Teamwork Quality and Project Success in a Software Engineering Capstone Course

A Case Study

Steffen Almås

Thesis submitted for the degree of
Master in Informatics: Programming and System Architecture
60 credits

Department of Informatics
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

Spring 2021

Teamwork Quality and Project Success in a Software Engineering Capstone Course

A Case Study

Steffen Almås

Teamwork Quality and Project Success in a Software Engineering Capstone Course

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

Background: Teamwork is essential in agile software development. Due to the COVID-19 pandemic, student teams in a software engineering capstone course at the University of Oslo were affected as they suddenly had to work digitally instead of meeting physically at the campus.

Aim: This thesis aims to investigate and recognize the success factors of student teams working virtually. The study contributes to the research field by exploring how teamwork quality and the relation to project success in student teams working digitally were affected by the COVID-19 pandemic.

Method: A mixed-methods case study was conducted; the qualitative data give context to the quantitative findings. The quantitative data were analyzed from comprehensive surveys representing 595 individual respondents in 126 student development teams. The qualitative data includes literature reviews, eight semi-structured interviews, and observations.

Results: The results showed that the virtual teamwork worked well. A positive relationship between teamwork quality and project success was found. Well-performing student teams had a facilitator providing good communication and progress in the teamwork.

Conclusion: The student teams seemed to find adequate collaboration tools to help facilitate the teamwork virtually. The teams were able to carry out the teamwork, and the final products were satisfactory. The motivation among the students remained high, although they had to collaborate virtually. However, making friendship between the team members seemed to be difficult.

Acknowledgements

Writing this master thesis has been a rewarding yet challenging experience. First of all, I would like to thank my supervisor, Yngve Lindsjørn, for his valuable guidance, support, and encouraging discussions. I would also like to thank the Software Engineering research group at IFI, particularly Viktoria Stray, for their exceptional input.

I want to thank my great friend Erlend Stenlund for five fantastic years at IFI; thank you for sharing this exciting adventure with me (and the memes).

Further, I would like to thank all participants of the study, both the survey respondents and especially the interviewees. Lastly, I would like to thank my family for their continuous support.

Contents

List of Tables

List of Figures

Abbreviations

API = Application Programming Interface

Design = The study program “Design, Use, Interaction”

Digec = The study program “Digital Economics and Leadership”

IN2000 = The course id for the capstone course called “Software Engineering with project work” which is studied in this thesis

MET = The Norwegian Meteorological Institute

PO = Product Owner

Prosa = The study program “Programming and System Architecture”

Robotics = The study program “Robotics and Intelligent Systems”

TA = Teaching Assistant. They were responsible for supervising the student teams

TL = Team Leader

TWQ = Teamwork Quality

UiO = The University of Oslo

1 Introduction

Teamwork is crucial in software development (Chow & Cao, 2008), especially in agile teams where "individuals and interactions" are more important than "processes and tools" (Manifesto, 2001). A team comprises three or more diverse people who collaborate to achieve a common goal (Hoegl & Gemuenden, 2001; Rubin, 2012, p. 209-210). Agile methods and practices provide values and principles for producing software rapidly while effectively responding to changes in requirements. Agile software development is now the common way of working. Starting in 2018, the University of Oslo offers a 20 ECTS software engineering capstone course where the students work on a project in agile teams. The students experience relevant aspects of software development throughout the project work, such as requirement collection and analysis, design, programming, testing, and maintenance (IN2000, n.d.). The students participate in multidisciplinary teamwork and deal with requirements from a real customer. During the project work in the spring of 2020, the teamwork became virtual overnight as the Norwegian Government closed down the society to prevent the Coronavirus from spreading. The students had to swiftly adapt to the new work environment and use tools to collaborate virtually.

1.1 Motivation

Through searches in the academic literature, some previous research was found on teamwork in software engineering capstone courses in higher education (Bastarrica et al., 2017; Paasivaara et al., 2018). The research conducted so far in this field mainly focuses on the importance of capstone courses, reporting that students are better prepared for the industry after participating in one (Radermacher et al., 2014). A case study conducted on the capstone course offered by the University of Oslo reports that high-performing student teams are multidisciplinary and have meetings regularly (Tegelaár, 2020). Some research has been conducted on online courses, and capstone courses offered as online courses (Richardson & Swan, 2003; Tappert & Stix, 2010); however, these focus on courses that are supposed to be online in the first place.

There has been little research on student teams in software engineering capstone having to adapt their workday from physically collaborating to suddenly working remotely from home. This master thesis is carried out as a case study and uses data gathered before and during the COVID-19 pandemic. This thesis investigates how the collaboration within students' teams in a software engineering course emerged from the sudden change to virtual collaboration.

Some previous research on software development teams working in the industry has been conducted (Lindsjørn et al., 2016). Investigating whether or not student teams are comparable to teams in the industry is highly relevant. This thesis also explores to what degree the team aspects reported to be important in industry teams also apply to student teams.

1.2 Research Questions

This thesis aims to investigate the teamwork in a software engineering capstone course offered by the University of Oslo. The purpose of this thesis is to understand what characterizes a well-performing student team. The thesis investigates how the teams were affected when the collaboration became virtual as a result of the COVID-19 lockdown. Exploring how the teams were able to utilize tools to aid the virtual teamwork is essential. The first research question of this thesis addresses key aspects within well-performing student teams working virtually:

- RQ1: ***"What characterizes a well-performing student team in a software engineering capstone course working virtually?"***

The second research question addresses how the use of digital collaboration tools affected the virtual teamwork:

- RQ2: ***"What role do digital collaboration tools play in student teams working virtually in software engineering capstone?"***

Both research questions address teams working virtually. An assumption is that the result of this study will be adequate in the future, though the student teams will be able to work physically again. The teamwork of agile software teams in both student and professional teams will likely be more flexible regarding collaboration within the teamwork. For example, many student teams in such capstone courses as in this study will conduct many meetings digitally instead of always have meetings physically on campus.

1.3 Thesis Structure

The rest of the thesis is structured as follows:

Section 2: Background outlines the relevant theoretical background within the scope of this thesis.

Section 3: Research Context presents the case studied, which is the software engineering capstone course offered by the University of Oslo.

Section 4: Research Method describes the chosen case study design of this thesis, data collection methods, validity, and reliability.

Section 5: Results presents the findings with regard to the research questions presented in Section 1.

Section 6: Discussion discusses the results in relation to the research questions and previous findings.

Section 7: Conclusion concludes the thesis and suggests rationales for future research.

2 Background

This section presents the background and previous findings relevant to this study. First, software engineering capstone courses are presented. Secondly, agile methodologies and popular agile practices. Then, team performance models and the Sociability scale are presented. Finally, a brief overview of some findings on the effect of the COVID-19 on higher education in Norway is given.

2.1 Software Engineering Capstone Courses

To reduce the skill gap between graduating students and industry expectations, the use of capstone courses in higher education software engineering has seen steady growth over the last few decades (Paasivaara et al., 2018; Radermacher et al., 2014). A capstone course, sometimes referred to as 'capstone project' or 'final year project', serves as a culminating and unifying course where the students demonstrate acquired knowledge throughout their degree. In software engineering capstone courses, the students often deal with real problems similar to those they will face in the industry (Bastarrica et al., 2017). The most common way to offer a software engineering capstone course is to offer it as a final project to finish a degree (Mahnic, 2012; Umphress et al., 2002). The capstone course offered at the University of Oslo differs from this, as it is offered for students in their fourth semester, and the students still have two more semesters left before finishing their degrees. More on the capstone course at the University of Oslo in Section 3.

The general findings on software engineering capstone courses are that they provide students with desirable knowledge and experiences which will be helpful later in their careers. Teamwork and involving the industry are essential aspects that students experience in capstone courses (Dzvonyar et al., 2018). A reported finding is that the software companies hiring graduate students look for the following qualities: programming knowledge, the willingness to learn new things, and communication skills (Stålhane et al., 2020). Therefore, most capstone courses expose students to such experiences. In addition, experiencing communication is found to be especially important in capstone courses (Majanoja & Vasankari, 2018).

A case study on the software engineering capstone course offered by the University of Oslo was conducted in 2020. The study mainly focused on the student teams' performance related to what score they received in the project, reporting that teams that perceived their teamwork as good also received a higher project score (Tegelaár, 2020). Tegelaár (2020) mainly applied data gathered during the 2019 semester when the course and teamwork were carried out physically on campus.

A study conducted on an online software engineering capstone course reports some key attributes to successfully facilitate remote teamwork: having digital photos of students, meeting physically three times, and having a website containing relevant updated information (Tappert & Stix, 2010). The study investigated a capstone course being offered as the final course in a master's degree. The teamwork in the course is carried out digitally, but Tappert (2010) still emphasizes the importance of the team members meeting a few times physically throughout the semester.

2.2 Agile Methodology

Leading developers made the Agile Software Development Manifesto as a response to dissatisfaction with plan-driven software development (Manifesto, 2001). The Agile Manifesto consists of four values and twelve principles that are made to help software development teams preserve the agile mindset. The use of agile methodologies in software development has increased drastically since its first appearance in software development around the turn of the millennium. Agile software development is now common practice and provides values and principles for producing working software rapidly while effectively responding to changing product specifications. One of the main traits with agile methodology is to have a product backlog – a list of prioritized features that are to be developed (Sommerville, 2019, p. 32-36). Having a list of features supports incremental development and delivery. Incremental development is to develop the end product piece by piece instead of developing the entire software at once and merging all functionality at the end (Rubin, 2012, p. 33). Incremental development makes the customer quickly see if the developed product is actually what they want. Several agile methods have been developed, and the choice of agile method should be carefully selected from project to project based on what is most suitable.

2.2.1 Scrum

Scrum is an agile method where the increments are developed during sprints, a small timespan that usually lasts between a week and a month (Rubin, 2012, p. 1-2). The product owner, a customer representative, decides what functionality is the most important and priorities items in the product backlog. At the start of each sprint, during the sprint planning meeting, the team picks tasks based on complexity and priority from the product backlog and creates a sprint backlog. The sprint backlog consists of time-estimated items to be finished before the end of the current sprint (Rubin, 2012, p. 417). A typical role in Scrum is the Scrum Master, which facilities the daily stand-up meetings and ensures Scrum as a process is correctly followed. Even though Scrum is reported as the most popular agile method (StateOfAgile, 2020), it is found that less than 50% of the Scrum users developed in incremental iterations (scrum.org, n.d.). Using Scrum but not following its core principles is called "Flaccid Scrum" or "Scrum-ish."

2.2.2 Kanban

Kanban is an agile method that focuses on workflow and continuous improvements (Rubin, 2012, p. 10). Kanban is task-boxed, which means increments are not developed on a schedule like with Scrum. When a task is finished, the developers proceed with a new one. A key aspect of Kanban is visualizing the tasks on a Kanban board and not exceeding the amount of work done in parallel. A number called WIP (Work in Progress) is used to specify how many tasks can be carried out at once. Kanban can be used when developing new software but is best suited for maintenance and support tasks as it supports a more "interrupt-driven work" approach (Rubin, 2012, p. 10). Kanban also emphasizes 'Kaizen,' which is continuous improvements; to constantly improve. The popularity of Kanban is slim, but it has increased a little over the years (StateOfAgile, 2020).

2.2.3 Scrumban

Scrumban is the second most popular agile method (StateOfAgile, 2020), and is a hybrid between Scrum and Kanban, hence the name. Scrumban is not a static method like Scrum and Kanban are, and one can freely select which aspects one wants from each of the methodologies. Scrumban provides a "best of both worlds" experience where teams can customize which aspects they want from project to project. Scrumban is suitable to apply in projects where following true Scrum or Kanban is not suitable.

2.3 Team performance models

In Section 1, the term 'team' was defined as a collection of three or more diverse people that collaborates to achieve a common goal. An also frequently used definition of a team is *"a small number of people with complementary skills who are committed to a common purpose, set of performance goals, and approach for which they hold themselves mutually accountable"* (Katzenbach & Smith, 1993). Often, the terms 'team' and 'group' are used interchangeably. A group is *"a collection of people with a common label"* (Rubin, 2012, p. 210). Members of a team are responsible for fulfilling the teams' vision, whereas members of a group will not adequately fulfill any responsibilities. A group can be transformed into a team which will improve efficiency and quality, but this requires resources such as time and cost (Tuckman, 1965). In this thesis, the term 'team' is used consistently. It should not be misinterpreted as 'group.' To measure the quality of teamwork, some constructs have been proposed over the years, such as the model tested on self-directed work teams (Janz, 1999) and the framework introduced by Dickinson and McIntyre (Dickinson & McIntyre, 1997).

2.3.1 Input-Process-Outcome

The input-process-outcome (IPO) model is a framework for studying team effectiveness and was initially conceptualized in 1964 (McGrath, 1964). The IPO model consists of three antecedent attributes: Inputs, Processes, and Outcomes (Mathieu et al., 2008). Inputs are factors that directly impact the interaction between the team members. These factors can be anything from team members' personalities to organizational design features. All inputs combined drive the processes, which is the team members' interactions towards the common goal. Outcomes are result from the team activities and can be everything from performance (i.e., quality of the product) and team members' perception, such as satisfaction and viability (McGrath, 1964). The IPO model has served as a foundation of teamwork quality models for researchers over the years (Mathieu et al., 2008).

2.3.2 The Big Five

The Big Five model describes five core components that have to be present to promote team effectiveness (Salas et al., 2005). Team effectiveness, as interpreted by Salas et al. (2005), is how the team performs and how the team interacts in order to achieve the team outcome. The five core components are: *team leadership, mutual performance monitoring, backup behavior, adaptability,* and *team orientation*. Further, three coordinating mechanisms to support the five core components are highlighted (Salas et al., 2005): *shared mental models, mutual trust,* and *closed-loop communication*.

Psychologists conceptualized the Big Five model to promote team effectiveness in general, not targeting software development teams. Also, the model was conceptualized before agile methodologies were frequently used in software development. A revised version of the Big Five, called STRAP, has been conceptualized (Strode et al., 2021) but has not been published yet. STRAP also includes five core components that promote team effectiveness, represented by each letter in the model's name: *Shared leadership, Team orientation, Redundancy, Adaptability,* and *Peer feedback*.

2.3.3 TWQ and Project Success

The TWQ (teamwork quality) construct was initially conceptualized in 2001 (Hoegl & Gemuenden, 2001). TWQ is a concept that measures the collaboration within teams, more specifically, the quality of interaction between the team members. The TWQ construct consists of the following six variables: *communication, coordination, balance of member contribution, mutual support, effort,* and *cohesion*. The TWQ construct aims to measure the effect its six attributes on Project Success, which consists of the following dependent variables: *Team Performance* and *Team Members' Success*. *Team Performance* consists of the variables *effectiveness* and *efficiency*, and *Team Members' Success* consists of *work satisfaction* and *learning* (see Figure 1).

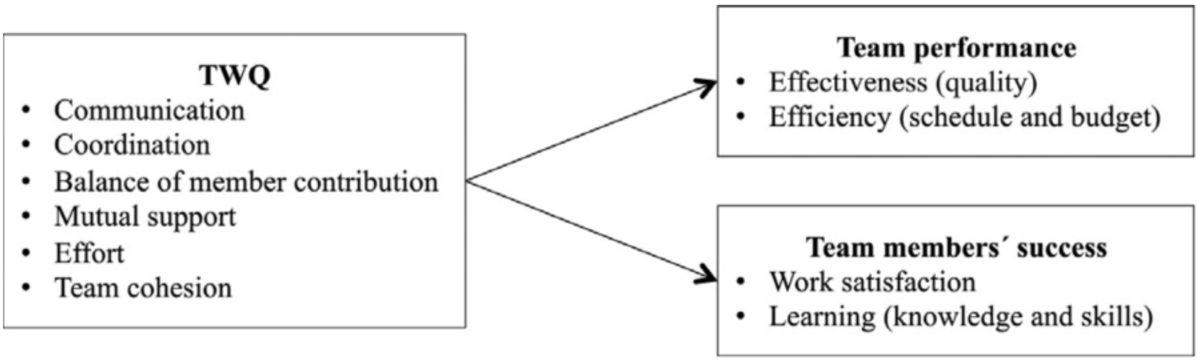


Figure 1: The TWQ model and Project Success (Hoegl & Gemuenden, 2001)

In more detail, the TWQ model and the dependent variables consists of the following variables (Hoegl & Gemuenden, 2001; Lindsjørn et al., 2018; Lindsjørn et al., 2016):

Communication

The attribute *communication* refers to how much time is spent communicating and how often it occurs. Communication in this sense includes formal communication, i.e., planning, status report, and informal communication, such as chatting, talking in front of the coffee machine, etc. Agile teams are often working closely together in open-plan offices to encourage informal communication. Communication in agile teams should be of high bandwidth (efficient communication with little overhead) and transparency. Transparency is important as it provides a clear understanding to all team members of what is actually happening (Rubin, 2012, p. 205-206).

Coordination

Coordination refers to what degree a team can structure task-related goals that are clear and assigned to each team member. Compliance between these individual tasks is essential for TWQ (Tannenbaum et al., 1992).

Balance of member contribution

Balance of member contribution refers to whether all team members' expertise is utilized to benefit the team. It emphasizes the importance of all team members participating in e.g., discussions to keep motivation up and leaving no one behind.

Mutual support

Mutual support refers to what degree to which team members are willing to assist other team members when needed, which benefits the entire team.

Effort

Effort is the amount of workload each team member spends on the team's task. Prioritizing the teams' tasks over other tasks is a good sign of effort.

Cohesion

Team *cohesion* is to what degree a group tend to stick together to achieve its goals and objectives. Team cohesion can be split up in to three parts:

1. Commitment of team tasks.
2. Interpersonal attraction of team members.
3. Team spirit.

Team Performance

One way to define this term is to which extent a team can meet (established) quality, cost, and time objectives. Team Performance consists of the following variables:

- *Effectiveness* refers to the expectation regarding product quality, while efficiency refers to the expectations regarding project quality, such as time and cost.
- *Efficiency* whether or not the team meets its expectations regarding project quality.

Team Members' Success

Team Members' Success consists of the attributes work satisfaction and learning. These attributes refer to the degree to which a team is motivated to work together in future projects using the same team.

In this thesis, the TWQ construct, as defined by Lindsjørn et al. (2016), is used. TWQ is limited to measuring the internal interactions within a team. TWQ aims to measure the effect of its six variables on Team Performance and Team Members' Success. As described by earlier findings applying this construct, the six TWQ attributes are highly related to Team Performance and Team Members' Success (Hoegl & Gemuenden, 2001; Lindsjørn et al., 2018; Lindsjørn et al., 2016).

2.3.4 Comparison of Teamwork Models

Model	**Description**	**Original source**
Input-Process-Outcomes (IPO)	Consists of antecedent factors: Inputs (factors that impact the interaction between team members), which forms the Processes (interactions towards a common goal) and results in Outcomes (team performance).	(McGrath, 1964)
The Big Five	The Big Five model consists of five core components that induce team effectiveness: team leadership, mutual performance monitoring, backup behavior, adaptability and team orientation. The five components are supported by the three coordination mechanisms: shared mental models, mutual trust and closed-loop communication.	(Salas et al., 2005)
Teamwork quality (TWQ)	TWQ consists of the following six variables that combined focuses on the quality of collaboration within teams: communication, coordination, balance of member contribution, mutual support, effort and cohesion. These variables affect the Project Success.	(Hoegl & Gemuenden, 2001)

Table 1: Comparison of teamwork models. Table based on (Strode et al., 2021)

2.4 Virtual Teamwork

In this thesis, the term "virtual teamwork" is often used. *Virtual teams* are teams in which the team members are scattered across different locations (Jabangwe et al., 2016). The terms 'digital', 'distributed' and 'virtual' are used interchangeably. When either of these terms are used in this thesis, read "virtual" (teamwork) as described by Jabangwe et al. (2016).

2.4.1 The Sociability Scale

Sense of community is crucial when working in teams since it increases information flow, support, and cooperation (Chatterjee & Correia, 2020). Feeling like a part of a community is essential, especially in a virtual context where the team members are co-located. The Sociability scale was conceptualized to measure the sociability within a CSCL (Computer-Supported Collaborative Learning) environment (Kreijns et al., 2007). The Sociability scale refers to how a CSCL is perceived to help cope with the distributed teamwork and learning. The concept of sociability is defined as "to what extent a CSCL environment is perceived to be able to facilitate the emergence of a sound social space" (Kreijns et al., 2007). Some key attributes of sociability are trusting within the team, belonging, and relationship, which is crucial in virtual teamwork.

No. Item	Item
1	This CSCL environment enables me to easily contact my team mates
2	I do not feel lonely in this CSCL environment
3	This CSCL environment enables me to get a good impression of my team mates
4	This CSCL environment allows spontaneous informal conversations
5	This CSCL environment enables us to develop into a well performing team
6	This CSCL environment enables me to develop good work relationships with my team mates
7	This CSCL environment enables me to identify myself with the team
8	I feel comfortable with this CSCL environment
9	This CSCL environment allows for non-task-related conversations
10	This CSCL environment enables me to make close friendships with my team mates

Table 2: The one-dimensional 10-item sociability scale (Kreijns et al., 2007)

The application of sociable CSCL environments can be a crucial success factor in a university context where project work has students work in teams with no history, and the team members do not initially know each other (Kreijns et al., 2007). Activities such as team composition, establishing team structures, and maintaining relationships make the team members feel part of the team, resulting in an effective complete learning experience (Kreijns et al., 2003). As the CSCL environments do not provide adequate opportunities for social interactions, making close friendships is hard (Kreijns et al., 2007). A study reviewing the Sociability scale criticizes it for the limited controlled study sample on which the model was tested (Yu et al., 2010).

De Lucia et al. (2008) conducted an experiment on university students in a fixed CSCL environment using the Sociability scale, combined with other conceptual models. The Sociability scale was adapted to fit the context of the experiment, resulting in removing five of the items compared to the original model. The study reports the Sociability scale worked well to measure perceived sociability in the experiment (De Lucia et al., 2009).

This thesis uses the sociability scale. To better fit the study of this thesis, the term "computer-supported collaborative learning" (CSCL) environment is reformulated to "virtual learning environment" in the questionnaire. The reformulation is done to capture that the students had to cooperate and learn using digital collaboration tools during the coursework as they had to work remotely.

2.4.2 Enterprise Social Networking

An ESN (Enterprise Social Network) can be defined as a platform that provides the following four attributes (Leonardi et al., 2013; Stray et al., 2019):

1. Send and receive messages either individually between members or to a group.
2. Choose and show coworkers as communication partners.
3. Share files.
4. View all previous conversations.

Examples of ESNs are Slack, Discord, and Mattermost. ESNs act as a forum where team members can communicate, supporting agile practices. A study conducted on virtual agile teams using Slack reported using it increased team awareness and communication flow (Stray et al., 2019). A challenge using ESN is the unbalance of messaging. The study reports that 33% of the team members wrote 86% of the messages, which is undesirable as everyone should communicate with everyone in agile teams. Even though the Slack logs examined in the study reported an imbalance of communication, the researchers believe the communication, in reality, was more balanced. The team members most likely also communicated using other platforms such as Facebook, which was outside the scope of the study (Stray et al., 2019).

2.5 Effects of Lockdown in Higher Education in Norway

The COVID-19 pandemic has lasted a year barely as of the writing of this thesis, so naturally, not much research on the effect of the lockdown in higher education has been conducted yet. However, some early findings have still been reported. As a result of the lockdown and the fact that the students suddenly had to adapt to a digital work environment, the motivation among students dropped drastically (Ahmed et al., 2020; Lindsjørn et al., 2021; Raaen et al., 2020). There was not one particular reason why the motivation dropped. However, several factors were pointed out, e.g., not physically meeting other students and easier getting distracted at home (Raaen et al., 2020). Ahmed et al. (2020), presents the impact of the coronavirus from two perspectives: the learning perspective and the teamwork perspective. They report that the motivation dropped, but the students could find good collaboration tools, making the teamwork somewhat manageable. Raaen et al. (2020) report that a higher threshold for asking team members for help when stuck and lack of communication between team members negatively affected teamwork in capstone courses. Raaen et al. (2020), further report that students felt stressed but could handle it well.

As a result of the COVID-19 pandemic, The Faculty of Mathematics and Natural Sciences decided to change the grade from a letter (A-F) to passed/not passed. The change of grade deeply impacted the motivation among students at the Department of Informatics (Lindsjørn et al., 2021).

3 Research Context

This section presents the research context the case study was conducted in, the capstone course offered by the University of Oslo. The course is elaborated on, including the learning objectives, student team composition, and how this capstone course differs from other capstone courses. Also, a brief overview of how the different semesters have been carried out is given.

3.1 IN2000

The capstone course studied in this thesis is called "IN2000 – Software Engineering and Project Work", herby also referred to as "IN2000" (IN2000, n.d.). The course was first offered in 2018 when it went as a pilot with 25 students representing four student teams. The year after the pilot, in 2019, IN2000 went full-scale with 201 students enrolled. The number of enrolled students has grown linearly from 2019 to 2021, with approximately 40 additional students per semester (see Table 3).

Year	2018	2019	2020	2021
Enrolled students	25	201	240	281
Number of teams	4	39	42	48

Table 3: Number of enrolled students and student teams for each semester the course has run.

IN2000 is an extensive course (20 ECTS), and there are many aspects regarding agile software development that are touched upon. The first eight weeks are filled with intensive lectures (see Table 4) and group sessions. After eight weeks, the students must work in agile teams and develop a mobile weather app on the Android platform using data from the Norwegian Meteorological Institute's API (MET, n.d.). The project runs for approximately twelve weeks. At the end of the project period, the students deliver the source code of the project and a written report. The report describes the process leading up to the final product, including agile practices applied and how the teamwork went—the source code and report combined count 50% towards the final grade in the course. A final written exam is held at the end of the semester, which also counts 50% towards the final grade.

3.1.1 Prerequisites

IN2000 is mandatory for students in their fourth semester on the Design, Digec, and Prosa study programs (see Abbreviations for full study program names). Most students enrolled in IN2000 are Prosa students as Prosa is the study program with the most students. Students from other study programs also get to enroll in IN2000 if they meet the required prerequisites. The students from Design, Digec, and Prosa have different courses leading up to IN2000.

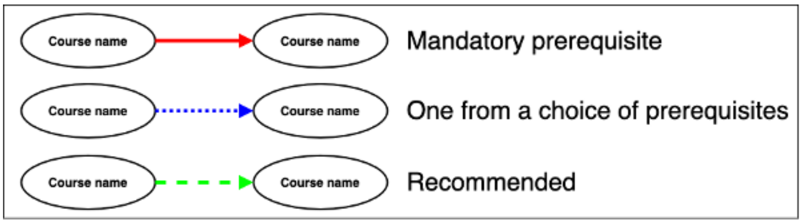


Figure 2: Explaining how to interpret Figure 3 and 4

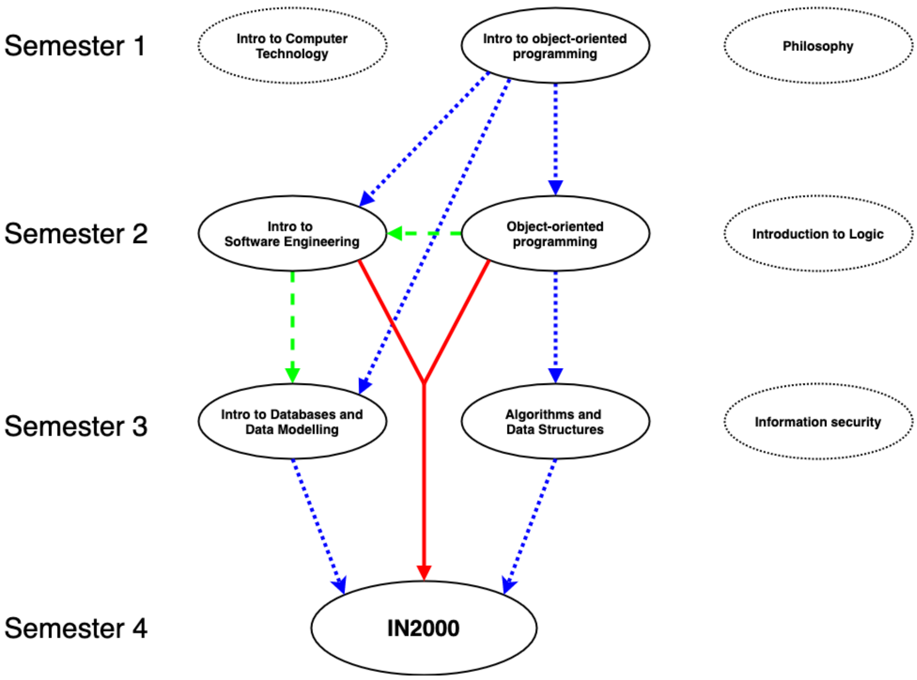


Figure 3: Courses leading up to IN2000 on the Prosa study program

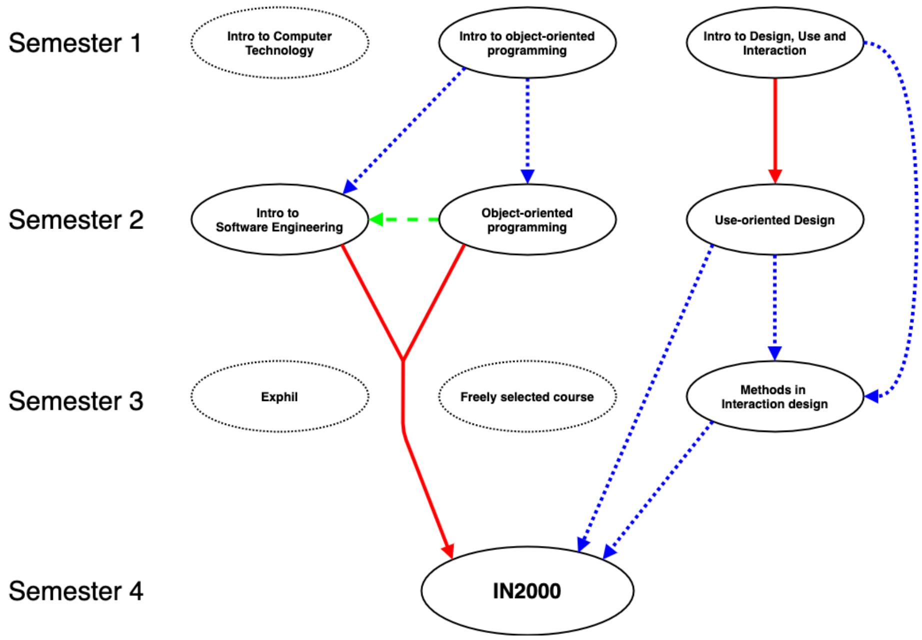


Figure 4: Courses leading up to IN2000 on the Design study program

Prosa students have courses such as *Databases and Data Modeling* and *Algorithms and Data structures* before IN2000 (see Figure 3). In contrast, Design students have *User-oriented Design* and *Methods in Interaction design* (see Figure 4). Having students with different backgrounds and expertise working together in IN2000 makes the student teams multidisciplinary. Students from the study programs often do what they know best. For example, Prosa students usually do programming, and Design students conduct usability testing. The course administration emphasizes that all students should experience and try out all aspects.

3.1.2 Learning Objectives

The learning objectives of IN2000 are specified on the courses' website and are the following (IN2000, n.d.):

- Knowledge of the most crucial system development methods, including their strengths and weaknesses.
- Knowledge of processes and actors in a project- and teamwork that applies agile principles.

- Knowledge of the following activities within system development: requirements collection and analysis, design, programming, testing, maintenance, and further development.
- Knowledge of professional system development methods, techniques, and tools.
- Have the competence to work in teams and reflect on own and the team's work in system development projects.
- Knowledge about the methods and principles for built-in safety and universal design.

From the learning objectives, it is clear that the course is ambitious, and the students get to experience industry-like software development. Looking at the learning objectives, we can see that participating in and reflecting on teamwork is an essential part of the course.

3.1.3 Course execution

During the first eight weeks of the course, there are two lectures per week (see Table 4). The first few lectures are about the technical aspects of the course to make the students familiar with the technology, such as Android Studio and Kotlin. The rest of the lectures are theoretical about software engineering, agile practices and prepares the students for the project work. During the eight weeks of lectures, the students have to deliver and pass two mandatory assignments on the technical aspects of the course to ensure that the students are at a specific technical skill level before the project work. In the first mandatory assignment, the students develop a simple Android app with basic functionality. In the second mandatory assignment, the students work with advanced aspects with Android and perform API calls to a dummy endpoint. More specifically, the students must retrieve data represented in JSON and XML and parse the response, respectively.

	No.	Lecture
Week 1	1	Introduction to the course
	2	The Basics of Android Studio and Kotlin
Week 2	3	More on Android Studio
	4	More on Kotlin
Week 3	5	RESTful API, Dataformats and How to parse response
	6	Teamwork, agile methodologies and project work
Week 4	7	Agile practices
	8	Basic principles of Testing
Week 5	9	Secure System Development
	10	Modelling and object-oriented principles
Week 6	11	Architecture and Technical Debt
	12	From theory to practice - the project from A to Z
Week 7	13	Introduction to APIs from MET
	14	Development of Android apps and How to use design patterns
Week 8	15	Universal Design
	16	Research Methods

Table 4: Overview of lectures in IN2000

Team Composition

During the first eight weeks of lectures, before the project starts, the students must submit a form regarding team formation. The students can submit the form alone, in pairs, or as a group of three. The form includes questions regarding the study program, motivation, and life situation. Based on the responses, the course administration collaborates with the teaching assistants and creates the project teams. Most student teams have between four and six team members, as having too large teams is undesirable (Rodríguez et al., 2012). Ideally, each team has students from at least two different study programs and gender equality to make them multidisciplinary and heterogenous (Løvold et al., 2020; Tafliovich et al., 2016). For example, in the spring of 2020, most teams were created by the course administration, with some minor exceptions. 35 student teams out of 42 had students from at least two different study programs. Due to many students from the Prosa study program, it is difficult to make all teams multidisciplinary. After the teams are formed, each team is assigned a supervisor, one of the course's teaching assistants.

With the current procedure of forming student teams, the students are faced with a more industry-like team formation, as they are not free to select all the team members themselves. The way the teams are formed consumes much time for the course administration. It is hard to account for everyone's wishes, and some teams end up being unbalanced due to the uneven number of students from the different study programs.

Kickoff

After the teams have been announced, a kickoff event is held to encourage the newly formed teams to meet. During the kickoff event, the teams participate in team-building exercises and start discussing the project cases. Previous students from IN2000, now working in the industry, attend the kickoff to presents how IN2000 is relevant in their professional work life. Demos of apps from earlier years are shown to motivate the students in what is possible to develop in twelve weeks. Usually, the kickoff has many attendees, and the students are relatively positive about it. The kickoff event was first offered in the 2020 semester.

Project Period

During the project period, the students get in touch with various aspects related to software development. Working in teams is one of the most crucial skills the students get to acquire. In addition, the students get to experience agile practices, developing a product with requirements from a real customer. The students are developing apps using Android Studio and the programming language Kotlin. The teams have to develop a comprehensive app throughout the project period and write a 50-pages report. In the report, the students are required to elaborate on the process leading up to the final product, write about the app's technical solutions, and reflect on the teamwork. The report is quite comprehensive and should reflect the work of twelve weeks.

Student Presentations

At the end of the project work, the students must hold a short presentation for the course administration, fellow student teams, and representatives from MET. The presentations include a demo of developed apps and reflection on the teamwork. Often, the presentations determine which teams are nominated for a prize of best-developed products.

Final Exam

At the end of the semester, the students have a final written exam. There are several learning outcomes in the course, which often makes the exam contain various questions. As the course administration emphasizes trying out different aspects throughout the project (e.g., programmers involve themselves in design-related tasks and designers tries out programming), the final exam often rewards those students.

Evaluation

The students are awarded a grade on a scale from A (being the best) to F (fail). The project work and the final exam both contribute 50% towards the final grade. In more detail, the project consists of the report (35%) and the source code (15%). Correcting the reports, project files, and final exams consumes much time. The students receive their final grades based on the total score combined with the project and the final exam.

3.1.4 The Customer

The Norwegian Meteorological Institute, also referred to as MET, is the 'customer' of IN2000. In reality, MET is not a customer like one traditionally defines it. MET has an extensive selection of weather data APIs, which the project cases in IN2000 are based on. The course administration and MET collaborate to create the project cases based on what is manageable within twelve weeks and which endpoints MET wants the students to use. The positive outcome for MET is that the students test their endpoints and report back if they discover flaws with endpoints or within the documentation. As the Norwegian State funds MET, they do not need economic gain from collaborating with IN2000. At the end of the semester, MET is involved in selecting the winner among the nominated apps. MET often selects the winning team based on how well their weather data is utilized, emphasizing the importance of combining at least two endpoints that produce something original.

Regarding the project cases, an example from the spring semester in 2020, the student teams could choose from five different cases, plus an 'open case' to design their own case. The cases were titled as follows:

- Case 1: Water movements in the oceans
- Case 2: Forecasts of landslides and avalanches
- Case 3: Air quality in municipalities
- Case 4: Predictions of climate and climate change
- Case 5: Drones and airspace
- Case 6: Open case – use weather data and design your own case

Figure 5 shows the distribution of the cases in the spring of 2020. As Figure 5 shows, the most popular case was Case 1. Case 1 was divided into sub-cases, and most teams selected *sea temperature*. Case 4 was the least popular, presumably since it required some prior knowledge in statistics. Case 4 required more heavy data processing and analysis, whereas Case 1 was more straightforward, technically speaking.

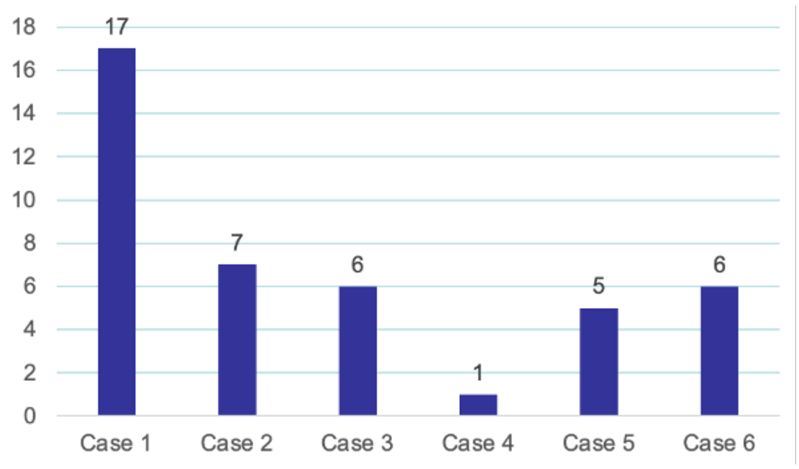


Figure 5: Number of teams on the different cases in the 2020 semester

Each semester, one case often stands out, having many student teams selecting it. In 2019, *air quality* was the most popular case, and in 2021, the *sea temperature*-case was the most popular.

3.1.5 Tools

The students must develop an Android app using the Kotlin programming language. They are also required to use GitHub for Version Control. Other than a short lecture where some useful tools were presented, the students select collaboration tools independently.

3.1.6 "New" Course

IN2000 was first offered in 2018 as a pilot with only 25 students. The course code then was IN2001. Being such a new course, drastic changes from year to year have been implemented due to students' feedback. E.g., in 2019, the students were given a single lecture on all the technical aspects of the course. After receiving the feedback from students, the course administration decided to give an additional three lectures on the technical aspects in 2020.

3.2 The investigated semesters

This sub-section explains the three investigated semesters of IN2000.

3.2.1 The 2019 semester

In 2019, the course was offered physically at the campus. This semester was the first time the capstone course was offered at full scale with over 200 enrolled students. The teaching assistants used the designated timeslots of their seminars to meet with the teams they supervised. The student teams had stand-ups physically at the campus. As the course was relatively new, many aspects could have been better. Most of the teaching assistants did not participate in the pilot course the year prior, which also made this a new experience for them as well.

3.2.2 The 2020 semester

In 2020, the first eight weeks of lectures went as usual. The students were physically at the campus, participating in lectures and group seminars. This was the first semester the kickoff event was offered, and it was held on February 29th, which means most teams had the opportunity to meet physically. On March 12th, 2020, the University of Oslo closed down to stop COVID-19 from spreading in society. The lockdown had a significant impact on IN2000 as the project work on short notice had to be carried out virtually. At this point, most teams had a written and signed "team contract" stating how they should collaborate throughout the project. Many teams had written physical activities such as dining together, physical standups at the campus, and weekly meetings with their supervisor. The seminars were canceled for the rest of the semester as a result of the lockdown. The teaching assistants had already been assigned to student teams, which the teams could directly contact if they needed assistance. Many teams reflected during the presentations on how they were affected by the Coronavirus situation.

Another impact of the lockdown was the change in grading. The Faculty of Mathematics and Natural Sciences decided that due to the uncertain situation the lockdown has caused, all evaluations will be on the passed/not passed scale instead of the A-F. The decision to change the grading seems to be one of the main reasons why the motivation among many teams dropped (Lindsjørn et al., 2021).

3.2.3 The 2021 semester

In 2021, the entire course was digital - everything from the lectures, seminars, kickoff, and teamwork. In 2021, the pandemic had lasted for around a year, meaning the students were used to working digitally from home. The grading was A-F instead of passed/not passed this semester as well. One new aspect this semester was the more extensive mandatory assignment on retrieving data from an API and parse the data.

4 Research Method

This section describes the research methods applied in this thesis and the reasoning behind them. Further, a description of how the data was collected and analyzed is given. Finally, a brief overview of data types, data validity, and ethical considerations are presented.

4.1 Case Study

This thesis has been conducted as a case study. Case studies are an empirical research method to investigate phenomena in their natural context (Yin, 2009, p. 17-19). Thus, they have a high degree of realism, often at the expense of the level of control (Runeson & Höst, 2008). Central methods of collecting data include interviews, observations, archival data, and metrics (from, e.g., surveys). Runeson & Höst (2008) emphasizes the importance of using multiple data sources within a case study to limit concluding from a single source. Triangulation is vital in case studies, as it provides a broader perspective on the investigated phenomena (Runeson & Höst, 2008). Triangulation addresses threats to validity, such as using multiple data sources, elaborated on in Section 4.2.

4.1.1 Case Study Designs

Yin (2009) describes four basic types of case study designs (see Figure 6). Yin (2009) distinguishes between holistic case studies and embedded case studies. In holistic case studies, the case is studied as a whole, and in embedded case studies, multiple units of analysis are studied within the case. Yin (2009) also distinguishes between single-case and multiple-case designs. Single-case case studies have one case investigated, whereas multiple-case designs may contain more than a single case. Yin (2009) still emphasizes that the single-case and multiple cases are variants of the same methodological framework and that the notable difference is between holistic and embedded designs (Yin, 2009, p. 46-62).

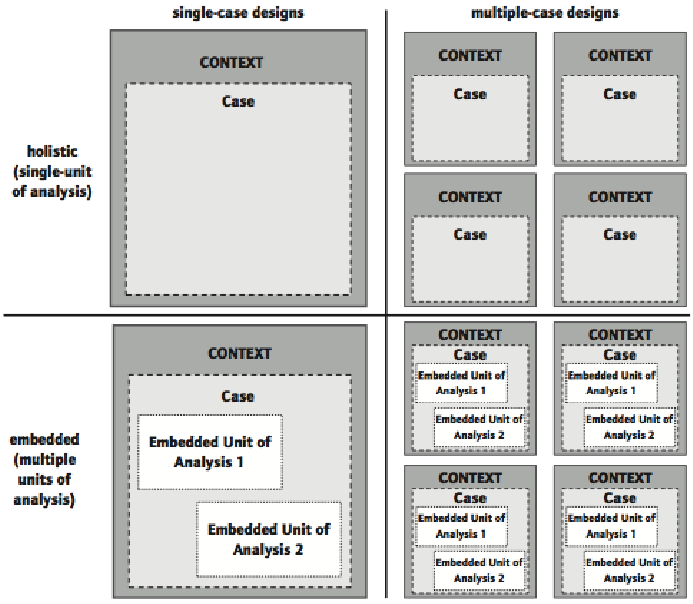


Figure 6: The four basic types of case study designs (Yin, 2009)

Selecting the case study design for this thesis was not trivial. This thesis explores and compares student teams from three different semesters (2019, 2020, and 2021). For this thesis, both single-case embedded design and multiple-case holistic design (see Figure 6) were considered as they both could fit depending on how the case and context are formulated.

Single-case embedded design could fit if the data from the investigated semesters were treated as separate units of analysis. A problem with this is that the units of analysis have to be within the same case. All datasets in this thesis were gathered in three completely different semesters. In 2019, the lectures, seminars, and project work were physical on campus. In 2020, the students had to adapt to the sudden closure of campus and start working virtually in the middle of the semester. In 2021, the COVID-19 pandemic had lasted for a year. In 2021, the students expected digital teamwork as the University of Oslo announced all educational activities were digital from the beginning of the semester. As all these three semesters (contexts) were completely different, it was decided not to merge all the semesters into a single case and conduct a single-case embedded design case study with one unit of analysis per semester.

4.1.2 Chosen Design

This thesis has applied the holistic multiple-case study design. Supporting the choice of case study design, Yin (2009) states selecting multiple-case design over single-case design, the chances of doing a good case study will increase (Yin, 2009, p. 60-61). As the three semesters that are studied were different, splitting them into separate cases is suitable. Multiple-case design should support replication, carefully selecting cases one predicts to produce similar or contrasting results (Yin, 2009, p. 60). Having cases that are similar or contrasting makes it easier to compare them analytically. As mentioned previously, 2019, 2020, and 2021 differ in how they were carried out. Considering each investigated semester as separate cases emphasizes rationale for comparing them as they would become a single unit of analyses within different cases. In the holistic multiple-case design of this thesis, the following attributes was defined: (based on Figure 6):

- **Contexts:** The three different semesters of IN2000.
- **Cases:** The students' teamwork within the contexts.

Yin (2009) mentions some pitfalls with the holistic multiple-case design. Multiple-case designs are time-consuming as the researcher must conduct one study for each case. As replication is crucial in multiple-case designs, the researcher should select the cases deliberately and conduct a sufficient number of replications and support them with a comprehensive theoretical framework (Yin, 2009, p. 53-60).

Another rationale for selecting a multiple case study design is that the investigated semesters differed even if the COVID-19 pandemic had not struck. Since the course is new, it is under constant evolution as course material and activities are changed from year to year. E.g., in 2019, most teams were formed by the students themselves, while in 2021, nearly all student teams were created by the course administration. New mandatory assignments, additional lectures, and new activities made the investigated semesters different.

4.1.3 Qualitative and Quantitative Data

This thesis applies both qualitative and quantitative data. Qualitative data are expressed as words, descriptions, pictures, etc., and are analyzed using categorization and sorting (Runeson & Höst, 2008). Quantitative data, on the other hand, express numbers and is analyzed using statistics (Runeson & Höst, 2008). Case studies are often based on qualitative data, but combining it with quantitative data may provide a better understanding of the investigated phenomenon (Seaman, 1999). This thesis combines qualitative and quantitative methods, also referred to as *mixed-methods* (Yin, 2009, p. 62-64).

4.2 Validity and Reliability

4.2.1 Construct Validity

Construct validity refers to the extent to which the measures investigated in a study are not a result of the researcher's subjective view of the phenomenon (Yin, 2009, p. 41-42). Yin (2009) and Runeson & Höst (2009) emphasizes the importance of using multiple data sources to address construct validity. By using multiple data sources, the concept of triangulation within case studies is also addressed.

4.2.2 Internal Validity

Internal validity refers to *"the extent to which a study establishes that a factor or variable has actually caused the effect that is found"* (Robson, 2002, p. 549). If the researcher fails to be aware of a "hidden factor", it could be a threat to the internal validity of a study (Runeson & Höst, 2008).

4.2.3 External Validity

External validity concerns the degree to which one can generalize the findings (Runeson & Höst, 2008). Yin (2009) states that critics of case studies often emphasize that external validity is a common threat as the researcher may generalize findings based on an insufficient basis.

4.2.4 Reliability

Reliability addresses the integrity of the study (Yin, 2009, p. 45). If another researcher were to conduct the same study, following the same steps, they should arrive at the same findings and conclusions.

4.3 Data Collection

This sub-section explains the processes of data collection and how it was analyzed. Ethical considerations regarding data collection are also addressed.

4.3.1 Literature Research

Literature research is a great way to get an overview of what is found on the topic; however, it is easy to confine what others have found (Robson, 2002, p. 50-54). Initial literature research was conducted in this thesis to get an overview of the topic. Some academic papers are often behind a paywall. Fortunately, some of the search engines used in this thesis offered access through login with a student account from the University of Oslo. The following search engines were used: Oria, Scopus, and Google Scholar.

As with all searching on the Internet, simply searching for a keyword results in a vast number of results. To narrow down the search, one can generate a search string containing relevant words. The following are some examples of the optimization of search strings applied in this thesis when searching for literature:

- "capstone course" AND "software engineering" AND "teamwork"
- "student teamwork" AND "software engineering capstone course" OR "final year project"
- "virtual teamwork" OR "teamwork" OR "digital teamwork" OR "collaboration" AND "higher education" OR "students" OR "capstone course" AND "software engineering"

To manage the relevant literature for this thesis, Zotero was used. Zotero is a tool to store, categorize and make notes of the literature digitally. Zotero has

LaTeX-integration, which means the bibliography was easily synchronized into the LaTeX-project, automatically updating the reference list when using a source.

Written report from MET

A representative from MET wrote a report on how the students in the 2020 semester could utilize the technical parts of the course, such as parsing the response, which APIs they used, and elaboration on the cases. As students in IN2000 are supposed to retrieve data from a specific proxy server, MET can extract statistics on how they use their weather data. The report describes how well the student teams could implement the cases and how well they were able to utilize the endpoints. The written report from MET has also been used in this study.

4.3.2 Surveys

The survey is a data collection method in which respondents (raters) fill out a form, often answering questions with text and rates statements on a scale from 1-5 (Likert scale). Surveys provide a simple approach to gather subjective data, such as attitudes, values, beliefs, and motives. As surveys are easily distributed, one can reach many respondents with little effort. With a sufficient data sample, it might be possible to generalize information from any human population (Robson, 2002, p. 227-235). Surveys provide subjective data that are affected by the characteristics of the respondents. Some disadvantages with surveys are that respondents' memory might differ from what occurred. They might also put themselves in a good light by responding inaccurately, even though that is not what their beliefs actually are (Robson, 2002, p. 227-235).

This thesis consists of three surveys representing the students' and teaching assistants' (TA) perception of student teamwork over the three investigated semesters. The surveys are based on teamwork quality models used earlier when measuring teamwork quality and project success on professional teams (Hoegl & Gemuenden, 2001; Lindsjørn et al., 2018; Lindsjørn et al., 2016). The teaching assistants (TA) represents the product owners (PO). The approach of assuming the TAs are POs has some flaws, elaborated on in Section 6. All items in the teamwork quality (TWQ), project success, and items in the questionnaire's virtual learning environment (Sociability scale) were statements. The respondents indicated their personal views for each statement on a Likert scale from 1 (strongly disagree) to 5 (strongly agree).

The student survey consisted of a total of 91 items. Ten questions measure Sociability (virtual learning environment), and 61 questions measure TWQ and project success. In addition, 20 items were regarding the tools and background questions like gender and age, study program, and previous experience in agile development. The supervisors (TAs) also answered a survey, and they evaluated the performance for each team they supervised during the project. In more detail, the teaching assistants answered 15 questions on Team Performance (see Appendix A). They also answered some questions regarding how frequently they had meetings and communicated with their teams.

2019

The first survey was conducted in May and June in the spring of 2019 before the project was delivered. The response rate was high (around 98%) as the students had to fill in the form physically at the university during the student teams' presentations. In 2019, the ten items regarding the Sociability scale were not included since the virtual learning environment was not relevant that year. The teaching assistants answered the PO survey during the summer of 2019. Some teaching assistants were relatively slow and waited until August 2019 before submitting a response.

2020

In June 2020, the survey was conducted after the students had delivered their projects. The ten items measuring sociability (digital learning environment) were added this semester due to the virtual teamwork. 151 students out of 240 answered this survey, which made the response rate 63%. The students had to submit the form digitally. The form was open for submissions from June 1st to July 1st. The TAs answered the PO survey throughout the summer, with the last response in August 2020.

2021

Like in 2019, the student survey was conducted during the students' presentations but was held digitally instead. The survey was voluntary to participate in, but as a timeslot to answer the survey was allocated during the presentations, the response rate increased compared to the year prior. The survey was conducted from May 18th to May 21st. The 2021 survey mainly consisted of the same questions as 2020, but a few were reformulated to avoid having to reverse-code the items. To ensure the teaching assistants (POs) responded rapidly to the survey compared to previous years, they were given a limited time to answer the PO survey.

	2019	**2020**	**2021**
Enrolled students	201	240	281
Student respondents	198	151	246
Student response rate	98,5%	62,9%	87,5%
PO respondents	5	9	10

Table 5: Number of respondents for each of the surveys

Parsing of survey data

All the responses (both students and POs) were imported into Excel, and the data were aggregated into team level, which means merging all the team members' responses for each team. To determine which team member acted as the team leader (TL), a question in the survey was about the student's role as a leader. The student within a team that perceived their role as the Scrum Master / Team Leader the highest were selected as the TL for that team. In practice, the team leaders called themselves Scrum Masters (SMs), but they are referred to as team leaders (TL) in this thesis. If more than one student perceived their role as team leader equally (e.g., two students from the same team selected 4 out of 5 on this statement), the student that answered the survey first was selected as the TL for that specific team. There are some academic deliberations with this approach, and they are elaborated in Section 6. The TLs and POs were extracted into separate sheets in Excel. Finally, a final sheet in Excel was created where data from all the different sheets were included. The final sheet is the one that has been imported into Python and SPSS for analysis. The same procedure was replicated for the data from all of the semesters, meaning there were three Excel files representing each of the investigated semesters. Python version 2.7.10 and 3.9.2, and SPSS version 27 was used to process and analyze the datasets.

Example: Measuring the effect of TWQ using SPSS

In SPSS the following options were selected: Analyze → Regression → Linear... This opened a window where one can select the dependent and the independent(s) variable(s) (see Figure 7). Figure 7 shows measuring the effect of TWQ on the dependent variable team members' effectiveness.

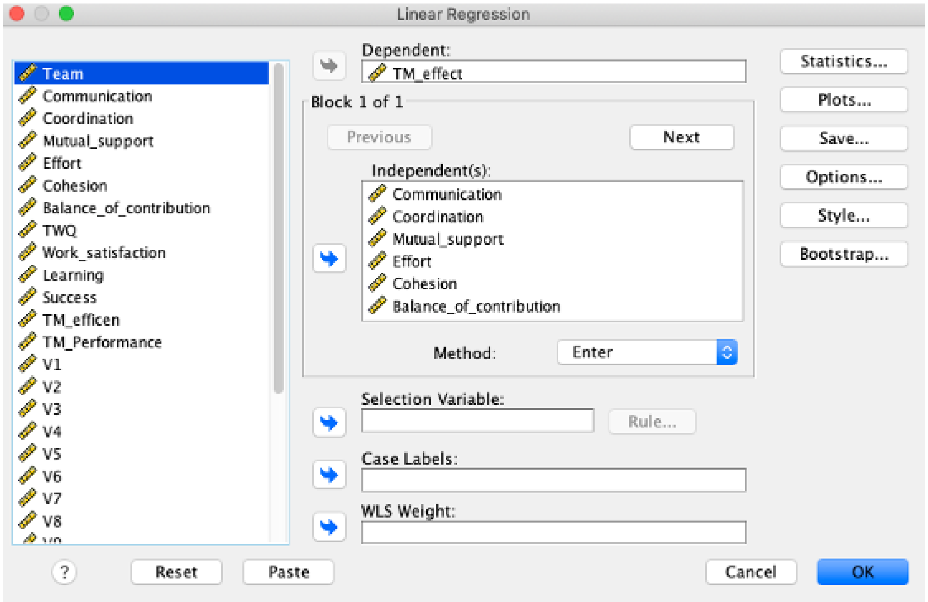


Figure 7: Screenshot from SPSS showing linear regression analysis between TWQ and team members' effectiveness

4.3.3 Interviews

Interviews are one of the most important data collection methods in a case study (Yin, 2009, p. 106). In an interview, the researcher has a dialogue with subjects and is guided with a set of interview questions (Runeson & Höst, 2008). During interviews, the respondents get to share what they know, think, and feel about specific topics, resulting in a lot of in-depth subjective data. Robson (2002) mentions three different types of interviews: fully structured interview, semi-structured interview, and unstructured interview.

Fully structured interviews are similar to questionnaires as it consists of a set of predetermined questions. Semi-structured interviews have a set of predetermined questions to be answered, but the order can be changed, and the interviewer might sometimes deviate from the predetermined questions. Unstructured interviews are informal, and the interviewer has an area of interest, but the conversation and questions develop along the way (Robson, 2002, p. 270).

Robson presents some advantages and disadvantages using interviews (Robson, 2002, p. 272-273). Interviews are great at getting deep insight into a subject's thoughts and feelings about a specific topic. They have the potential to provide data that are rich and enlightening. Interviews are flexible, which means the interviewer can adapt how the questions are formulated to the subjects. However, adapting the questions along the way might lead to some reliability issues as the questions might be formulated differently, and the interviewees may answer differently from interview to interview. Another drawback is that interviews are time-consuming. All involved activities such as planning, reaching out to respondents, scheduling, conducting the interviews, transcribing, and analyzing take much time.

In the spring of 2021, eight semi-structured interviews were conducted online. Seven students that completed IN2000 in 2020 and one representative from MET were interviewed. The interview guide (see Appendix B) for this thesis was created to make the interviews last between 30 minutes and an hour, as this is the desired length of interviews (Robson, 2002, p. 273). The average time spent for all the interviews was approximately 39 minutes.

Interviews Study Sample

Respondent	Study program	Semester	TA	Duration
S1	Prosa	2020	Yes	43:05
S2	Prosa	2020	Yes	35:39
S3	Design	2020	Yes	42:21
S4	Design	2020	No	32:01
S5	Design	2020	Yes	34:54
S6	Prosa	2020	Yes	42:32
S7	Robotics	2020	No	43:33
C	Not relevant	2019, 2020, 2021	Not relevant	35:48

Table 6: The interviewees of this study

Table 6 shows all interviewees of this study. 'S' is student, 'C' is customer representative, a respondent from MET. The semester column shows the semester(s) the respondents were affiliated with. TA shows whether or not the interviewees were teaching assistants the following semester (in 2021). Duration is how long the interview lasted.

Parsing of interview data

The interviews were recorded and transcribed. NVivo 12 was used for analysis. In NVivo, nodes were created to make it easier to code the data. The nodes mainly consisted of the theoretical frameworks presented in Section 2, such as the TWQ model and the Sociability scale.

4.3.4 Observations

Observation is a method where the researcher watches how specific tasks are conducted by subjects (Robson, 2002, p. 309). Robson (2002) describes some advantages and disadvantages with observations. Observations provide a deep understanding of the phenomenon that is studied. The observer gets to see how the subjects behave in their natural context. As opposed to an interview where the respondents may say they do something, the observer can actually verify what they are doing in observation. The observed subject may act unnaturally due to the presence of the observer. The observer should not interfere with the objects they observe, e.g., start participating in the observed activity or talk with the subject as this disturbs the context.

The observations conducted as a part of this thesis are informal through my job as a TA (teaching assistant). I worked as a TA in IN2000 in 2019 and 2020, where I supervised a total of 13 student teams combined. When supervising, I had weekly meetings with most teams to ensure that the teamwork went well. In some cases, I also acted as product owner and got access to some teams' internal documents and were added to their Slack channels. I have participated in the project presentations in 2019, 2020, and 2021. During these presentations, I have been exposed to the students reflecting on their teamwork. In 2021, I have been involved with the weekly course administration meetings where all course-related decisions were made. When issues regarding teamwork within the teams arose, we have discussed how to solve them in plenary.

The observations conducted by myself were informal as they represent my experiences of working as a TA in IN2000. In this thesis, it is unavoidable not to introduce some aspects of which I have observed.

4.3.5 Ethical considerations

In this thesis, the following actions have been pursued in order to acknowledge ethical deliberations:

- Before the data gathering, a consent form was sent to NSD (Norwegian Centre for Research Data), an ethical committee (Robson, 2002, p. 69-70) that ensures ethical precautions are taken before gathering data. NSD approved that this thesis does not violate any privacy policies.

- The surveys were anonymized and did not include personally identifying questions.

- The interviews were recorded and transcribed. When transcribing the interviews, all transcriptions were assigned a random ID. When analyzing the transcriptions in NVivo, only the IDs were visible.

- All data collected as a part of this thesis were stored on an encrypted research database operated by the University of Oslo.
- It was entirely voluntary to participate in the study. All participants were able to withdraw their responses at any time.

See Appendix F for the full NSD consent form, which was agreed on before participating in this study.

4.4 Statistics

This sub-sections presents which aspects from statistics that are applied in this thesis.

Descriptive statistics

"Descriptive statistics are ways of representing some important aspects of a dataset by a single number" (Robson, 2002, p. 407). This includes values such as the standard deviation, the mean value, the median, and the variance.

Cronbach alpha

Cronbach alpha is statistic for internal-consistency reliability alpha values and should be between 0,7 and 0,9 to be satisfactory (Nunnally & Bernstein, 1994). Cronbach alpha is used on a dataset to investigate to what degree the items in a questionnaire are associated while providing new data that still is relevant for the given instrument. In this thesis, Cronbach alpha values were calculated using SPSS with the following selections: "Analyze → Scale → Reliability analysis."

Statistical significance

Statistical significance, represented with the symbol p, tells what degree the probability of an outcome is due to chance; the lower the p-value, the less likely the result is due to chance (Robson, 2002, p. 400-401). For example, if the p-value is below 0,05 ($p < 0.05$), there is less than a 5% chance the result is due to chance.

Correlation

A correlation is a relationship between two variables (Colin, 2002, p. 420). There are several ways to measure the correlation between two variables. However, the most popular one is the Pearson product-moment correlation coefficient (Goodwin & Leech, 2006). The correlation is represented with the symbol r, which ranges between the values -1 (negative relationship) to 1 (positive relationship). If the correlation is 0, there is no relationship between the two variables. *"The squared correlation (r^2) indicated the proportion of the variance for a dependent variable that is predictable from a regression model"* (Goodwin & Leech, 2006). The correlations in this thesis were calculated and visualized using Python 3.9, and was verified using SPSS.

5 Results

This section presents the findings of the analyzed collected data and relates them to the research questions of this thesis. First, findings on TWQ and project success are presented. Then the results from the Sociability scale are given. The results from all the investigated semesters are correlated with one another. Students' motivation is also elaborated on.

5.1 Teamwork Quality and Project Success

This subsection presents the findings on teamwork quality (TWQ) and project success (team members' success and team performance).

5.1.1 Descriptive Statistics from the surveys

				2019		2020		2021	
Construct	**Rater**	**Variable**	**No. items**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Team Quality (TWQ)	Team member	Communication	10	4,17	0,37	3,99	0,51	4,15	0,35
		Coordination	4	4,05	0,40	3,98	0,45	4,14	0,33
		Mutual support	7	4,42	0,37	4,33	0,47	4,50	0,33
		Cohesion	10	4,26	0,45	4,20	0,45	4,37	0,37
		Effort	4	3,86	0,65	3,73	0,62	3,97	0,53
		Balance of contribution	3	4,25	0,41	4,24	0,50	4,32	0,46
Team members' success	Team member	Work satisfaction	4	4,28	0,44	4,37	0,40	4,41	0,43
		Learning	4	4,41	0,47	4,42	0,45	4,55	0,31
Team performance	Team member	Effectiveness_TM	10	3,86	0,42	4,03	0,34	4,07	0,42
		Efficiency_TM	5	3,81	0,60	3,98	0,55	4,07	0,51
	Team leader	Effectiveness_TL	10	3,83	0,65	4,30	0,54	4,15	0,58
		Efficiency_TL	5	3,83	0,74	4,20	0,75	4,12	0,72
	Product owner	Effectiveness_PO	10	3,90	0,72	4,07	0,56	4,11	0,63
		Efficiency_PO	5	3,71	0,87	3,92	0,73	4,12	0,72

Table 7: Descriptive Statistics from the surveys

Table 7 shows and compares the descriptive statistics for all the investigated semesters on team level, which is the aggregated values of all team members' evaluation for each team. From Table 7, we can see that the results are relatively similar for all the semesters. The mean values of the TWQ variables are somewhat higher in 2021 compared to both 2020 and 2019. The 2020 semester overall has the lowest evaluation of all the TWQ variables. Taking a closer look at, e.g., *communication*, we can see that it dropped (-0,18) from 2019 to 2020 but increased (+0,16) from 2020 to 2021. The mean value of *learning* was very similar in 2019 and 2020 but increased from 2020 to 2021 (+0,13). When looking at *effectiveness* under Team Performance, it was significantly higher in 2020 than in 2019 (+0,17) and even higher in 2021 (+0,03).

Data gathered during the 2021 semester shows that the students' evaluation of all variables was perceived higher than in 2020 and 2019, except *communication*. The standard deviation is also significantly lower in 2021 overall compared to the other semesters for all variables.

The mean values of team members' the evaluation for TWQ were 4,17, 4,08, and 4,24 in 2019, 2020, and 2021 respectively. Further, the mean values of Project Success were 4,09 in 2019, 4,20 in 2020, and 4,28 in 2021.

The Cronbach alpha values (not shown in Table 7) were satisfactory for all variables (between 0,7 and 0,9) except for the TWQ variable *balance of member contribution* in all of the semesters.

5.1.2 Interviews

The interviewees are students that enrolled in the 2020 semester.

Communication

As we can see from Table 7, *communication* was the attribute that was perceived as the third-lowest of the investigated variables in 2020 and was lowest compared to the other semesters. Most of the interviewed students expressed maintaining well communication through the project work was hard. One student expressed the poor communication was a result of the lack of physical presence at the campus:

S6: *"I feel like there were two completely different worlds: before the lockdown and after the lockdown. When we physically met at the campus, we maintained good communication and communicated using Slack between the physical meetings. However, after the lockdown, we only used Slack, and the communication became slow and sporadical."*

One student emphasized that the meetings became more formal, making the informal communication vanish:

S4: *"Before the lockdown, the social aspects were good as we could talk before and after the meetings. However, after COVID-19 hit, we had to carry out the meeting on Zoom, and the meetings suddenly became way more formal. When turning on the laptop camera, we only worked with the project, and we never had any breaks where we just talked like we normally would if we could physically meet."*

Teams with students knowing each other from before reported communication went well. S2: *"We were a team where everyone knew each other from before, so the communication was good."* S7: *"Four out of five team members knew each other from before, so it was easy for us to communicate properly."*

One respondent expressed the threshold of asking team members for help increased due to the lockdown:

S2: *"Personally, I felt like the threshold of asking teammates for help increased. I did not want to disrupt others with small problems, and in the beginning, starting a Zoom-meeting to solve a problem was a quirk. I think the teamwork would have been better if the threshold for reaching out to teammates were lower".*

Coordination

The evaluation of *coordination* was the second-lowest of all evaluated variables in the 2020 semester (See Table 7). All interviewees reported they used some form of tool to keep track of the tasks. Some said they coordinated tasks on sprint planning meetings: S5: *"We decided who did what during the spring planning meetings. If someone finished their task before the sprint was over, they selected a new one and announced it to the others".* Furthermore, S2: *"We had weekly sprint planning meetings, and we selected tasks based on own wishes. We used Trello to visualize the tasks on a board".*

Another respondent expressed:

S1: *"We had no formal process; we picked tasks based on what we thought was the best for the time being. (...) The Design students made user-stories while the Prosa students implemented the app."*

One respondent reported that coordination was challenging due to the lack of an overall plan:

S6: *"We did not have a well-established backlog. We never assigned specific roles to the team member. We never maintained the backlog, and it was never changed throughout the entire project work. (...) I presumed the design students in our team should do 'design-related'-tasks, and the students from 'Prosa' should do most of the programming. This did not happen, I tried to suggest it, but I never received any response from the others".*

Mutual support

All interviewees emphasized that the team members were helpful if they asked the other for help: S5: *"If someone was stuck at a task and asked for help, we jumped into a Discord-room and solved it right away. I felt like the team members were always ready to help."*

If a disagreement came up, the interviewees overall said they solved it suitably. One said they never rushed any major decisions but slept on it until the next day:

S3: *"If we had a discussion, which we rarely had, we did not conclude before the next day. If we could not agree the day after, we voted, which meant the majority decides. We tried to make sure no one felt like any ideas were bad".*

Another interviewee said that lack of engagement in the team made the discussions and meetings demotivating. S6: *"Everyone said their opinion, we did not 'fight' during discussions. The problem, however, was that not many were engaged, many team members did not have an opinion at all. This was demotivating".*

Effort

The evaluation of the variable *effort* was lowest among the raters in 2020 and compared to 2019, it was significantly lower (see Table 7). Effort was the only variable below 4 in 2021. The interviewees emphasized the effort were affected by the lack of motivation after the lockdown and change in grade:

S2: *"After the announcement of passed/not passed instead of A-F, the team members worked less with the project as everyone knew we would get a 'pass' anyways. Furthermore, I think we worked less with the course than we would have if it was physical instead".*

One student said the effort was varying throughout the course:

S6: *"The effort varied. I felt like the design student was not prioritizing the project; they worked with other courses instead. (. . .) I also feel like when we assigned tasks, the effort was high as some engagement was built up, but then, no one actually produced anything for a whole week or two weeks".*

Some said the effort was high despite the lockdown:

S3: *"The effort remained high as we adapted and thought that we can still learn something from this process. The app we are developing can also be shown in future job interviews. We could not do anything else due to the lockdown, so why not just work with the app? So we enjoyed ourselves and kept the motivation up using Discord."*

Cohesion

One interviewee said they ate dinner together physically before the lockdown, which made the cohesion within the team better:

S3: *"We were lucky since we were able to eat dinner together before the lockdown. I have worked in student teams before, and I knew we should get to know each other, get along, have the same goals, the same thoughts, and ambitions. (. . .) I think this helped our teamwork to be better than it otherwise would have been."*

Another respondent thought it was easier to be a part of a team where everyone knew each other from before:

S4: *"I did not know anyone on the team from before; they all knew each other from before. Despite this, I felt it was easy to 'become a part of the gang.' It was easier for me to fit in with a team where the team members knew each other from before than if we all were strangers. (...) But today, I have no connection with any of the team members anymore".*

Balance of member contribution

The interviewees were divided regarding the *balance of member contribution.* Some felt like team members' contribution was unsatisfactory:

S6: *"I felt like the most motivated members worked the most; which was me. However, I did not encourage the others to work either. We lacked a leader figure, and nobody took that role, which resulted in a very unbalanced workload".*

Another interviewee emphasized that the teamwork was unbalanced but that everyone was okay with it:

S7: *"All in all, there was unbalance of member contribution. Two team members only had this course, and they said they could work a lot with the project. The rest of the team had other courses and periodically worked on those as well. The ones that did more were never annoyed with the others as they accepted they were working more on the project since they had more time".*

One student said they delegated workload evenly:

S5: *"I feel we delegated the tasks as fair as possible, but it was hard to estimate how comprehensive a task was going to be. But we always delegated tasks in plenary and discussed who should do what. And everyone did their parts."*

Team Performance

Regarding the Team Performance, most interviewees were satisfied with their developed products: S1: *"I am happy with our app, and I think it is nice to show off during future job interviews."* S2: *"I feel like the developed app backend-wise was good but compared to the other apps we saw during the presentations, our app was not that appealing."*

Due to not good estimation, their team had to prioritize and cut down on the functionality: S6: *"We did not have that much time in the end, and therefore, we had to make an app with less functionality than we first intended."*

Team Members' Success

One interviewee emphasized the importance of learning and testing out new skills:

S3: *"I am a design student, but I did some heavy back-end programming. I learned so much from this process, daring to jump into the unknown. And if I was stuck, I got help, especially from one team member. We spent a lot of time on Discord together.*

One interviewee expressed they had fun working together: S1: *"We had a good time with each other and enjoyed ourselves a lot at Zoom. Everyone had camera on, and we were a really good group of friends really."*

5.1.3 The effect of TWQ on Project Success

Figure 8 shows the effect of TWQ on the dependent variables for each of the investigated semesters. The results are from the surveys from 2019 to 2021.

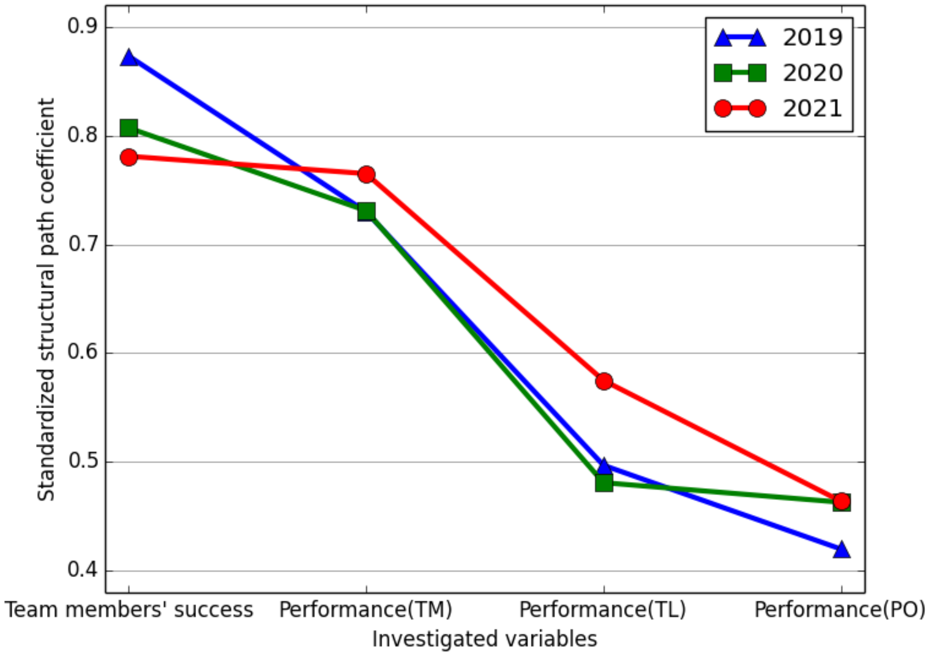
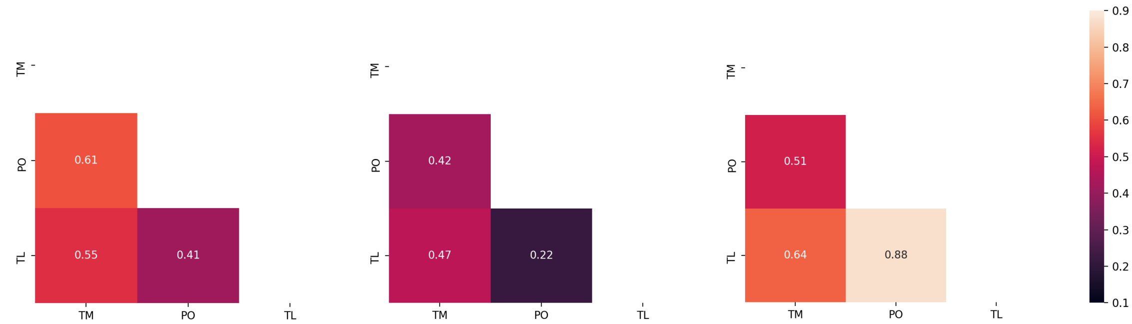


Figure 8: Standardized path coefficients from TWQ to the four dependent variables

Studying Figure 8, we can see that the effect of TWQ on Team Members' Success was significantly higher in 2019 than in both 2020 and 2021. The team member perception of Team Performance was slightly higher in 2021 than in 2019 and 2020. Looking at the team leaders' perceptions of Team Performance, we can see that in 2021, it was significantly higher. The POs (teaching assistants) perceived Team Performance equally in 2019 and 2020 and lower in 2019.

5.1.4 Correlation between the raters



(a) 2019 **(b)** 2020 **(c)** 2021 **(d)**

Figure 9: The correlation between the evaluation of the raters and the dependent variables for each of the investigated semesters

In Figure 9 the following applies: TM = Team Members, PO = Product Owner, TL = Team Leaders. The figure shows the correlation between the rater groups and the dependent variables.
The most notable differences are the correlations between PO (Product Owner) and the two other raters. The correlations between PO and TM and TL and TM were similar in all the investigated semesters. The correlation between PO and TL varied significantly. In 2019, the correlation was 0,41. In 2020, it was 0,22, and in 2021 it was 0,88. Considering the correlations in 2021, we can see that the correlation between TM and TL and PO and TL are the highest of all the semesters. The correlation between TM and PO is somewhat lower than in 2019 but significantly higher than in 2020. The p-values for all correlations in all semesters are below 0,05 ($p < 0,05$), except from the TL and PO correlation in 2020 which was above 0,05 ($p > 0,05$).

5.1.5 Teamwork is essential

During the interviews, each interviewee was asked what they think was the most valuable learning outcome from IN2000. Five out of the seven interviewed students explicitly stated that ‘teamwork’ was the most crucial aspect they learned. The two other students briefly mentioned teamwork but emphasized experiencing the process leading up to the final product.

The MET representative was also asked the same question. The representative emphasized the importance of "real" project work, similar to what the students will face in the industry. The representative also highlighted that learning how to utilize their weather data (API) and learn how to read documentation is desirable.

5.1.6 Teamwork before and after the lockdown

The interviewees were asked how they perceived the teamwork before and after the lockdown. Most students reported the teamwork became worse as a result of the lockdown.

S2: *"We met physically at the university prior to the lockdown. We had a plan to conduct daily standup meetings at the university, but we did not have the number of meetings per week which we agreed to before the lockdown. We worked less with project after the lockdown occurred."*

One interviewee mentioned that the teamwork became more formal, but they were still able to cooperate:

S4: *"The teamwork became more formal after the lockdown. We were able to work after the lockdown, but we missed the informal interactions. I think the teamwork would have been better if we worked physically since we lost 'social happenings'."*

One of the students expressed that the teamwork probably became better as a result of the lockdown:

S3: *"I think the lockdown resulted in our team conducted a better project than we would have, if the semester went as normal. Cancellation of spare time activities and people getting laid of work ended in us having a lot of time to work with the project. I think the lockdown also did something in favor for our teamwork, not only affected us negatively."*

An interviewee pointed out a noticeable experience regarding when the team met physically to collaborate:

S3: *"At the end of the project work, we could physically meet. One day, we decided to gather all team members at my house to work together. We were not able to work physically as we got more easily tired. You were not alone, so you could not take breaks, walk around and make food alone. Suddenly we were all cluttered into my small living room. I think one reason could be that people were used to their setups at home, but this day, they had to sit on a stool in my small living room. So, despite us complaining throughout the entire project work that we had to work remotely, we were not able to work when we actually first met."*

5.2 Sociability

This subsection presents findings from the sociability scale.

5.2.1 Descriptive statistics

		2020		2021	
No.	**Item**	**Mean**	**SD**	**Mean**	**SD**
1	The virtual learning environment enables me to easily contact my teammates	3,90	0,83	3,90	0,66
2	I do not feel lonely in the virtual learning environment	3,65	0,78	3,58	0,71
3	The virtual learning environment enables me to get a good impression of my teammates	3,31	0,75	3,38	0,72
4	The virtual learning environment allows spontaneous informal conversations	3,57	0,89	3,45	0,75
5	The virtual learning environment enables us to develop into a well performing team	3,54	0,70	3,67	0,64
6	The virtual learning environment enables me to develop good work relationships with my teammates	3,48	0,70	3,48	0,69
7	This virtual learning environment enables me to identify myself with the team	3,52	0,67	3,45	0,62
8	I feel comfortable in the virtual learning environment	3,86	0,73	3,95	0,56
9	The virtual learning environment allows for non-task-related conversations	3,37	1,00	3,33	0,81
10	The virtual learning environment enables me to make close friendships with my teammates	2,73	1,00	2,70	0,76

Table 8: Descriptive statistics of the sociability items

Table 8 shows the mean and standard deviation values of the evaluation of the virtual learning environment items from the 2020 and 2021 semesters. All the items are calculated at the team level: each raters' evaluation aggregated into their team. In 2020 and 2021, all the nine first items' mean values were between 3 and 4 (3,6 on average) and are relatively similar for both years. The evaluation of item 10 (making close friendship) was significantly lower than the other items' evaluation in both years (below 3). Item 5 (develop into a well-performing team) increased the most from 2020 to 2021 among all the items. The standard deviation was overall lower in 2021 compared to 2020.

Both in 2020 and 2021, the Cronbach alpha values were relatively high, 0,94 and 0,95 respectively.

5.2.2 Interviews

As item 10 in the Sociability scale was evaluated significantly lower than any other items, this aspect was further investigated during the interviews. One interviewee stated that they would probably know the other team members better if the teamwork was physical:

S6: *I think I would have known the others better if there were physical collaboration throughout the semester. In the beginning, when we met physically, we discussed and got to know each other. It is easier to joke and get to know each other in the same room than in a formal chat".*

Another student pointed out the formal aspects of the digital meetings made it hard to get to know each other:

S4: *"I feel like I connected with the team members, but more on a colleague-level and not 'friend'. The social aspects disappeared when the teamwork went digital and as all activities became formal.*

The lack of physical interaction made establishing a close friendship with team members digitally hard.

S5: *"I was not able to make close friends with the others, more a colleague-relation. (...) I think getting close to someone digitally is very hard. (...) Making friendship is based on prosperity, doing stuff together, but also being comfortable while disagreeing* *and discussing. Digitally, I think the human aspects are watered out, hence making close friendship hard."*

One student stated that getting to know each other digitally was not that hard as they did a lot of informal activities:

S1: *"We got to know each other primarily through Zoom. I think what helped was having the camera on. I heard from other teams where the cameras were off, and they talked to black screens. I think having the camera on is crucial in other to establish closeness with team members. And we also used Zoom not only for formal meetings, but we also did a lot of 'fun-stuff together as well."*

One interviewee did not find making friendship relevant as all team members knew each other from before. However, the student expressed they missed working with other students: S2: *"I missed working with students I did not know from before."*

5.2.3 Correlation between TWQ and Sociability

Figure 10 shows the correlation between the six TWQ constructs and sociability as a construct. The sociability constructs in Figure 10 represents the ten sociability items (see Table 8) aggregated. Project Success is omitted as this correlation compares teamwork quality and not its effects, with Sociability.

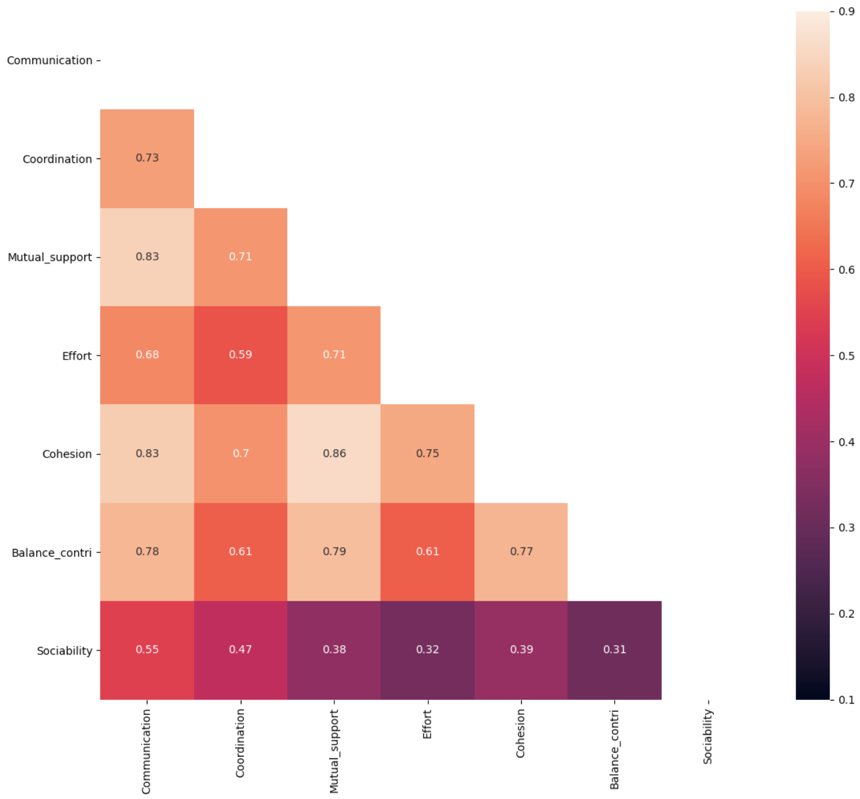


Figure 10: Correlation between the TWQ constructs and Sociability from 2021

Figure 10 shows the correlations with a lower-triangle mask. The p-values for all correlations are below 0,05 ($p < 0,05$). The TWQ constructs Communication, Mutual support, and Cohesion correlates the most with Sociability. Balance of member contribution correlates the least. The TWQ construct Cohesion correlates the most with the other TWQ constructs in general. When comparing the correlations in Figure 10 with data from 2020, the same tendency is present. The TWQ constructs Communication, Coordination, and Balance of member contribution also correlate the most with Sociability in 2020.

5.3 Students' Motivation

The team members' motivation was a significant factor in how well the teams performed. During the presentations after the project in 2020, all the teams were asked the same question: *"How did the coronavirus situation (closing down the Campus) and the fact that the grade was only passed/not passed influence your motivation in the course?"* The most common reply was:

"The motivation became lower at once, but when we really started to work together as a team, we wanted to make a good app, write a good report and learn how to use some agile practices during the teamwork, the motivation came back."

All interviewees reported that the motivation dropped after the lockdown in the 2020 semester. There was not one particular reason why the motivation dropped. Some interviewees mentioned the changing in grade from A-F to passed/not passed was the main reason:

S2: *"It was demotivating when the grading was changed since we knew our product was sufficient. We were less motivated to work with the course, and as a result, I think we ended up putting in less effort than we normally would if the grade had not changed."*

Another student pointed out that change in grade affected their admission to a master's degree program:

S7: *"... Four out of five team members in my team wished to apply for a master's degree program. As IN2000 was not a mandatory course in our current study programs, we took it since it was 20 ECTS which would heavily count towards the master admission, and we had a lot of interesting ideas before the beginning of the course. When the grading was changed, a lot of value was lost. My previous experience was that passed/not passed courses did not require a lot of work. Why should we put in a lot of effort if we get a 'pass' anyways"?*

Even though the motivation dropped at first, the students were able to increase it throughout the project work. During the presentations, most students reported that their chosen project case was engaging, which boosted the motivation. One interviewed student pointed out that the motivation increased as they wanted to make the best out of the situation:

S1: *"The team members agreed that we should do the best we can, even if there are no letter grades anymore. (. . .) If we put in enough effort, we will still have a great outcome of this course; we can show the product during job interviews. The motivation rose when we changed the mental focus from grade to personal gain".*

Some students still pointed out that they were not able to recover from the drop in motivation:

S2: *"I felt like the motivation was low throughout the project period after the lockdown. It became harder to motivate oneself to work with the course. Even though we never stopped having meetings and produced code, I felt like the drive was gone. (. . .) The team did never discuss how we could raise the motivation. We just accepted the fact that this is how the course is going to be".*

5.4 Tools

Table 9 shows the tool usage of student respondents from the surveys. As all teams had to use Android Studio and GitHub to develop the app, these tools were not options in the survey. Table 9 presents the tools in which the students teams chose themselves.

Tool	Count	Percentage	Tool	Count	Percentage
Slack	122	80,0%	Zoom	156	63,4%
Zoom	111	73,5%	Trello	149	60,6%
FB Messenger	64	42,4%	Discord	129	52,4%
Discord	56	37,1%	Microsoft Teams	104	42,3%
Google Services	19	12,6%	FB Messenger	103	41,9%
Microsoft Teams	11	7,3%	Slack	75	30,5%
Trello	11	7,3%	Google	42	17,1%
Monday	4	2,7%	Miro	35	14,2%
Notion	3	2,0%	Jira	32	13,0%
Skype	2	1,3%	Figma	10	4,1%

Table 9: Usage of tools. The left and right sub-tables represents 2020 and 2021 respectively

The most popular tool was Slack, followed by Zoom and Facebook Messenger. Most students used a combination of some of them. The total percentage exceeds 100 in both tables as respondents could submit more than one tool. Note: Google Services include Google Drive and Google Hangouts. The survey was somewhat changed in 2021 to capture that the students often use more than one tool in parallel. Microsoft teams were used significantly more than in 2020.
Even though Table 9 shows the tool used on an individual level, it reflects the tool usage of that team. E.g., If a respondent said they used Zoom, we can assume that the rest of the team used it as well.

During the interviews, the students were asked which tools they used and how they found them. Most respondents underlined that combining tools was beneficial:

S1: *"We mainly used Slack, Zoom, and Google Drive. We did not want to use Facebook to separate work and spare time. We used Slack for messaging, Zoom for meetings, and Google Drive for sharing files. Slack was nice since we could create specific channels regarding the different aspects of the project."*

One respondent emphasized they used tools to communicate, often orally instead of typing in a chat:

S5: *"We ended up using Discord because of voice channels. In this project, we found it useful to have a platform where one could easily 'jump' in and out of rooms. We spent much time talking, and we did not write a lot of messages. So having the voice channels of Discord really aided our teamwork."*

One respondent expressed their team was not able to utilize the tools fully:

S6: *"I felt Trello was hard to relate to. This resulted in Trello never being prioritized, and we had a messy backlog. We mostly used Slack to discuss which tasks to choose, but as we chose tasks, we did not update the board in Trello. (...) We had never scheduled meetings. We just randomly met on Zoom from time to time. I feel like finding timeslots and breaking the ice using the tools was hard."*

5.5 Summary

To conclude the Section 5, the reported findings are related to both research questions of this thesis.

A well-performing student team in a virtual environment is a team that handled the transmission from physical collaboration to virtual well. They had frequent meetings, followed an agile framework, and utilized the digital collaboration tools they found helpful. Well-performing student teams might have had some initial drop of motivation after the lockdown but were able to see the benefits of the learning outcomes despite not meeting physically. Well-performing student teams also scored higher on TWQ, Project Success, and Sociability.

A non-well-performing student team was not able to the same degree to adapt to the virtual environment. They were not able to see the learning outcomes and did not put enough effort into project work. They could not get back the motivation to a lesser degree and instead focused on the misery they found themselves in. A common characteristic with such teams was the lack of a facilitator role, someone taking the initiative to advance with the teamwork.

Most teams used some sort of combination of multiple tools. The most popular tool in 2020 was Slack, which was often combined with Zoom. In 2021, Zoom and Trello were the two most popular. Student teams that were able to utilize the functionality of the tools overall performed better than those that did not. Using digital collaboration tools has been essential to helping aid teamwork. Without the collaboration tools, the virtual teamwork would nearly be impossible.

6 Discussion

In this section, the findings from the Section 5 Results, are discussed in relation to the research questions and previous findings in the area.

6.1 Well-performing student team working virtually

This sub-section discusses the first research question of this thesis:

RQ1: ***"What characterizes a well-performing student team in a software engineering capstone course working virtually?"***

The findings show that the effect of TWQ on Project Success did not drop significantly in 2020 (See Figure 8). As reported by the respondent from The Norwegian Meteorological Institute (MET), the quality of the developed products did not decrease during the pandemic (the 2020 and 2021 semesters) compared to 2019. One reason for this could be that during the COVID-19-outbreak semester (2020), the students were given a more thorough introduction to the course's technical aspects. Compared to 2019, the students were given three additional lectures on the technical aspects: Android Studio, Kotlin, and API. In addition, the students had to deliver an extra mandatory assignment. The students had a better precondition in 2020 and 2021 than in 2019 regarding the technical aspects.

Another reason why the teamwork and the effect of TWQ on Project Success did not drop in 2020 could be that some improvements to the course were implemented from 2019 to 2020. In 2019 some teaching activities were done suboptimal, and some of the TAs had no experience with the project work themselves and no experience in supervising student teams. Improving the teaching activities in 2020, such as introducing the kickoff event, likely made an impact.

In 2021, the students were more used to work virtually. The University of Oslo announced early that the spring semester of 2021 would be digital several months before IN2000. A combination of consistency and clear guidelines from the University made the semester more reliable compared to 2020. Also, as the students were used to working digitally from the previous semesters, they did not have to adapt to change of work situation mid-project like in 2020.

6.1.1 Motivation

The findings presented in the previous section reveal that motivation was a huge factor regarding how well the student teams performed. Teams that were unmotivated found it hard to collaborate, hence making the teamwork worse. As the surveys were not directly measuring motivation (only indirectly in some questions of the TWQ construct cohesion), it is hard to compare motivation between the investigated semesters. The interviews, however, addressed motivation more in-depth. Both during the project presentations in 2020 and the interviews, the students reported that the motivation was lower in the 2020 semester when the lockdown occurred. Factors that affected the motivation were that students were temporarily laid off from work, and the campus closed down. The students were unfamiliar with working in teams virtually and using digital collaboration tools. Another factor that impacted the motivation negatively was the change in grade from A-F to passed/not passed.

A study was done on students in a project work course offered by the University of Southeastern Norway. Data collected during the lockdown reports that lack of motivation on an individual level negatively affected teamwork (Ahmed et al., 2020). As students were dealing with other issues introduced by the lockdown, such as mental disturbance and lack of focus, less effort was put into the teamwork. The findings reported by Ahmed et al. (2020) show that the direct consequence of worse collaboration in student teams resulted in worse developed products. However, this was not the case with IN2000. Even though the motivation dropped immediately after the lockdown, the motivation came back, and the quality of the developed products did not decrease. Many teams expressed that the motivation came back sometime after the lockdown during the project presentations and interviews. The student teams that were able to raise the motivation focused on the learning outcomes and personal gain of putting in an effort. As reported by interviewees and during the student presentation, engaging cases to work with was also a factor (Lindsjørn et al., 2021).

Another study reports that when the students in their capstone project suddenly had to work remotely (due to the lockdown), the team members' collaboration was negatively affected (Raaen et al., 2020). The findings of this thesis suggest that well-performing student teams working virtually were good at adapting to the sudden changes caused by the lockdown. Student teams that had a more 'agile mindset' going into the course could better adapt and accomplish the virtual teamwork well.

6.1.2 Communication

Findings reported in Section 5 reveal that the threshold for asking team members for help increased due to the lockdown. One interviewee mentioned that it was harder to ask ad-hoc because they did not want to bother the others too much. Another interviewee also expressed that they did not want to bother the team members when they were stuck on minor problems, hence not reaching out to them.

Other research conducted on a software engineering capstone course also reports a higher threshold for asking team members for help when stuck. Raaen et al. (2020) reported that due to lack of communication, the students in their capstone course felt a higher threshold for asking for help (Raaen et al., 2020). In agile teams, communication should be of high bandwidth and transparency (Rubin, 2012, p. 205-206). If team members feel a higher threshold for communicating, these crucial aspects are neglected, resulting in poor quality non-agile teamwork. Not being able to ask fellow team members for help can result in much idle time, which means tasks can be delayed or not finished at all.

One reason why some students reported that the threshold of asking for help increased might be that they were not used to work in a virtual environment. They applied collaboration tools to support their teamwork, but when the interactions became virtual instead of physical, it became harder to ask for help. Some mentioned they felt they were disrupting the other teammates when they asked for help. Another reason could be the absence of a distinct team leader (or coach), resulting in teams having no one taking the initiative to advance with the teamwork.

In the spring of 2020, students were heavily affected by the consequences that the COVID-19 lockdown caused. Some experienced a new living situation; these were uncertain times, making could have made it hard to put much effort into the teamwork in the course.

Another interesting finding from the 2020 semester is that the teams that were able to physically meet before the lockdown or participate on the kickoff event were overall more positive regarding the teamwork than those that did not. Having informal interactions and communication between team members is essential in agile software development, and enhances the quality of the teamwork (Nerur et al., 2005).

6.1.3 TWQ and Project Success

Descriptive statistics

The findings suggest that the investigated semesters' descriptive statistics were relatively equal (see Table 7). There was a notable decrease of all items in the 2020 compared to the other semesters. One reason the students perceived their teamwork worse in 2020 than the other semesters might be a result of the direct impacts caused by the lockdown. Interestingly, the student's perception of nearly all the TWQ variables was the highest in 2021 compared to the other semesters. One reason could be that some students might have felt "pressure" to submit a "good" result since the surveys were conducted during the presentations, where course administration was present. Other factors could be that some teams in 2021 were psychically located in the same room during the presentations. They could have communicated with each other when answering the survey, discussing what to answer on each of the questions and not daring to submit their genuine perceptions. Compared to the 2019 semester, the surveys were also conducted during the presentations, but the students were not allowed to talk while answering the survey.

Comparing the descriptive statistics of this study to the study conducted on professional teams (Lindsjørn et al., 2016), the mean values of the student teams were slightly higher in all investigated semesters (2019, 2020, and 2021) than the values of the study of professional teams. The variance among the team members' evaluation of all variables was significantly higher in the students' teams than in the professional teams, with a standard deviation of 0,45 on average in the student teams and 0,30 on average among the professional teams (Lindsjørn et al., 2021). Interestingly, among all investigated variables, Effort is the only variable that is evaluated to higher by professional teams than in student teams (all semesters). This could result from the industry having a facilitator for daily standup and performing daily standups, whereas student teams might lack the role, making the students' rate Effort lower.

The effect of TWQ on Project success

Figure 8 shows the effect of TWQ on the four dependent variables. Studying the figure shows that the effect of TWQ on Team members' Success was significantly higher in 2019 than in both 2020 and 2021. This is probably because the entire teamwork was conducted physically at the campus. The team member perception of Team Performance was slightly higher in 2021 than in 2019 and 2020. Overall, the effect of TWQ on Project Success (Success and Performance combined) as rated by the team members is high in all semesters. Looking at the Team Leaders' perceptions of Team Performance, we can see that in 2021, it was significantly higher. The reason for this might be that the students who perceived their role as team leaders were closer to their teams this year. In 2021, a more thorough team composition process could have resulted in teams being happier with the team; hence, they perceived the Team Performance better.

Figure 8 shows the POs (teaching assistants) perceived Team Performance were equal in 2019 and 2020 and lower in 2019. It makes sense that the effect of TWQ on Team Performance as evaluated by the PO is the lowest among all the raters. Among all the raters, POs are the roles that are the most distant from the teamwork.

Looking at the effect of TWQ on the dependent variables in professional teams, we can see that the effect is overall higher in student teams than in professional teams (Lindsjørn et al., 2016). One reason might be that in this study, all students (both TL and TM) answered both the TWQ and Team Performance questions in the same survey. In the survey conducted on professional teams, the TLs only answered the Team Performance questions, and not TWQ. This plays a role because the respondents often answer similarly. For example, if a respondent answered, e.g., "4" on average on TWQ, they likely answered "4" on Team Performance, making the answers similar.

Correlation between the raters

Figure 9 shows the correlations between the raters' evaluation of Team Performance in the investigated semesters. The most significant difference in correlation is between TL (team leader) and PO (teaching assistant). In 2019 the correlation was 0,41, it was 0,22 in 2020 and 0,88 in 2021. There are several reasons why the correlation varied this much. One reason might be that the surveys in 2019 and 2020 were answered way after the project had ended. The respondents might have had forgotten specific aspects of teamwork when they answered. In addition, the survey in 2020 had less responses (compared to the other semesters), meaning if few respondents answered within a team, a student that was not actually the team leader might have been selected as the TL in the dataset.

A reason why the correlation between the TL's and PO's evaluation of Team Performance was low in 2020 compared to the other semesters could be that several teams did not have a connection with their supervisor (PO) the entire semester. The POs themselves were also affected by the lockdown, meaning their work environment was changed, affecting their ability to supervise. When it comes to the 2021 semester, the correlation between TL and PO was higher, probably because the teaching assistants took a more active role in reaching out to the student teams they supervised. In 2021, all TAs sent out a personal email at the beginning of the project to the teams they supervised. This established both immediate connections and provided the teams with an email to communicate with their supervisor. In 2020, the supervisors were listed on a webpage, making the threshold for initial contact higher.

Another reason why the correlation was low in 2020 compared to 2021 could be the point in time the survey was answered. In 2020, the respondents answered from late June to August, including both students and teaching assistants. Waiting until August before answering the survey could have affected their memory of the teamwork aspects. In 2021, all respondents answered within a week in May.

Comparing the raters' evaluation of Team Performance of student teams with professional teams (Lindsjørn et al., 2016), the correlation between PO and TL is significantly lower in professional teams. One reason why students have a higher correlation might be that the POs are "closer" to the students. The POs are students themselves, as most of them have done the project work themselves. The POs are more involved with the student teams, like having weekly meetings. In professional teams, the PO is often more distant from the development teams. The teaching assistant POs and professional teams POs most likely have a different view of Team Performance. Teaching assistants have been involved in the teamwork, whereas professional POs consider Team Performance on how well the team was able to deliver and stay within budget. Efficiency is not relevant in IN2000 as the students are not required to deliver anything within a budget. The professional POs only see the end-product, as the student POs also see the teamwork and helps if a conflict arises.

Another reason could be POs in student teams are not doing the works of an actual PO. POs in professional teams prioritize the backlog and represent the customer. The POs in IN2000 is a supervisor and is not responsible for the functionality of the developed products. However, they do to some degree represent the course administration.

6.1.4 Sociability - friends or colleagues?

When using the Sociability scale, a finding was that the evaluation of item ten (making close friendship with team members) was significantly lower than other items' evaluation (see Table 8 in results). The respondents in the interviews expressed that making friendship virtually is more challenging than physically. The mean value of all sociability items combined was 3,51 in 2020 and 3,39 in 2021 (out of 5).

Kreijns (2007) reports that "making friendship" scored the lowest among the participants (Kreijns et al., 2007). However, the investigated study sample was constrained to a limited controlled group. In practice, students find their own ways to reach out to each other and do creative activities to promote learning, communication and establish a bond between team members (Yu et al., 2010). Furthermore, even performing such activities seems to strengthen the "colleagues bond" instead of the friendship. Student teams that were able to do such activities overall report that the teamwork was fun, resulting in better performance.

An experiment on students where the Sociability scale was applied averaged out at 3,87. This number represents the average of all sociability items, so investigating the items' separate values is unfortunately not possible. Some items from the sociability scale were removed to fit the study better (De Lucia et al., 2009). Comparing the number to the findings of this study, it is higher than in student teams (3,51 and 3,39 in 2021 and 2021, respectively). Kreijns (2007) reported a mean value of 3,00, which includes all the items in the conceptual model. The mean value reported by Kreijns (2007) is probably because it represents the mean value on an individual level, whereas this study represents the mean values from 2020 and 2021 on team level.

Even though the students were not able to establish close friendships in the virtual environment, the teams were still able to perform well. As long as the students were able to work together as a team, whether or not they were friends with their teammates seems to not have affected the performance. A sense of community within the teams seems to be more important than friendship.

Sociability was conceptualized for studying an individual's perception of a CSCL (Kreijns et al., 2007). When applying the sociability scale in this study, it has been used on a team level.

Correlation between TWQ and Sociability

Figure 10 shows the correlation between the six TWQ variables and the Sociability scale aggregated into a single construct (combining all items into a latent variable).

The results show that Sociability correlates the most with the TWQ variables Communication and Coordination. The correlation for both of these is significant ($p < 0,05$).

These two variables, especially Communication, correlate the most with Sociability, likely because the feeling of perceived Sociability is tightly related to Communication. Many items in the sociability scale contain items regarding contact with teammates. Proper Communication and Coordination are essential, especially when working virtual where there is a lack of informal interactions (Nerur et al., 2005).

6.2 Tools

This sub-section discusses the second research question of the thesis:

RQ2: ***"What role do digital collaboration tools play in student teams working virtually in software engineering capstone?"***

Tools seem to be essential when working in teams virtually. Combining tools and using them proficient seems to be a success factor. Teams that were able to utilize tools to their benefit performed better than teams that did not. There was no one specific 'go-to' tool but a combination of more. The student teams were overall able to find suitable tools to carry out the teamwork virtually. Most respondents answered they used some form of ESN, either Slack or Discord, in either of the years (see Table

9). Interestingely, Trello was significantly more popular in 2021 compared to 2020. One reason for this could be that many teams visualized their tasks on a board, which many teams expressed they did during the presentations in 2021.

6.2.1 Enterprise Social Networking

As presented in Table 9, Slack and Discord were the most popular ESNs in 2020 and 2021, respectively. Discord was significantly more prevalent in 2021 compared to 2020. One reason for this could be that many students use Discord in their everyday lives, meaning it is easier to facilitate teamwork on a platform already being used instead of downloading and setting up a new one. In 2020, the students were not used to virtually, which means there where "trying and failing" different tools in 2020. As reported in the interviews, many students started using Slack but moved to Discord. In 2021, as the students were used to work remotely from home, they had likely already found their desired collaboration tools before the teamwork and used it in the teamwork in 2021.

When comparing the two semesters where the students worked virtually, we can see that Slacks' popularity decreased, and Discords' and Microsoft Teams' increased. Microsoft Teams can be considered an ESN as it meets the criteria to be considered one (Leonardi et al., 2013). Microsoft Teams were more prevalent in 2021 than 2020, probably due to the course administration using it. The University of Oslo provided all enrolled students with an account, and IN2000 actively used it to communicate to the students. The course administration communicated more directly with students using an ESN in 2021 than in 2020 when only the courses' web page was used. This could also have positively impacted the students' perception of the course, responding more positively to both Sociability and TWQ in 2021.

A study on agile teams working virtually investigated the usage of ESNs, the importance of such tools, and potential obstacles (Stray et al., 2019). The study was done on global teams, which are teams that are scattered across different countries. Compared to IN2000, the students were mainly in the same country, but they still had to use tools to collaborate. One reported disadvantage with an ESN is unbalanced activity, a few members responsible for a majority of the exchanged messages (Stray et al., 2019). This issue is highly related to the TWQ attribute *balance of member* *contribution*. The investigated student teams in this thesis seemed not to be an issue. *Balance of member contribution* was evaluated relatively equal in all semesters; when the teamwork physical and virtual. Unbalance was not an issue in the student teams, probably because the number of members within the student teams were smaller compared to the global teams studied (Stray et al., 2019).

6.2.2 Proficient use of Tools

Proficient use of collaboration tools is crucial when working virtually. Some teams reported that using tools for the sake of it was not beneficial. Teams should carefully select a few tools that suit their teamwork. When using too many tools at once, it is hard to keep all tools up-to-date, and one should consider maybe dropping a tool.

6.3 Implications for Theory

The two main theoretical frameworks applied in this thesis are 1) the TWQ construct and relation to project success adapted by (Lindsjørn et al., 2016) and 2) the Sociability scale by (Kreijns et al., 2007). They were both used to measure the virtual teamwork and tool usage in a software engineering capstone course.

6.3.1 TWQ and Project Success

The TWQ construct and the relation to project success have been previously proven to be suitable for measuring teamwork quality and its effect on project success, both in traditional development (Hoegl & Gemuenden, 2001) and agile (Lindsjørn et al., 2018; Lindsjørn et al., 2016). The concepts of TWQ and project success worked well in this study as well. However, one discovery made was that the attribute of *Efficiency*, which is a part of Team Performance, seems to not be that important in student teams. *Efficiency* refers to whether or not the teams meets expectations regarding project quality, such as time and cost (Lindsjørn et al., 2016). The students have to deliver within the deadline, but otherwise, there are no formal requirements regarding the project quality. They are not required to deliver on budget, for instance. Therefore, when using the Team Performance in the future on students teams, I suggest removing *efficiency* when researching teamwork quality in a software engineering capstone course where a budget is not relevant.

When rephrasing some items for the 2021 survey, a discovery was made. The purpose of rephrasing some items was to avoid having to reverse-code the items after the data collection. Many of the items that had to be rephrased were items containing the word *conflict*.

6.3.2 The Sociability Scale

The concept of Sociability was found to be applicable to student teams.

One interesting finding when using the Sociability scale was that the Cronbach alpha value for both 2020 and 2021 was unsatisfactory (0,94 and 0,95 respectively). Having such high alpha values on multiple datasets might hint that some of the Sociability scale items are redundant. The Sociability scale should be revised in the future, where some items should be removed or rephrased to lower the Cronbach alpha value.

6.4 Implications for practice

The students should be given a thorough introduction to the useful tools to help increase the quality of interactions between the team members. The students are only given an introduction to the tools they have to use, such as Android Studio and GitHub. GitHub is a useful collaboration tool for version control but lacks the functionality to communicate directly with the other team members. As of now, many teams use different tools sub-optimal. Having a lecture, seminar, or workshop where different tools are showcased, such as, e.g., Slack would be beneficial, even if the teamwork were to be carried out physically in the future. The students should not be forced to use a specific tool, but showcasing some would be helpful.

Implement mandatory physical workshops for all teams early in the semester and during the project work. Meeting and establishing relations with team members is crucial, even if the teamwork is either physical or virtual. Having some physical meetings is desirable as team members in distributed teams that have met each other physically ensures better communication (Dorairaj et al., 2012; Stray et al., 2019).

The kickoff event offered in IN2000 is already a massive success, as reported by the students and teaching assistants, so having more than one event that gathers the student teams would be an excellent addition to the course.

The teams should be having a fixed number of mandatory meetings with their supervisor throughout the project work. This would come at the cost of the teams' freedom, and the teaching assistants would have to work more. However, making sure that the students' teamwork is working out is essential.

Introduce peer-reviewing. This proposal might be a little controversial as teams can plagiarize each other. Nevertheless, this could be an excellent opportunity for the teams to share experiences and success factors of teamwork with the other teams.

6.5 Limitations

This sub-section presents and reflects upon the limitations of this thesis.

6.5.1 Data collection

Others have collected some of the data applied in this thesis. My supervisor and Tegelaár (2020) gathered and parsed the survey data from 2019 (Tegelaár, 2020). I, myself, have collected the survey data from 2021, parsed the survey data from both 2020 and 2021, carried out a total of eight interviews, and made observations when I worked as a teaching assistant in IN2000.

Point in time of Data Collection

The interviews conducted as a part of this thesis were conducted nearly a year after the students finished their projects (the students delivered in May 2020; interviews were conducted in February/March 2021). A disadvantage with the long timespan from project to interviews is that some might have forgotten certain aspects. This was clear as some respondents sometimes stated they did not remember the answer to some of the questions. Fortunately, most respondents eventually remembered more the longer the interviews lasted. An advantage of waiting nearly a year before interviewing is that the respondents had time to process their thoughts. If the interviews were conducted just after the semester had ended, the students would probably not yet have seen the value and learning outcome of participating in a software engineering capstone course. In addition, the survey data were collected at the end of the semester, and the summer break had just started. Reaching out to students and asking them for participation during the summer break might have ended in fewer respondents.

Assumptions in the Data

Both in the original article on TWQ and in the artcile on agile teams, the term 'Team Leader' refers to the leader of the team, which is an existing role in the investigated teams (Hoegl & Gemuenden, 2001; Lindsjørn et al., 2016). As the team leader role is not directly transferable to the students' teams, some assumptions were made in the data collection of this thesis. During the surveys, the respondents had to evaluate (on a scale from 1 to 5) to which degree they were the Scrum Master / Team Leader of the team. When the data was analyzed, the students that perceived their role as the Scrum Master / team leader the highest were selected as the team leader (TL) in the dataset. If two or more students on the same team responded that they were team leaders to the same degree, one of them was selected as a leader for that specific team. This could have resulted in respondents who were not actually the team leaders of their team being chosen as the team leader in the dataset. Most teams had only one respondent who perceived their team leader's role the highest, so this edge-case did not often occur.

A drawback with approach is that the student who perceived their role as the team leader the highest might not have been the actual team leader. Another drawback is that if several students on the same team responded they were team leader to the same degree, randomly selecting one of them as the team leader might not be representative.

If there were few responses from a specific team, all teams had to be represented with at least one role. This means a TM, a TL, and a PO. If teams in the dataset missed one of these roles, that team was removed from the dataset. Removing some teams from the dataset means that not all student teams are represented. In 2019 only one team was removed from the dataset. In 2020, due to the low response rate, four teams were removed. Removing underrepresented teams was also the procedure in the study done on professional teams (Lindsjørn et al., 2016).

6.5.2 Study Sample

The response rate of surveys in 2020

The survey data gathered in 2020 had a lower response rate than the two other semesters. The main reason for this is a combination of several factors. In 2020, the students were not "forced" to answer like they were in 2019 and 2021, where the surveys were held during the presentations. In 2020, the students were sent invitations to the survey by e-mail after the semester had ended. This resulted in a lot fewer responses.

Interviews

Five out of the seven interviews were conducted with students that were teaching assistants in the software engineering capstone course the year after they enrolled in it (see Table 6). A disadvantage with this could be that the data is somewhat biased as teaching assistants often are more engaged than regular students. Many requests were sent out to students that took the course in 2020, but the most response came from teaching assistants. When receiving little response, the e-mails were generated in a "spear-phishing"-way, but little response was still received. It seems like teaching assistants are more likely to participate in interviews compared to regular students. Another reason why some students did not answer the inquiry might be because they were unsatisfied with their teamwork and the developed products and did not want to share their experiences.

2021 Surveys

The surveys in 2021 were announced during the students' presentations. It was clear that some teams were located in the same room as they had turned on their cameras. Being located in the same room while answering the survey might have affected the result somehow, as they could have discussed what they should answer on each of the questions.

6.5.3 Own Participation

From the early days of IN2000, I have been much involved in the course. In 2018 I took the pilot, from 2019 to 2021, I worked as a teaching assistant in the course. I have produced course material throughout the years, corrected mandatory assignments and exams, and held guest lectures. The fact that I have been such involved in the course have made me come up with opninions on the course, which could have led to some bias.

6.6 Validity and Reliability

This sub-section briefly reflects upon how this thesis has addressed the aspects of validity and reliability.

6.6.1 Construct Validity

Construct validity refers to what extent the result of a study reflects the researchers' own subjective view on the investigated phenomenon (Yin, 2009, p. 41-42). To address construct validity in this thesis, multiple sources of evidence have been used, such as: surveys, interviews and observations. Among the multiple sources there were surveys from three different semesters, eight interviews and informal observations.

6.6.2 Internal Validity

Internal validity refers to the extent to which the effect found in a study actually results from the investigated factors and not some hidden factor (Robson, 2002; Runeson & Höst, 2008, p. 542). Pattern matching has been used to strengthen the internal validity of this study. The patterns relate to the predictions of dependent and independent variables. Either basing the predictions on previous findings or come up with proposals before data collection to figure out what could have caused the effect is desirable (Yin, 2009, p. 136-139). Specifically for this study, it includes how TWQ affects Project Success.

6.6.3 External Validity

External validity is concerned with the degree to which the findings of a study are generalizable (Runeson & Höst, 2008). The use of theories is a way to strengthen the external validity of a study (Yin, 2009, p. 41-44). The initial TWQ construct as introduced by Hoegl & Gemeunden (2001) and the revised version by Lindsjørn et al. (2016) have been used in this study. The Sociability scale by Kreijns et al. (2007) have also been applied.

6.6.4 Reliability

Reliability addresses the integrity of the study (Robson, 2002, p. 551). If another researcher were to conduct the same study, following the same steps, they should arrive at the same findings and conclusions (Runeson & Höst, 2008; Yin, 2009, p. 45). To strengthen the reliability of this thesis, a case study database has been used. All relevant material was stored in the database.

7 Conclusion

A case study has been conducted on a software engineering capstone course consisting of surveys, interviews, and observations. The investigated semesters were 2019, 2020, and 2021. In 2019 the teamwork was carried out physically; all teaching and team activities were conducted at campus with all team members present. In 2020 the teamwork started physically at the campus but became digital in the middle of the semester. Finally, in 2021, all the teamwork was carried out digitally. An essential part of this study has been the comprehensive surveys measuring teamwork quality (TWQ) and project success for all three semesters. The surveys were evaluated by three rater groups: team members (TM), team leaders (TL), and product owners (PO).

The first research question addressed key characteristics of well-performing student teams working virtually. The results showed a positive relationship between the evaluation of teamwork quality and project success for all three rater groups. The analysis also shows a positive relationship between the evaluation of TWQ and how they could adapt to the virtual work environment. Another finding was that well-performing teams had a clear facilitator role, ensuring frequent communication and better coordination of tasks. The survey of the teams working virtually is also compared to teams working physically. The findings show that teams working physically evaluated communication slightly higher than virtual teams. However, all other teamwork quality aspects were evaluated similarly or lower. Compared to studies conducted on professional teams, the mean values of all raters were overall higher in student teams for all years than in professional teams. The effect of TWQ on team performance was also higher in student teams compared to professional teams for all raters. The most significant difference was the effect of TWQ on POs' evaluation of team performance. One reason for this could be that the POs in the student teams also supervised the teams with a more close relation to the teamwork.

The second research question aimed to understand what role virtual collaboration tools play in virtual student teams. Most teams used several collaboration tools to facilitate their virtual teamwork. Well-performing student teams had a common understanding of how the tools could effectively benefit the teamwork. In particular, the integration of data between the different tools - both freely selected tools (e.g., Slack and Trello) and tools they were required to use (e.g., GitHub and Android Studio). Further, well-performing teams recognized the teams' norms which is essential when collaborating using tools.

Future work

Future work suggests changing the teaching assistant's role from a product owner (PO) role into a facilitator role (team leader, coach). Instead, the PO role should be more similar to professional teams where the PO role represents the customer and works with the backlog. Implementing this change would make the basis of comparison between student teams and professional teams more reliable.

Investigating the Sociability instrument, used in 2020 and 2021, would be interesting. The Cronbach alpha values for both semesters were unsatisfactory (above 0,9). Future work should revise the Sociability scale and explore which instrument items should be rephrased or removed.

It would also be interesting to investigate the teams with the lowest and highest perception of TWQ and team performance to discover characteristics of the "best" and "worst" teams regarding their teamwork.

8 References

Ahmed, S. U., Nguyen-Duc, A., & El-Gazzar, R. (2020). IT Students Project Group Work in the Day of COVID-19: Understanding the Impact and Attitudes [Number: 4]. *Norsk IKT-konferanse for forskning og utdanning*, (4). https://ojs.bibsys.no/index.php/NIK/article/view/820

Bastarrica, M. C., Perovich, D., & Samary, M. M. (2017). What Can Students Get from a Software Engineering Capstone Course? *2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering Education and Training Track (ICSE-SEET)*, 137–145. https://doi.org/10.1109/ICSE-SEET.2017.15

Chatterjee, R., & Correia, A.-P. (2020). Online Students' Attitudes Toward Collaborative Learning and Sense of Community. *American Journal of Distance Education*, *34*(1), 53–68. https://doi.org/10.1080/08923647.2020.1703479

Chow, T., & Cao, D.-B. (2008). A survey study of critical success factors in agile software projects. *Journal of Systems and Software*, *81*(6), 961–971. https://doi.org/10.1016/j.jss.2007.08.020

De Lucia, A., Francese, R., Passero, I., & Tortora, G. (2009). Development and evaluation of a virtual campus on Second Life: The case of SecondDMI. *Computers & Education*, *52*(1), 220–233. https://doi.org/10.1016/j.compedu.2008.08.001

Dickinson, T. L., & McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. *Team performance assessment and measurement* (pp. 19–43).

Dorairaj, S., Noble, J., & Malik, P. (2012). Understanding Team Dynamics in Distributed Agile Software Development. In C. Wohlin (Ed.), *Agile Processes in Software Engineering and Extreme Programming* (pp. 47–61). Springer. https://doi.org/10.1007/978-3-642-30350-0_4

Dzvonyar, D., Alperowitz, L., Henze, D., & Bruegge, B. (2018). Team composition in software engineering project courses. *Proceedings of the 2nd International Workshop on Software Engineering Education for Millennials*, 16–23. https://doi.org/10.1145/3194779.3194782

Goodwin, L. D., & Leech, N. L. (2006). Understanding Correlation: Factors That Affect the Size of r. *The Journal of Experimental Education*, *74*(3), 249–266. https://doi.org/10.3200/JEXE.74.3.249-266

Hoegl, M., & Gemuenden, H. G. (2001). Teamwork Quality and the Success of Innovative Projects: A Theoretical Concept and Empirical Evidence. *Organization Science*, *12*(4), 435–449. https://doi.org/10.1287/orsc.12.4.435.10635

IN2000. (n.d.). IN2000 – Software Engineering med prosjektarbeid - Universitetet i Oslo. https://www.uio.no/studier/emner/matnat/ifi/IN2000/index.html

Jabangwe, R., Šmite, D., & Hessbo, E. (2016). Distributed software development in an offshore outsourcing project: A case study of source code evolution and quality. *Information and Software Technology*, *72*, 125–136. https://doi.org/10.1016/j.infsof.2015.12.005

Janz, B. D. (1999). Self-directed teams in IS: Correlates for improved systems development work outcomes. *Information & Management*, *35*(3), 171–192. https://doi.org/10.1016/S0378-7206(98)00088-3

Katzenbach, J., & Smith, D. (1993). The Discipline of Teams. *Harvard Business Review*, *71*(2), 111–120.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, *19*(3), 335–353. https://doi.org/10.1016/S0747-5632(02)00057-2

Kreijns, K., Kirschner, P. A., Jochems, W., & van Buuren, H. (2007). Measuring perceived sociability of computer-supported collaborative learning environments. *Computers & Education*, *49*(2), 176–192. https://doi.org/10.1016/j.compedu.2005.05.004

Leonardi, P. M., Huysman, M., & Steinfield, C. (2013). Enterprise Social Media: Definition, History, and Prospects for the Study of Social Technologies in Organizations. *Journal of Computer-Mediated Communication*, *19*(1), 1–19. https://doi.org/10.1111/jcc4.12029

Lindsjørn, Y., Almås, S., & Stray, V. (2021). Exploring motivation and teamwork in a large software engineering capstone course during the coronavirus pandemic

[Number: 1]. *Nordic Journal of STEM Education*, *5*(1). https://doi.org/10.5324/njsteme.v5i1.3938

Lindsjørn, Y., Bergersen, G. R., Dingsøyr, T., & Sjøberg, D. (2018). *Teamwork Quality and Team Performance: Exploring Differences Between Small and Large Agile Projects* [Accepted: 2018-06-07T07:53:35Z Publication Title: 267-274]. https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2500705

Lindsjørn, Y., Sjøberg, D. I. K., Dingsøyr, T., Bergersen, G. R., & Dybå, T. (2016). Teamwork quality and project success in software development: A survey of agile development teams. *Journal of Systems and Software*, *122*, 274–286. https://doi.org/10.1016/j.jss.2016.09.028

Løvold, H. H., Lindsjørn, Y., & Stray, V. (2020). Forming and Assessing Student Teams in Software Engineering Courses. In M. Paasivaara & P. Kruchten (Eds.), *Agile Processes in Software Engineering and Extreme Programming – Workshops* (pp. 298–306). Springer International Publishing. https://doi.org/10.1007/978-3-030-58858-8_31

Mahnic, V. (2012). A Capstone Course on Agile Software Development Using Scrum [Conference Name: IEEE Transactions on Education]. *IEEE Transactions on Education*, *55*(1), 99–106. https://doi.org/10.1109/TE.2011.2142311

Majanoja, A.-M., & Vasankari, T. (2018). Reflections on Teaching Software Engineering Capstone Course. *CSEDU*. https://doi.org/10.5220/0006665600680077

Manifesto, A. (2001). Manifesto for Agile Software Development. http://agilemanifesto.org/

Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team Effectiveness 1997-2007: A Review of Recent Advancements and a Glimpse Into the Future [Publisher: SAGE Publications Inc]. *Journal of Management*, *34*(3), 410–476. https://doi.org/10.1177/0149206308316061

McGrath, J. E. (1964). *Social psychology, a brief introduction* [OCLC: 255591]. Holt, Rinehart; Winston.

MET. (n.d.). WeatherAPI. https://api.met.no/

Nerur, S., Mahapatra, R., & Mangalaraj, G. (2005). Challenges of migrating to agile methodologies. *Communications of the ACM*, *48*(5), 72–78. https://doi.org/10.1145/1060710.1060712

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory [Accepted: 2020-06-02T13:27:37Z Publisher: McGrawHill.]. http://vlib.kmu.ac.ir/kmu/handle/kmu/84743

Paasivaara, M., Voda, D., Heikkilä, V. T., Vanhanen, J., & Lassenius, C. (2018). How Does Participating in a Capstone Project with Industrial Customers Affect Student Attitudes? *2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*, 49–57.

Raaen, K., Sørum, H., & Gonzalez, R. (2020). IT bachelor capstone project during lockdown: Student experiences [Number: 4]. *Norsk IKT-konferanse for forskning og utdanning*, (4). https://ojs.bibsys.no/index.php/NIK/article/view/818

Radermacher, A., Walia, G., & Knudson, D. (2014). Investigating the skill gap between graduating students and industry expectations. *Companion Proceedings of the 36th International Conference on Software Engineering*, 291–300. https://doi.org/10.1145/2591062.2591159

Richardson, J., & Swan, K. (2003). Examing Social Presence in Online Courses in Relation to Students' Perceived Learning and Satisfaction [Accepted: 2011-03-15T15:37:11Z]. https://www.ideals.illinois.edu/handle/2142/18713

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers* (2nd ed). Blackwell Publishers.

Rodríguez, D., Sicilia, M. A., García, E., & Harrison, R. (2012). Empirical findings on team size and productivity in software development. *Journal of Systems and Software*, *85*(3), 562–570. https://doi.org/10.1016/j.jss.2011.09.009

Rubin, K. S. (2012). *Essential Scrum: A practical guide to the most popular agile process*. Addison-Wesley.

Runeson, P., & Höst, M. (2008). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, *14*(2), 131. https://doi.org/10.1007/s10664-008-9102-8

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in Teamwork? [Publisher: SAGE Publications Inc]. *Small Group Research*, *36*(5), 555–599. https://doi.org/10.1177/1046496405277134

scrum.org. (n.d.). Scrum.org. https://www.scrum.org/about

Seaman, C. B. (1999). Qualitative methods in empirical studies of software engineering [Conference Name: IEEE Transactions on Software Engineering]. *IEEE Transactions on Software Engineering*, *25*(4), 557–572. https://doi.org/10.1109/32.799955

Sommerville, I. (2019). *Engineering software products* (First edition). Pearson.

Stålhane, T., Deraas, B., & Sindre, G. (2020). What competence do software companies want from university graduates? [Number: 1]. *Nordic Journal of STEM Education*, *4*(1), 1–15. https://doi.org/10.5324/njsteme.v4i1.3296

StateOfAgile. (2020). 14th Annual State of Agile Report. https://stateofagile.com/#ufh-i-615706098-14th-annual-state-of-agile-report/7027494

Stray, V., Moe, N. B., & Noroozi, M. (2019). Slack Me If You Can! Using Enterprise Social Networking Tools in Virtual Agile Teams. *2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE)*, 111–121. https://doi.org/10.1109/ICGSE.2019.00031

Strode, D. E., Dingsøyr, T., & Lindsjørn, Y. (2021). A Team Effectiveness Model for Agile Software Development. *Unpublished article.*

Tafliovich, A., Petersen, A., & Campbell, J. (2016). Evaluating Student Teams: Do Educators Know What Students Think? *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, 181–186. https://doi.org/10.1145/2839509.2844647

Tannenbaum, S. I., Beard, R. L., & Salas, E. (1992). Chapter 5 Team Building and its Influence on Team Effectiveness: An Examination of Conceptual and Empirical Developments. In K. Kelley (Ed.), *Advances in Psychology* (pp. 117–153). North-Holland. https://doi.org/10.1016/S0166-4115(08)62601-1

Tappert, C. C., & Stix, A. (2010). The Trend Toward Online Project-Oriented Capstone Courses. *Computers in the Schools*, *27*(3-4), 200–220. https://doi.org/10.1080/07380569.2010.523882

Tegelaár, L. E. (2020). Teamwork Quality in Software Engineering Education; A case study of the course IN2000 at the University of Oslo [Accepted: 2020-09-21T23:47:52Z]. https://www.duo.uio.no/handle/10852/79602

Tuckman, B. W. (1965). Developmental sequence in small groups [Place: US Publisher: American Psychological Association]. *Psychological Bulletin*, *63*(6), 384–399. https://doi.org/10.1037/h0022100

Umphress, D., Hendrix, T., & Cross, J. (2002). Software process in the classroom: The Capstone project experience [Conference Name: IEEE Software]. *IEEE Software*, *19*(5), 78–81. https://doi.org/10.1109/MS.2002.1032858

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed). Sage Publications.

Yu, A. Y., Tian, S. W., Vogel, D., & Chi-Wai Kwok, R. (2010). Can learning be virtually boosted? An investigation of online social networking impacts. *Computers & Education*, *55*(4), 1494–1503. https://doi.org/10.1016/j.compedu.2010.06.015

A 2021 Survey

Background Questions (13)	
General	1. Team number 2. Age 3. Gender 4. Highest completed education 5. Year of first encounter with software engineering 6. Year of first encounter with programming
Team Information	7. Approximately how many hours per week did you spend on the project? 8. Approximately how many times per week did your team meet during the project? 9. How happy are you with the use of stand-up meetings? 10. To what degree have you acted as a Scrum Master in the project? 11. What has been your primary function in the team? 12. What digital platforms did you use in the teams?
Technical Skills	13. How was it to use METs API?
Teamwork Quality (38)	
Communication	1. There is frequent communication within the team 2. The team members communicate often in spontaneous meetings, phone conversations, etc. 3. The team members communicate mostly directly and personally with each other 4. The communication in the team does NOT go through central persons* 5. Relevant ideas and information relating to the teamwork is shared openly by all team members 6. Important information is rarely kept away from other team members in certain situations* 7. In the team there are few conflicts regarding the openness of the information flow* 8. The team members are happy with the timeliness in which they receive information from other team members 9. The team members are happy with the precision of the information they receive from other team members 10. The team members are happy with the usefulness of the information they receive from other team members
	11. The work done on subtasks within the team is closely harmonized 12. There are clear and fully comprehended goals for subtasks within our team 13. The goals for subtasks are accepted by all team members 14. There are NOT conflicting interests in our team regarding subtasks/subgoals*
	15. The team members help and support each other as best they can 16. If conflicts come up, they are easily and quickly resolved

	17. Discussions and controversies are conducted constructively 18. Suggestions and contributions of team members are respected 19. Suggestions and contributions of team members are discussed and further developed 20. The team is able to reach consensus regarding important issues 21. The team cooperate well
	22. Every team member fully pushes the teamwork 23. Every team member makes the teamwork their highest priority 24. The team put(s) much effort into the teamwork 25. There are rarely conflicts regarding the effort that team members put into the teamwork*
	26. The teamwork is important to the team 27. It is important to team members to be part of the team 28. The teamwork positively means something to me 29. The team members are strongly attached to the team 30. All team members are fully integrated in the team 31. There were few personal conflicts in the team* 32. There is mutual sympathy between the members of the team 33. The team sticks together 34. The members of the team feel proud to be part of the team 35. Every team member feels responsible for the team
	36. The team recognizes the specific characteristics (strengths and weaknesses) of the individual team members 37. The team members contribute to the achievement of the team's goals by their specific potential 38. Imbalance of member contributions does rarely cause conflicts in our team*
Team members' Success (8)	
Work satisfaction	39. So far, the team can be pleased with its work 40. The team members gain from the collaborative teamwork 41. The team members will like to do this type of collaborative work again 42. We are able to acquire important know-how through this teamwork
Learning	43. We consider this teamwork as a technical success 44. The team learn important lessons from this teamwork 45. Teamwork promotes one personally 46. Teamwork promotes one professionally
Team Performance (15)	
Effectiveness	47. Going by the results, this teamwork can be regarded as successful 48. All demands we have set for the project was realized 49. From the project case description, the team has reached its goals 50. The performance of the team increases the understanding of methods in software engineering 51. The teamwork result is of high quality

	52. The instructors are happy with the quality of the result of the teamwork 53. The team is satisfied with the teamwork result 54. The product produced in the team, requires little rework 55. The product proves to be stable in operation 56. The product proves to be robust in operation
Efficiency	57. The instructors are satisfied with the progress of the teamwork 58. Overall, the team works in an efficient way 59. Overall, the team works in a time-efficient way 60. The team is within schedule 61. The team stays within the scheduled time
The Sociability Scale (10)	
	1. The virtual learning environment enables me to easily contact my teammates 2. I do not feel lonely in the virtual learning environment 3. The virtual learning environment enables me to get a good impression of my teammates 4. The virtual learning environment allows spontaneous informal conversations 5. The virtual learning environment enables us to develop into a well performing team 6. The virtual learning environment enables me to develop good work relationships with my teammates 7. This virtual learning environment enables me to identify myself with the team 8. I feel comfortable in the virtual learning environment 9. The virtual learning environment allows for non-task-related conversations 10. The virtual learning environment enables me to make close friendships with my teammates

* rephrased in 2021 to prevent reverse-coding.

B Interview Guide

Interview guide for students

Estimated duration: 30-45 minutes

Opening

- Informal chatting with the interview object.
- Present my project
  - Inform participants regarding confidentiality and anonymity.

Background

- What was your team number?
- Which case did your team choose? (If you chose open case, please specify)
- Have you ever worked in the IT industry?
  - IF YES: What are the main differences between projects in the industry and the student project in IN2000?

Teamwork

- How would you describe the communication in your team?
  - How often did you communicate with your team?
  - What kind of communication? (meetings, chats, etc.).
- How was information shared within the team?
- How did your team coordinate the tasks?
- What was the usual answer if the team needed help?
- How did you distribute the workload?
  - IF UNCLEAR ANSWER: Did you feel the team members evenly distributed the workload?
- How did you come to an agreement when the team had discussions?
- What did you contribute the most to the team?
- How did the team members prioritize the team and project work?

Covid-19 related questions & virtual learning environment

- How did the change in grade from A-F to Passed/Not passed affect the teamwork?
  - IF THE MOTIVATION DECREASED:
    - Did it come back?
    - What did you do to raise it again?
- Was there anyone on your team that you did not know from before?
  - Did you feel it was easy to establish close friendships with these students?
    - Why? / why not?
- How was the teamwork before closing down compared to after closing down?
- Do you think the teamwork would have been better if you had not worked digitally?
  - Why? / why not?

Tools

- What digital tools did you use?
  - How did you find these?
  - Why did you use these?
- How do you think the use of digital tools has affected teamwork?
- What tools would you recommend for informal interaction between team members?

The course

- What did you think about the way teams were put together before / after the project work?
- What do you think was the most important thing you learned in IN2000.
- Have you applied some of the knowledge you acquired in IN2000 in later courses/work contexts?

Closing

- Is there something you thought I should ask that I did not? / Is there anything you want to add?

C Python script - Coefficient

```
import matplotlib.pyplot as plt
import numpy as np
import itertools
import csv

dataset = {}
marker = itertools.cycle(('^', 's', 'o'))

# Read data
with open('data.csv') as csvfile:
        reader = csv.reader(csvfile, delimiter=',')
        for row in reader:
                dataset[row[:1][0]] = row[1:]

# Initialize x-axis
x = np.array([0,1,2,3])

# Fill x-axis with values and plot data from year on y-axis
for key, value in dataset.iteritems():
        if key == 'Variable':
                plt.xticks(x, value)
        else:
                plt.plot(x, value, label=key,linewidth=3,marker=marker.
                    next(),markersize=10,zorder=3)

# Set properties for plot and display figure
plt.xlim([-0.1, 3.1])
plt.xlabel('Investigated variables')
plt.ylim([0.38, 0.92])
plt.ylabel('Standardized structural path coefficient')
plt.grid(color='grey', which='major', axis='y', linestyle='solid',
    zorder=2)
plt.legend()
plt.show()
```

D Python script - Correlation

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import seaborn

# Note: This script is run with only one semester at a time

# 2019
data_all = pd.read_excel('twq_2019.xlsx', sheet_name='Averages(TM,SM,PO)
    ')
data_i_will_use = data_all[['TM', 'PO', 'TL']]

# 2020
data_all = pd.read_excel('twq_2020.xlsx', sheet_name='
    Statistics_teams_WO_SM')
data_i_will_use = data_all[['TM', 'PO', 'TL']]

# 2021
data_all = pd.read_excel('twq_2021.xlsx', sheet_name='
    Statistics_teams_WO_SM')
data_i_will_use = data_all[['TM', 'PO', 'TL']]

# Correlate and create lower triangle mask
corr = data_i_will_use.corr(method='pearson')
matrix = np.triu(corr)

# Plot
seaborn.heatmap(corr, annot=True, mask=matrix, vmin=0.1, vmax=0.9)
plt.show()
```

E MNT Article

MNT konferansen 2021 - UiA

Exploring motivation and teamwork in a large software engineering capstone course during the coronavirus pandemic

Yngve Lindsjørn, Steffen Almås, Viktoria Stray

Department of Informatics, University of Oslo, Norway {ynglin, steffa, stray}@ifi.uio.no

Keywords: Teamwork, Agile practices, Software Engineering, Covid-19, Capstone course

ABSTRACT: In the spring of 2020, the Department of Informatics covered a 20 ECTS capstone course in Software Engineering, mainly focusing on developing a complex application. The course used active learning methods, and 240 students were working in 42 cross-functional, agile teams. The pandemic caused by the coronavirus had a significant impact on the teaching given by the University of Oslo, as all physical education and collaboration among the teams had to be digital from March 12. At the end of the semester, we conducted a survey that focused on 1) aspects of teamwork (e.g., communication and coordination in the teams) and the relation to team performance (e.g., the application product) and 2) the students' motivation and ability to cooperate through digital platforms. A total of 151 respondents in 41 agile student teams answered the survey. This study aimed to investigate how the teamwork and motivation of the students were affected by having to work virtually. The results are compared to results from the same course in 2019 and a similar survey on 71 professional teams published in 2016. Our results show that the teamwork was evaluated similarly to both the evaluation of survey conducted in 2019 and on the professional teams in 2016. The motivation among the students remained high, even though they had to collaborate virtually.

1 INTRODUCTION

Working in teams in software development is essential (Chow & Cao, 2008). Developers work in teams because, "in software development specifically, the speed, frequency, complexity, and diversity of changes needed for modern software-rich systems mean that teams are essential." (Skelton & Pais, 2019). Agile software development is now the common practice and provides values and principles for producing working software rapidly while responding effectively to change.

The Department of Informatics at the University of Oslo offers a 20 ECTS capstone software engineering course where working in teams is a central part. Due to the coronavirus situation, the Campus was closed down on March 12th, 2020. This course's consequence was that the project work (which started ten days before the lockdown) and collaboration between the team members had to be carried out using digital tools.

Other research on the effect of the coronavirus situation in higher education in Norway includes some relevant articles from the 2020 NIKT conference (Norwegian conference for ICT-research and education). One of the findings in (Lorås et al., 2020) is that "informal learning spaces are essential to students yet challenging to transfer effectively to the online environment." Hjelsvold et al., 2020 reported that some educators found certain aspects of online teaching to be better than when teaching physically on Campus. When facing such sudden changes, the educators were good at collaborating and exchanging pedagogical experience.

This article aims to investigate the effect of not being able to physically meet when developing software in student teams. The research question is formulated as follows: *How did the coronavirus situation and the Campus shutdown affect the teamwork and the motivation of the students in a large software engineering capstone course?*

This study uses data from the course in spring 2020 (42 teams) and data from the same course in 2019 (39 teams) for comparison. Central in the study is a survey given to the students at the end of the semester. In 2019 and 2020, there were questions about teamwork using the teamwork quality (TWQ) concept and the relationship to team performance. TWQ was initially developed and used by Hoegl and

Gemuenden (2001) and further applied by Lindsjørn et al. (2016) and Lindsjørn et al. (2018). The TWQ constructs measure the quality of interactions within a team and consist of six variables; *communication*, *coordination*, *balance of member contribution*, *mutual support*, *effort*, and *cohesion*. Team Performance consists of the variables *effectiveness* and *efficiency*, and the Team Members' Success consists of *work satisfaction* and *learning*. Effectiveness refers to the expectation regarding product quality, while efficiency refers to the expectations regarding project quality, such as time and cost (Lindsjørn et al., 2016).

Due to the coronavirus situation in 2020, we decided to include additional questions regarding the digital learning environment using the concept of sociability, which refers to the extent to which digital tools are perceived to help cope with the distributed teamwork (Kreijins et al., 2007). Some key attributes of the concept of sociability are trust within the team, belonging, and relationship.

2 THE COURSE

The 20 ECTS capstone software engineering course is mandatory for most students at the Department of Informatics and has approximately 250 students each semester. The course starts with eight weeks of intensive lectures (see Table 1) and group sessions. During these weeks, the students also prepare for teamwork, and teams are formed with ideally six students in each team (Løvold et al., 2020). The students submit a survey regarding their motivation, background, and up to two fellows they want as team members. Based on the survey response, the students are assigned to a team by the course lead.

		Lecture
Week 1	1	Introduction to the course
	2	The Basics of Android Studio and Kotlin
Week 2	3	More on Android Studio
	4	More on Kotlin
Week 3	5	API, data formats, HTTP-requests and Proxy-servers
	6	Teamwork, agile methodologies and project work
Week 4	7	Agile practices
	8	Basic Principles of Testing
Week 5	9	Secure System Development
	10	Modelling and object-oriented principles
Week 6	11	Architecture and Technical Debt
	12	From Theory to Practice – the project from A to Z
Week 7	13	Application Programming Interface (API)
	14	Development of Android apps and use of patterns
Week 8	15	Universal Design
	16	Evaluation Method / Research Methods

Table 1 – Lectures in the Software Engineering capstone course

During the project period of 12 weeks, the student teams were assigned to make a mobile weather app on the Android platform using data from The Norwegian Meteorological Institute´s API (MET, 2020). There were six weather cases made in collaboration with the Norwegian Institute of Meteorology. The cases were titled as follows:

- Water movements in the oceans
- Forecasts of landslides and avalanches
- Air quality in municipalities
- Predictions of climate and climate change
- Drones and airspace
- Open case – use weather data and design your own case

The students were given introductions to agile methods and software engineering practices, and the teams were free to select development methodology and practices in their teamwork. Practically all the teams used their own adaptations of the agile process models Scrum and Kanban and applied agile practices such as sprints, daily meetings, sprint planning, and retrospective meetings. The majority of the teams had a designated Scrum master, and some teams rotated the Scrum master role during the period.

The students were free to select collaboration tools to work virtually, except GitHub, which was mandatory for Version Control.

An essential part of this course is the mandatory presentations held at the end of the project. During these presentations, the students share their thoughts and reflect on their experience of working in teams. They also presented their final product by demonstrating their application online, where everyone was invited. Several teaching assistants and lecturers were present during these presentations, making it possible for the students to share their opinions directly with the course lead. As a part of the grading, the students delivered a comprehensive report (together with the app) at the end of the project. In the reports, the team members elaborated on the process leading up to the final product and described both the product (technical and non-technical aspects) and the process. They also reflected on the coronavirus situation and how they were affected.

3 METHOD

In June 2020, a survey was conducted after the student teams had submitted and presented their project. The survey was published on June 1st, 2020, and was open for one month (the last response was received on July 1st). The survey consisted of a total of 91 items. We included 10 questions measuring sociability (digital learning environment) and 61 questions measuring TWQ and team performance. In addition, we included 20 items about the tools and background questions like gender and age, study program, and previous experience in agile development. A total of 151 students out of 240 responded, making the response rate 63%. In May and June 2019, when the same survey was conducted, the response rate was higher as the students answered the forms physically at the university during the mandatory presentations (the response rate in 2019 was 98%). Each item in the TWQ and team performance model, and items in the digital learning (Sociability model) of the questionnaire were statements. The respondents indicated their personal views for each statement on a Likert-scale from 1 (strongly disagree) to 5 (strongly agree).

Cronbach alpha is a statistic for internal-consistency reliability alpha values and should be higher than 0.7 to be satisfactory (Nunnaly and Bernstein (1994). All alpha values in all constructs were satisfactory, except the TWQ construct Balance of member Contribution (3 items). In a similar study conducted on professional teams, the same construct was the only construct with the alpha value below 0.7 as well (Lindsjørn et al. (2016).

4 RESULTS AND DISCUSSION

All teams found appropriate tools useful in teamwork. Zoom was used during the digital lectures and supervision, so naturally, this tool's usage among students was high. The most popular tool used was Slack (used by 81 %). The use of Slack has been found valuable in agile distributed teams because it increases team awareness (Stray, 2020). Next, the survey respondents reported using Zoom (74 %), Facebook Messenger (42 %), and Discord (37 %). Other tools were also used, such as Google Disk, Microsoft Teams, Trello, Monday, Notion, and Skype. Most of the teams used more than one tool. Some teams started chatting on Slack but realized that video meetings were more effective: «... *Since we could not meet physically and discuss at the University, we met digitally instead. Through trial and error, we eventually found a balance that worked for everyone. In the start, we mostly used Slack to discuss, but gradually we discovered meeting over video (Zoom) was more effective*." Some teams reported that they were not affected by the situation as stated in one of the reports: "*Generally, we feel that the project was not very affected by the coronavirus situation, and we managed to work well together with the help of digital tools and good communication*."

Table 2 shows that the results are similar for both 2019 and 2020. We see that the mean values of the TWQ variables are slightly higher in 2019 than in 2020, in particular, *Communication* with a difference of 0,18 in 2020 and Effort with a difference of 0,13. The mean values for the team performance variables, however, are higher in 2020 than in 2019, both Effectiveness (product quality) and Efficiency (project quality). Compared to the study conducted on professional teams (Lindsjørn et al., 2016), the mean values of the student teams were slightly higher (both in 2019 and 2020) than the values of the study of professional teams. The variance among the team members' evaluation of all variables was significantly higher in the students' teams than in the professional teams, with a standard deviation of 0,45 on average in the student teams and 0,30 on average among the professional teams.

All values in Table 2 are calculated on the team level: the aggregated values of all team members' evaluation, while the values in Table 3 are calculated on the individual level. This does not affect the mean values, only the standard deviation values. For the TWQ and team performance model, the values are presented on the variable level (the variable communication has, e.g., 10 items – See Table 2). In contrast, the only construct in the Sociability model (digital learning environment) is presented on the item (question) level (see Table 3).

			2019		2020		Difference	
Construct	**Variable**	**No. Items**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev**	**Mean**	**Std. Dev**
Team Quality (TWQ)	Communication	10	4,17	0,37	3,99	0,51	-0,18	0,14
	Coordination	4	4,05	0,40	3,98	0,45	-0,07	0,05
	Mutual support	7	4,42	0,37	4,33	0,47	-0,09	-0,10
	Cohesion	10	4,26	0,45	4,20	0,45	-0,06	0,00
	Effort	4	3,86	0,65	3,73	0,62	-0,13	-0,03
	Balance of contribution	3	4,25	0,41	4,24	0,50	-0,01	0,09
Team members' success	Work satisfaction	3	4,28	0,44	4,37	0,40	0,09	-0,04
	Learning	5	4,41	0,47	4,42	0,45	0,01	-0,02
Team performance	Effectiveness	10	3,86	0,42	4,03	0,34	0,17	-0,08
	Efficiency	5	3,81	0,60	3,98	0,55	0,17	-0,05

Table 2 – Descriptive statistics of investigated variables

No.	Item	Mean	Standard deviation
1	The virtual learning environment enables me to easily contact my teammates	3,96	1,07
2	I do not feel lonely in the virtual learning environment	3,62	1,18
3	The virtual learning environment enables me to get a good impression of my teammates	3,33	1,15
4	The virtual learning environment allows spontaneous informal conversations	3,57	1,24
5	The virtual learning environment enables us to develop into a well performing team	3,60	1,03
6	The virtual learning environment enables me to develop good work relationships with my teammates	3,44	1,15
7	This virtual learning environment enables me to identify myself with the team	3,51	1,11
8	I feel comfortable in the virtual learning environment	3,87	1,02
9	The virtual learning environment allows for non-task-related conversations	3,44	1,19
10	The virtual learning environment enables me to make close friendships with my teammates	2,78	1,28

Table 3 – Descriptive statistics of virtual learning environment items

Table 3 shows the mean and standard deviation values of the evaluation of the virtual learning environment items. All the 9 first items' mean values are between 3 and 4 (3,6 on average). The evaluation of item 10 (making close friendship) was significantly lower than the other items' evaluation. Similar findings were found in Kreijns et al., 2007.

It is remarkable how the student teams have adapted so quickly to virtual teamwork and that they evaluated both aspects of the teamwork, satisfaction of work, and product similar to the student teams in 2019 were they met physically. The results presented in Table 3 (virtual learning environment) also support the fact that the student teams could collaborate well regarding the project. However, it shows that it was hard to make close friendships with teammates. This indicates that working remotely has a more negative impact on social aspects than the teamwork among students in capstone courses like this.

Most of the teams reported during the presentations that the case they had chosen was engaging, which raised the motivation. All the teams were asked the same question after the presentation of the project: "*How did the coronavirus situation (closing down the Campus) and the fact that the grade was only passed/not passed influence your motivation in the course*"? The most common answer was: "*The motivation became lower at once, but when we really started to work (digitally) together as a team, we just wanted to make a good app, write a good report and learn to use some agile practices during the teamwork.*"

Most of the student reports reflected this fact, and some teams also reported advantages. Here is an example: *"... We discovered some advantages of having digital meetings, i.e., fewer excuses to skip, and using digital entertainment mediums such as "Jackbox" was a nice way to replace physical Game nights."* Another example: "... *We were able to collaborate well despite the social distance. The digital meetings went well and were probably more effective than what psychical meetings would have been. Since we met digitally, we delegated the tasks more clearly than we would if we met physically*".

5 CONCLUSION

This article has addressed the following research question: How did the Corona situation and the Campus shutdown affect the teamwork and the motivation of the students in a large capstone course working in teams? Our findings show that the teamwork worked well despite working remotely. Though many students' motivation dropped just after the lockdown, the motivation increased when they started to work (digitally) together as a team. One of the reasons that the teamwork worked well and the motivation increased is that the students found good digital collaboration tools. Another reason was that they found the cases to be exciting and challenging.

REFERENCES

Bastarrica, M., Perovich, D., & Samary, M. (2017). What can students get from a software engineering capstone course? Proceedings of the 39th International Conference on Software Engineering, 137–145. https://doi.org/10.1109/ICSE-SEET.2017.15

Chow, T. & Cao, D. (2008). A survey study of critical success factors in agile software projects. The Journal of Systems and Software, 81(6), 961–971. https://doi.org/10.1016/j.jss.2007.08.020

Dzvonyar, D., Alperowitz, L., Henze, D. & Bruegge, B. (2018). Team composition in software engineering project courses. Proceedings of the 2nd International Workshop on Software Engineering Education for Millennials, 16–23. https://doi.org/10.1145/3194779.3194782

Hjelsvold, R., Nykvist, S. S., Lorås, M., Bahmani, A & Krokan, A. (2020). Educators' Experiences Online: How COVID-19 Encouraged Pedagogical Change in CS Education. *Norsk IKT-konferanse for forskning og utdanning*, 2020.

Hoegl, M., & Gemuenden, H.G. (2001). Teamwork Quality and the Success of Innovative Projects: A Theoretical Concept and Empirical Evidence. Organization Science (Providence, R.I.), 12(4), 435–449. https://doi.org/10.1287/orsc.12.4.435.10635

Iacob, C., & Faily, S. (2019). Exploring the gap between the student expectations and the reality of teamwork in undergraduate software engineering group projects. The Journal of Systems and Software, 157, 110393. https://doi.org/10.1016/j.jss.2019.110393

Kreijns, K., Kirschner, P. A., Jochems, W. & van Buuren, H. (2007). Measuring perceived sociability of computer-supported collaborative learning environments. Computers and Education, 49(2), 176–192. https://doi.org/10.1016/j.compedu.2005.05.004

Lindsjørn, Y., Sjøberg, Dag I.K, Dingsøyr, T., Bergersen, G.R. & Dybå, Tore. (2016). Teamwork quality and project success in software development: A survey of agile development teams. The Journal of Systems and Software, 122, 274–286. https://doi.org/10.1016/j.jss.2016.09.028

Lindsjørn, Y., Bergersen, G.R., Dingsøyr, T. & Sjøberg, Dag I.K. (2018). Teamwork Quality and Team Performance: Exploring Differences Between Small and Large Agile Projects. Springer. https://doi.org/https://doi.org/10.1007/978-3-319-91602-6_19

Lorås, M., Hjelsvold, R., Nykvist, S. S., Bahmani, A. & Krokan, A. (2020). The Hidden Benefits of the Campus - What the Covid-19 Pandemic Can Teach Us About the Computing Learning Environment. *Norsk IKT-konferanse for forskning og utdanning*, 2020.

Løvold, H., Lindsjørn, Y., Stray, V.: Forming and assessing student teams in software engineering courses. In: International Conference on Agile Software Development, XP2020 Companion. Springer (2020)

Nunnally, J, C., Bernstein, I.H., 1994. Psychometric Theory, third ed. McGraw-Hill, New York.

Skelton, M., & Pais, M. (2019). *Team Topologies: Organizing Business and Technology Teams for Fast Flow*: It Revolution.

Stray, V., & Moe, N. B. (2020). Understanding coordination in global software engineering: A mixed-methods study on the use of meetings and Slack. *Journal of Systems and Software*, *170*, 110717.

The Norwegian Meteorological Institute. MET. *Weather API*. http://api.met.no/.

F NSD Consent Form

Are you interested in taking part in the research project "Teamwork Quality and Project Success in a Software Engineering Capstone Course"?

This is an inquiry for you to participate in a research project where the aim is to investigate teamwork in the project work in the course IN2000 - Software Engineering and project work at the University of Oslo (UiO). This paper contains information about the goals of the project and what participation will mean for you.

Project description
In my master's thesis, I research teamwork in IN2000 at UiO and how COVID-19 affected this in the spring of 2020 (and 2021).
More specifically, I look at quality aspects related to teamwork, such as communication in the teams, coordination, effort, and mutual support. I will also examine how the teams worked distributed and how the teams were affected by COVID-19. The semesters examined in my thesis are the years 2019, 2020, and 2021.

Why are you asked to participate?
The study sample consists of involved actors in IN2000 including students, teaching assistants, and partners.

What does participation involve for you?
If you choose to participate in the project, you either answer a survey or are interviewed. Both activities revolve around your thoughts on the teamwork and project work in IN2000. The survey takes around 15 minutes to answer, and an interview lasts approximately 40 minutes.

Participation is voluntarily
It is voluntary to participate in the project. If you choose to participate, you can withdraw your consent at any time without giving any reason. All your personal information will then be deleted. It will not have any negative consequences for you if you do not want to participate or later choose to withdraw.

Anonymity
We will only use the information about you for the purposes we have described in this paper. We treat the information confidentially and by the privacy regulations.

- Only the student and supervisor at UiO will have access to the data during the project.
- I will replace your name and contact information with a code that is stored in a separate name list separate from other data. The data material is stored encrypted on a research server operated by UiO.

Participants will not be recognized in the publication. If quotes from the interview are used, they will be anonymized.

What happens to your information when we end the research project?
The information and recordings will be deleted when the assignment has been approved, which is according to the plan by June 30th, 2021. The information will be deleted by December 31st, 2021, at the latest.

Your rights
As long as you can be identified in the data material, you have the right to:

- gain access to which personal information is registered about you, and to receive a copy of the information,
- to have personal information about you corrected,
- to have personal information about you deleted, and
- send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

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

- University of Oslo via Yngve Lindsjørn (ynglin@ifi.uio.no)
- Our Data Protection Officer: Roger Markgraf-Bye (personvernombud@uio.no)
- NSD – The Norwegian Center for Research Data AS (personverntjenester@nsd.no , +4755582117)

Yours sincerely,
Steffen Almås (student) – steffa@ifi.uio.no
Yngve Lindsjørn (supervisor) – ynglin@ifi.uio.no

Consent form

I have received and understood information about the project *Teamwork Quality and Project Success in a Software Engineering Capstone Course* and have been given the opportunity to ask questions. I consent:

☐ to participate in interview
☐ to participate in survey

I hereby give consent for my personal data to be processed until the end date of the project.

(Signed by participant, date)