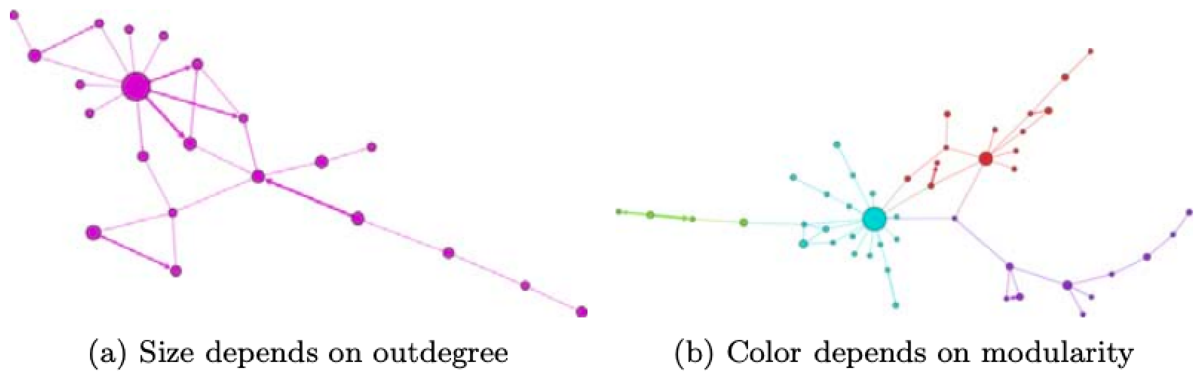
Once the papers were identified, they were analysed and categorised through topics which emerged. These papers are presented in the following subsections which are divided into the identified topics, these being Learning Analytics in Online Labs, Learning Analytics in Discussion Forums, Learning Analytics Visualisations and Learning Analytics with Human Computer Interaction (HCI) Theory.

2.4.1 Learning Analytics in Online Labs

Learning analytics techniques allow processing of student activity data to gain an understanding of learning performances within online learning environments, including environments like online labs. S. Romero et al. (2015) show an example of this in the Weblab-Deusto Remote Laboratory platform, using a VISIR remote experiment. The VISIR remote experiment allows practice of digital and analog electronics. A software layer over the Weblab-Deusto platform registers great amounts of data regarding the students' interactions with the experiment, recording clicks and traces within the system and storing it in a database. This data includes the time and duration of each experiment, and errors occurring. This provides insight into the process of each student's performance, and not only the final result. This allows teachers to facilitate their learning based on each student's needs. The analysed data were presented to both teacher and student through a software showing differences and similarities between the exercises performed by the student compared to the teacher's execution to show a summative evaluation. It is stated that such data can be processed for both individual students and groups of students, however this study focuses primarily on the performance of the individual student.

Qvist et al. (2015) present a similar example of employing learning analytics in the LabLife3D virtual laboratory environment, where they store data of student mouse clicks and time spent on tasks. The analysed data are presented through timelines of data trails from the executed experiments, enabling teachers to identify occurring errors and students to reflect on their learning process. Students found that seeing a visualisation of their own performances was interesting and supported them in focusing on future learning activities. Some students were also motivated by seeing how they had done compared to other students. However, other students could feel threatened by the comparison among students as it made the exercise feel too competitive. The analysis is somewhat limited by the sample size and data set being considerably small for quantitative analysis. These laboratory experiments also do not yet provide collaboration amongst students as it focuses primarily on the individual student.

The collaboration of students has been further investigated by Orduña et al. (2014) in the same Weblab-Deusto remote laboratory platform as in S. Romero et al. (2015), demonstrating a method of extracting social networks from datasets not fitting for such extractions. Students individually test experiments in the remote lab, specifically a Complex Programmable Logic Device (DPLD) remote lab. The students compile code on their own computer, then submit their file to the remote device through Weblab-Deusto. The platform does not store data of social interaction between students, so in order to gain knowledge of social connections outside of the system, data about the uploaded files were used for social network analysis. The uploaded files are compared to check for similarities in code, name and time of compilation. They have used an approach for data collection in sociocentric networks, a method within social network analysis which focuses on the interactions between students within a network. The results are presented in a directed graph with relations between students, as shown in Figure 2.7(a), displaying which files are shared among them and who are the bigger sources. By interpreting the degree centralities of each node in the directed graph, teachers can gain insight into the roles of each student. Students with a high outdegree work and solve the assignment and share the file with others, whereas students with high indegree receive those shared files. This information can help teachers identify the students in need of help, this being the students with high indegree. Additionally, the social network is presented in a modularity graph, as shown in Figure 2.7(b), displaying different communities in the network. This representation can assist the teacher in discovering which students enjoy working together, guiding the teacher in the case of forming groups.



(a) Size depends on outdegree (b) Color depends on modularity

Figure 2.7: Representations of group networks in Orduña et al. (2014).

There are few remote laboratory environments that allow collaboration within the system, however, Teng et al. (2016) developed one that does. The NetLab remote laboratory allows students to use the system at the same time as other students and provide them with a built-in chat window for communication. Additionally, all actions made by students online are reported and broadcasted in its own window. Students express a strong satisfaction with the ability to collaborate with other students from anywhere in the world. The system records the students' actions and is currently used only for usage statistics, but planned for future work is to put learning analytics methods into use to analyse those data.

2.4.2 Learning Analytics in Discussion Forums

As the research on learning analytics in online laboratories and collaboration is dearth, the inclusion criteria of the literature review were expanded to online learning environments as well as online laboratories. Several research papers regarding discussion forums have been identified. Kent & Cukurova (2020) have developed an approach for measuring the process of collaboration called Collaborative Learning as a Process (CLaP), using social network analysis on data from online discussion communities in the online discussion tool Liglio. The CLaP approach means to analyse collaboration based on the coordination costs and interactivity gains of the student communities. Coordination costs refers to the flow of information within a network, where group structures preventing effective flow of information leads to high coordination costs. Interactivity gains refers to the knowledge obtained from interactions with other learners, where high interactivity gains are interactions amongst learners which contributes to new information. The relationship between coordination costs and interactivity gains are referred to as a CLaP index, where a high CLaP index indicates high likelihood of efficient collaboration amongst learners. Feedback from an instructor implied that the CLaP index provided insight into collaboration which can not be seen with regular participation analysis. For instance, the CLaP index showed that some students were uncomfortable with the first collaborative tasks, whereas they felt more comfortable with the second collaborative task which led to improvement in collaboration within the groups. This indicates that CLaP provides valuable information for the instructor to reflect on the dynamics within groups, in addition to facilitating the design of collaborative tasks.

In a more recent study, Doleck et al. (2021) measures the performance of social learning networks in Massive Open Online Course (MOOC) discussion forums through an algorithm which offers to optimise these networks by connecting users with similar tendencies. The network only consisted of 1/20 of the registered users, suggesting that the MOOC's discussion forum in question did not play a major role in the student learning. The results from the algorithms show a sparse matrix, meaning few students took part in discussion. The benefits of the social learning network appear restricted due to the limited use of the discussion forum. The discussion itself is also restricted to administrative topics. The algorithm does however prove useful in assessing the efficiency of social learning, and testing of methods which encourage social learning in MOOCs.

Pillutla et al. (2020) demonstrate a different example of learning analytics in MOOC discussion forums through text classification. Through assessing the nature of interactions between students using a supervised learning technique of classification on discussion posts, they demonstrate the potential of said classification model to inform about the student-student interactions. The classification model is found to be robust, as 4 out of 5 posts were classified correctly. The model supports educators in assessing the interaction between students, and the course of constructing knowledge. This work-in-progress does also address some limitations. The dataset obtained is from a single MOOC, and relied solely on the text within the discussion posts. The direction for long-term research involves the deployment of monitoring dashboards within the learning systems, which provide the instructors with information about student progress, enabling them to discover students who are struggling.

2.4.3 Learning Analytics Visualisations

A central part of learning analytics is the visualisation of student data. In monitoring collaborative groups in the learning management system Moodle, Sáiz-Manzanares et al. (2021) have chosen a Heat Map visualisation to analyse student behaviour in a Health Science course. The measurements of student interactions are provided by the UBUMonitor tool, a desktop application running on the client, which is connected to the Moodle server and obtains data from the server. The UBUMonitor is an open-source application and includes different modules, including visualisations, clustering, comparison, and risk of dropping out. The visualisation module is used in this study, and more specifically a Heat Map graph as it provides results in colour and numerical visualisations. The visualisation module also provides other sorts of graphs presenting an analysis of course frequencies within Moodle. An example of the heat map is presented in Figure 2.8.

The UBUMonitor tool proved to be useful for the purpose of discovering interaction patterns within each collaborative group, as it supports uncomplicated monitoring of each student in the Moodle environment, at different periods of a course. It was found through the behavioural profiles of each group presented in the Heat Map that the interaction patterns are not homogenous. Within a collaborative group, students work differently, and there seems to always be one or two students who take the lead. The authors, therefore, draw to the conclusion that it is essential to monitor the learning process of each student through the entire course for best detection of students at risk.

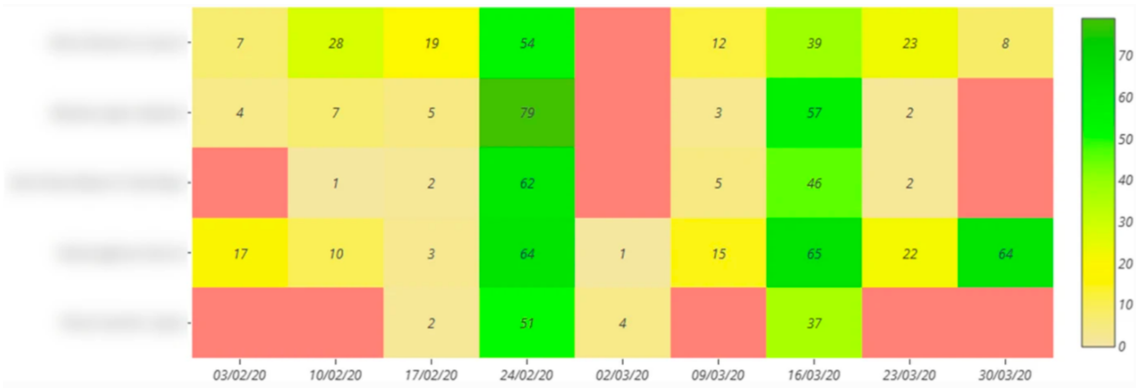


Figure 2.8: The Heat Map presented in Sáiz-Manzanares et al. (2021).

As previously stated, learning analytics dashboards are a widely accepted way of visualising learning analytics (Khalil & Ebner, 2015). An example of such a dashboard is presented in Tarmazdi et al. (2015), where the authors have developed one for teamwork in an online computer science course. The dashboard is based on the analysis of students' online discussion data, combining Natural Language Processing, information retrieval techniques and sentiment analysis to support teachers in monitoring online teamwork. The dashboard, illustrated in Figure 2.9, features information about a specific team, including a graph of roles within the group, the group's sentiment chart and their discussion. At the bottom of the screen is a summary of all groups, in which the instructor may click and select which group to investigate further.

The dashboard was found to be valuable to the lecturer as it provided information about students and groups in need of attention, in terms of how the work was executed and how the groups participated in the work. It allowed the lecturer to give feedback on tone used within the discussion, and members who needed to participate more. The feedback from the evaluation of the dashboard is however limited to a single lecturer and a single course, a limitation addressed for further improvement. Future work intends to improve the dashboards efficiency by investigating further its usability with several lecturers of various subjects and different types of teamwork discussions. Additionally, further implementation of automated warnings could assist the lecturer in more timely interventions, alerting when issues are occurring within groups.

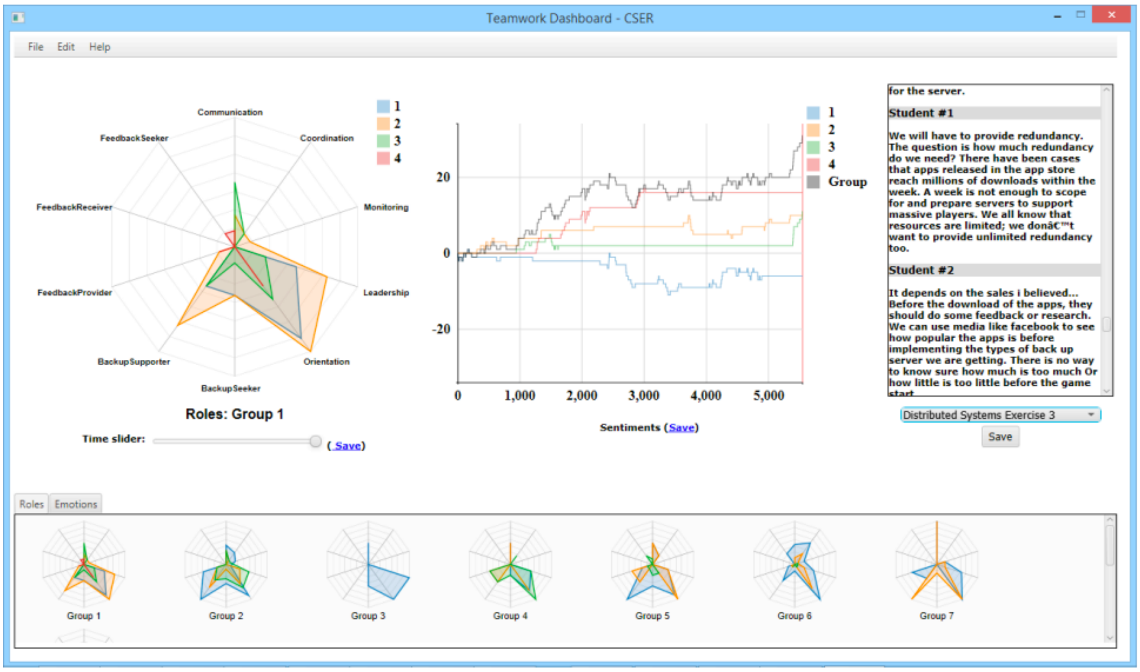


Figure 2.9: The Teamwork Dashboard in Tarmazdi et al. (2015).

It is essential that such dashboards present relevant information so that teachers can provide assistance grounded on the students' needs. van Leeuwen et al. (2019) further investigate how teachers interpret information about collaboration among students on learning analytics dashboards in a CSCL environment. They investigate three different aids for teachers to interpret the processes of collaboration, these being *mirroring*, *alerting* and *advising*. *Mirroring* dashboards contain information about the students gathered from the digital learning environment, where the interpretation is up to the teacher. The *alerting* dashboards display information about students, in addition to providing alerts of groups needing attention or help. *Advising* dashboards displays both information and alerts, in addition to providing supplementary advice as to what a teacher should do about a given event. The results show that the different types of aids did influence the teachers' interpretation of the CSCL situation. Concerning the detection of events, there was no significant distinctness between the different aids as all teachers managed to identify the problematic groups. The advising teacher dashboards are, however, found to be preferred over mirroring and alerting dashboards, as it provides the teacher with a higher understanding of the CSCL situation, making their actions more effective in supporting collaborative work. It was found that teachers using advising dashboards spent the longest time inspecting the dashboards, indicating that teachers took the time to read and consider whether they agree on the dashboard's recommendation.

2.4.4 Learning Analytics with Human Computer Interaction (HCI) Theory

As previously mentioned, prediction models are commonly used methods within learning analytics in higher education (Leitner et al., 2017). Xing et al. (2015) present an example of a student performance prediction model using data from a CSCL environment, namely Virtual Math Teams with Geogebra. They combine learning analytics approaches, educational data mining and HCI theory in development of the prediction model. The theory in question, activity theory, is applied first to reduce the dimensionality of the data and contextualise it. The activity theory system allows the exploration of interactions and collaboration within the CSCL environment, specifically focusing on the learning of each individual student in a collaborative group. The structure of the activity system involves 6 variables of interacting components, namely Subject, Object, Tools, Division of Labour, Community, and Rules. Secondly, Genetic Programming (GP) is applied to build the prediction model. The results demonstrate that the GP-based prediction model proves to be interpretable, and outperforms traditional prediction models. It is argued that in order to build a student prediction model, learning analytics must provide actionable information, which requires that the model is understandable for teachers. Activity theory is applied to achieve more understandable analytics, as the use of theory can provide teachers with a familiar language from which they can draw conclusions from. Narrowing it down to the 6 variables in the activity theory system reduces the data dimensionality, simplifying the interpretations of the analytics. Through the data which is organised theoretically, the model built with GP presents the data to the end users. Still, there are several limitations to the method. The work considers solely quantitative aspects of the data. Future work intends to include qualitative data by integrating natural language processing to students' chat logs. Additionally, the model is yet to be evaluated through user studies. Evaluating the teachers' interpretations of different model representations and testing its effectiveness are also proposed for future studies.

2.5 Chapter Summary

This chapter presented the relevant background literature for the research problem of this thesis, introducing the concepts of learning analytics, collaboration and virtual labs. Through the scoping review, the current state-of-the-art was explored, revealing a knowledge gap within the research area of learning analytics in collaboration and virtual labs.

The findings of the literature review provided valuable insight into current research, and what to further consider in the development of a learning analytics framework, which is presented in Chapter 4.

Chapter 3

Methodology

A design science research methodology will guide the research. This chapter presents an overview of the concepts of design science, and the other methods used within this research.

3.1 Design Science

Design science is a research methodology which aims to resolve some given issue through the design of artefacts (Dresch et al., 2015). It is a method focused on problem solving, and through the understanding of the problem, the construction and evaluation of artefacts, the method can contribute to solving the identified problem.

A design science research methodology was chosen for this research to seek understanding of the area of learning analytics and collaboration in virtual lab, and through the construction of an artefact in the form of a learning analytics framework, potentially resolve new knowledge which may help advance theories and lessen the gap between theory and practice.

A general outline of the design science research methodology is presented in Figure 3.1, where the factors of *relevance* and *rigour* are central. *Relevance* refers to the means in which design science needs to consider the relevance of a particular research within a specific environment, so that organisations within these environments may utilise the results from research of relevance in problem solving (Dresch et al., 2015). The environment section in Figure 3.1 refers to such an environment where a problem is recognised. The environment includes organisations, its people and its technologies. Based on the organisational needs within the environment, the development of artefacts through design science research can strengthen the knowledge base within an organisation.

Rigour too is an essential factor of design science research as it ensures validity and reliability, and can expand the knowledge base within a given field (Dresch et al., 2015). The knowledge base refers to all existing knowledge, theories and artefacts within a given research area. The existing knowledge base is used as a basis for further development of new research and artefacts. The existing knowledge is however often not sufficient enough for the development of new artefacts. Therefore, researchers may build on the existing knowledge by exploring new strategies and ideas during the development of a solution, in which the process itself will contribute to the knowledge base.

Once the organisational needs and the existing knowledge base have been identified, artefacts are developed and further assessed and evaluated to justify its importance. Methods are used to evaluate the quality of the artefact, which may be executed in an iterative process several times in order to meet the organisational needs, where correlations and improvements are made in each iteration.

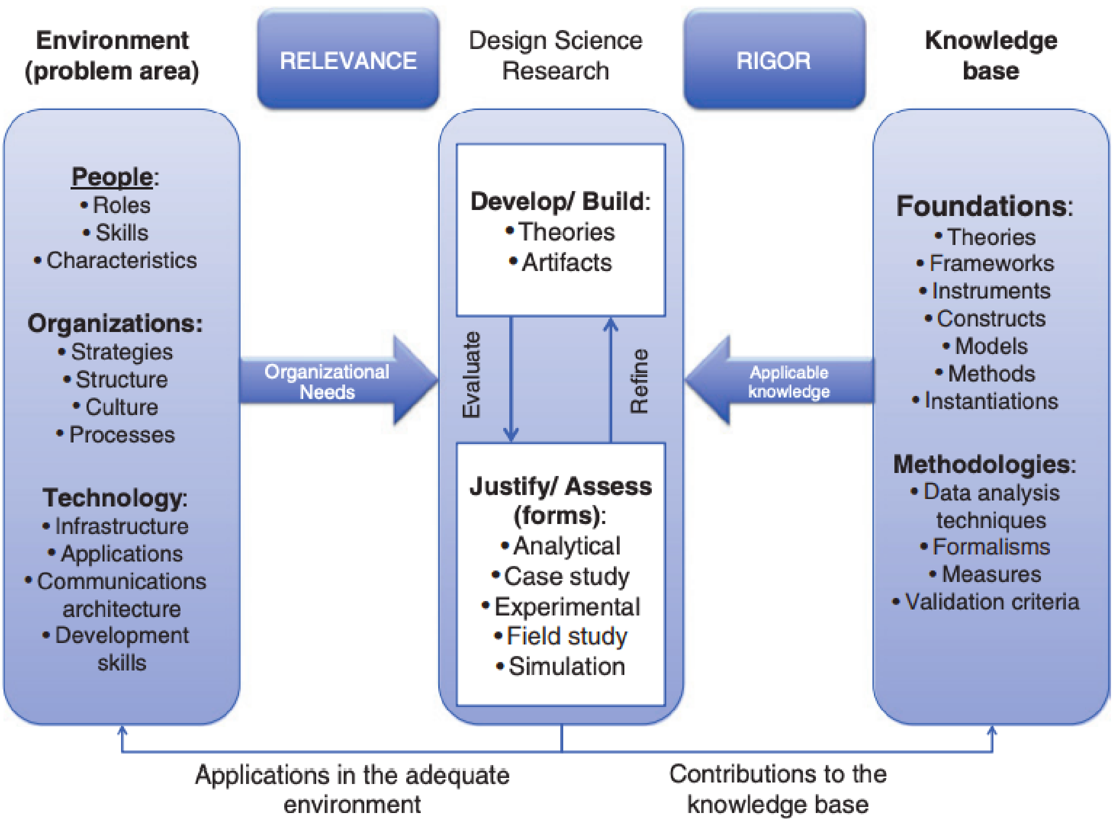


Figure 3.1: Outline of Design Science Research (Dresch et al., 2015, p.69).

3.1.1 Design Science Guidelines

To guide researchers in effectively applying design science to their research, Hevner et al. (2004) proposes a set of seven guidelines with descriptions of how to produce a viable artefact. These guidelines are further elaborated below.

Design as an Artefact. *"Guideline 1: Design-science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation"* (Hevner et al., 2004).

The design-science research should result in a purposeful artefact that addresses important organisational issues. As stated in the guideline, an artefact can take the form of constructs, models, methods or instatiations. A construct is a conceptual object that means to describe some phenomena in the world, in terms of knowledge that are of significance to business or social processes (Williamson & Johanson, 2017, p.269). A model is also a conceptual object which uses constructs to describe and present some real-world phenomena, and aid understanding of problem and solution in a given context. A method is a set of processes which provides instructions on how to solve a specific problem. Lastly, an instantiation is a concrete instance of something. It can be a hardware or software system which is the result of utilising a method to implement a construct or a model (Williamson & Johanson, 2017, p.270). Constructs, models, methods and instantiations are all considered equally crucial in the creation of IT artefacts (Hevner et al., 2004). Artefacts constructed in design science research are not often fully developed products ready to use, but are rather innovations that determine how ideas, practices and products can be accomplished through the process of development of the artefact.

Problem Relevance. *"Guideline 2: The objective of design-science research is to develop technology-based solutions to important and relevant business problems."* (Hevner et al., 2004).

The main objective of research within information systems is the development of technology-based solutions that intend to solve relevant and existing business problems. Design science research addresses this goal through the construction of an artefact which aims to resolve the occurring problem (Hevner et al., 2004).

Design Evaluation. *"Guideline 3: The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods."* (Hevner et al., 2004).

The evaluation is an important factor of the research process, as evaluation contributes to justify the artefact's usefulness and relevance (Hevner et al., 2004). The choice of evaluation method must conform with the developed artefact as well-selected evaluation methods help to demonstrate the quality of the artefact.

Research Contributions. *"Guideline 4: Effective design-science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies."* (Hevner et al., 2004).

Effective design science research is required to contribute to the knowledge base. There are three different types of research contributions in design science based on generality, novelty and significance of the artefact, where at least one of them should be delivered in the research (Hevner et al., 2004):

1. **The Design Artefact.** Commonly, the contribution to design science is the artefact itself. The artefact must provide a solution to unsolved issues, either by extending the existing knowledge base or by adding existing knowledge through new practices and ideas.

2. **Foundations.** The development of artefacts that expand and improve the foundations of the knowledge base within design science research are also essential contributions. Examples of such foundations can be found in problem and solution representations, ontologies, modelling formalisms and innovative information systems.

3. **Methodologies.** Lastly, the use of creative evaluation methods also provides important contributions to design science research.

Research Rigour. *"Guideline 5: Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact."* (Hevner et al., 2004).

Hevner et al. (2004) underlines the importance of constructing an artefact in a rigorous way. Rigorous research is achieved through effective use of theories and research methodologies from the knowledge base. Selection of appropriate development and evaluation techniques in order to justify the artefact is therefore paramount. Even so, an overemphasis on rigour can result in lower relevance. It is essential that research is both rigorous and relevant (Hevner et al., 2004).

Design as a Search Process. *"Guideline 6: The search for an effective artefact requires utilising available means to reach desired ends while satisfying laws in the problem environment."* (Hevner et al., 2004).

A search process involving actions used to reach a desired end result should be reflected through the artefact (Williamson & Johanson, 2017, p.273). Design science is an iterative search process in which its intention is to discover solutions to a given problem. Problem solving can be regarded as the employment of available *means* to achieve desirable *ends* while satisfying the *laws* which are set by the environment (Simon, 1996). *Means* refer to the actions and resources accessible for the construction of the solution (Hevner et al., 2004). *Ends* are the goals and constraints of the solution. *Laws* are the principles within the environment for which the artefact is developed. The representation of applicable means, ends and laws are essential to design science research. Through iterations, means, ends and laws are refined, which will then induce a valuable and relevant artefact.

Communication of Research. *"Guideline 7: Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences."* (Hevner et al., 2004).

The design science research and the process of developing the artefact must be well documented, so that the audience understands the construction and evaluation processes of the artefact (Hevner et al., 2004). The technology-oriented audience calls for sufficient details in order to implement and put to use the developed artefact within the organisation. Presentation of sufficient details of the processes used can also enable such audiences to further extend the knowledge base. It establishes repeatability and further extension of the research by other researchers.

Management-oriented audiences call for sufficient details in order to decide if the organisation should commit to purchase and use of the artefact. Design science research should therefore provide knowledge of how to effectively implement the artefact within the organisational context. It may also be essential to provide details which allows managers to understand and acknowledge the artefact (Hevner et al., 2004).

3.2 Desk Research

The desk research, also known as secondary research, includes the literature review presented in Chapter 2. Literature reviews are essential to any research process as the mapping and assessment of previous, relevant literature can help motivate the aim of the study in question (Snyder, 2019). There are different methods within literature review research, depending on different contexts. A systematic scoping review was chosen for this work as it provides an overview of the volume and nature of existing literature on the given topic (Peters et al., 2015).

3.3 Conceptual Design

Conceptual design involves the development of the conceptual model, a model which intends to describe how a system is organised and how a user can interact with it (Sharp et al., 2015, p.434). A conceptual model provides the benefit of a framework for concepts involved in a system, giving a clear idea of a systems outline before starting the development. The development of the learning analytics framework resulting from the conceptual design is presented in Chapter 4.

3.4 Evaluation Methods

Evaluation is an essential part of development in design science research. The artefact should be evaluated in order to determine its validity and whether it brings value and new knowledge to the field (Dresch et al., 2015). Even if the results from evaluation of an artefact proves to not provide value to the field, this result in itself contributes to new knowledge within the area of research. A qualitative approach has been considered for the evaluation methods, as this would be most beneficial in getting meaningful results. The methods used for evaluation are further elaborated below.

3.4.1 Data Gathering

Data gathering is an essential part of evaluation (Sharp et al., 2015, p.260). The aim of data gathering is to collect relevant and sufficient data so that design can be carried out and proceed, as well as capturing the opinions and experiences of the participants contributing to the evaluation. There are various techniques in which to gather data. The method relevant for the evaluation in this thesis is interviews.

3.4.2 Expert Evaluation

For this research, seven experts were consulted for the evaluation of the artefact. The experts all had experience in teaching within higher education, and all had to a varying degree experience with virtual labs and learning analytics. The intention of the expert evaluation was to get feedback on the developed framework and what aspects to consider when utilising learning analytics to facilitate learning in virtual labs and collaboration. By consulting experts within these fields, valuable feedback was provided by those who have experience within the relevant areas. Two iterations of evaluations were conducted, where four experts were consulted in the first round, and three experts in the second round.

3.4.3 Semi-structured Interviews

Semi-structured interviews were chosen as a method for guiding the expert evaluation. Such interviews consist of both open and closed questions, where the interviewer uses a script for guidance so that for each interviewee, the same subjects are investigated (Sharp et al., 2015, p.269). The interviewer uses probing, a technique which involves asking follow up questions, encouraging the interviewee to go more in-depth. The advantage of semi-structured interviews is the opportunity it provides to gather more information and explorations of opinions though probing.

3.5 Chapter Summary

This chapter has described the method guiding this research, a design science research methodology, and other methods utilised in design and evaluation such as conceptual design and semi-structured interviews.

Chapter 4

Artefact Development

This chapter describes the iterative development process of the artefact that is the learning analytics framework. The development process comprised of two iterations. The development in the first iteration was based on the results from the desk research, and was evaluated with experts through semi-structured interviews. A second version was constructed in the second iteration, based on the feedback from experts in the first iteration. A similar evaluation was then performed to further explore the framework. First, an overview of the data collected in this process is presented following the tools that were used. Next, the two iterations of development are described.

4.1 Data Collection

Data was gathered through the systematic scoping review in Chapter 2 and through evaluations in the artefact development iterations. An overview of the data collected in this research is presented in Table 4.1.

4.2 Tools

Draw.io

Draw.io is a free diagram tool integrated with Google Drive[1]. Draw.io is a valuable tool for making flowcharts, network diagrams, Unified Modelling Language (UML), process models and organisational maps. The tool was used to create the learning analytics framework.

[1] Google Workspace Marketplace. https://gsuite.google.com/marketplace/app/diagramsnet/671128082532 (accessed 09.05.2022).

Table 4.1: Overview of data collection.

	What data	What method	Use
Systematic scoping review	What does current research provide on the use of learning analytics in virtual labs and collaboration?	Systematic scoping review	Map current research within the use of learning analytics in virtual labs and collaboration, in order to decide which factors to consider when developing the learning analytics framework.
Expert Evaluation 1st iteration	Expert opinions on the learning analytics framework	Semi-structured interviews	Explore opinions and improve the learning analytics framework
Expert Evaluation 2nd iteration	Expert opinions on the learning analytics framework	Semi-structured interviews	Explore opinions and improvements of the second version of the framework

Zoom
Zoom is a video platform that allows you to arrange video meetings online[2]. It also supports recording of sessions and sharing of home screens. The platform was used for the expert evaluation.

4.3 First Iteration

Based on the findings in the literature review, a learning analytics framework for collaboration in virtual labs has been developed. The objective of the framework is to show how learning analytics can be used to support and improve collaboration in virtual labs. The framework is presented in Figure 4.1. It consists of five different sections: stakeholders, learning environments, data, data analysis and visualisation. The concepts and the processes within the framework are further elaborated in the following subsections.

[2] https://zoom.us/

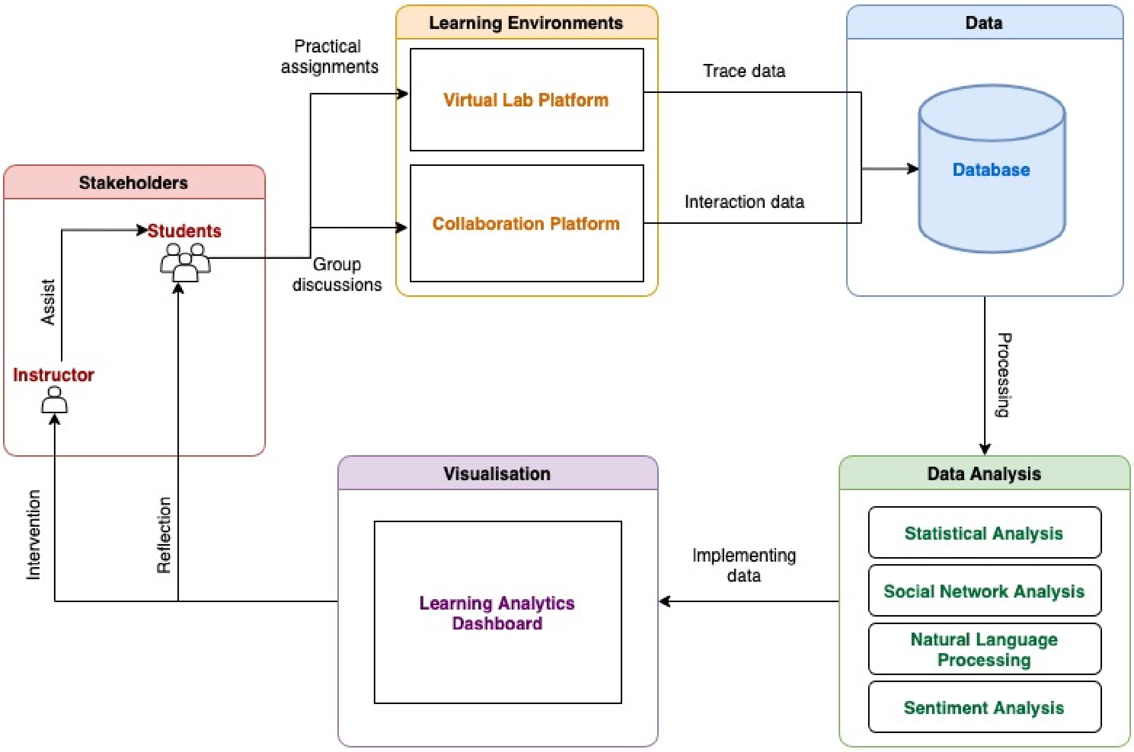


Figure 4.1: First version of the Learning Analytics Framework.

Stakeholders

The main stakeholder is the students. The reviewed research in the literature review is mostly aimed at teachers, with the focus of guiding teachers in facilitating collaboration among students. There is limited focus on students and providing them with analytics. This is somewhat contradicting to the objective of *reflection* in learning analytics, which supports improvement of learning based on reflection of previous work (Khalil & Ebner, 2015). Therefore, this framework aims to fill that gap by concentrating on the students as the main stakeholder. The instructor, which can be a teacher or lab assistant, will act as an assisting stakeholder. The intention is that learning analytics is integrated for the purpose of enhancing students' learning performances, and for the instructor to assist students in this process.

Learning Environments

The learning environment includes a virtual laboratory platform and a collaboration platform, where the students will perform practical course assignments in a virtual lab platform and participate in group discussions in a collaboration platform. The literature review revealed that few online laboratory environments allow collaboration within the environment. It is therefore assumed that collaboration will take place in a separate platform. Additionally, in relation to the ENVISION_2027 project, their objective of innovating e-learning modules to facilitate teamwork and laboratory exercises is based on these two activities taking place on separate platforms. However, this does not mean that platforms integrating both laboratory exercises and collaboration do not exist (Hu et al., 2018; Jara et al., 2009). Nevertheless, this thesis will focus on these two activities taking place in separate platforms.

The literature review further revealed that few students actively take part in discussion forums, which suggest that this is not the most preferred tool for collaboration (Doleck et al., 2021). An alternative collaborative tool is therefore worth considering. Instant messaging (IM) tools could be valuable in collaborative settings, as these are one of the most popular forms of communication tools among university students (Quan-Haase, 2008). IM systems are applications that support near-synchronous, text-based communication between two or more users. Some examples of these systems are Discord, Slack, Facebook Messenger and WhatsApp. IM has proven to be a useful tool in enhancing communication among groups in higher education (Lauricella & Kay, 2013). Exactly what platforms are used will be based on how the instructor designs the course.

Data

While students are using learning platforms, large amounts of data are being generated. In this section, the student data is being extracted and retrieved from the learning platforms and stored in a database. The data retrieved must be of relevance to the course in question and to improving collaboration. The data of relevance are interaction data and trace data. The interaction data will be extracted from the collaboration platform. This will be data involving the interaction between students. As previously stated, collaborative learning is a broad term which can include many different kinds of interactions and contexts (Dillenbourg, 1999). The kind of interaction data would therefore depend on how the collaboration is structured. If students are divided into groups, one might wish to see how each group is working together. If students are not divided into groups, but there is for instance a common collaboration channel for the whole class, one would need data offering insight into how the class is collaborating as one whole group.

Trace data will be extracted from the virtual lab, which will involve the traces of students within the platform. These traces can be the number of mouse clicks, number of logins, number of finished assignments and number of accessed documents and videos (Khalil & Ebner, 2015). The database would then presumably consist of different datasets with data of various formats. It is then essential for the data to be processed and cleaned in order to retrieve meaningful information from it.

Data Analysis

After the data has been processed, data analysis can be carried out. There are various data analysis methods within education which seek to identify interesting patterns in educational datasets (Khalil & Ebner, 2015). Statistical analysis is to be applied to the trace data. By applying statistics to the trace data, numeric computations can be carried out and reveal knowledge out of the data. This could be calculation of time spent on tasks and analysis of mouse click.

Discovered through the literature review, social network analysis is a valuable method for identifying patterns within collaboration. Social network analysis will therefore be employed as it allows investigation of relationships between different entities (Khalil & Ebner, 2015). Such analysis is particularly used to identify students who are isolated, which students are active in communication and how well-established a learning environment is within a group.

It could also be valuable integrating qualitative analysis based on students' chat logs. That could be integrating natural language processing to gain further insight into the discussion between students. Additionally, as a way of discovering dynamics within a group, sentiment analysis could be beneficial, as was shown in Tarmazdi et al. (2015). The choice of data analysis methods would, similarly to the choice of data, depend on the structure of the collaboration. That being whether students are divided into groups or if the class as a whole is collaborating on a common channel.

Visualisation

The results from the previous section, the analysed data, is then visualised to the stakeholders through a learning analytics dashboard. Learning analytics dashboards are a widely accepted way of visualising learning analytics as they offer a simple insight into learning processes of students and are easy to understand (Khalil & Ebner, 2015). The information presented on the learning analytics dashboard again needs to reflect the intention of this framework, which is to support and improve collaboration.

The learning analytics objectives of relevance in which to achieve this, may relate to *reflection and iteration* and *intervention* as described by Khalil & Ebner (2015). *Reflection and Iteration* involve self-evaluation based on past work in order to improve future learning experiences. The learning analytics dashboard will then need to provide students with information about their efforts in the group work and in the virtual lab, so that they can reflect and adapt based on this information. Additionally, the information within the dashboard can facilitate *interventions*. Interventions performed by teachers can contribute to improving student success and hinder drop-out based on students' needs and assist accordingly (Khalil & Ebner, 2015).

The literature review revealed that advising dashboards were preferred over mirroring and alerting dashboards, guiding teachers in effectively supporting collaborative work (van Leeuwen et al., 2019). Therefore, in addition to presenting relevant information which promotes reflection and interventions within the dashboard, recommendations could be applied according to the stakeholder's needs.

This first version of the framework was evaluated by four experts via zoom.

4.3.1 Evaluation

The goal of the evaluation was to get feedback on the learning analytics framework, and to discuss what aspects to consider when aiming to best facilitate learning in virtual labs and collaboration. Four experts participated in the first round of evaluation. All evaluations were performed separately as semi-structured interviews via zoom, guided by the interview guide in Appendix B. Through agreement with the participants, the session was recorded and analysed. The session was structured so that the framework was presented first, then background questions were asked, following questions regarding the framework and the different sections within it. A description of the participants' background and professions are presented below, together with the feedback categorised in the following subsections: suggestions, stakeholders, learning environments, data, data analysis, visualisation and learning analytics.

Participants

The expert team consists of two females and two males. The first female expert is an education coordinator, with a background in molecular technology, innovation policy, and molecular endocrinology. The second female expert is a university lecturer within biomedicine. The first male expert is a senior research specialist, with a background in molecular biology. The second male expert is a senior researcher who teaches in biomedicine. All experts had experience with teaching using virtual labs. Some experts had more experience with learning analytics than others, but all were aware of the concept.

Suggestions

The general opinion amongst all experts was that the framework was well-presented and understandable. Two experts suggested adding another step regarding student feedback. This would involve an evaluation where students give feedback regarding their own efforts within the group work, in addition to evaluating how the group worked together as a whole. Two experts suggested adding more images or graphs.

Stakeholders

Three experts stated that students and instructors are the stakeholders who should be addressed by the framework. One expert also stated that ' *'the main people who will benefit are not the students who have generated the data, but the future students"*. Another expert stated that course leaders and teachers would benefit the most from the learning analytics. Additionally, three experts suggested additional stakeholders above the instructor, these being faculty leaders and those responsible for the course (if different than the instructor).

Learning Environments

All experts had experience with the virtual lab Labster. None of them had experienced a virtual lab with collaborative features. One of the experts commented on the way in which students communicate online, and how communication depends on whether the students do the virtual lab synchronously or not. He stated that students might like to do the lab on their own time, and an instant messaging platform such as Slack would work as a discussion forum anyhow because students might not be in the same stages in the lab. To improve the communication in such a scenario, the expert suggested to arrange it so all students do the virtual lab at the same time and this way they could all discuss through an instant messaging platform such as Slack along the way.

The same expert also stated that Labster, which is the virtual lab they use, does not promote much collaboration. The reason for this was that Labster allows students to perform techniques in experiments and do quizzes along the way, which might only invoke discussions about how to get past certain levels if you get stuck. The expert suggested another lab which could be more suitable for collaborative work, namely LabBuddy[3]. LabBuddy focuses on the design of experiments and allows students to make flow schemes of experiments before performing it in a physical lab. The reason for this suggestion was that this platform *"would work much better as a synchronous event with a Slack channel"*, where students could

[3]https://www.labbuddy.net/

work by themselves first and then compare what they have done with a group, discussing the different steps and solutions online.

Data

In regards to what data is of relevance in supporting collaboration in virtual labs, most experts suggested data regarding the different roles within a group and how well students within a group work together. One expert also suggested data regarding collaboration between different groups. If collaboration is not structured into groups, one expert suggested data of who interacts with who and if anyone is working in pairs. Another expert suggested data of how often students ask questions and how well they respond to each other's questions with constructive answers.

Data Analysis

Most experts had little experience with data analysis methods, but all stated that the presented methods seemed satisfactory. One expert added that some sort of predictive model could be relevant also, in order to see which groups will work out. This way, students would avoid conflicts where the focus would be on getting along rather than on learning.

Visualisation

One of the experts shared a doubt in regards to providing students with data through visualisations, seeing that teachers might benefit more from this information than students. The reason for this was that students might be confused by the meaning of it, whether it is a part of assessment or not, and also compare themselves to other students. Another expert shared that opinion, stating that course leaders and teachers would benefit most from the learning analytics dashboard. He stated that learning analytics would be useful for teachers in reflecting and preparing the course for next year, based on the students' performances the current year.

One of the experts commented on the act of interventions performed by instructors. He stated that in a course of 100 to 200 students, interventions might not be performed by teachers, as there is a great number of students to keep track of.

In regards to what information should be presented on the learning analytics dashboard, one expert stated that the different stakeholders would need to have different amounts of information, which would involve dividing the dashboard into a student view and an instructor view. The student view would include information about how they each have performed and the instructor view would include information about all students. The expert also stated that as an instructor *"you may want to know individually how students have done across everything, but also how well a question has performed in terms of all students."*

Regarding specifically what kind of information to be presented on the dashboard, the experts suggested time spent on different tasks within the virtual lab, specific questions the students are struggling with and what part of the platforms that contributes to learning, which was suggested to be measured with follow-up quizzes. In terms of collaboration, the experts wished to see how groups communicate with each other, the nature of the students' discussion and different roles within groups.

Learning Analytics
One expert stated that collaboration is *"such an elusive area to be able to analyse, that any tool that could give any data on it I think would be useful,"* and that through learning analytics, students can become more aware of how they are performing, what they are not that good at and therefore work on it. If you don't know what you need to work on, *"it takes a lot of your own self-reflection and self-awareness to be able to improve it."* He also stated that if such feedback regarding the collaboration could be given for each course and kept track of during a program, it could give a powerful dataset which could stimulate multiple rounds of improved collaboration.

Two other experts stated that simply the act of making students aware of their data being monitored and gathered for learning analytics can make them more engaged. One of them stated that perhaps not everything has to come from the learning analytics itself, although it is through the learning analytics one becomes aware of how students are performing, which then facilitates interventions, change and improvement.

4.4 Second Iteration

Based on the evaluation of the first iteration, a second development of the framework was performed where several adjustments were made. The second version is presented in Figure 4.2. A second round of evaluation was performed in order to gain a broader perspective on the different aspects within the framework and further explore what other aspects to consider when facilitating learning in virtual labs and collaboration.

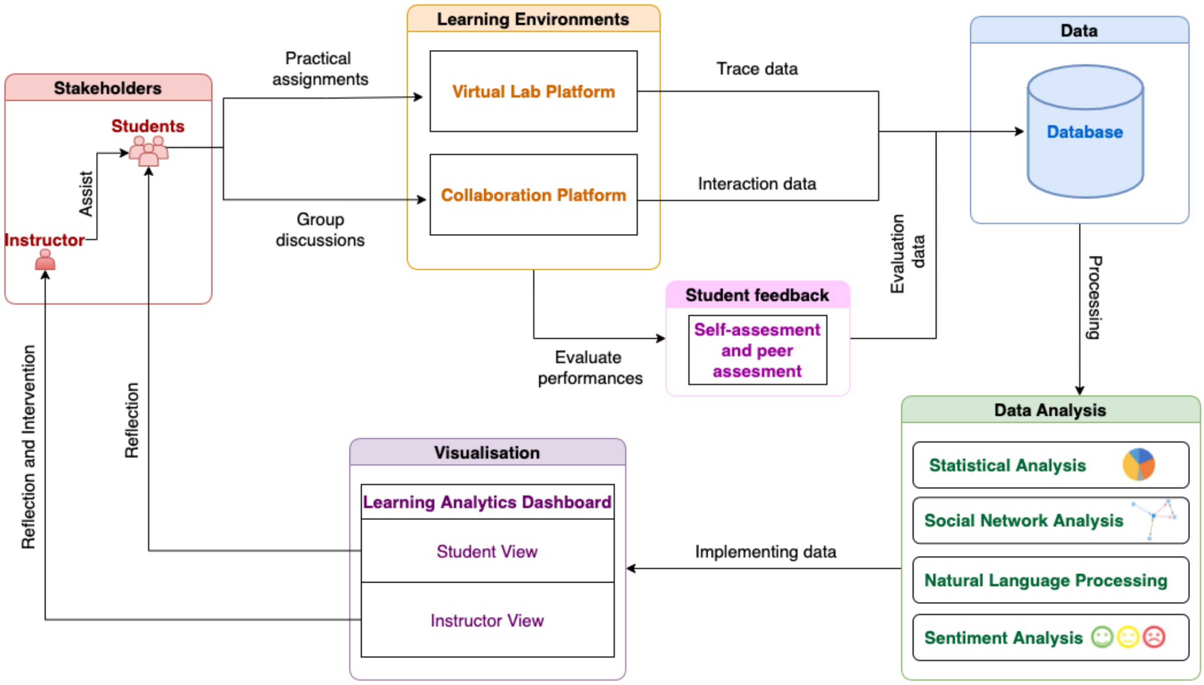


Figure 4.2: Second version of the Learning Analytics Framework.

4.4.1 Adjustments

Based on the feedback from the evaluation, an additional section was implemented in the second version in Figure 4.2, namely *Student Feedback*. This step involves students evaluating their own performances and/or that of their fellow students. This would involve the performances in the virtual lab, as well as the performances in the collaboration. This could be done through self-assessment and peer assessment. Peer assessment (PA) is commonly defined as “an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners” (Topping, 2009, p.20-21) and self-assessment is the act or process of analysing and evaluating one’s actions or performance to improve their learning (McDonald & Boud, 2003). It can encourage students to reflect on their own performances as well as their peers’, which can help them to identify their strengths and weaknesses, and thereafter learn from it. Peer assessment can be used for evaluation of each students’ individual effort but also their contributions to the group work[4]. In this context, it might be difficult for students to know how their fellow students have performed in the virtual lab, as they might like to do this on their own time. Peer assessment would therefore be most useful in evaluating fellow students’ contributions to the group work and how well they worked together, instead of their efforts in the virtual labs.

[4] University of Reading. Peer assessment. https://www.reading.ac.uk/engageinassessment/peer-and-self-assessment/peer-assessment/eia-peer-assessment.aspx (accessed 28.04.2022)

The *Data Analysis*-section was updated with some images and graphs related to the different analysis methods. The *Visualisation*-section was also updated. Based on the feedback, the learning analytics dashboard were divided into two views: student view and instructor view. The instructor would need more information compared to the students. This could be information regarding performances of each individual student, but also of each group. As suggested by one of the experts, the instructor also might wish to see how a question has performed in terms of all students. Students would need for instance information about their own performances within the virtual lab and in the collaboration to promote reflection.

Lastly, the act of *Reflection* was added to the arrow pointing from the learning analytics dashboard to the instructor. The feedback from the experts implied that instructors can use the information on the learning analytics dashboard to reflect on how to improve their course and what changes to make in order to make the course better for the students. This could be based on what themes most students are struggling with, then altering the course by discussing these themes more in class.

4.4.2 Evaluation

The goal of this round of evaluation was to get feedback on the second version of the framework, and further explore different aspects within the use of learning analytics in collaboration and virtual labs. The second round of evaluation was performed similarly to the first evaluation described in Section 4.3.1, through a semi-structured interview via zoom, guided by the same interview guide (Appendix B). Three experts participated in this round of evaluation, using the same structure as in the first evaluation. A description of the participants' background and professions are presented below, together with the feedback categorised in the following subsections: suggestions, stakeholders, learning environments, data, data analysis, visualisation and learning analytics.

Participants

The second iteration expert team consists of two females and one male. The first female expert is a senior researcher and course coordinator with a background in rare diseases, teaching in genetics. The second female expert is a professor, with a background in computer science and technology enhanced learning. The male expert is a professor emeritus with a background in biomedical engineering and pedagogics. All experts had some experience with virtual labs, either in teaching or research. Likewise, all experts had experience with learning analytics in routine work or research.

Suggestions
The general opinion about the presentation of the framework was in this iteration similar to the last, where all experts stated that it was understandable and easy to follow. One expert had a suggestion about including steps regarding learning goals and learning design. This would involve how the course is set up, what tools to use, how collaboration is structured and what it is the students are to do and learn in the virtual labs. These goals could be set by the instructor and would be stored in the database.

Stakeholders
In addition to the instructor and students, one expert suggested including next year's students as stakeholders. Another expert suggested facilitators, which are often former students who are responsible for different groups of students in the course, helping them with different tasks. The last expert suggested researchers or designers of the virtual lab or collaboration platform. Designers might want insight into how their tools have been used in order to improve them.

Learning Environments
One of the experts had experienced virtual labs through different projects where collaboration was possible within the same platform. The two other experts had no experience with such a platform. One of the experts had used Labster in teaching, stating it is not designed for group discussions.

Data
Regarding what data is of relevance in supporting collaboration in virtual labs, the experts suggested time stamps, the text which students write to each other, who is writing what and how much each student contributes. One expert stated that one might like to connect the data from the collaboration platform with data from the virtual lab to see how different events could be related to each other, which could be achieved through timestamps.

Another expert raised a question regarding whether the data are used for real-time or post-processed purposes. He stated that most of the data is gathered after the course is done, and questions how the students of the current year are to benefit from the learning analytics. From his experience, the gathered data is used to help next year's students.

One expert raised concerns regarding privacy issues related to the students, stating that it might be too intrusive gathering student discussion data. Another expert stated that students need to be aware that their data is being collected and analysed, and why it is being done. Showing the students how the collection and analysis of their data can help them would be advantageous, and she specifies the importance of assuring the students that their data is secure and will not be shared with unauthorised personnel.

Data Analysis
Two of the experts stated that the chosen methods were good examples. Another expert was on the other hand sceptical regarding the necessity of data analysis based on how students may feel uncomfortable knowing their data is being traced and analysed.

Visualisation
One expert suggested dividing the learning analytics dashboard into two different sections: one section which presents information about how students perform in the virtual lab, and another section presenting information about collaboration, which would be different for each student and for the instructor. Regarding specifically what information, the expert suggested that for the teachers this could be students' activity levels, the different goals and domain concepts students have been working with and whether they are struggling with the different domain concepts or collaboration aspects. This could help the teacher reflect and improve practical assignments for the next round. For students, the expert suggested information about which aspects within the course they are struggling with. Additionally, some students might want to see how they are doing compared to the rest of the class, but the expert specifies that this should be an option.

Another expert stated that they would appreciate having the results from the self-assessment and peer assessment presented in the dashboard for the teachers, as this would show how the students are reflecting on their own performances. The expert also suggested for the students' dashboard to be provided with motivating sentences to help engage them.

Learning Analytics
One expert stated that learning analytics can help students become more aware of where they need to put in effort, in which the students can focus on the particular skill they seem to not have mastered. Additionally, the expert specified that learning analytics can help students become more aware of their role in collaboration, providing information about the degree of contribution for each student.

This could show that some students might not be contributing enough, and others too much.

4.5 Chapter Summary

This chapter described the development process of the learning analytics framework. The process consists of two iterations, where a first version of the framework was built in the first iteration based on the desk research, and a second version was built in the second iteration based on feedback from the evaluation. The evaluation phase and results were described for each iteration. A final version of the framework is presented in the next chapter.

Chapter 5

Final Artefact

This chapter presents the final version of the learning analytics framework, where some alterations were made after the second iteration. The framework is presented in Figure 5.1. The different sections within the framework are described below.

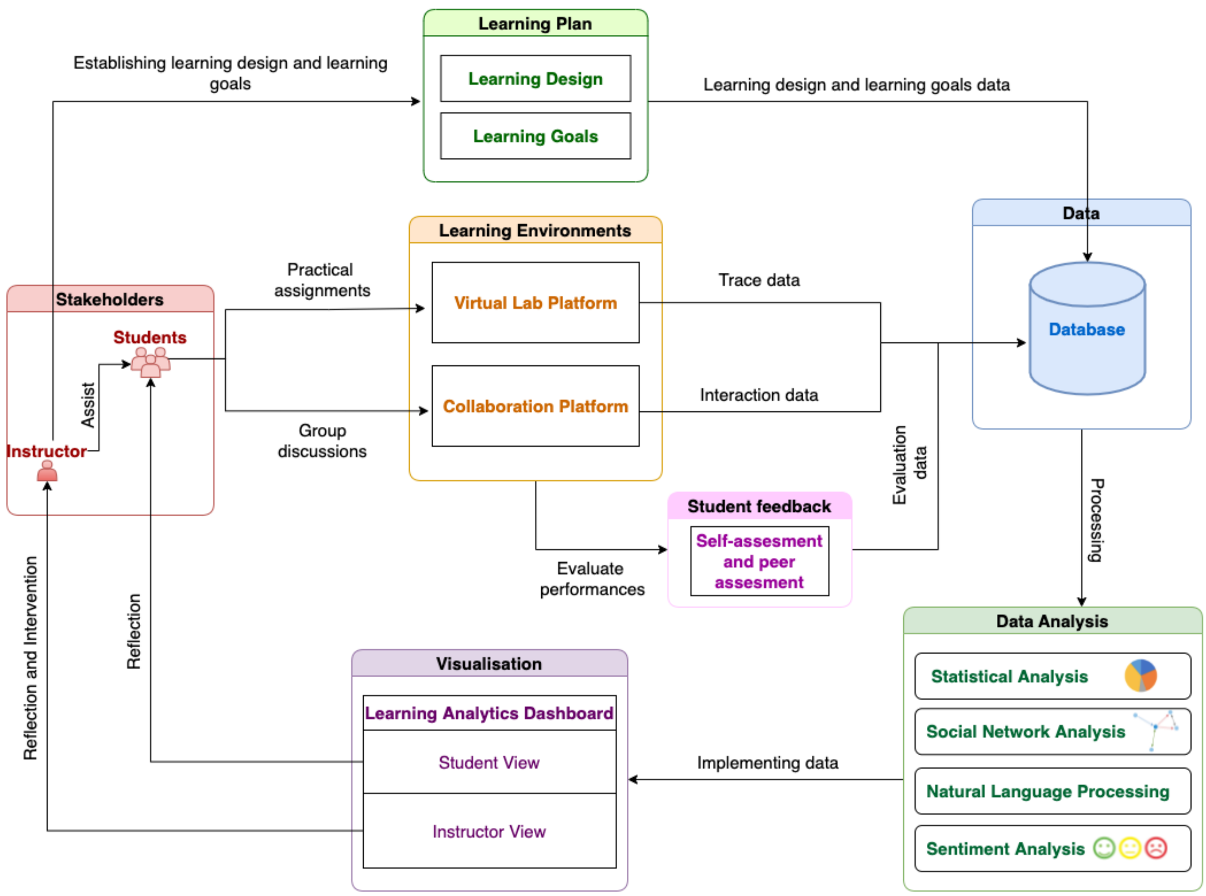


Figure 5.1: Final version of the Learning Analytics Framework.

5.1 Stakeholders

The instructor and the students remain as the stakeholders, with the students as the main stakeholder and instructor as an assisting stakeholder. Although other stakeholders like faculty leaders, researchers and platform designers were suggested as potential stakeholders, they have not been included based on the decision to keep the focus of the framework within the course and on the instructor and the students. Including a facilitator was also proposed in the evaluation. It has not been implemented in the framework, but is implied in the instructor category, as the instructor represents both teachers, lab assistants and facilitators.

5.2 Learning Plan

A *Learning Plan*-section was added based on feedback from the evaluation described in Section 4.4.2. The learning plan is established by the instructor, and includes the learning design and the learning goals of the course. The learning design should describe how the course is organised, that includes what kind of activities are planned, what platforms should be used and how the collaboration should be structured. The learning goals are descriptions of the knowledge and skills students should achieve in the course, which will apply to both course material and collaboration. The learning plan will influence the other sections within the framework as it describes what students are to learn in the course and what platforms are to be used. The learning plan will be stored in the database.

5.3 Learning Environments

Based on the feedback from the evaluation, no changes were made to the framework in the *Learning Environments*-section. Practical assignments will take place in the virtual lab and group discussions will take place separately in the collaboration platform. Exactly what platforms will be used, however, depends on the learning plan. The learning plan will also decide how the collaboration is structured. The choice of platforms may again affect the type of collaboration that takes place. From the evaluation in Chapter 4, two of the experts stated that the virtual lab they had experience with, namely Labster, was not suitable for group discussions. It might therefore be worth considering alternative virtual lab platforms which are more suited for collaboration. A virtual lab that could be relevant is LabBuddy, suggested by one of the experts.

The collaboration platform would also be based on the learning plan, and what is preferred by the instructor and students. As found in the literature review, students often don't take an active part in discussion forums. An alternative tool could therefore be IM tools as these are popular forms of communication tools among university students (Lauricella & Kay, 2013).

5.4 Student Feedback

In the *Student Feedback*-section involving self-assessment and peer assessment, students will assess their own performances and that of their peers. The self-assessment will be based on how the students themselves perform in the virtual lab as well as in the collaborative work. The peer assessment will be based on their fellow students' contributions to the group work. The learning plan will have an influence on the assessment. When doing these assessments, their performances would be measured against the learning goals from the learning plan set for the course, as the learning goals are what the students should aim to achieve through the course. The self-assessment and peer assessment can be performed either throughout the course or at the end of the course. The data from the assessments will be stored in the database.

5.5 Data

The database will consist of different datasets with various formats from the other sections within the framework, these being the learning plan-dataset, student feedback-dataset, trace data from the virtual lab platform and interaction data from the collaboration platform. Based on the feedback from the evaluation in Chapter 4, the data of relevance in this context would be time stamps from both virtual lab and collaboration platform, time spent on tasks within the virtual lab, attempts used on different tasks within the virtual lab, the text in which students write to each other, who is writing what and how much each student contributes in collaboration. Additionally, data regarding what roles students take within a group and how well the group works together are of importance. The data from the virtual lab and from the collaboration platform could also be connected through time stamps in order to see how different events within the different platforms relate to each other.

Regarding the raised concerns of privacy in Section 4.4.2, the students need to be aware that their data is being gathered and analysed. They also need to be assured that their data is safe and will not be shared with unauthorised personnel. Presenting to the students information on exactly how and why data collection and analysis is performed is therefore important. Based on other feedback in Chapter 4, simply making students aware of their data being gathered for learning analytics can encourage them and make them more engaged.

Whether the data is used for real-time or post-processed purposes could depend on when the students perform the virtual labs and do the student feedback, as well as when the learning analytics dashboard is implemented. As stated in the evaluation in Section 4.3.1, students like to do the virtual lab on their own time, meaning they would have to coordinate their collaboration based on this. If the learning analytics dashboard is implemented at the start of the course, it could be updated for each student once they have performed group work and practical assignments in the virtual lab, instead of being updated at the same time for all students. The data could then be used for real-time purposes.

5.6 Data Analysis

The data is to go through the process of cleaning and mining before data analysis is applied. Statistical analysis is to be applied to carry out numeric computations from the data, revealing for instance analysis of time spent on tasks, mouse clicks and succession rate within the virtual lab. Social network analysis will be applied to investigate relationships between students and discover patterns within collaboration. Natural language processing based on students' text can help gain further insight into the discussion between students. Lastly, sentiment analysis is to be applied to discover dynamics within a group. It could also be worth considering a predictive model as suggested in the evaluation in Section 4.3.1, to see which students will work well together and possibly avoid conflicts within groups. The results from the analysis will then be implemented and visualised to the stakeholders through the learning analytics dashboard.

5.7 Visualisation

The learning analytics dashboard will be composed of two different views: student view and instructor view, where they would receive information of different extent. The dashboard would hold two categories of analytics, one regarding student performance in the virtual lab and another regarding student performance in collaboration. Their performances would be measured against the learning goals, as these are descriptions of what the students are to achieve in the course.

Instructor View
The instructor would need a broader extent of information than the students. Based on feedback from the evaluation in Chapter 4, valuable information regarding the virtual lab would involve time spent on tasks, measurements of performance of each individual student in the virtual lab, if there are different concepts the students are struggling with in the lab, and how specific questions and concepts have performed in terms of all students. Regarding collaboration, valuable information would be the nature of students' discussion, the different roles within each of the groups and how they communicate with each other within the groups. The results from the students' self-assessment and peer assessment will also be presented in the learning analytics dashboard.

Student View
The student view will contain individual information for each student, which will involve information about their own performances within the virtual lab, which concepts within the course they are struggling with and how they have contributed to the collaboration. Some students might also want to see how they are performing compared to the rest of the class, whether that is compared to individuals or a class average. This should however be optional as other students might not appreciate this information. Additionally, one could provide students with other students' peer assessment of them. However, this might not be appreciated by all students, therefore the class should come to an agreement about how much information should be available. As a way of engaging students, the dashboard could additionally provide motivating sentences, suggested by one of the experts in Section 4.4.2.

How to act on Learning Analytics?
From the evaluation in Section 4.3.1 there were shared doubts about providing students with learning analytics through the dashboard as it might cause confusion. However, it was also stated that through learning analytics, students can become aware of how they are performing and what they need to work on. Students are therefore provided with the learning analytics dashboard because it can promote reflection regarding how they have performed in both the virtual lab and collaboration, and if there are aspects they need to work on.

For instructors, the information on the learning analytics dashboard can help them reflect on the practical assignments and course content based on how the students have performed. If there are specific tasks or concepts students seem to be struggling with, the instructor can alter the course for the next round, making sure the specific concepts are thoroughly covered. This would be beneficial for next year's students. The learning analytics dashboard can additionally promote interventions, assisting students based on where help is needed. This could be regarding issues in collaboration, if certain students are isolated from the group, or regarding the virtual lab and certain aspects that students are struggling with. There were, however, concerns raised regarding interventions as reported in Section 4.3.2, where it was stated that interventions might not be performed in a course with 100 to 200 students. A solution to this could be to share the responsibility amongst teachers and facilitators. Also, if the instructors are equipped with an advising dashboard which provides them with instructions based on the student data, interventions might be more easily accomplished.

As pointed out in the evaluation in Section 4.4.2 all data might not be available until the end of the course, meaning students will not be able to reflect on their performances during the course and interventions might be difficult for instructors to perform. The learning analytics will nonetheless help instructors improve the learning plan and course content for next year's students and support students in identifying their strengths and weaknesses, helping them reflect on what they need to work on for future work, both in collaboration and virtual labs. As suggested in the evaluation in Section 4.3.1, keeping track of collaboration performances throughout an entire program within various courses could be worth considering as well. This could provide a powerful dataset which could stimulate multiple rounds of improved collaboration.

5.8 Chapter Summary

This chapter described the final version of the learning analytics framework and the different sections it contains. Still, there are aspects that are in need of further investigation as it is a complex research topic composed of multiple fields. This is further discussed in the next chapter.

Chapter 6

Discussion

This chapter discusses the guidelines of design science, the findings from the literature review and the artefact development, as well as answering the research questions and presenting the limitations of this research.

6.1 Design Science Guidelines

This research was steered by the guidelines of Design Science Research as described in Chapter 3. *Design as an artefact* indicates that design science should result in a purposeful artefact which aims to solve organisational issues (Hevner et al., 2004). The artefact produced in this research is a learning analytics framework which intends to describe how learning analytics can be implemented to better support learning performances in collaboration and virtual labs. As online learning is expanding, the need for such a framework can inspire the improvement of collaboration and laboratory work in online environments. This again reflects the *problem relevance* in this research.

The guideline describing *design as a search process* indicates that design science is an iterative process which aims to discover solutions to a given problem (Williamson & Johanson, 2017, p.273). This is demonstrated in this research through a preliminary desk research which was conducted to map the current research on the use of learning analytics in a collaborative virtual lab context. This further inspired the development of the learning analytics framework which then went through an iterative artefact development process with *evaluations* involving expert interviews. The evaluation helped demonstrate the relevance and usefulness of the artefact by discussing various aspects of learning analytics and collaborative learning in virtual labs with experts within the field.

Research rigour is essential when constructing an artefact and can be achieved through appropriate use of theories and research methodologies (Hevner et al., 2004). The learning analytics framework was developed through conceptual design and evaluated using a qualitative approach with semi-structured interviews. The semi-structured interviews were chosen as they are guided by a script (see interview guide in Appendix B) and therefore ensured that the same subjects were covered for each interviewee. It also allowed the exploration of opinions through probing.

The research contributions guideline explains how design science has to contribute to the knowledge base either through the artefact, design foundations and/or design methodologies (Hevner et al., 2004). The main contribution of this research is the design artefact, which extends the existing knowledge base by providing a framework for the implementation of learning analytics in a collaborative virtual lab context. The scoping review presented in Chapter 2 is another contribution of this research, providing an overview of the current research within the area.

Lastly, *communication of research* expresses the importance of a detailed and understandable documentation of the artefact development process that enables implementation in organisations as well as further research (Hevner et al., 2004). The literature provides descriptions of the concepts involved in learning analytics, collaboration and virtual labs which provides details for the technology-oriented and management-oriented audiences, helping them decide whether to invest in such an artefact. The processes involved in the development of the artefact has been thoroughly documented which provides understanding of the process as well as repeatability and further extension of the artefact.

6.2 Findings in Literature Review

There is an apparent knowledge gap within learning analytics research with regard to online laboratories and collaboration in labs. Many of the identified studies have not sufficiently developed learning analytics approaches, collaboration, or both. The lack of research in this area could be as a result of collaborative online laboratory environments scarcity. Six years ago, Teng et al. (2016) stated that few such environments allow collaboration within the system. This lack seems to persist, given that the results of this scoping review support this claim. Only one study by Orduña et al. (2014) has sufficiently developed a learning analytics approach for collaboration using social network analysis but doing so on connections outside of the online laboratory environment, as the system itself does not store interaction data between students. Looking at the current work of learning analytics and collaboration in other online learning environments was therefore essential as they provide examples of how it might be implemented in an online lab environment.

With respect to other online learning environments, several studies concerning discussion forums were identified, where the common method used to analyse collaboration amongst students was social network analysis. The use of natural language processing and sentiment analysis were also found to be applied as a way of gaining insight into the discussion and dynamics between students. An issue arising in discussion forums is that few students take an active part in the discussion (Doleck et al., 2021), suggesting that discussion forums are not essential to students' learning. Another possible explanation could be that discussion forums were not required by the instructor or the learning design of the course. An alternative collaborative tool more preferred by the students is worth considering.

Regarding presenting the analysed data to the target group, learning analytics dashboards are valuable tools. Learning analytics dashboards providing the aid of advising proves to be preferred, although teachers have no trouble identifying problematic groups no matter the aid (van Leeuwen et al., 2019). Xing et al. (2015) proposes to combine learning analytics with HCI theory in order to achieve more understandable analytics. Through the use of theory, teachers are provided with a familiar language from which they can more easily draw conclusions.

It was found that the stakeholders identified in the selected studies are often teachers, where the work is aimed at guiding teachers in facilitating collaborative work. There is limited work aimed at presenting collaboration data for students to encourage self-reflection in group work, suggesting more research is required on this subject. Additionally, there seems to be a lack of evaluation amongst some of the learning analytics methods used, where the methods have not been evaluated on the target group, or the evaluation is limited to few subjects. To conclude, the knowledge gap in this scoping review identifies learning analytics to support collaboration in virtual labs as an area for further investigation, supporting the need for this research.

6.3 Findings in Artefact Development

The framework development process consisted of two iterations and a final development based on the findings of both. The literature presented in Chapter 2 inspired the first version of the framework which consisted of the stakeholders (instructor and students), learning environments (a virtual lab platform and a collaboration platform), data, data analysis (statistical analysis, social network analysis, natural language processing and sentiment analysis), and visualisation through a learning analytics dashboard. Based on the feedback from experts in the first iteration, an additional section regarding student feedback was implemented in the framework as this would provide the instructors with the students' own reflections regarding their own performances and that of their peers. It would also promote the act of self-reflection. In addition to this, the learning analytics dashboard was divided into student view and instructor view, as the two stakeholders would need different amounts of information. The act of reflection was also implemented in regards to instructors, as the information on the learning analytics dashboard could help instructors reflect on the student performances and alter the course accordingly. After the second iteration, a learning plan-section was added, containing the learning design and learning goals of the course.

In the first iteration, doubts were shared in regards to providing students with information through the learning analytics dashboard, as this might cause confusion regarding assessment and students comparing themselves to fellow students. This result could correspond to the findings in the literature review, where the identified studies were mainly aimed at guiding teachers in facilitating collaborative work rather than displaying collaboration data to encourage students to self-reflect on group work. On the other hand, this claim is contrary to other feedback from the evaluation which showed that learning analytics can help students become more aware of their performances, and henceforth work on that which needs to be worked on. This is supported by the objectives of learning analytics as described in Khalil & Ebner (2015), which includes the act of reflection as beneficial for all stakeholders involved. There is also a lot of research which supports the benefits of learning analytics dashboards for students as it contributes to self-reflection and self-regulated learning (Corrin & de Barba, 2014; Govaerts et al., 2012; Grann & Bushway, 2014). Therefore, in order to avoid confusion amongst students, it is important to be clear about the purpose of the learning analytics dashboard so that it contributes to the right outcomes. The concern regarding students comparing themselves to other students could be solved by not making class performance data available to the student view, and instead including only each individuals' performances. Another solution could be to illustrate a class average performance instead of individual student performances. In spite of this concern, other feedback suggested that some students may like to see how they have done compared to others. This issue is also presented in Qvist et al. (2015), where they found that some students felt motivated by comparing their efforts to other students, whereas others felt threatened. It is therefore essential that all students agree on whether this is an option or not.

Feedback showed that information on the learning analytics dashboard illustrating time spent on different tasks within the virtual lab, specific questions the students are struggling with, how groups communicate with each other, the nature of the students' discussion and different roles within groups would be appreciated by the instructors. This feedback is consistent with that of Tarmazdi et al. (2015), where the teacher acknowledged the ability to see how groups participated in the group work and tones used in discussion, which allowed the teacher to give feedback to those who needed to participate more. Another interesting finding from the evaluation was the concern regarding interventions and how it might not always be performed by teachers in a class of many students. This concern relates to the challenges reported by Wong & Li (2020), where problems regarding interventions often lie in the restricted time for instructors to handle the problems of students. Interventions are also dependent on the capacity of personnel, the infrastructure and the urgency of the problems of students. As suggested as a solution in Section 5.7, the responsibility of interventions could be shared amongst teachers and facilitators as well as providing advising dashboard making it easier for instructors to know where to assist. Choi et al. (2018) suggest utilising different methods for interventions in accordance with the urgency and importance of the student problems. This way, less costly but more easily distributed methods like email reminders can be used for students who do not have substantial issues, whereas approaches that are expensive and not easily distributed, for instance face-to-face consulting, are reserved for students who are in need.

Feedback from some of the experts showed that future students are those that will benefit most from the learning analytics as it enables the instructor to alter the course based on the performances of the students that current year. This feedback is somewhat contradicting to the objectives of stakeholders in learning analytics presented in Khalil & Ebner (2015), which states that learning analytics are to enhance students' learning performances by informing them about their progress, in addition to providing instructors with the ability to provide real-time feedback to the students. Nonetheless, such analytics is still important as it promotes the ability to map and identify the learning design decisions that instructors are making, as well as how students perform based on these decisions (Macfadyen et al., 2020). This is supported by the objective of *benchmarking* in learning analytics which aims to identify best practices in teaching in order to improve performances, also presented in Khalil & Ebner (2015). Additionally, findings from Wasson & Kirschner (2020) states that more research on how to best support instructors in learning design work is needed.

In the second iteration, a concern was raised regarding the privacy of students in this context, pointing out that students might not be comfortable with their discussion data being collected and analysed. Botnevik (2021) has further explored how students perceive privacy in learning analytics, and found that privacy is perceived as highly important and that data security should be of high priority in learning analytics. Most students would accept learning analytics to be implemented at their higher education institution, but consent should be required and their privacy respected. Providing transparency is therefore essential. Still, this thesis lacks evidence of how students perceive the collection and analysis of their data in a collaborative context as it is beyond the scope of the current research. There is also lack of evidence as to what students would like to gain from learning analytics in a collaborative virtual lab context, which is an important issue for future research.

It was found through the evaluation that only one expert out of seven had experience with virtual labs that allow collaboration within the same platform. These results corroborate with the findings in the literature review, which showed that few online laboratory environments allow collaboration within them. Since most experts have experienced that collaborative work and practical work in virtual labs take place separately, a plan for how to best facilitate such a scenario is beneficial.

6.4 Answers to Research Questions

The purpose of this research was to gain an understanding of how to support and improve collaboration in virtual labs through learning analytics. The findings from the literature review and the artefact development process made it possible to answer the research questions.

RQ1: What is the current state of research on the use of learning analytics to understand and support collaboration in virtual labs?

The current state of research on this area was found to have an apparent gap of knowledge, indicating the need and motivation for this research. Merely one study included all elements of interest: learning analytics, collaboration, and virtual labs. Other learning environments were therefore explored to discover ways of facilitating collaboration which could be implemented in a collaborative virtual lab context. It showed that data analysis methods such as social network analysis, natural language processing and sentiment analysis are valuable methods for gaining an understanding of collaboration. Learning analytics dashboards were discovered to be a good way of visualising the analysis to the stakeholders. A more thorough discussion of the findings is presented in Section 6.2.

RQ2: How can learning analytics support collaboration in virtual labs?

Based on the findings in the literature review, social network analysis is a valuable method in discovering patterns within collaboration among students. This was shown in Orduña et al. (2014) who developed a learning analytics approach for collaboration using social network analysis on connections outside of the online laboratory environment, as the system itself does not store interaction data between students. The goal was to present an analysis of the social networks which could support instructors in identifying social dynamics within courses. Other ways of discovering patterns in collaboration proved to be achieved through natural language processing, text classification and sentiment analysis. The literature review further revealed that learning analytics dashboards are a valuable tool for visualising the analysis to stakeholders, which can promote self-reflection and intervention based on the performances of students in collaboration and virtual labs. These findings together with the feedback from the evaluations helped inspire a framework for how the use of learning analytics can support collaboration in virtual labs (see Figure 5.1). To summarise the framework: Students perform practical assignments in the virtual lab platform and participate in group discussion separately in a collaboration platform. From these learning environments the student performance data will be gathered and stored in a database together with the learning plan and student feedback-dataset. The data is then processed and cleaned for various data analysis methods to be performed and for patterns to be revealed. These analysis methods include statistical analysis, social network analysis, natural language processing and sentiment analysis, as these was shown through the literature review to be of value in supporting collaboration in a virtual lab environment. The analysis will then be implemented and visualised to the stakeholders (students and instructor) through a learning analytics dashboard, containing two views of information: student view and instructor view. The information on the dashboard should inform instructors about all students' performances in the virtual lab and in the collaborative work, helping them reflect on their teaching methods and provide feedback to students in need of assistance. Students should be provided with their own performances within the virtual lab and collaborative work, helping them reflect on their work in order to improve future learning in both areas (e.g., how to collaborate better, redoing the lab to improve results, etc.).

RQ3: What aspects are important when designing learning analytics implementations for collaboration in virtual labs?

In order to design a learning analytics framework which intends to support and improve collaboration in virtual labs, background literature and previous research had to be investigated. Khalil & Ebner (2015) presents a learning analytics life cycle of the processes that are involved in learning analytics, which helped inspire the first version of the learning analytics framework. The essential elements involved in the processes of learning analytics include the stakeholders, learning environments, data produced in the learning environments, different analytical methods, and the objectives of learning analytics (Khalil & Ebner, 2015). What needed to be considered was who are the stakeholders involved, what are the learning environments, what data needs to be collected, what analysis methods should be applied and how should the results of the analysis achieve the objectives of learning analytics in a collaborative virtual lab context. Through the literature review and the evaluation of the framework, these aspects were further investigated.

The stakeholders needed to be identified as they affect other aspects such as what data to collect and how to utilise this data. For this research, the stakeholders in question are students and instructors, where the students act as a main stakeholder and the instructor as an assisting stakeholder. What learning environments to utilise is an important aspect as well. Found through the literature review, few online lab environments allow collaboration within the online lab. It was therefore assumed that practical assignments and collaboration work will take place separately. Further discovered in the evaluation is that the choice of virtual lab platform will affect the collaborative work. This feedback is consistent with what was found in Kent & Cukurova (2020), which showed that when students felt comfortable with the collaborative tasks, the collaboration within groups were improved. It is therefore essential that the choice of platforms facilitate both learning in practical assignments within virtual labs as well as promote collaborative work and discussion among students. It was also found that few students take active part in discussion forums (Doleck et al., 2021), suggesting that other collaboration platforms more prefered by students should be considered.

The data retrieved from these learning environments needs to be of relevance for the course in question, and it must support the improvement of collaboration in virtual labs. Therefore, interaction data involving the engagement between students needs to be collected from the collaboration platform, together with trace data involving trails of the students within the virtual lab platform such as mouse clicks and number of finished assignments. Based on the feedback from the evaluation, the data of relevance would be time spent on tasks within the virtual lab, attempts used on tasks within the virtual lab, the text students write to each other in the collaboration platform, who is writing what and how much each student contributes in collaboration. The data from the virtual lab and from the collaboration platform could also be connected through time stamps to see how different events within the different platforms relate to each other. Privacy is also an important aspect, in which students perceive to be highly important (Botnevik, 2021). It is therefore a necessity to provide transparency and privacy of student data when learning analytics is implemented.

The choice of data analysis methods also needs to support the improvement of collaboration in virtual labs. It was found through the literature that statistical analysis, social network analysis, natural language processing and sentiment analysis are valuable methods for discovering patterns within collaboration in virtual labs. Further, how the analysis is visualised to the stakeholders are of great importance as it will affect how the analysis is interpreted and used to achieve the objectives of learning analytics. A learning analytics dashboard was chosen for visualisation in this context as it offers an easy insight and provides good visibility of the analysis (Khalil & Ebner, 2015). The dashboard itself will consist of two different views: one student view and one instructor view, and each view will hold two categories of analytics: one regarding student performance in the virtual lab and the other regarding student performance in collaboration. Based on feedback, valuable information for the instructor regarding the virtual lab would be time spent on tasks, measurements of performance of each individual student in the virtual lab, if there are different concepts they are struggling with, and how specific questions and concepts have performed in terms of all students. Regarding collaboration, valuable information would be the nature of students' discussion within each group, the different roles within groups and how they communicate with each other. The student view would contain information about their own performances within the virtual lab, which concepts within the course they are struggling with and how they have contributed to the collaboration. Some students might wish to compare themselves to the rest of the class. This should however be optional (i.e., the learning analytics dashboard could allow students to select this view if they wanted) as other students might not appreciate this information. The information on the dashboard could help achieve the objectives of reflection and intervention. The dashboard can support students in reflecting on their previous performances to improve future learning. It can support instructors in performing interventions and providing assistance to students, as well as reflecting on teaching methods based on how students have performed.

The evaluation also revealed the need for student feedback, where students will carry out self-assessment and peer assessment. This would also provide the instructors with more rich data about the students' own experience within the course as well as promoting self-reflection for the students. Another important aspect discovered was the need for a learning plan. The learning plan will decide how the course is set up and affect the other aspects, such as what learning environments will be used, how the collaboration is structured and what the students are to learn.

6.5 Limitations

There are several limitations in this research. The decision to use one database for the literature review implies that there is a possibility that valuable sources of information have been overlooked. Also, the publications used for the literature review did not always extensively address all the desired topics, that being learning analytics, collaboration, and virtual labs. An extensive search through other various databases could possibly yield more satisfying results.

The evaluations in the artefact development also had its limitations. The expert evaluations rendered the perspective of instructors who had experience with virtual labs and teaching. A broader scope of experts within various areas could perhaps yield more detailed feedback regarding each section within the framework. Furthermore, this research is missing the perspective of students and how they would perceive the implementation of learning analytics in a collaborative virtual lab context. Involving students could broaden the scope of this research and provide a more extensive view.

6.6 Chapter Summary

This chapter discussed the design science guidelines within this research, the findings of the literature review and the artefact development, as well as answering the research questions and stating the limitations of this thesis.

Chapter 7

Conclusion

This research has explored how learning analytics can be utilised to support collaboration in virtual labs. This was carried out within the frames of design science research, where the main contribution was an artefact in the form of a learning analytics framework (Figure 5.1). Preliminary to the artefact development, a literature review was performed in order to map current research on the topic and inspire the development of the artefact. The literature review is another contribution of this thesis, and a contribution to the existing knowledge base. It revealed that there is a gap of knowledge within the area of learning analytics in collaborative virtual lab environments, in which this research has worked towards reducing. This thesis is also a contribution to the European ERASMUS+ ENVISON_2027 project outputs.

The artefact development consisted of two iterations in which experts were included in each iteration to help evaluate and identify important and challenging aspects within the framework. The framework firstly consisted of five main parts: stakeholders, learning environments, data, data analysis and visualisation. The framework describes the following: the stakeholders include students and instructors, in which students act as the main stakeholder and the instructor as the assisting stakeholder. The learning environment includes a virtual lab platform and a collaboration platform, where the students perform practical assignments in the virtual lab platform and participate in group discussion separately in a collaboration platform. The datasets produced in these learning environments are stored in a database. The data is then processed and cleaned before going through data analysis. The data analysis methods include statistical analysis, social network analysis, natural language processing and sentiment analysis. The analysed data are then visualised to the stakeholders through a learning analytics dashboard. The dashboard should present information about students which will help them reflect on their performances in virtual labs and collaboration. Additionally, it should help instructors reflect on their teaching methods and provide feedback to students in need of assistance.

Through the artefact development process, it was found that implementing *student feedback*, which involves the act of self-assessment and peer assessment, is a valuable extension as it promotes reflection with students as well as providing instructors with more rich student data. Concerns were raised regarding how collaboration will be structured, as this will be affected by which learning environments are utilised. By implementing a *learning plan*, this issue could be solved as it will define how the course will be structured, what platforms will be utilised and how collaboration will be organised. Feedback regarding the learning analytics dashboard indicated that students and instructors should have different views of information. Students should be presented with information regarding their own performances in the virtual lab and in the group work. Instructors should get information regarding all students' performances in the virtual lab and in the group work, as well as how course content and assignments have performed in terms of all students. Challenges regarding privacy were discovered, indicating that transparency and privacy needs to be provided in this context. Students need to be aware of how and why their data is being collected and analysed.

The learning analytics framework proved to be understandable for all experts involved in the evaluation. The framework provides a plan for how learning analytics can be utilised to support and improve collaboration in virtual labs in practice. It is however a complex model involving various aspects which would need even further investigations.

7.1 Future Work

The contribution of this research provides opportunities for future work. An essential part of future work is to put the framework into practice. This would involve developing an infrastructure for the gathering and analysing of data from a course which utilises a virtual lab and a collaboration platform as described in the framework. Future research should investigate the utilisation of the aforementioned data analysis methods on data from virtual labs and collaboration platforms by the various stakeholders. As suggested by one of the experts, a predictive model could also be implemented to help discover which students will have success in collaboration and who will work well together. Such information could help instructors identify which students need to work on their collaborative skills and assist thereafter. It could also help the instructor with pairing students who are compatible, which can support students in focusing more on learning rather than getting along.

Another important component for future research will be the creation of a learning analytics dashboard. Substantial research will be needed in terms of what data and data analysis should be presented on the dashboard to support the objectives of learning analytics. This would require consultation and evaluation with students and instructors to discover what they would prefer from such a dashboard.

Future work should also consult students about their opinions on privacy in such a collaborative virtual lab context. Additionally, other stakeholders could be involved to explore the benefits for them, in addition to students and instructors. These stakeholders could include faculty leaders, researchers, and designers of the learning platforms.

References

Alam, F., Hadgraft, R. G., & Subic, A. (2014). Technology-enhanced laboratory experiments in learning and teaching. In *Using technology tools to innovate assessment, reporting, and teaching practices in engineering education* (pp. 289–302). IGI Global. doi: https://doi.org/10.4018/978-1-4666-5011-4.ch021
2.3

Alkhaldi, T., Pranata, I., & Athauda, R. I. (2016). A review of contemporary virtual and remote laboratory implementations: observations and findings. *Journal of Computers in Education*, *3*(3), 329–351. doi: https://doi.org/10.1007/s40692-016-0068-z
1, 2.3, 2.3.2

Andress, J. (2014). Chapter 1 - what is information security? In J. Andress (Ed.), *The basics of information security (second edition)* (Second Edition ed., p. 1-22). Boston: Syngress. doi: https://doi.org/10.1016/B978-0-12-800744-0.00001-4
2.1.3

Atif, A., Richards, D., Bilgin, A., & Marrone, M. (2013). Learning analytics in higher education: a summary of tools and approaches. In *Ascilite-australian society for computers in learning in tertiary education annual conference* (pp. 68–72).
2.1.1

Birkeland, H., Khalil, M., & Wasson, B. (2022). Learning analytics in collaborative online lab environments: A systematic scoping review. *Companion Proceedings of the 12th*, 92.
(document), 2.4, 2.4, 2.4.1, 2.4.3, 2.4.3

Botnevik, S. (2021). *Student perceptions of privacy in learning analytics: A quantitative study of norwegian students* (Unpublished master's thesis). The University of Bergen.
6.3, 6.4

Choi, S. P., Lam, S. S., Li, K. C., & Wong, B. T. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Journal of Educational Technology & Society*, *21*(2), 273–290.

6.3

Corrin, L., & de Barba, P. (2014). Exploring students' interpretation of feedback delivered through learning analytics dashboards. In *Proceedings of the ascilite 2014 conference.*
6.3

Dillenbourg, P. (1999). *What do you mean by collaborative learning?* Citeseer.
2.2, 4.3

Doleck, T., Lemay, D. J., & Brinton, C. G. (2021). Evaluating the efficiency of social learning networks: Perspectives for harnessing learning analytics to improve discussions. *Computers Education*, *164*, 104124. doi: https://doi.org/10.1016/j.compedu.2021.104124
2.1, 2.4.2, 4.3, 6.2, 6.4

Dresch, A., Lacerda, D. P., & Antunes, J. A. V. (2015). Design science research. In *Design science research* (pp. 67–102). Springer.
(document), 3.1, 3.1, 3.4

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5-6), 304–317. doi: http://dx.doi.org/doi:10.1504/IJTEL.2012.051816
2.1

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., ... LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Phys. Rev. ST Phys. Educ. Res.*, *1*. doi: https://doi.org/10.1103/PhysRevSTPER.1.010103
2.3.2

Given, L. M. (2008). *The sage encyclopedia of qualitative research methods.* Sage publications.
2.1.1

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). The student activity meter for awareness and self-reflection. In *Chi '12 extended abstracts on human factors in computing systems* (pp. 869–884). New York, NY, USA: Association for Computing Machinery. doi: https://doi.org/10.1145/2212776.2212860
6.3

Grann, J., & Bushway, D. (2014). Competency map: Visualizing student learning to promote student success. In *Proceedings of the fourth international conference on*

learning analytics and knowledge (pp. 168–172). New York, NY, USA: Association for Computing Machinery. doi: https://doi.org/10.1145/2567574.2567622
6.3

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 75–105. doi: https://doi.org/10.2307/25148625
3.1.1, 3.1.1, 6.1

Hu, X., Le, H., Bourgeois, A. G., & Pan, Y. (2018). Collaborative learning in cloud-based virtual computer labs. In *2018 ieee frontiers in education conference (fie)* (pp. 1–5).
4.3

Jara, C. A., Candelas, F. A., Torres, F., Dormido, S., Esquembre, F., & Reinoso, O. (2009). Real-time collaboration of virtual laboratories through the internet. *Computers & Education*, *52*(1), 126–140. doi: https://doi.org/10.1016/j.compedu.2008.07.007
4.3

Jones, K. M., Thomson, J., & Arnold, K. (2014). Questions of data ownership on campus. *EDUCAUSE Review*.
2.1.3

Kent, C., & Cukurova, M. (2020). Investigating collaboration as a process with theory-driven learning analytics. *Journal of Learning Analytics*, *7*(1), 59–71. doi: https://doi.org/10.18608/jla.2020.71.5
2.1, 2.4.2, 6.4

Khalil, M., & Ebner, M. (2015). Learning analytics: principles and constraints. In *Edmedia+ innovate learning* (pp. 1789–1799).
(document), 2.1.1, 2.1, 2.1.1, 2.1.3, 2.2, 2.1.3, 2.4.3, 4.3, 6.3, 6.4

Lauricella, S., & Kay, R. (2013). Exploring the use of text and instant messaging in higher education classrooms. *Research in Learning Technology*, *21*. doi: https://doi.org/10.3402/rlt.v21i0.19061
4.3, 5.3

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher educationa literature review. *Learning analytics: Fundaments, applications, and trends*, 1–23. doi: https://doi.org/10.1007/978-3-319-52977-6_1
2.1.2, 2.4.4

Macfadyen, L. P., Lockyer, L., & Rienties, B. (2020). Learning design and learning analytics: Snapshot 2020. *Journal of Learning Analytics*, *7*(3), 6–12. doi: https://doi.org/10.18608/jla.2020.73.2

6.3

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in education: principles, policy & practice*, *10*(2), 209–220. doi: https://doi.org/10.1080/0969594032000121289
4.4.1

Orduña, P., Almeida, A., Ros, S., López-de Ipina, D., & Garcia-Zubia, J. (2014). Leveraging non-explicit social communities for learning analytics in mobile remote laboratories. *J. Univers. Comput. Sci.*, *20*(15), 2043–2053.
(document), 2.1, 2.4.1, 2.7, 6.2, 6.4

Park, Y., & Jo, I.-H. (2015). Development of the learning analytics dashboard to support students learning performance. *Journal of Universal Computer Science*, *21*(1), 110.
2.1.1

Peters, M. D., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Chapter 11: scoping reviews (2020 version). *JBI manual for evidence synthesis, JBI*.
2.4

Peters, M. D., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *JBI Evidence Implementation*, *13*(3), 141–146.
2.4, 3.2

Pillutla, V. S., Tawfik, A. A., & Giabbanelli, P. J. (2020). Detecting the depth and progression of learning in massive open online courses by mining discussion data. *Technology, Knowledge and Learning*, *25*(4), 881–898. doi: https://doi.org/10.1007/s10758-020-09434-w
2.1, 2.4.2

Prinsloo, P., Slade, S., & Khalil, M. (2022). The answer is (not only) technological: Considering student data privacy in learning analytics. *British Journal of Educational Technology*. doi: https://doi.org/10.1111/bjet.13216
2.1.3

Quan-Haase, A. (2008). Instant messaging on campus: Use and integration in university students' everyday communication. *the information society*, *24*(2), 105–115.
4.3

Qvist, P., Kangasniemi, T., Palomäki, S., Seppänen, J., Joensuu, P., Natri, O., … Nordström, K. (2015). Design of virtual learning environments: Learning analytics

and identification of affordances and barriers. *International Journal of Engineering Pedagogy (iJEP)*, *5*(4), 64–75. doi: https://doi.org/10.3991/ijep.v5i4.4962
2.1, 2.4.1, 6.3

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12–27. doi: https://doi.org/10.1002/widm.1075
2.1

Romero, S., Guenaga, M., Garca-Zuba, J., & Orduña, P. (2015). Automatic assessment of progress using remote laboratories. *Int. J. Online Eng.*, *11*(2), 49–54. doi: https://doi.org/10.3991/ijoe.v11i2.4379
2.1, 2.4.1

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In *2013 international conference on collaboration technologies and systems (cts)* (pp. 42–47).
2.1

Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., Rodríguez-Díez, J. J., Rodríguez-Arribas, S., Díez-Pastor, J. F., & Ji, Y. P. (2021). Improve teaching with modalities and collaborative groups in an lms: An analysis of monitoring using visualisation techniques. *Journal of Computing in Higher Education*, *33*(3), 747–778. doi: https://doi.org/10.1007/s12528-021-09289-9
(document), 2.1, 2.4.3, 2.8

Sharp, H., Rogers, Y., & Preece, J. (2015). *Interaction design: beyond human-computer interaction*. Wiley. 5th edition.
3.3, 3.4.1, 3.4.3

Simon, H. A. (1996). *The sciences of the artificial (3rd ed)*. MIT Press.
3.1.1

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 235–244). doi: https://doi.org/10.1145/3303772.3303796
2.1.3

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of business research*, *104*, 333–339. doi: https://doi.org/10.1016/j.jbusres.2019.07.039
3.2

SoLAR. (2021). *What is learning analytics?* Retrieved from https://www

.solaresearch.org/about/what-is-learning-analytics/ (Accessed: 2021-05-20)
2.1

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In *Cambridge handbook of the learning sciences* (p. 409-426). Cambridge University Press.
2.2

Tarmazdi, H., Vivian, R., Szabo, C., Falkner, K., & Falkner, N. (2015). Using learning analytics to visualise computer science teamwork. In *Proceedings of the 2015 acm conference on innovation and technology in computer science education* (pp. 165–170). doi: https://doi.org/10.1145/2729094.2742613
(document), 2.1, 2.4.3, 2.9, 4.3, 6.3

Teng, M., Considine, H., Nedic, Z., & Nafalski, A. (2016). Current and future developments in the remote laboratory netlab. *International Journal of Online Engineering*, *12*(8). doi: https://doi.org/10.3991/ijoe.v12i08.6034
1, 2.1, 2.4.1, 6.2

Topping, K. J. (2009). Peer assessment. *Theory into practice*, *48*(1), 20–27.
4.4.1

van Leeuwen, A., Rummel, N., & Van Gog, T. (2019). What information should cscl teacher dashboards provide to help teachers interpret cscl situations? *International Journal of Computer-Supported Collaborative Learning*, *14*(3), 261–289. doi: https://doi.org/10.1007/s11412-019-09299-x
2.1, 2.4.3, 4.3, 6.2

Wasson, B., & Kirschner, P. A. (2020). Learning design: European approaches. *TechTrends*, *64*(6), 815–827. doi: https://doi.org/10.1007/s11528-020-00498-0
6.3

Wiesner, T. F., & Lan, W. (2004). Comparison of student learning in physical and simulated unit operations experiments. *Journal of Engineering Education*, *93*(3), 195–204. doi: https://doi.org/10.1002/j.2168-9830.2004.tb00806.x
2.3.2

Williamson, K., & Johanson, G. (2017). *Research methods: Information, systems, and contexts*. Chandos Publishing.
3.1.1, 3.1.1, 6.1

Wong, B. T.-m., & Li, K. C. (2020). A review of learning analytics intervention in

higher education (2011–2018). *Journal of Computers in Education*, *7*(1), 7–28. doi: https://doi.org/10.1007/s40692-019-00143-7
6.3

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, *47*, 168–181. doi: https://doi.org/10.1016/j.chb.2014.09.034
2.1, 2.4.4, 6.2

Appendix A - Consent Form

Are you interested in taking part in the research project

"Understanding Collaboration in Virtual Labs: A Learning Analytics Framework"?

This is an inquiry about participation in a research project where the main purpose is to gain understanding of how learning analytics can be used to understand collaborative work in virtual labs. In this letter we will give you information about the purpose of the project and what your participation will involve.

Purpose of the project

This research project is part of a master's study at the Department of Information and Media Science at the University of Bergen. The purpose of this study is to gain an understanding of how learning analytics can be used to understand collaborative work in virtual labs.

During this research, an artefact in the form of a framework for the use of learning analytics in virtual labs and collaboration will be developed. We therefore find it necessary to get expert feedback in order to evaluate the framework. Feedback from experts within learning analytics are important as they may provide valuable insight into essential parts of the framework.

Who is responsible for the research project?

The department of Information and Media Science at the University of Bergen is the institution responsible for the project.

Why are you being asked to participate?

We would like to gain knowledge of the use of learning analytics in virtual labs and collaboration. We would therefore like to know what experiences you have with learning analytics and the use of virtual labs. Based on your experience and background within this area, we would like to know your opinions on the developed framework. This way we can evaluate the framework, and further develop it for the purpose of facilitating learning in said environment based on expert opinion.

Why are you being asked to participate?
We would like to gain knowledge of the use of learning analytics in virtual labs and collaboration. We would therefore like to know what experiences you have with learning analytics and the use of virtual labs. Based on your experience and background within this area, we would like to know your opinions on the developed framework. This way we can evaluate the framework, and further develop it for the purpose of facilitating learning in said environment based on expert opinion.

What does participation involve for you?
The participants will participate in an interview involving questions about their knowledge on learning analytics, virtual labs, and collaboration, in addition to questions regarding their opinions on the developed framework. The interviews will be recorded on audio tape and written notes might be taken along the way. The interviews will take approximately 30-40 minutes.

Participation is voluntary
Participation in the project is voluntary. If you chose to participate, you can withdraw your consent at any time without giving a reason. All information about you will then be made anonymous. There will be no negative consequences for you if you chose not to participate or later decide to withdraw.

Your personal privacy – how we will store and use your personal data
We will only use your personal data for the purpose(s) specified in this information letter. We will process your personal data confidentially and in accordance with data protection legislation (the General Data Protection Regulation and Personal Data Act).

All personal information will be treated confidentially. Only the student and supervisor of the project have access to personal information. Personal information (both in writing and in the form of audio recordings) is not stored directly by name, names will be replaced with a reference number. List of names with reference numbers will be stores on an external storage device. Participants in the study will not be recognized in the publication.

What will happen to your personal data at the end of the research project?
The project is scheduled to end June 1st 2022. The data will be stored up until one month after the project has finished. After this time, all files with personal information will be deleted. List of names and reference numbers will be deleted and all audio recordings together with notes taken during the interviews will be deleted and shredded.

Your rights
So long as you can be identified in the collected data, you have the right to:
- access the personal data that is being processed about you
- request that your personal data is deleted
- request that incorrect personal data about you is corrected/rectified
- receive a copy of your personal data (data portability), and
- send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

What gives us the right to process your personal data?
We will process your personal data based on your consent.

Based on an agreement with the Institution of Information Science and Media Studies, Data Protection Services has assessed that the processing of personal data in this project is in accordance with data protection legislation.

Where can I find out more?
If you have questions about the project, or want to exercise your rights, contact:

- Institution of Information Science and Media Studies via supervisor *Mohammad Khalil, email: mohammad.khalil@uib.no, telephone: +47 41 20 92 72.* Student *Hanna Birkeland, email: hbi008@uib.no, telephone: +47 97 42 22 40*
- Our Data Protection Officer: *Janecke Helene Veim, email: janecke.veim@uib.no , telephone: +47 55 58 20 29*
- Data Protection Services, by email: *personvernombud@uib.no* or by telephone: +47 55 58 20 29.

Yours sincerely,

Mohammad Khalil (Researcher/supervisor)

Hanna Birkeland

Consent form

I have received and understood information about the project *Understanding Collaboration in Virtual Lab: A Learning Analytics Framework* and have been given the opportunity to ask questions. I give consent:

- ☐ to participate in the interview
- ☐ to provide feedback on the developed framework
- ☐ for information about my educational background and profession to be published in the thesis

I give consent for my personal data to be processed until the end date of the project, approx. 1. June

----

(Signed by participant, date)

Appendix B - Interview Guide

INTERVIEW GUIDE

Introduction
The framework will first be presented to the participants, then the following questions concerning the framework and different sections within the framework will be asked:

Background questions
1. What is your educational background?

2. What is your title?

3. Have you used learning analytics in your routine work or research, or both?

Key questions regarding the framework
1. Based on your background and experience, is there anything missing in the framework?

2. Results from the literature review suggest that few learning platforms integrate both virtual labs and collaborative features. Have you experienced a virtual lab platform which includes collaboration features?

3. What kind of data is generated in virtual labs?
   a) What kind of data needs to be collected to support collaboration in virtual labs?

   b) What is your opinion on privacy and ethics regarding collecting student data?

4. What kind of virtual labs are relevant for such an environment?

5. What is your opinion on the choice of data analysis methods?

6. Based on your experience, what should be presented on the learning analytics dashboard?

7. Why are learning analytics needed for virtual labs and collaboration?
   a) How may learning analytics and data help engage the students?

   b) How may learning analytics and data prevent students dropping out?

   c) How may learning analytics and data improve collaboration between students?

8. What kind of stakeholders do you think should be involved in this context?

9. What is your opinion on the representation of the framework in such a visual diagram?

Closing question

1. Do you have any questions, or anything you would like to add?