Understanding the Teamwork Challenges of Software Engineering Students

L. Dorić*, N. Luburić*, J.Slivka* and A. Kovačević*
*Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
{luka.doric, nikola.luburic, slivkaje, kocha78}@uns.ac.rs

***Abstract* - Developing collaborative skills in students is nontrivial. The fact that students work in teams does not mean they become skilled in teamwork. Students face varied challenges when working in teams that harm their skill development and attitude towards teamwork. To prepare students for the collaboration-intensive workplace, we researched and designed a catalog of challenges present in the teamwork of undergraduate software engineering students on 3-month projects.**

We created an initial catalog of 10 challenges by examining the literature, surveying 15 teaching assistants, and coding their opinions regarding the problems faced by student teams. Using the catalog, we crafted a survey for students nearing the end of their team project to assess which challenges were present in their teamwork. We surveyed students from multiple contexts, including teams of 3, teams of 4, and teams of 16 students.

We analyzed 155 answers to determine the prevalence and intensity of the 10 challenges in student teams. We discuss our findings and best practices for resolving the most prevalent challenges. The catalog and recommendations are directly valuable for software engineering educators and can inform the broader community of collaborative learning researchers and instructional designers.

***Keywords* - teamwork challenges, collaboration, computing education, software engineering, higher education**

I. INTRODUCTION

The need for collaboration is a natural consequence of the growing complexity of our society, where modern projects surpass the individual's capabilities [1][2]. Collaboration skills, such as communication, team task management, and solution brainstorming, are crucial aspects of most contemporary professions [3]. Software engineering is one such profession [4][5]. Software engineering involves code reviews, software design brainstorming, and meetings with engineers, managers, and clients. Consequently, the *ACM Computing Curricula* [6] and more general learning engineering literature [7-9] emphasize the need to develop collaborative skills in young software engineering professionals.

Many software engineering programs require students to develop software in teams [10] with the goal of developing their collaborative skills. Developing collaborative skills in students can be challenging [11][12], and the fact that students work in teams does not mean that they become skilled in teamwork. These challenges prevent teamwork skill development and degrade students' attitudes towards them [13-16].

The first step in solving this problem is understanding the challenges students face when working in teams. In this paper, we research and define a catalog of challenges present in the teamwork of software engineering students on projects that last 3 months. We frame our methodology using the empirical cycle of design science [17]. First, we surveyed 15 teaching assistants using open-ended questions to collect their opinions on the problems in student teamwork. We coded their responses to create quantitative data and combined the findings with related work on teamwork challenges [13][14] to create an initial challenge catalog. Using the catalog, we created a survey with 25 questions and distributed it to 155 students. The students came from three different contexts, including (1) teams of 3 students, (2) teams of 4 students, and (3) teams of 16 students. Finally, we analyzed the results to determine the least and most prevalent challenges.

Section 2 lists the related work. Section 3 explains our methodology in detail. Section 4 presents the results. Section 5 discusses the findings and limitations of our study. Section 6 summarizes our conclusions.

II. RELATED WORK

Much research has been conducted examining what challenges students encounter in teamwork [13][15][16].

Kazemitabar et al. [13] examined how participants collaborate in teams for two days in the context of a hackathon. The goal of this hackathon was for each team to develop a new computer program to demonstrate a physics phenomenon of their choice artistically. The participants were undergraduate or graduate students with expertise in physics, computer science, software, electrical, mechanical, or civil engineering. Teams consisted of 2 to 4 participants. Their results revealed 16 general challenges that hamper teamwork in a hackathon.

Järvenoja et al. [15] explored what type of challenges trigger team-level emotion regulation. First-year teacher education students attended a six-week course on mathematics and worked on six collaborative mathematical tasks in three-to-four-member teams. The

This research was supported by the Science Fund of the Republic of Serbia, Grant No 6521051, AI-Clean CaDET.

results showed that in collaborative learning situations, a wide range of challenges emerge, including cognitive, motivational, and emotional issues and different socially- and contextually-oriented challenges.

Piia et al. [16] explored cognitive, motivational, and socio-emotional challenges experienced in collaborative learning. The participants were 22 adult students attending an education science program at a Finnish university. The participants worked in teams of four to five students. Groups were required to complete a project by the end of the semester (3 months). Overall, all teams experienced challenges in their collaborative tasks. Socio-emotional challenges were exhibited the most frequently, followed by motivational and cognitive challenges.

The aforementioned research had some limitations. They did not examine teams of more than five members in terms of long-term collaboration spanning over two months. Their research subjects were students from various fields, whereas our focus was on software engineering students.

III. METHODOLOGY

Our knowledge goal is to define a catalog of challenges present in the teamwork of software engineering students. We decompose this goal into several research questions we seek to answer to provide a more comprehensive catalog. The questions include:

1. Which challenges manifest themselves in students' teamwork, and to what degree? Answering this question determines the challenge's likelihood of occurrence, helping educators prioritize their attention.
2. How do students perceive the challenge's impact on teamwork quality? Answering this question determines the challenge's significance when it occurs, helping educators prioritize their attention.
3. How do familiarity and cohesion between students in one team affect these challenges? While indirectly related to the catalog, we explore the correlation between familiarity and challenge occurrence. As some student teams are made from a group of friends and others from strangers, we hypothesize that this difference can have an impact on challenge manifestation.

To find answers to these questions, we follow the empirical cycle of design science [17]. Our methodology entails two phases. The first phase consists of consulting the literature and the opinions of the teaching assistants to compile the initial catalog of challenges. Based on this catalog, we created a survey for the second phase, which involved surveying students and analyzing their responses.

A. Creating the initial challenge catalog

At our university, the teaching assistants working with software engineering undergraduates are in direct contact with the student teams and follow their progress on projects regularly. We created a survey that consisted of a brief introduction and a single open-ended question "What are the problems you observed in student teamwork?" 15 teaching assistants gave comprehensive answers.

We coded [18] their answers based on a schema of teamwork challenge categories defined in Adaptive Instrument for Regulation of Emotions (AIRE) [14]. AIRE is an instrument aimed at accessing students' experiences of individual and socially shared regulation of emotions in a socially challenging learning situation. We explored which categories of challenges were perceived by teaching assistants to determine if the AIRE schema was sufficient or if we needed to remove or add some challenge category. The results of this phase are presented in detail in section 4.D.

The output of this phase was an initial set of challenge categories relevant to our contexts, which include:

a. Team members have different project goals,
b. Communication is inefficient because members prioritize not offending other members,
c. Members have different work habits/styles,
d. Communication is inefficient because a member dominates the conversation with their opinions,
e. Members do not maintain friendly relations,
f. Members do not complete their weekly tasks,
g. Members have different work quality standards,
h. Members do not know how to resolve conflicts due to differences in opinions,
i. Members misunderstand tasks, and
j. Members have trouble organizing due to personal or non-faculty-related commitments.

B. Creating the student survey

We used the initial challenge catalog to create a survey for the second phase. The survey consists of 24 questions divided into three sections, each related to one of our research questions.

The first section contains one Likert-scale question for each challenge (10 questions in total), where students mark if a challenge was: not present in their team (0), if it presented a minor issue (1), if it was a notable issue (2), if it was a significant issue (3), or if it presented an overwhelming issue (4).

The second section has one Likert-scale question for each challenge (10 questions in total), where students denote their opinion on the impact that a challenge has on teamwork results on a scale of 1 (has little impact) to 5 (has a major impact).

The third section has one Likert-scale question that determines the students' overall satisfaction with their team, one Likert-scale question to denote how frequently they had business meetings, and one Likert-scale question to define how frequently they organized teambuilding activities. The final question is an open-ended question where students can highlight any additional challenges in collaboration they may have encountered or observed.

C. Administering the survey

The surveyed student population was composed of software engineering students from two study programs (we denote them as Program A and Program B), working in teams to develop a software solution over three months.

Program A students are third-year undergraduate students attending a study program that focuses on software engineering. Working in teams of three members, Program A students were tasked with developing a large software solution through five courses during a single semester.

Program B students are fourth-year undergraduate students attending a study program that covers computer engineering, control engineering, and software engineering. Working in teams of four, Program B students developed a software component through one course during a single semester. Additionally, every four teams formed one larger team (i.e., a team of sixteen members) that integrated the four components into a large software solution.

Despite their differences, students from both programs possessed similar software engineering knowledge and skills (e.g., programming, software design, and working with databases). Also, students from both study programs have experience in working in teams. The resulting solutions were web applications with roughly the same size and complexity (16 people x 1 course ~= 3 people x 5 course). Students in both programs received roughly the same amount of feedback on their work, which translates to roughly 6 minutes per student per week[1].

We examine the challenges in student collaboration in the three different contexts:

- The experience of Program A students working in teams of three (Context A),
- The experience of Program B students working in teams of four (Context B1),
- The experience of Program B students working in teams of sixteen (Context B2).

We collected responses from 155 students. 66 students from Program A supplied answers relevant to Context A. 89 students from Program B supplied answers relevant to contexts B1 and B2[2].

D. Processing the answers

To answer our first research question (*Challenge occurrence likelihood*), we use the Likert-scale answers to calculate the average value, the median, and the mode [19]. We derive the same values for the second research question (*Challenge impact perception*).

To answer our third research question (*Team cohesion and challenge occurrence correlation*), we calculate Spearman's rho correlation coefficient between each challenge occurrence value and the three values given for the questions in the third section (i.e., team satisfaction, work meeting frequency, teambuilding frequency).

Finally, we encoded [18] the students' answers to the open-ended question to determine if they encountered or observed any additional challenges, apart from the initial 10 categories.

[1] In both programs, 16 students attended sessions dedicated to monitoring their progress once per week. The session lasted for one and a half hour, which roughly translates to 6 minutes per student.

[2] The answers from students to the survey can be found at the following link - https://zenodo.org/record/7602921#.Y90SX3bMLGJ.

IV. Results

Here we present the results of our survey and coding efforts, grouped around our three research questions.

A. To what degree is each challenge present in a team?

Table I presents the mean, median, and mode values of students' responses from the survey based on the initial catalog of challenges. We observe that challenges *c* and *g* are minor issues in context A (median = 1), while the others are not present (median = 0). In context B1, challenges *a, c, g, and j* have a minor presence, while the others are not present. In context B2, challenges *b, e, f,* and *j* are minor issues, challenges *g* and *c* are moderate

TABLE I. Mean, median, mode values of students' responses (the presence of challenges)

Context	*Challenge*	*Mean*	*Median*	*Mode*
A	a. Different goals	0.45	0	0
	b. Avoiding offending	0.5	0	0
	c. Different work styles	0.86	**1**	0
	d. Dominating member	0.45	0	0
	e. No friendship	0.32	0	0
	f. Tasks not completed	0.61	0	0
	g. Different quality standard	0.77	**1**	0
	h. Poor conflict resolution	0.41	0	0
	i. Task misunderstanding	0.35	0	0
	j. Work-life organization	0.7	0	0
B1	a. Different goals	1	**1**	0
	b. Avoiding offending	0.84	0	0
	c. Different work styles	1.52	**1**	1
	d. Dominating member	0.60	0	0
	e. No friendship	0.42	0	0
	f. Tasks not completed	0.75	0	0
	g. Different quality standard	1	**1**	0
	h. Poor conflict resolution	0.51	0	0
	i. Task misunderstanding	0.42	0	0
	j. Work-life organization	1.11	**1**	0
B2	a. Different goals	1.55	**2**	2
	b. Avoiding offending	1.06	**1**	0
	c. Different work styles	1.93	**2**	2
	d. Dominating member	0.82	0	0
	e. No friendship	0.96	**1**	0
	f. Tasks not completed	1.27	**1**	1
	g. Different quality standard	1.62	**2**	2
	h. Poor conflict resolution	0.9	0	0
	i. Task misunderstanding	0.54	0	0
	j. Work-life organization	1.07	**1**	0

issues (median = 2), and the others are not present.

B. How impactful is each challenge to students?

Table II presents the mean, median, and mode values of students' responses to the survey. Since the students from Program B gave the same answers to this part of the survey for Contexts B1 and B2, we group the data around Program A and B. We observe that students from Program B rated the impact of all challenges on teamwork quality with a median of 4 or 5. The students from Program A gave similar answers, rating challenges *c e* and *j* with a median value of 3.

C. How does team cohesion affect challenge occurrence?

To measure associations among Likert item answers, we used the nonparametric Spearman correlation. According to [19], this is an appropriate way to analyze Likert item questions.

Table III presents Spearman's rank-order correlation matrix. Statistically significant correlations (significance level $\alpha = 0.05$) are bolded. Correlations higher than 0.29 indicate a medium-to-strong relationship between items [20].

The results show a moderate-to-strong negative correlation between team satisfaction and challenges *a, b, c, e, f, g,* and *h* in all three contexts. Additionally, in all three contexts there is a moderate-to-strong negative correlation between work meeting frequency and the occurrence of challenges *a*, *e, f* and *g*. Finally, in all three contexts, there is a moderate-to-strong negative correlation between team building frequency and challenge *e*.

TABLE II. MEAN, MEDIAN, MODE VALUES OF STUDENTS' RESPONSES (IMPACT OF EACH CHALLENGE)

Context	*Challenge*	*Mean*	*Median*	*Mode*
A	a. Different goals	4.3	5	5
	b. Avoiding offending	4.09	4	5
	c. Different work styles	3.55	3	3
	d. Dominating member	3.65	4	4
	e. No friendship	3.27	3	3
	f. Tasks not completed	4.3	5	5
	g. Different quality standard	3.71	4	4
	h. Poor conflict resolution	3.73	4	4
	i. Task misunderstanding	3.73	4	5
	j. Work-life organization	3.38	3	4
B1, B2	a. Different goals	4.35	5	5
	b. Avoiding offending	4.12	4	5
	c. Different work styles	3.93	4	5
	d. Dominating member	3.84	4	5
	e. No friendship	3.63	4	5
	f. Tasks not completed	4.4	5	5
	g. Different quality standard	4.13	4	5
	h. Poor conflict resolution	3.76	4	4
	i. Task misunderstanding	3.83	4	5
	j. Work-life organization	3.79	4	4

D. Coding results

Based on the challenge categories from AIRE, we coded the responses from teaching assistants on a single open-ended question "What are the problems you observed in student teamwork?".

TABLE III. SPEARMAN CORRELATIONS BETWEEN TEAM SATISFACTION, MEETING FREQUENCY, AND CHALLENGE PRESENCE

Context	*Challenge*	*Team satisfac.*	*Work meeting freq.*	*Team building freq.*
A	a. Different goals	**-0.461**	**-0.331**	-0.166
	b. Avoiding offending	**-0.466**	-0.118	**-0.363**
	c. Different work styles	**-0.378**	-0.127	-0.225
	d. Dominating member	**-0.474**	**-0.454**	**-0.267**
	e. No friendship	**-0.402**	**-0.383**	**-0.510**
	f. Tasks not completed	**-0.527**	**-0.371**	**-0.393**
	g. Different quality standard	**-0.546**	**-0.313**	**-0.408**
	h. Poor conflict resolution	**-0.323**	**-0.261**	-0.051
	i. Task misunderstanding	-0.165	0.042	-0.059
	j. Work-life organization	**-0.319**	-0.140	-0.181
B1	a. Different goals	**-0.609**	**-0.482**	**-0.225**
	b. Avoiding offending	**-0.578**	**-0.469**	**-0.309**
	c. Different work styles	**-0.572**	**-0.414**	**-0.577**
	d. Dominating member	**-0.379**	**-0.234**	**-0.379**
	e. No friendship	**-0.427**	**-0.401**	**-0.377**
	f. Tasks not completed	**-0.493**	**-0.529**	**-0.227**
	g. Different quality standard	**-0.524**	**-0.367**	**-0.265**
	h. Poor conflict resolution	**-0.416**	-0.184	**-0.214**
	i. Task misunderstanding	-0.147	-0.122	-0.114
	j. Work-life organization	**-0.404**	**-0.495**	**-0.287**
B2	a. Different goals	**-0.502**	**-0.269**	-0.79
	b. Avoiding offending	**-0.262**	**-0.276**	-0.178
	c. Different work styles	**-0.437**	**-0.293**	**-0.211**
	d. Dominating member	-0.110	-0.033	-0.158
	e. No friendship	**-0.504**	**-0.332**	**-0.309**
	f. Tasks not completed	**-0.346**	**-0.235**	-0.183
	g. Different quality standard	**-0.392**	**-0.240**	-0.027
	h. Poor conflict resolution	**-0.402**	-0.190	-0.139
	i. Task misunderstanding	-0.007	-0.005	-0.154
	j. Work-life organization	-0.181	-0.110	0.24

We identified 10 responses that belong to category *a* (Our goals for the project were different), 11 responses that belong to category *c* (We seemed to have incompatible styles of working), 1 response that belongs to category *d* (We seemed to have different styles of interacting), 5 responses that belong to category *e* (People in our team did not connect very well with one another), 2 responses that belong to category *f* (One/some people were not fully committed to the team project) and 5 responses that belong to category *h* (Team members were not equal). Other categories had 0 answers that belonged to them.

Based on the challenge categories from AIRE, we have composed our survey for students. The first step was to eliminate challenge categories from AIRE that are not relevant to our context or to expand existing ones.

1. We removed the category "different priorities" because it relates to a context where students collaborate over a shorter period. An example of such a challenge would be "some people were more interested in socializing than getting on with the task.",
2. We combined the categories "some members were rather shy and others very outspoken" and "some members' opinions were not taken into account" into *b. Communication is inefficient because members prioritize not offending other members* and added another category, *d. Communication is inefficient because a member dominates the conversation with their opinions* because we have noticed this kind of behavior in student collaboration in previous projects,
3. We removed the category "some people were easily distracted." because it relates to a context where students collaborate over a shorter period. An example of such a challenge would be "students made and received phone calls on their mobiles during meetings.".

The final question in the students' survey was an open-ended question where they could highlight any additional challenges in collaboration they may have encountered or observed. After coding the students' responses, we didn't identify any new categories of challenges.

V. Discussion

Here we discuss our findings, present implications for practice, and describe the limitations of our study.

A. Results

Based on the results listed in Table I (Presence of challenges), we observe that, in general, challenges occur more frequently in larger teams. Table I shows that mean, median, and mode values rise with the number of team members (e.g., Students from Context A with teams of 3 members have significantly fewer challenges than students from Context B2 with teams of 16 members).

We observe that challenge *c* (*Different work styles*) changes across all contexts (Context A – Median 1, Mode 0; Context B1 – Median 1, Mode 1; Context B2 – Median 2, Mode 2). We conclude that as the number of team members grows, so does the number of different work styles. For example, in software engineering, different work styles can be related to code size per commit or frequency and schedule for PR reviews. Interestingly, students do not perceive that this challenge significantly impacts team results relative to other challenges (Context A – Median 3, Mode 3; Context B1 and B2 - Median 4, Mode 5). However, Table III shows a moderate-to-strong negative correlation between the level of teamwork satisfaction and different work styles which means that teams with varying work styles had members that were less satisfied with their teamwork. This finding, coupled with the perception that most challenges significantly impact results (Table II), indicate that our students were not great evaluators of team performance.

We also observe that challenges *g* (*Different quality standards*) changes across all contexts (Context A – Median 1, Mode 0; Context B1 – Median 1, Mode 0; Context B2 – Median 2, Mode 2). One explanation for this result is the lack of an accurate metric to determine quality. When paired with different coding styles, work styles, and poor communication, team members might perceive the work of others as of low quality simply because it differs from their way of developing software.

Finally, challenge *a* (*Different goals*) changes across all contexts (Context A - Median 0, Mode 0; Context B1 - Median 1, Mode 0; Context B2 - Median 2, Mode 2). This result can be a consequence of the fact that students from both B1 and B2 contexts have four courses apart from this project, and students from context A have just one additional course. Also, students from contexts B1 and B2 are in the final year of their studies and, in large part, have started internships thus their focus is divided.

According to students in all contexts, the two challenges perceived as most impactful on team results were *a* (*Different goals* - Median 5, Mode 5) and *f* (*Tasks not completed* - Median 5, Mode 5). These results align with the correlation of these challenges with overall team satisfaction (Table III) and the broader literature on shared goals and accountability as a necessary component of strong teams [21].

Challenges *i*, *h*, and *d* have median 0 and mode 0 in all three contexts. Given that students have received clearly written tasks throughout the project, it is not unusual for the challenge *i* to not have manifested. However, the results for challenges *h* and *d* are unusual, considering that the students have not been formally trained in conflict resolution and similar soft skills.

B. Implications for practice

The results of the study we conducted on the students of the third and fourth years can aid higher education teachers who incorporate team projects as part of their courses. Since the study primarily refers to software engineering students that are nearing the end of their studies, it has the potential to serve software vendors that

hire and train junior engineers, as well as to manage their current software engineering teams.

We recommend educators and team leads become aware of the general challenges teams face and share our findings with their teams. An awareness of hurdles goes a long way in resolving them. Furthermore, we recommend educators examine the more prevalent challenges and those that correlate with overall team satisfaction when developing teamwork training materials. Finally, educators can employ our survey or its derivative to examine which challenges their trainees face and make a step towards resolving these issues.

C. Limitations

In terms of construct validity, we do not guarantee that we have covered all the challenges students may encounter. We utilized existing tools (i.e., AIRE) and adapted them to our context. Additionally, we employed open-ended questions on teachers and students to discover any additional categories and mitigate this issue. However, a more systematic approach is needed to create a comprehensive catalog of teamwork challenges.

In terms of internal validity, it is possible that some of the students did not understand all the survey questions. To reduce their chance of misunderstanding, we provided an example with each question. It is also possible that there was an error in the open-ended question coding. To mitigate this threat, two authors examined the answers to verify the coding validity.

In terms of external validity, we believe that the results of this research are applicable to any domain where people work in teams and have prior experience with collaboration.

VI. Conclusion

In this paper, we examined challenges present in the context of third and fourth-year software engineering students working on 3-month projects. We covered three contexts: teams of 3 members, teams of 4 members, and teams of 16. The research results are the quantified presence of challenges in each context (median, mean, frequency), students' opinions on the impact of each challenge (median, mean, frequency), and the correlations between team satisfaction and challenges. This can be significant for teachers and IT companies. In future work, we plan to utilize the enumerated challenges and their presence to help students overcome them using an automated intelligent tutoring system.

Acknowledgment

We are supported by the Science Fund of the Republic of Serbia, Grant No 6521051, AI-Clean CaDET.

References

[1] Griffin, P. and Care, E. eds., 2014. Assessment and teaching of 21st century skills: Methods and approach. Springer.

[2] von Davier, A.A. and Halpin, P.F., 2013. Collaborative problem solving and the assessment of cognitive skills: Psychometric considerations. *ETS Research Report Series*, *2013*(2), pp.i-36.

[3] Graesser, A.C., Fiore, S.M., Greiff, S., Andrews-Todd, J., Foltz, P.W. and Hesse, F.W., 2018. Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, *19*(2), pp.59-92.

[4] Lingard, R. and Barkataki, S., 2011, October. Teaching teamwork in engineering and computer science. In *2011 Frontiers in education conference (FIE)* (pp. F1C-1). IEEE.

[5] Mason, S., 2020, August. Collaborative learning in computing education: Faculty perspectives and practices. In *Proceedings of the 2020 ACM Conference on International Computing Education Research* (pp. 136-146).

[6] Lunt, B.M., 2021, October. CC2020: An IT-Centric Overview. In Proceedings of the 22st Annual Conference on Information Technology Education (pp. 5-8).

[7] Iniesto, F., Sargent, J., Rienties, B., Llorens, A., Adam, A., Herodotou, C., Ferguson, R. and Muccini, H., 2021, October. When industry meets Education 4.0: What do Computer Science companies need from Higher Education?. In *Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21)* (pp. 367-372).

[8] Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M. and Rumble, M., 2010. Draft white paper 1: Defining 21st century skills. *Available online also at: http://cms. education. gov. il/NR/rdonlyres/19B97225-84B1-4259-B423-4698E1E8171A/115804/defining21stcenturyskills. pdf [accessed in Kota Kinabalu, Malaysia: April 15, 2015]*.

[9] Häkkinen, P., Järvelä, S., Mäkitalo-Siegl, K., Ahonen, A., Näykki, P. and Valtonen, T., 2017. Preparing teacher-students for twenty-first-century learning practices (PREP 21): a framework for enhancing collaborative problem-solving and strategic learning skills. *Teachers and Teaching*, *23*(1), pp.25-41.

[10] Raibulet, C. and Fontana, F.A., 2018. Collaborative and teamwork software development in an undergraduate software engineering course. *Journal of Systems and Software*, *144*, pp.409-422.

[11] Le, H., Janssen, J. and Wubbels, T., 2018. Collaborative learning practices: teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education*, *48*(1).

[12] Stenberg, M., Mangrio, E., Bengtsson, M. and Carlson, E., 2021. Formative peer assessment in higher healthcare education programmes: a scoping review. *BMJ open*, *11*(2), p.e045345.

[13] Kazemitabar, M.A., Lajoie, S.P. and Li, T., 2022. A classification of Challenges encountered in Complex Teamwork Settings. *International Journal of Computer-Supported Collaborative Learning*, pp.1-23.

[14] Järvenoja, H., Volet, S. and Järvelä, S., 2013. Regulation of emotions in socially challenging learning situations: An instrument to measure the adaptive and social nature of the regulation process. *Educational Psychology*, *33*(1), pp.31-58.

[15] Järvenoja, H., Näykki, P. and Törmänen, T., 2019. Emotional regulation in collaborative learning: When do higher education students activate group level regulation in the face of challenges?. *Studies in Higher Education*, *44*(10), pp.1747-1757.

[16] Näykki, P., Järvelä, S., Kirschner, P.A. and Järvenoja, H., 2014. Socio-emotional conflict in collaborative learning—A process-oriented case study in a higher education context. *International Journal of Educational Research*, *68*, pp.1-14.

[17] Wieringa, R.J., 2014. Design science methodology for information systems and software engineering. Springer.

[18] Seaman, C. B., 1999. Qualitative methods in empirical studies of software engineering. IEEE Transactions on software engineering, 25(4), 557-572.

[19] Subedi, B.P., 2016. Using Likert type data in social science research: Confusion, issues and challenges. *International journal of contemporary applied sciences*, *3*(2), pp.36-49.AA

[20] Cohen, J., 2013. *Statistical power analysis for the behavioral sciences*. Routledge.

[21] Coyle, D., 2018. The culture code: The secrets of highly successful groups. Bantam.