Addressing the Real Skill Issue: Exploring the Impact of Soft Skills on Software Engineering Team Performance.

Research internship report

by

Kelian Schekkerman
S3920852
k.schekkerman@student.rug.nl

Under the supervision of

Andrea Capiluppi, Prof Dr

and

Dimka Karastoyanova, Prof Dr

Faculty of Science and Engineering
University of Groningen

Abstract

Performance prediction of student groups is a thoroughly researched topic, often focusing on demographic data or historical traits. Soft skills such as communication, teamwork, leadership, and adaptability are considered essential by employers [17, 13], but have not yet been included in prediction studies of student group performance.

This study aims to address this gap by investigating the impact of soft skills on team performance, alongside technical skills, weekly statuses, and collaboration preferences. We used data from a real-life Software Engineering course at Rijksuniversiteit Groningen during the academic year 2023-2024.

Our methodology involved finding and analysing the correlations between variables and performing statistical tests to determine significance. We formed quadrant-based clusters based on group soft skills and grades. We then identified and investigated the movements of groups among these clusters over time.

We found that prior programming experience and the ability to extract the core problem are crucial for performance in block A. In block B, on the other hand, teamwork and communication skills are essential for success. Interestingly, we found that negative grade modifiers in block B, as a result of uneven workload, correlated negatively with many block B soft skills and the indicated collaboration preference.

These findings highlight the relevance of soft skills in determining group success. This study provides a foundation for future research, including supporting struggling groups, developing prediction models, or further exploring soft skills.

Contents

1 Introduction

Soft skills are the skills that represent the personal qualities and interpersonal skills of an individual. These skills are essential in today's job market due to increasing competitiveness. Employers will naturally be more likely to hire candidates with a comprehensive skill package, including soft skills [25]. This means that soft skills can positively influence career advancements [16, 25]. Since these skills are crucial, it is important that students develop a well-rounded skill set before they enter the job market. Educators can greatly influence the development of their students' soft skills [25], and soft skills are especially learned effectively in an educational setting through project-based learning [22]. Despite research highlighting the importance and benefits of soft skills, soft skills are not yet commonly considered in predicting the performance of (student) groups.

This paper explores the impact of soft skills, alongside soft skills, we also evaluate the impact of technical skills, weekly statuses, and collaboration preferences, on student group performance. This work contributes to a more comprehensive understanding of the effects of soft skills in collaborative settings, specifically in education. This research was carried out on real-life data from the Software Engineering course at the University of Groningen in the academic year 2023-2024. Our main goal is to explore the relationships within the data and determine the factors that could predict group success, as measured by academic grades. Our main research question can be phrased as follows:

Can group success be predicted, and which aspects influence this success?

To answer this question systematically, we formulate four subquestions. Each subquestion represents key variables in our dataset and will guide our exploration of crucial relationships in the data.

SQ1: To what extent do soft skills influence group performance?

SQ2: To what extent do technical skills influence group performance?

SQ3: To what extent do weekly statuses influence group performance?

SQ4: To what extent does collaboration preference influence group performance?

We answer these questions in this research paper, which is structured as follows. In section 2, we discuss the current state of the art in predictive models and soft skills. In section 3, we describe our methodology in which we explain our data collection and cleaning process and describe our experiments. Then in section 4, we describe the results of these experiments. After that, we discuss the implications of these results in section 5. Finally, we conclude our research in section 6.

2 Related work

This section discusses the current state of the art in the topics of performance prediction, group formation, and soft skills. Therefore, this section is divided according to these topics.

2.1 Performance prediction

Predicting student or group performance is commonly researched, often using demographic and external factors to predict this performance [5, 11, 12]. Other commonly used data are historical grades from previous courses to predict student grades in the next term [10], or for a final project [4]. We found various motivations for these predictive studies, ranging from the early detection of at-risk students [18], to determining the impact of previous courses to be able to strengthen the curriculum [14], and even to predicting the best learning environment for individual students [27].

A study by Umer et al. [28] examines and discusses popular methods for predictive analysis in higher education. They, too, found that demographic data and data on external pre-academic factors are commonly used in performance prediction. In addition, they found that Virtual Learning Environment data and assessment data are often used. They found that the most commonly used machine learning methods in performance prediction were rule-based classifiers, tree-based classifiers, function-based classifiers, ensemble classifiers and Bayes-based algorithms. The accuracy metric was often found to be too biased when evaluating the results of these prediction models. Instead, metrics such as confusion matrix, precision and recall, false positive rate, R-squared, mean absolute error, and root mean square error were used. The authors also determined a variety of challenges and limitations, including a lack of benchmark data and data-related issues, such as diversity in sample size, data quality, and data imbalance.

2.2 Group formation

Team formation is crucial in the education of software engineering. The approaches range from random assignment to skill-based allocation and student preferences [20]. Belbin's role-based team formation has been shown to improve team performance and social skills [2]. Ruiz and Wever studied the methods of student team formation. They found that observer-based methods were time-consuming, and methods based on self-rating methods are too reliant on preference instead of optimal team formation. They determined that mixing self-assessment, through a questionnaire, and observer-assessment, through exercises and student interaction, was the most preferred means of forming teams [23].

Rusticus and Justus [24] investigated the differences between 5-student groups formed by teachers versus by students themselves. They found that student-formed groups often found more balance in workload. However, the process of forming groups could be stressful for students, especially when students were unable to form a 5-person group themselves and needed to find a few additional students to fill their group. A transparent formation process by the teacher leads to less stress and is recommended by the authors if groups need to be made quickly. Although self-formed groups scored slightly higher, the authors report that these differences are minimal and suggest that teachers decide their group formation methods based on the context of the class [24].

2.3 Soft skills

Soft skills represent the personal qualities and interpersonal skills of an individual [25]. Soft skills are considered essential in software engineering, creating a beneficial balance between technical expertise and collaboration. Employers emphasise soft skills in communication, teamwork, leadership, and adaptability [17, 13]. Sharma & Sharma discovered that resilience within a team is an essential skill, resulting in more productive teams that are more agile and innovative during difficult times. However, this skill is difficult to measure as there is no valid or reliable measure scale for this [26]. This measurement difficulty translates to all soft skills, as their assessment is subjective. To navigate this challenge, various methods have been explored, including self-assessment questionnaires [7, 1, 15], peer assessment questionnaires [8], fuzzy expert systems [19], and AI-driven approaches [9]. Petkovic et al. developed a web-based electronic Teamwork Assessment Tool (e-TAT) to assess soft skills more efficiently and effectively [21]. Unfortunately, no progress updates regarding this tool could be found after the publication of the work in progress back in 2010.

Although many studies highlight the importance of soft skills, the effects of different personality compositions on the performance of groups are understudied, especially in educational settings [15]. Kwok Hung & Qian investigated the effects of personality composition on group work performance in undergraduate students in China. They found that aggregated personality traits do not affect team effectiveness, but that uniform emotional stability among group members positively impacts group performance [15]. Similarly, Beyer et al. [3] explored the influence of social and cognitive competences on team performance. They discovered a positive correlation between both competences, showing that cognitively competent groups are more likely to develop strong social competencies over time as well. In terms of performance, cognitive competences appeared to enhance short-term performance, whereas social competences seemed to enhance long-term performance [3].

Channon et al. set out to determine the characteristics defining a 'good group'. The authors found that task success is associated with group dynamics, demographic composition, the personalities of group members, and the attitudes between group members. They also found that high-performing groups in terms of grades cannot be assumed to have learned any collaboration skills. The authors suggest skills should be assessed directly if this is a desired learning outcome [5]. Dutt et al. researched the opposite, exploring the factors that make groups dysfunctional. An interesting finding that they uncovered was that perceived dysfunction by students can affect their attitudes, which significantly impacts the team's performance. Therefore, the authors highlight the importance of addressing negative attitudes in teamwork, as this could improve team performance [6].

2.4 Search terms and databases

This section was constructed through a literature review between February and June 2025. We consulted the databases IEEE Xplore, Web of Science and Google Scholar. In our search terms, we used combinations of the following keywords: 'Group/team', 'Student', 'Performance', 'Clustering', 'Soft skills', and 'Personality'.

3 Methodology

In this section, we discuss the steps we took to achieve the goal of this study. More specifically, we elaborate on how we collected and preprocessed our datasets and how we used this data. A visualisation of the process described in this section can be seen in Figure 1.

3.1 Data collection during the Software Engineering course

The data used for this study were collected from the Software Engineering course in the academic year 2023-2024, the process for this is shown in Figure 2. We collected data at various points in the course, which we call collection points (CP). The following collection points exist:

- **CP-a**: Initial data collection used for team formation. Data are collected on the preferences for projects and teammates, as well as students' technical skills.
- **CP-b**: Assessment at the halfway point of the course. Groups are assessed on their deliverables, and students are assessed on soft skills.
- **CP-c**: Assessment at the end of the course. Groups are assessed on their deliverables, and students are assessed on soft skills.

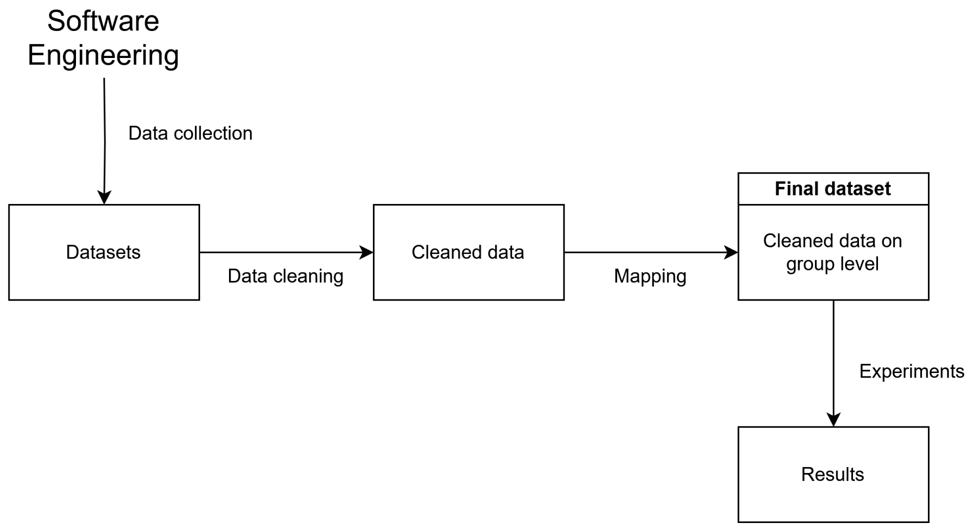

Figure 1: Visualisation of the pipeline from methodology to results, including data collection, data cleaning, and obtaining results.

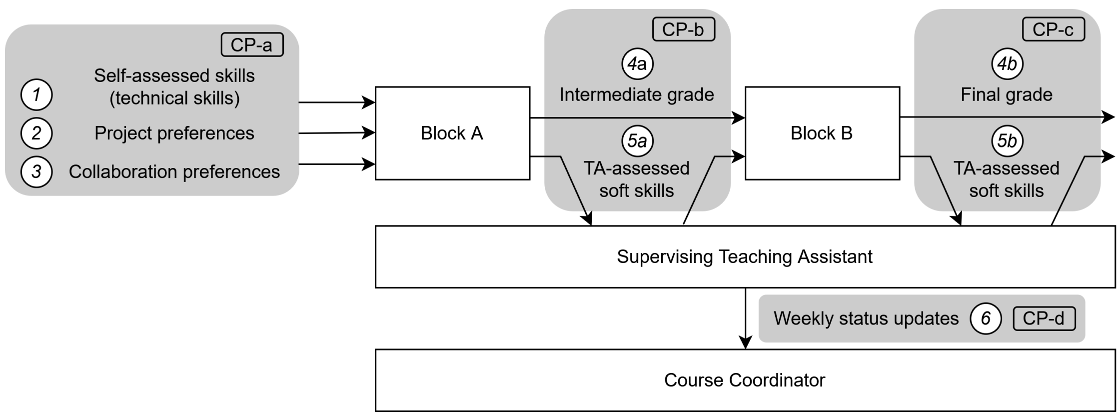

Figure 2: Visualisation of the Collection Points (CP) and the corresponding collected datasets (numbered) during the Software Engineering course.

- **CP-d**: Continuous monitoring. Groups are continuously monitored throughout the course by their Teaching Assistant (TA).

Software Engineering is a course in the second semester of the second year in the Computing Science Bachelor's degree at the University of Groningen. Software Engineering is a mandatory course for students in the Computing Science degree, and often also includes a few students from the Artificial Intelligence Bachelor's degree who have chosen the course as one of their electives. Our students come from various cultures, but demographic data were not available for this study.

Software Engineering students will have taken a variety of courses preceding the Software Engineering

Table 1: Soft skill attributes rated by Teaching Assistants.

Ability to benefit from constructive criticism	Managing Complexity
Ability to extract core problem	Managing Conflict
Ability to take initiative	Managing Relationships
Ability to understand instructions	Managing Stakeholders
Analytical Thinking	Motivating Team Members
Collaborating With Others	Negotiation Skills
Communicating Effectively	Prioritising Work
Continuous Improvement	Problem Solving
Delegating Tasks and Directing Others	Project Management
Driving Vision and Purpose	Presentation and Facilitation Skills
Giving and Receiving Feedback	Time Management
Growth Mindset	

course, such as Advanced Programming and Web Engineering, which serve as assumed prior knowledge.

The projects used in Software Engineering are real-life projects from our industry partners. The project topics depend on the needs of each industry partner and can range widely from creating websites to creating a tool to automate certain tasks, or improving an existing codebase. A short description of the projects in the form of project proposals is given to the students at the start of the course, so that they can get an idea of what each project is about and decide their preferences.

Group formation is based on the top 5 preferences for projects (dataset 2) of each student, as well as their preference to work together with specific students (dataset 3). In the academic year 2023-2024, 118 students took the Software Engineering course. These students were split into groups of 4 to 5 members, creating a total of 28 groups.

During the course, the group works for one of the industry partners (also referred to as the client) and regularly meets with their client (often a product owner or software developer) to discuss the precise product requirements and corresponding priorities. The client is also involved in the testing of the product, ensuring that the student group implemented the required features according to their wishes.

The progress of the student group is observed and guided by an assigned TA, whom they meet on a weekly basis. The TA determines the weekly status of each group they supervise (dataset 6). This status is shared and discussed with the course coordinator to determine if specific groups need extra support or intervention. In 2023-2024, the course counted 9 TAs, each supervising 3 to 4 groups.

Students within a group should work together using Scrum methodology to achieve their goals. Different groups usually do not work together, this only happens in exceptional cases where groups work on different parts of the same project. This was not the case for any of the groups in 2023-2024. The TAs also keep track of every student member's individual contributions and soft skills (datasets 5a and 5b). TAs also assess each student's soft skills. For this, we use a list shared with us by an industry collaborator, who uses this list internally to assess the skills of their employees. This list represents a variety of important skills in group work, such as communication and interpersonal skills, problem-solving and critical thinking skills, and leadership and management skills. An overview of the soft skills in this list can be found in Table 1.

Halfway through the course and at the end of the course, the group is graded on their deliverables by the course coordinator (datasets 4a and 4b). These deliverables are shown in Table 2. The average grade of these deliverables per block is the group grade, and the final total group grade is the average of the grade for each block. Individual students might receive grade modifiers if the teaching assistant observes differences in contributions between student members of a group. In 2023-2024, we gave a maximum of +1 modifier

Table 2: Grade deliverables per block for the Software Engineering course.

ID	Deliverable	Block A	Block B
D1	Requirements document	✓	✓
D2	Design document	✓	✓
D3	Presentation	✓	✓
D4	Source code	✗	✓
D5	Testing document	✗	✓

to students who contributed significantly more to the project than their group members. Students who contributed significantly less than can be expected from a student in the Software Engineering course could get a negative modifier, reducing their grade to possibly even a failing grade. The lowest modifier given in 2023-2024 was -3.

3.2 Datasets

As mentioned in subsection 3.1, the data for this study was gradually collected throughout the Software Engineering course in 2023-2024. The collected data includes information about grades, weekly statuses, soft skills, and technical skills. Some of this data is on group level, such as the grades on deliverables and weekly statuses. Other data is on student level, such as the grade modifiers, soft skills assessment, and technical skills evaluation. A full overview of all variables in the data as collected in the Software Engineering course in 2023-2024 can be found in Table 8, which has been placed in Appendix A.

3.2.1 Data preprocessing

In this section, we discuss how we cleaned the datasets described above and transformed them into an appropriate format for this study.

Missing and inconsistent values Our datasets contain missing or inconsistent values due to a range of reasons. Some of the missing values could be safely set to a default value, as we could retrieve the reasoning for the absence of a value. Other missing values needed to be set to `NaN`. By setting values to `NaN`, we are still missing the datapoints, but we ensure that they are ignored by mathematical operations, and will therefore not change the results as they would if we set them to 0 or another arbitrary value.

We inspected the data for inconsistencies and missing values manually, as well as through tests such as `isna()` from the Pandas library. Below, we list the specific instances of missing or inconsistent data that we found through inspection and our solutions to resolve these issues.

- **Inconsistent answers to Yes/No questions in questionnaires**: For some Yes/No questions, such as the collaboration preference, the answer needed to be entered in a text field instead of a multiple choice since the students needed to specify who they wanted to collaborate with. This resulted in a few students entering inconsistent answers such as '-' or longer sentences as an answer instead of 'No'. To resolve this, we mapped all answers containing 'Yes' to `Yes`, and any other answers to `No`.
- **Missing values for Yes/No questions in questionnaires**: For the Yes/No questions regarding students' experience with certain technologies or programming languages, some of the questions were accidentally not set to mandatory to fill out, resulting in some students skipping certain questions. For

these cases, we assume that students who skipped specific questions do not have experience with the technology in question. Therefore, all missing values were set to 'No'.

- **Inconsistencies in grades**: The grades were entered manually in a spreadsheet, and some grades of students who dropped the course or failed for other reasons were entered as 'F'. To resolve this, all F's were mapped to 1's to keep the dataset consistent and numerical.
- **Individual modifiers:** Individual modifiers were also manually entered into a spreadsheet. Since 0 is considered the default, often only non-zero modifiers were entered. This means that for any empty cells, we can safely assume no modifier was applied. Therefore, we replaced all empty cells and `NaN` values with 0.
- **Missing values in grades**: Additionally, students who dropped the course in block A were not considered for individual modifiers in block B since they would already receive a failing grade. For the dataset, we ensured that this was still counted as an additional negative modifier, as this would be the most likely outcome had they not dropped the course. If this were not done, the results could be skewed, as it would look like improvement from block A to block B concerning the individual modifiers.
- **Missing values in soft skills**: Some soft skills could not be assessed by the Teaching Assistant and were therefore left empty. A common example of one of these unassessed skills was presentation skills, as some TAs were unable to attend the presentation of the group in question. These values were set to `NaN` to ensure they are ignored in mathematical operations.

Mapping from student level to group level To achieve our goal of determining and predicting the success of a group, we need to have all our data at group level. We followed the following steps to map variables from student level to group level:

- **Student IDs**: This column was removed from the datasets.
- **Final grade & Rounded final grade**: Instead of these final grades on an individual level (including the grade modifier), we considered the group grades for each block and final group grade.
- **Grade modifiers**: A fraction of modifiers in the group was calculated. For example, two modifiers in a 5-person group would yield the value 0.4.
- **Soft skills:** For each soft skill attribute, the average of all group members was calculated and used as the group skill value.
- **Years of industry or programming experience**: Calculated by the average value of group members.
- **Technology experiences**: These columns were removed, given that all students have approximately the same experiences since they followed the same courses.
- **Collaboration preference**: This value was represented by a fraction of the team members who answered 'yes' to this question. For example, if two out of five students had a collaboration preference, the entered value would be 0.4.

For the soft skills and years of experience, we calculated the average/mean value of all group members. While the average is relatively sensitive to outliers, we found this to be the most suitable operation. Other operations were considered, but were found to be less intuitive and convenient. The reasoning for each considered operation is as follows:

- Median (middle number): Would not capture the full range of skills within the group
- Mode (most frequent number): Would be problematic if multiple repeated scores exist, or if there are no repeated scores at all.
- Summing: Would be dependent on group size, which makes it incomparable between groups of different sizes.

Final cleaned dataset The variables in the final resulting dataset can be found in Table 3. This data is ready for exploration and visualisation. It is worth noting that 'Communication skills' is a variable in the tech skills category. This question has been part of the technology questionnaire for several years before the soft skills assessment was introduced. Additionally, communication skills are a self-assessed skill in the technical questionnaire, as this questionnaire was filled out by the students.

3.3 Experiments

Given the large number of variables in our dataset, we want to focus on the most important ones. To determine these most important variables and the investigations we want to conduct on them, we phrased the following experiments.

1 - Correlation analysis This experiment is the first exploration of the data and the relationships between the variables. We calculated the correlation scores between all variables to determine the most important relationships in the data. Most of our variables in the final cleaned dataset are continuous variables. Therefore, we use Pearson correlation to determine the correlation between these variables. The weekly status variables are discrete and ordinal, so we used Spearman correlation instead, which is suitable for ordinal data.

We display the correlation scores in heatmaps so that they are easily interpretable. For most heatmaps, we excluded any weak correlations (any correlation score under 0.4) to prevent irrelevant results from cluttering the figures.

We decided whether a correlation was worth investigating further based on the correlation scores. The exact correlation values, along with the associated correlation strength and whether we decided to investigate further, are shown in Table 4.

2 - Statistical significance We performed statistical tests to determine if the discovered correlations were statistically significant. We split the student groups into two categories: below-median and above-median grades. We then determined if the differences in soft skills, technical skills and weekly status between the two categories are statistically significant. For the soft skill attributes, we performed the test on the individual soft skills, soft skills averaged by block, and the total average score of soft skills.

To determine whether the differences are statistically significant, we used the Wilcoxon rank-sum test (also called the Mann-Whitney U test). This is a non-parametric test that compares two independent samples. An alternative test we considered was an independent samples t-test. However, t-tests assume normally distributed data, which we cannot guarantee. Therefore, we decided that the Wilcoxon rank-sum test was more suitable, as these tests do not require normally distributed data. The statistical significance of our results depends on the p-value. The significance per range of p-values is shown in Table 5.

Table 3: Overview of the cleaned dataset. For each variable, we list a short description, variable type, and the possible values. We consider the grades the dependent variables, these rows are shown in teal. We consider the weekly statuses, soft skills and technical skills as the independent variables, these rows are shown in light grey.

Variable	Description	Variable type	Possible values
Group	Unique ID for each group	Nominal	G[01, 28]
Final grade	Final group grade at the end of the course	Continuous	[0, 10]
Rounded final grade	Rounded group grade at the end of the course	Continuous	[0, 10]
D1 grade in block A	Requirements grade	Continuous	[0, 10]
D2 grade in block A	Architecture grade	Continuous	[0, 10]
D3 grade in block A	Presentation grade	Continuous	[0, 10]
Modifier block A	Fraction of grade modifiers in the group in block A	Continuous	[0, 1]
Total grade block A	Total group grade of block A	Continuous	[0, 10]
D1 grade in block B	Requirements grade	Continuous	[0, 10]
D2 grade in block B	Architecture grade	Continuous	[0, 10]
D3 grade in block B	Presentation grade	Continuous	[0, 10]
D4 grade in block B	Source code grade	Continuous	[0, 10]
D5 grade in block B	Testing grade	Continuous	[0, 10]
Poster bonus block B	Bonuspoint on the presentation grade	Continuous	[0, 1]
Presentation grade in block B	Presentation grade + bonus point	Continuous	[0, 10]
Modifier block B	Fraction of grade modifiers in the group in block B	Continuous	[-10, 1]
Total grade in block B	Total group grade of block B	Continuous	[0, 10]
Weekly statuses block A	Weekly status assessed by TA	Ordinal	[1,4]
Weekly statuses block B	Weekly status assessed by TA	Ordinal	[1,4]
Soft skills assessment block A	Average rating of soft skills from the group members assessed by TA	Ordinal	[1,5]
Soft skills assessment block B	Average rating of Soft skills from the group members assessed by TA	Ordinal	[1,5]
Years of programming experience	Average experience over all group members	Continuous	[0, ...>
Years of industry experience	Average experience over all group members	Continuous	[0, ...>
Communication skills	Average self-assessed communication skills of the group members	Continuous	[1,5]
Collaboration preference	Fraction of collaboration preferences in the group	Continuous	[0, 1]

3 - Visualisation of very significant relationships After determining the statistically significant relationships in the data in experiment 2, we visualised the relationships with p-values of 0.01 (very significant relationships) and lower to investigate them further. The correlations were calculated using Pearson correla-

Table 4: Strength of a correlation based on the correlation value

Correlation value	Associated correlation strength	Further investigation
0-0.19	Very weak	No
0.2-0.39	Weak	No
0.4-0.59	Moderate	Yes
0.6-0.79	Strong	Yes
0.8-1.0	Very strong	Yes

Table 5: P-values and associated significance levels.

Value of p	Associated significance
$p > 0.05$	Not significant
$p \leq 0.05$	Significant
$p \leq 0.01$	Very significant
$p \leq 0.001$	Highly significant

tion, which indicates that a linear relationship is present. Therefore, we visualised these linear relationships using a regression plot.

4 - Cluster analysis Finally, we manually formed clusters by dividing the data into quadrants. These quadrants were formed by the median of the total grade per block, creating the horizontal axis, and the median of the average soft skill assessment per group, creating the vertical axis. This creates four quadrants, each representing a unique category of groups through the two different assessments. An overview of these quadrants and their characteristics is shown in Figure 3.

We created two plots, one for each block, showing the positioning of each group spread over the quadrants. We then compared the differences between the positioning of the groups and tracked the movement between the clusters. We visualised this movement in a Sankey plot to get an intuitive overview that is easy to analyse. We evaluated each possible movement to determine whether the movement is positive or negative, whether the movement occurred in the Software Engineering course in 2023-2024, and attempted to describe the movement by a key characteristic.

4 Results

In this section, we lay out the results of our experiments as proposed in the Methodology section.

4.1 Correlation analysis

In Figure 20, we can see a heatmap of the correlation matrix of the entire dataset. This figure was placed in the Appendix under subsection B.1 as the figure is enormous due to the large number of variables. From this figure, we can see a few very brightly coloured areas, indicating a high positive correlation. The brightest areas seem to be mostly between variables in the same category, specifically the grade variables and soft skill variables. Some weaker but still positive correlations seem to exist between the grade variables and soft skill variables. In addition, we notice some dark-colored pixels, indicating a high negative correlation. We can see this mostly for correlations with negative grade modifiers in block B and some technical skills.

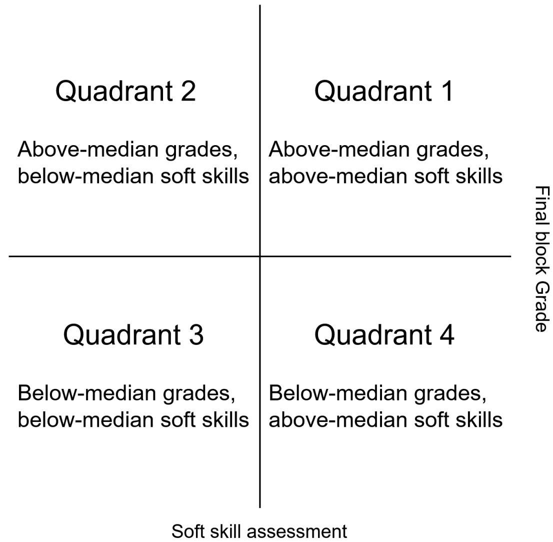

Figure 3: Visual representation of manual clustering of the data into quadrants. The data is split by the median value of the grades and soft skill assessment. This splits the data into below- and above-median values, creating four quadrants, each representing a unique category of group performance in terms of assessed soft skills and grades.

Figure 21 shows the correlations between soft skills and grade variables. This figure too was placed in the Appendix, under subsection B.2 due to the size of the figure. From this figure, we notice that the largest number of correlations is between the deliverables D1 and D2 of both blocks and the soft skills assessment in block B. Additionally, we can see that the block B negative modifier is negatively correlated with a good number of soft skills, most of which are in block B. Interestingly, we also find a negative correlation between the presentation skills in block A and the source code deliverable in block B.

Aside from the personality traits, we also investigated the correlations between the technical skills and grades. The heatmap of this correlation matrix can be found in Figure 4. We can see a moderate positive correlation between these self-assessed communication skills and the variable for bonus points on the posters and the block B presentation (D3 deliverable). Additionally, we can see that the number of years of programming experience is positively correlated with the grade for the requirements document (D1 deliverable) in block A. We also observe a positive correlation between collaboration preference and the testing document (D5 deliverable) in block B, and a negative correlation between the collaboration preference and negative grade modifiers in block B.

Finally, we visualised the correlations between the weekly statuses, assessed by the TAs, and the grade variables, as can be seen in Figure 5. We can see a positive correlation in Week 6 of block A with deliverable D5 in block B. But even more prominently, we can see that Week 3 in block B seems to have the highest number of correlations, both with variables from block A and with variables from block B. It is positively correlated with the architecture document (D2 deliverable) in both block A and B, along with the total grade of both block A and B, and the final total group grade. Additionally, the status of this week is negatively correlated with the negative grade modifiers in block B.

To summarise, we observed positive correlations between soft skills and the D1 and D2 grades, as well as the total and final grades. The years of programming experience, self-assessed communication skills and collaboration preferences also yield some positive correlations with various grade variables. The status of week 3 in block B seems to be positively correlated with both blocks' D2 grades and total grades, as well as the final grade. A negative correlation with the negative grade modifiers is observed in some weekly statuses, the collaboration preference and most of the personality traits.

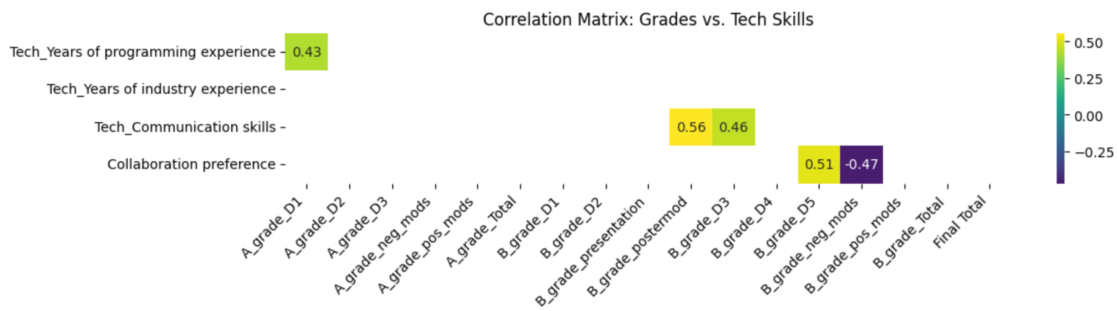

Figure 4: Heatmap of a correlation matrix showing the correlations between the average scores for the technical skills versus the grades. All values between -0.4 and 0.4 have been excluded, so as to only view the correlations of at least moderate strength.

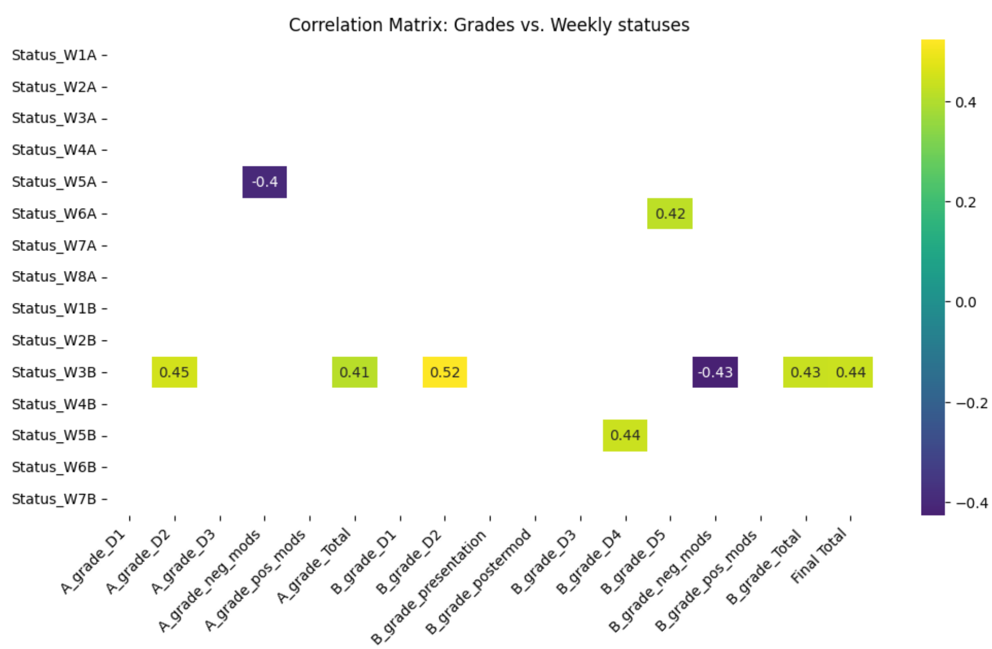

Figure 5: Heatmap of a correlation matrix showing the correlations between the weekly statuses versus the grade variables. All values between -0.4 and 0.4 have been excluded, so as to only view the correlations of at least moderate strength.

4.2 Statistical significance

The results of the Wilcoxon rank-sum test between the below-median and above-median grade categories for the soft skill attributes can be found in Figure 22, which has been placed in Appendix C due to the size of the figure. From this figure, we observe that in block A, the most significant difference between the two grade categories is the soft skill 'Ability to extract core problem', impacting all three deliverables and the total block A grade. The skill 'Managing complexity' seems to be important for the grade on the requirements document (D1). 'Communicating effectively' appears to be most significant in determining the negative modifiers in block A. In block B, the most important skills seem to be 'Collaborating with others', 'Communicating effectively', 'Managing complexity' and 'Managing conflict'. Additionally, more than half of all block B skills are also relevant towards the negative grade modifier in block B.

When considering only the averages of the soft skills assessment, as seen in Figure 6, we can see that no significant differences exist in the block A average soft skills for any of the grade variables. There are, however, significant differences in block B average soft skills for the block A total grade, block B negative modifiers, and the final total grade. For the average total soft skills, we find significant differences in the block A total grade, block B requirements document (D1) and block B negative grade modifiers.

Looking at the differences in technical skills in Figure 7, we observe the following things. There is a statistically significant difference in years of programming between the two categories when looking at the block A presentation grade and the block A total grade. Additionally, a significant difference is found in communication skills for the block A presentation grade (D3) and the poster bonus points in block B. Finally, we find a significant difference in collaboration preference for the negative grade modifiers in block B.

The significant differences in weekly statuses between the two grade categories are shown in Figure 8. From this, we observe statistically significant differences in status for week 3 in block B for two deliverables per block, making 4 deliverables in total. There also seems to be a significant difference in the status of week 5 in block A regarding the negative modifiers. Finally, we find a significant difference in the status of week 5 in block B for the grade of the testing document (D5).

To summarise, we found a significant difference in various variables with respect to grade modifiers. Another important difference seems to exist in the status of week 3 in block B, affecting 4 different grade variables.

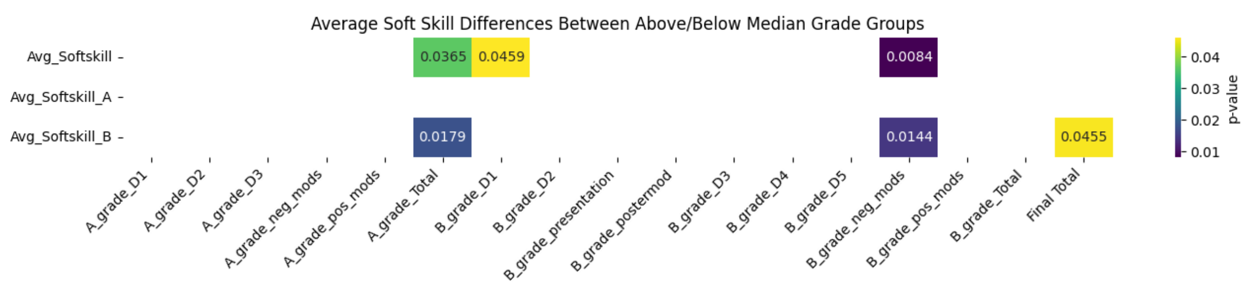

Figure 6: Heatmap of the p-values indicating the statistical significance between the below-median and above-median grade categories regarding the correlation between average soft skills and grades.

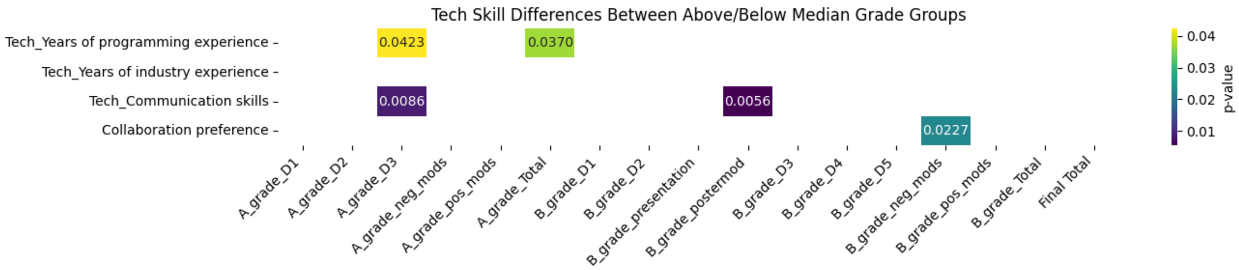

Figure 7: Heatmap of the p-values indicating the statistical significance between the below-median and above-median grade categories regarding the correlation between technical skills and grades.

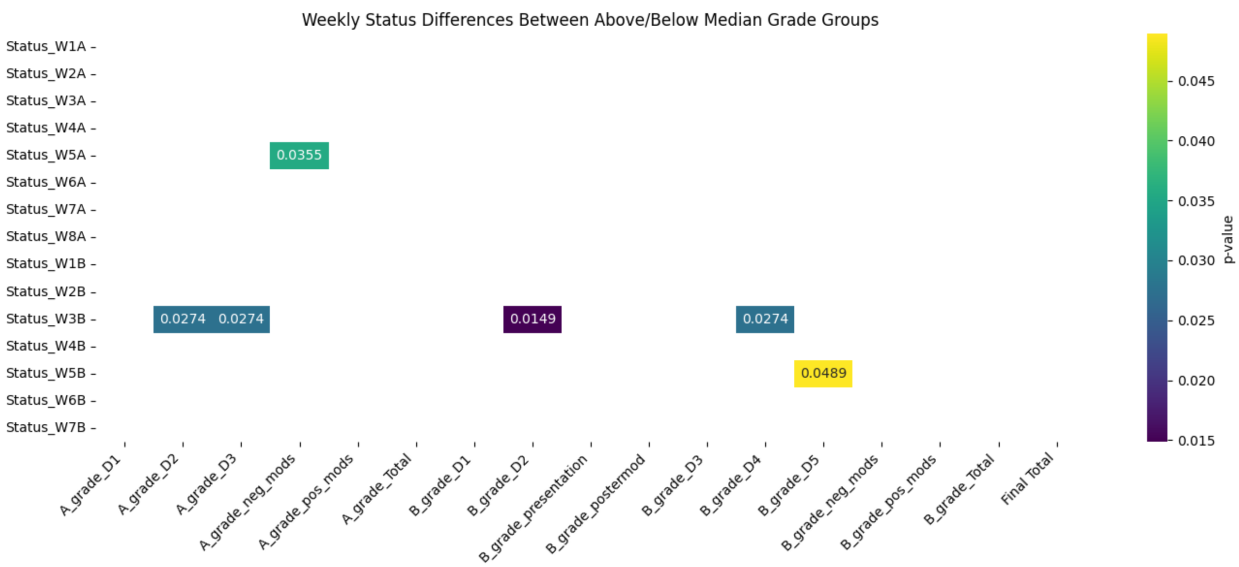

Figure 8: Heatmap of the p-values indicating the statistical significance between the below-median and above-median grade categories regarding the correlation between weekly status and grades.

4.3 Visualisation of very significant relationships

In Figure 9, we can see the relationship between a group's average rating for the soft skill 'Ability to extract core problem' as rated by their TA in block A, and the total grade they received in block A. We can see a positive linear relationship, where the higher the rating of this specific soft skill, the higher the total grade.

The correlation between growth mindset in block A and the presentation grade in block B can be seen in Figure 10. Here, we observe a negative linear relationship, where higher ratings for this soft skill seem to lead to lower presentation grades.

The correlations between the negative grade modifiers in block B and essential soft skills assessment from block A and B can be found in Figure 12 and Figure 14 respectively. We can see that all of these relationships are negative, where collaborating with others shows the steepest line, and managing relationships is the least steep line. From Figure 13, we can see that when considering the averaged soft skills instead of individual soft skills, the negative relationship with the negative grade modifiers in block B still exists.

Finally, in Figure 15, we can see a positive linear relationship between the communication skills as assessed by the students themselves in the technical questionnaire and the presentation grade in block A, as well as the poster bonus point in block B.

To summarise, we found a negative linear relationship between various individual soft skills, as well as the average soft skill rating and the block B negative modifiers. Besides that, we found various positive linear relationships, TA-assessed as well as self-assessed communication skills and various grade variables.

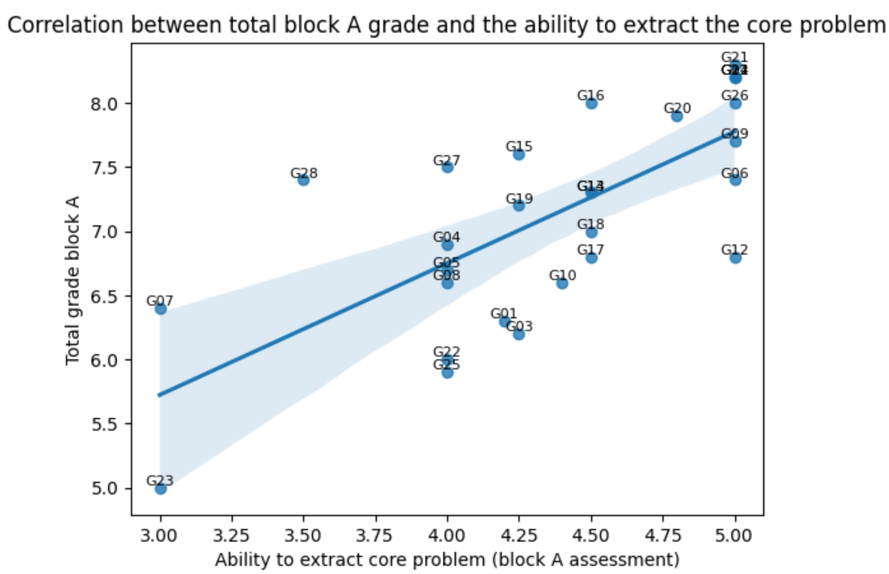

Figure 9: Visualisation of the correlation between the total block A grade and the soft skill 'Ability to extract the core problem' in block A.

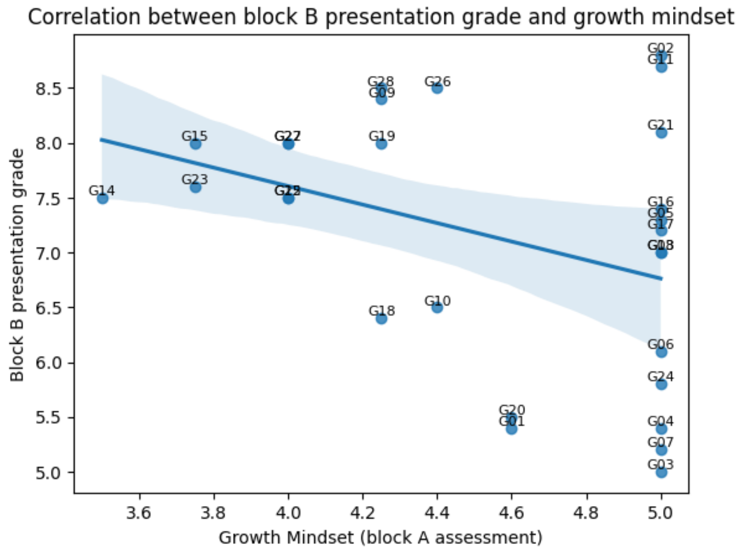

Figure 10: Visualisation of the correlation between the block B presentation grade and the soft skill 'Growth Mindset' in block A.

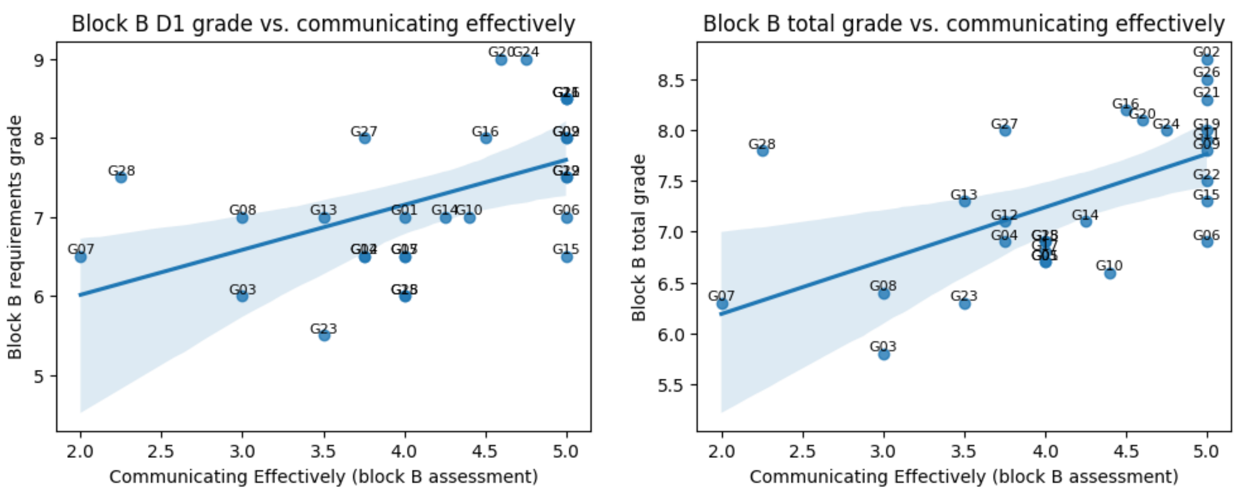

Figure 11: Visualisation of the correlation between the block B assessment of the soft skill 'Communicating effectively' and the requirements and total grades in block B.

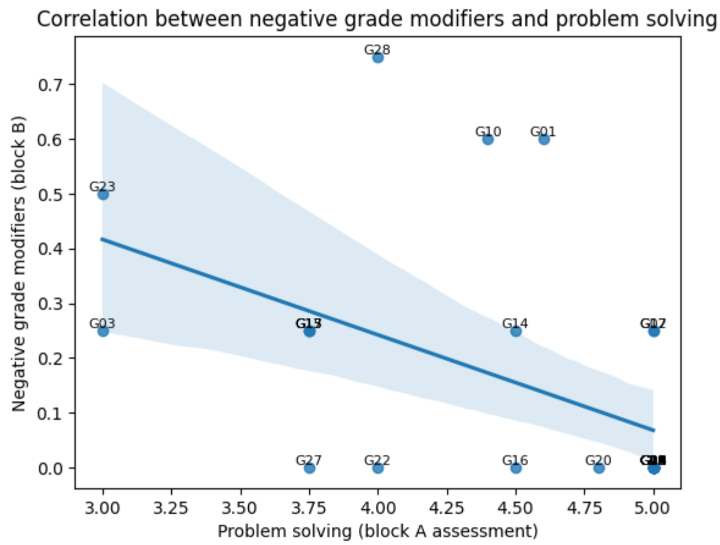

Figure 12: Visualisation of the correlation between the block B negative grade modifiers and the skill 'Problem solving' in block A.

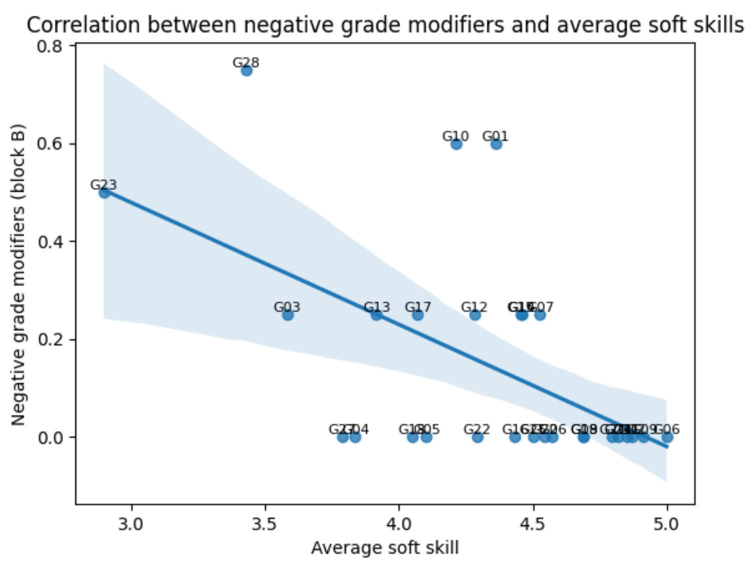

Figure 13: Visualisation of the correlation between the block B negative grade modifiers and the average soft skill rating per group throughout the course.

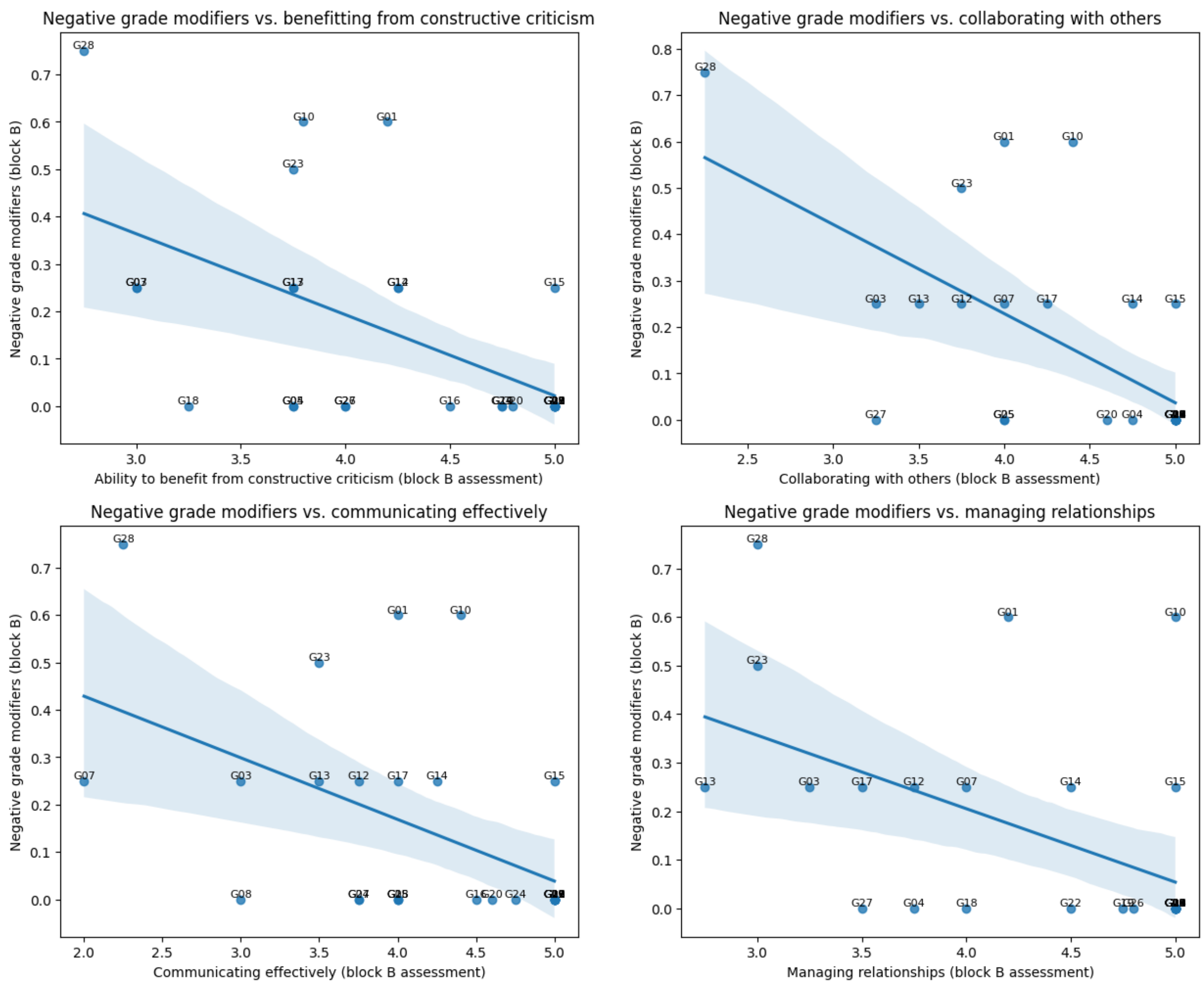

Figure 14: Visualisation of the correlation between the block B negative grade modifiers and the soft skills 'Benefitting from constructive criticism', 'Collaborating with others', 'Communicating effectively', and 'Managing relationships' as assessed in block B.

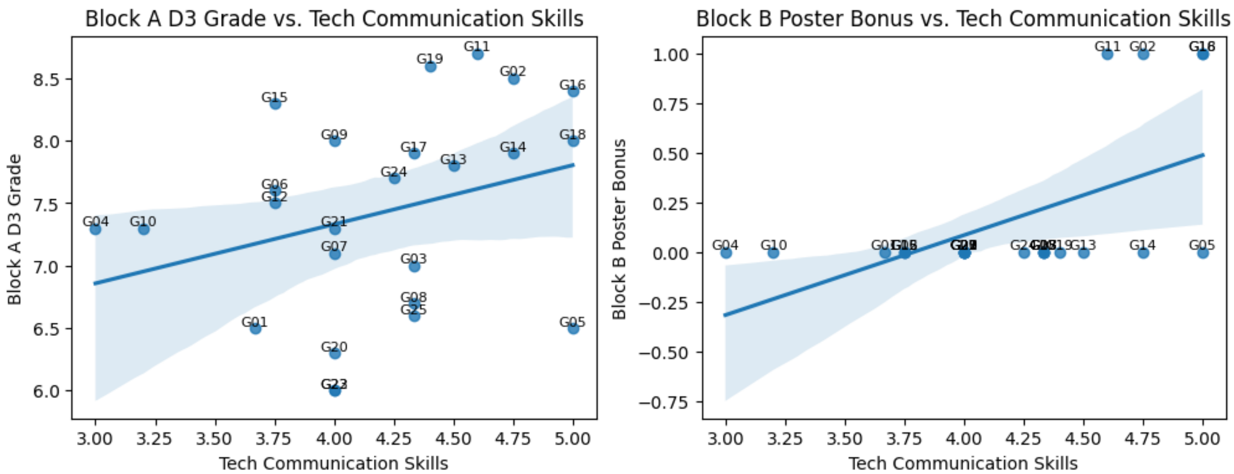

Figure 15: Visualisation of the correlation between the technical communication skills (self-assessed by students) and the presentation (D3) in block A, as well as the poster bonus points in block B.

4.4 Cluster analysis

The positioning of groups per block spread over the four quadrants can be seen in Figure 16. We can see that quadrants 2 and 4 are a bit sparser compared to quadrants 1 and 3.

Looking at the labels of the datapoints, indicating the group numbers, we can see that some groups move between the quadrants when comparing block A to block B. These movements are visualised in Figure 17. We can see that most of the groups positioned in quadrants 1 and 3 for block A remain in the same quadrant in block B, with only a few groups migrating to other quadrants. For quadrant 2, we can see that half of the groups remain in the same quadrant in block B, while the other groups migrate to quadrant 1. For groups in quadrant 4, we can see that most groups moved to quadrant 3 in block B.

To summarise, quadrants 1 and 3 are more densely populated compared to the other quadrants. Most groups within quadrants 1 and 3 for block A also remain within them for block B, whereas groups located in quadrants 2 or 4 for block A tend to migrate more towards other quadrants.

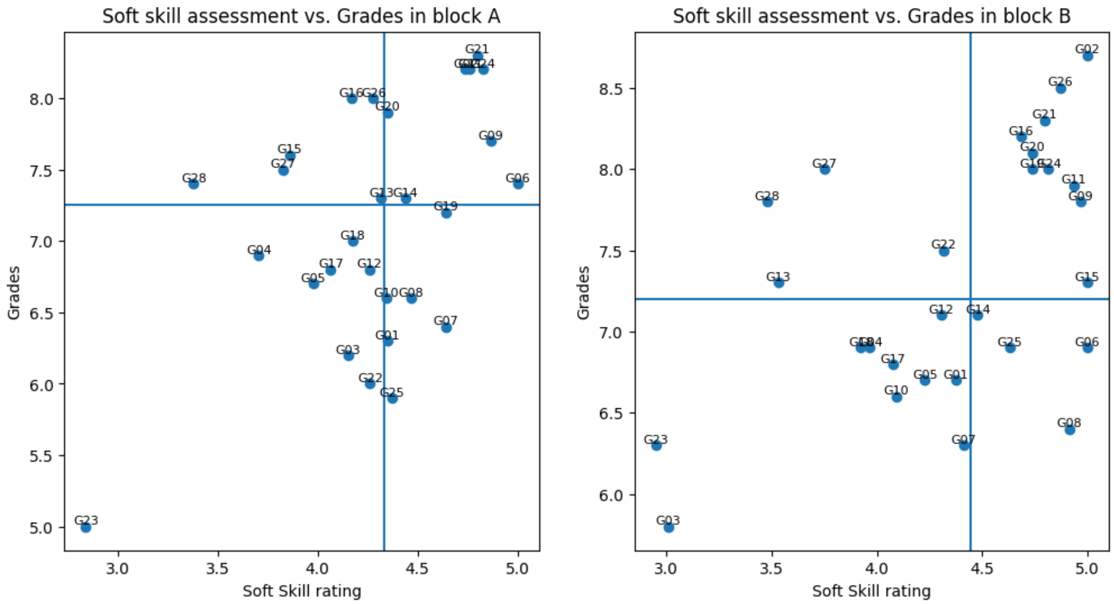

Figure 16: Results of manual clustering into quadrants, split between block A and block B.

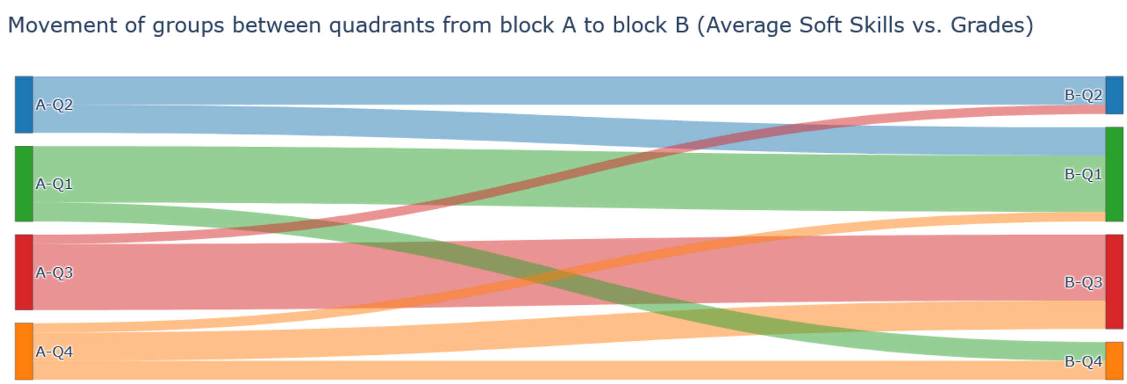

Figure 17: Movement of groups between the quadrants between block A and block B.

5 Discussion

In this section, we discuss the results and their implications. We evaluate what our findings mean concerning our research questions, and consider the future work

5.1 Correlation analysis

The results from the correlation heatmaps give us various insights. First of all, there are consistent negative correlations regarding the negative grade modifier in block B. We could see this with respect to the soft skill assessment, which indicates that groups with a lower average soft skill assessment have a higher chance of receiving one or more negative modifiers. This could be because the individual(s) receiving the negative modifiers receive lower soft skill ratings, bringing down the group average. This also works the other way around, that groups with members who consistently obtain high-rated soft skills have a lower chance of getting negative grade modifiers in the group.

We also noticed this negative correlation between collaboration preference and block B negative modifiers. This means that a group constructed based on collaboration preferences has a lower chance of receiving negative grade modifiers. This could indicate that these groups might be more functional and balanced. Finally, in week 3 of block B, the weekly status is the most correlated with the block B negative modifiers. This indicates that the status of this week might be a predictor for the negative modifiers.

We also saw correlations with various specific deliverables. For example, the soft skills are mostly correlated to the requirements and architecture documents (D1 and D2 deliverables, respectively) in both blocks. These two documents lay the foundation for the final product, so this result suggests that groups with high average soft skills are more likely to set up and document this foundation for the product successfully. From this, the high final grades also follow naturally, which explains the correlation between the soft skills and the total and final grades as well.

The requirements document (D1 deliverable) in block A is positively correlated with years of programming experience. This would indicate that more programming experience leads to better documentation of the requirements of an application. The self-assessed communication skills are positively correlated with the block B presentation grade and bonus points for the poster presentations. This indicates that groups where members find that they communicate well are more likely to obtain a higher presentation grade and a bonus point for the poster presentations. Whether the self-assessed skills are accurate representations of the actual skill level or not, this result makes sense given that students who rate themselves likely feel confident in this aspect, and confidence can be very beneficial in presenting.

Finally, we saw positive correlations between the block B testing document (D5 deliverable) and the week 6 status in block A, as well as the collaboration preference. This indicates that the testing document, which is the last task in the project, is more likely to be performed and graded well if the group is made based on the collaboration preferences of students. Combining this result with the correlation between the D5 grade and status of week 6 in block A, this could indicate that if a group is comfortable with each other at that point during the course (which is more likely to happen in groups based on collaboration preferences), they score better on the testing document.

5.2 Statistical significance

From the results, we saw a significant difference in the 'Ability to extract core problem' soft skill, affecting all grades of the block A deliverables as well as the block A total grade. This result is reasonable, as block A is focused mostly on retrieving the product requirements and laying out a plan for the architecture. These tasks focus on understanding what needs to be done and which components are most important, representing the ability to extract the core problem that needs to be addressed.

To determine the block A negative grade modifiers, the skill 'Communicating effectively' seems most significant. In block B, more than half of the block B assessed skills are relevant towards determining the negative grade modifiers. This would indicate that issues in teamwork and contribution in block A could be improved through targeted support in communication skills, which should reduce the negative grade modifiers. For block B, the issue seems more complex, with 14 significant soft skills impacting the modifiers. This means targeted support would be more difficult to apply here, since the issues are distributed among so many skills. It could be that the issues are somewhat present in block A and that they start to spiral out of control in block B. If that were to be the case, support in block A might also improve teamwork in block B.

In block B, we noticed that the skills 'Collaborating with others', 'Communicating effectively', 'Managing complexity' and 'Managing conflict' are most significant. This, too, is reasonable, as the foundation that was created in block A now needs to be realised. These tasks require consistent teamwork and communication to ensure that everyone understands the goal, the work is distributed equally, and all tasks get done on time.

When averaging the soft skills, we noticed that the block A skills seemingly had no impact on the grades. This can be explained by fewer skills being relevant in block A. Since 23 skills are rated, the few relevant ones get lost in the calculation of the mean. In block B, more than half of the skills are relevant towards the grades, making the mean more relevant as well.

In the technical skills, we found that the previously found positive correlation between self-assessed communication skills and presentation grades is statistically significant. This highlights the relevance of the impact, indicating that this is not likely occurring by chance. As mentioned before, this could mean that groups with more confidence obtain higher presentation grades.

We also found a statistically significant difference in collaboration preferences between groups with above- and below-median negative grade modifiers in block B. Combining this result with the negative correlation that we found before, we can say that groups with fewer indicated collaboration preferences are indeed more likely to obtain negative grade modifiers in block B. This aligns with results by Rusticus and Justus [24], who stated that self-formed groups find a more balanced workload distribution. These results could be explained by groups formed based on preferences being more comfortable with each other and naturally collaborating more easily. This does not automatically mean that students in randomly made groups cannot be comfortable with each other, but simply that this happens more naturally in groups based on preferences, as these students are already familiar with each other. Teambuilding exercises could make the atmosphere in the group more comfortable, bridging the gap between groups where members know and don't know each other.

Using the weekly statuses as assessed by the TAs, we found that week 5 in both blocks and week 3 in block B have the most significant relationships. Week 5 in block A is correlated with the negative modifiers in block A. In the previous experiment, we saw that this correlation is negative. This could be explained by considering that any issues within the team in week 5 can no longer be disregarded as the group simply having a slow start, therefore, issues become more apparent and worrisome at this point. Because of this, individual differences might also become apparent and will be considered for the negative modifiers. Therefore, this correlation seems reasonable.

In week 3 of block B, we see multiple significant differences for various deliverables, even for deliverables in block A. Earlier, we saw that these relationships are positive. This could indicate that groups who performed well in block A have an easier start in block B, whereas groups with relatively lower grades might have a more rocky start to the new block. It also makes sense that performing better at the start of the block results in higher grades for that block, which could explain the significant difference in the positive correlation with the grades for the architecture document (D2) and source code (D4).

Another positive correlation and statistically significant difference was found between the status of week 5 in block B and the grade of the testing document (D5). Around this week, the groups should be freezing their features and focusing their efforts on testing their product. A lower status could indicate delays or other complications with a group's progress. This means they might start their testing process later, therefore, it makes sense that groups with good statuses in this week obtain significantly higher grades on their testing document compared to groups whose status is lower.

Based on these findings, we can answer our research questions about the impact of each variable on the grades. Many soft skills seem to have a positive impact on the grades, especially in block B. We also found a negative correlation between various soft skills and the negative modifiers. Regarding technical skills, we noticed that the amount of programming experience positively correlates with the requirements grade. While a statistically significant difference could not be found here, we found a significant difference in the presentation grade and block A total grade based on the years of programming experience. From this, we can conclude that the technical skills indeed influence the grades, but mostly in block A. From the data on the weekly statuses, we found that week 5 in block A and weeks 3 and 5 in block B show significant correlations to the grades. Finally, we found that the collaboration preference in a group is a significant indicator of the negative grade modifiers in block B.

5.3 Visualisation of very significant relationships

The results from this experiment overlap with the results from the previous experiment, but we focus on visualising only the very significant relationships here (with p-values of 0.01 and lower).

We saw a positive relationship between the block A grade and the soft skill 'Ability to extract the core problem'. As we noted before, this relationship is reasonable since block A in software engineering is focused on determining the main problem and setting up a foundation to solve the problem. From the visualisation, we can see that the ratings for this skill ranged from 3.0 to 5.0. We can see that higher ratings yield higher grades at the end of block A, with average differences up to around 1.5 grade points. These findings indicate that the ability to extract the core problem might be a predictor for the total grade of block A.

We also found a negative correlation between the presentation grade of block B and the assessment of block A of the skill 'Growth Mindset', meaning that higher scores for this skill result in lower grades. The soft skill ratings for this skill range from slightly below 3.6 to 5.0, yielding average differences in grades of approximately 1.0 grade points. We are currently uncertain about the cause of this negative correlation. It could be that groups were over-confident, that certain high-scoring groups prioritised deliverables other than the presentation, or that we do not have enough data to determine whether some of these groups are outliers. To determine the actual reason, more investigation is necessary.

Other negative correlations of statistically significant difference are between the negative modifiers in block B and various soft skills. The average soft skills also show this negative correlation. We found that the skill 'Problem solving' in block A influences the block B negative grade modifiers. This means that this skill could be an early indication of workload imbalance in block B. Other important skills were 'Ability to benefit from constructive criticism', 'Collaborating with others', 'Communicating effectively', and 'Managing relationships'. The first skill could indicate that groups who disregard the feedback from the TA and course coordinator are more likely to achieve lower grades in block B. The latter three skills are skills representing the group's ability to work together and communicate with each other. Where the ability to extract the core problem was crucial in block A, teamwork and communication seem to be the key to the success of a group in block B. Groups lacking these skills are more likely to obtain negative grade modifiers.

In block B, the skill 'Communicating effectively' had a positive and significant effect on the requirements document (D1) and total grade. Self-assessed communication skills seem to influence performance during the presentations, affecting both the presentation in block A and the poster bonus point in block B significantly. We could not find any relevant correlation between the self-assessed communication skills and any communication-related soft skills, or even any soft skills at all. This could indicate that the self-assessed communication skills are more representative of the presentation skills. This could be related to confidence. Another explanation could be that the student feels confident in the ability to explain what they mean, which is highly relevant during presentations, but that this skill does not translate to discussions within a team where other group members are also involved, which might make it more difficult to get their point across.

It is clear that certain variables influence the grades. We see that various soft skills, technical skills and statuses of certain weeks positively impact the grades. We also noticed a significant negative correlation between certain soft skills and the negative modifiers, indicating that these negative modifiers contain information about the workload balance in a group. To determine whether these skills, as discussed above, are indeed predictors for certain grades, we would need to train and test a prediction model. This would give us further knowledge on how skills affect the grades or other variables, and might offer insight into why each of those variables is relevant. However, this is currently not possible due to the limited amount of data available.

5.4 Cluster analysis

Using our results and the definition of each quadrant, we can describe the groups in each quadrant. In the overview below, we define a descriptive name for each quadrant and argue why this name characterises the groups in this quadrant.

- **Quadrant 1**: *All-rounders*
  Groups in this quadrant are characterised by their above-median grades and above-median soft skills. They show a well-rounded skillset containing skills that lead to high performance and enable smooth teamwork, making the name 'All-rounders' suitable to describe their comprehensive skillset.

- **Quadrant 2**: *Hidden gems*
  Groups in this quadrant are characterised by their above-median grades but below-median soft skills. At first sight, they might struggle a bit with skills such as teamwork or motivation. But ultimately, they perform well and obtain high grades, earning the name 'Hidden gems'.

- **Quadrant 3**: *Developing learners*
  Groups in this quadrant are characterised by their below-median grades and below-median soft skills. These groups are running into issues and might still need to find their flow. To solve these issues, these groups will need to develop the necessary skills to do so, earning the name 'Developing learners'.

- **Quadrant 4**: *Social butterflies*
  Groups in this quadrant are characterised by their below-median grades and above-median soft skills. The name 'Social butterflies' refers to the high soft skill ratings in these groups, indicating that these groups consist of social and teamwork-oriented members.

By understanding the characteristics of each quadrant, we can start to understand what happens when groups move to a different quadrant between blocks. An overview of the evaluation of these movements is shown in Table 6.

Table 6: Overview of all possible movements between the four quadrants. For each movement, an indication of improvement is given, along with an indication of whether this movement was observed in the 2023-2024 cohort of Software Engineering. Finally, each movement is described through a unique characteristic.

Block A	Block B	Positive change?	Occurrences	Characteristic
Q1	Q2	✗	0	Soft skill regression
Q1	Q3	✗	0	Major setback
Q1	Q4	✗	2	Academic regression
Q2	Q1	✓	3	Skill refinement
Q2	Q3	✗	0	Unrealised academic potential
Q2	Q4	✗	0	Skill prioritisation
Q3	Q1	✓	0	Major breakthrough
Q3	Q2	✓	1	Academic gain
Q3	Q4	✓	0	Collaborative gain
Q4	Q1	✓	1	Academic refinement
Q4	Q2	✗	0	Academic prioritisation
Q4	Q3	✗	3	Unrealised collaborative potential

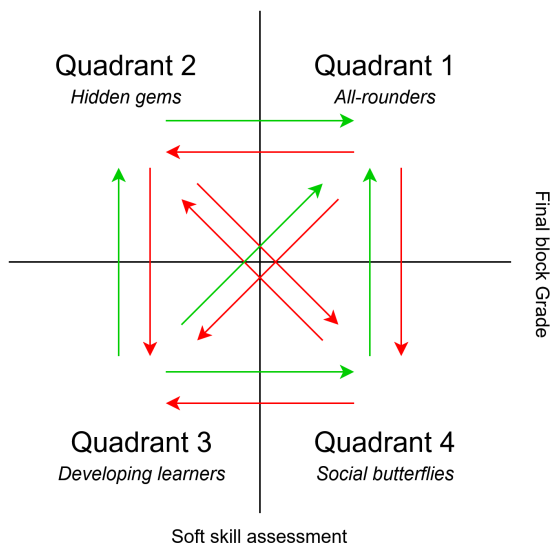

Figure 18: Overview of the four quadrants, created by the median values for grades and soft skills. Each quadrant has a name, describing the characteristics of the groups within it. Every possible movement between the quadrants is visualised too. Positive changes are indicated by green arrows, whereas negative changes are indicated by red arrows.

A new overview of the quadrants, updated with the descriptive names and movements indicating a positive or negative change, can be found in Figure 18.

Looking at the setbacks occurring in Software Engineering 2023-2024, we have $Q1 \rightarrow Q4$ and $Q4 \rightarrow Q3$. These movements give us insight into the support some of our groups are currently missing. Movement $Q1 \rightarrow Q4$ indicates that the group still functions in the same way, maintaining their soft skill ratings, but that their grades decreased, resulting in academic regression. To avoid this from happening, groups might benefit from support in the form of a discussion with the TA or course coordinator on how to maintain their high grades in block B of the course. Movement $Q4 \rightarrow Q3$ represents social and functioning groups with high soft skill ratings and the potential to collaborate and become all-rounders in $Q1$, but ultimately, we see their soft skill ratings deteriorate instead, making these groups end up in $Q3$. The movement $Q4 \rightarrow Q1$ does occur, so groups ending up in $Q3$ instead could use extra support to guide them in the other direction. The soft skill ratings in these groups might deteriorate due to conflicts or other issues within the group. Therefore, once the TA notices that conflict arises, these teams could use extra support in the form of a mediator or help developing their conflict resolution skills. This might help groups recover from conflict quickly and keep them on track towards success.

Vice versa, looking at the possible improvements that did not occur in Software Engineering 2023-2024, we see $Q3 \rightarrow Q1$ and $Q3 \rightarrow Q4$. While the movement $Q3 \rightarrow Q1$ would be ideal, it is unlikely that a team improves so much in every single aspect in a short amount of time. Regardless, these groups in $Q3$ should be offered additional support similar to the preventive measures discussed above. We believe these groups would benefit from support in developing important collaboration skills, as well as support towards understanding the expectations regarding the deliverables.

Having mentioned these ideas for additional support, we can include our previous findings showing that certain collaboration and communication skills are crucial in determining the block B negative grade modifiers, grades of various deliverables, and the total grade. This highlights the importance of supporting groups in $Q2$ and $Q3$ in developing these skills to prevent conflicts from resulting in poor grades.

5.5 Addressing the research questions

With our results, we can answer our main research question: *Can group success be predicted, and which aspects influence this success?*. We found that group success cannot currently be predicted with the available data. However, we found many aspects that influence success. We answer this question through our subquestions SQ1, SQ2, SQ3, and SQ4, each focusing on a different data category:

- *SQ1: To what extent do soft skills influence group performance?* We found that critical thinking skills such as 'Ability to extract core problem' significantly influence group performance in terms of grades in block A. In block B, the grades were influenced by teamworking skills such as 'Collaborating with others' and 'Communicating effectively'. Additionally, we found that these communication and teamwork skills are highly significant in determining the negative grade modifiers. The lower the average soft skill rating, the higher the chance of negative grade modifiers, especially in block B.

- *SQ2: To what extent do technical skills influence group performance?* The self-assessed communication skills appear to be very significant towards presentation grades. We also found notable correlations between the years of programming experience and various block A grades.

- *SQ3: To what extent do weekly statuses influence group performance?* The statuses of most weeks do not seem to influence the grades. However, there seem to be three crucial weeks. Week 5 in block A appears to be an early indication of negative modifiers, groups that struggle in this week are more likely to obtain one or more negative grade modifiers. Week 3 in block B seems to correlate positively and significantly with various deliverable grades. Finally, week 5 in block B is the best indication of the progress on the testing phase of students. This status positively and significantly influences the grade of the testing document.

- *SQ4: To what extent does collaboration preference influence group performance?* We found a negative and significant correlation between the collaboration preference and negative grade modifiers in block B.

5.6 Threats to validity

Given our findings and corresponding implications, we must be aware of the following limitations or threats to validity.

Limited data This study works with limited data, as we only use the data of Software Engineering 2023-2024. Expanding the dataset was not possible as previous runs of the course did not include the soft skill assessment, and the newest run of the course has not finished yet. As a result of the limited data, we noticed larger confidence intervals, as it's difficult to determine outliers. Additionally, we could not train or otherwise assemble a predictor, simply because insufficient data was available. However, despite these circumstances, we could still find statistically significant results.

Personal circumstances We found many negative correlations around the block B negative grade modifiers, and all these correlations are explainable and sensible. However, we need to keep in mind that while these results make sense, they don't explain the cause. Negative modifiers could occur because of a variety of reasons, including toxicity in a group or personal circumstances. Since these circumstances aren't known from our dataset, we cannot be sure if each possible situation affects the variables in the same way.

TA bias in weekly status For the correlations discussed above, we need to keep in mind that the weekly status as indicated by the TA is not always entirely objective. TAs base this status on their idea of what good progress and teamwork should look like, which can be influenced by, for example, how their own team operated previously or how they have seen other teams operate.

Collaboration preference The experiments in this study were performed under the assumption that all indicated collaboration preferences were honoured. While most collaboration preferences were indeed honoured, we cannot guarantee that all of them were honoured.

Missing values in soft skill assessment by TAs In some cases, TAs are unable to assess certain skills. A common example of this is presentation skills when the TA was unable to attend the presentation of the group they supervise. In these cases where the TA was not able to assess the skills, the skill is simply left unassessed, leading to missing values in the data. Through our data cleaning process, we ensured that these missing values would be excluded so as not to affect the results. However, the gap in data is still present and might have some impact on the results, for example, through larger confidence intervals.

Bias in TA observations The output from the TA observations, both the weekly statuses and the TA assessment, are subject to comparison. While there is a rough indication of what the progress in each week should be, and how students should behave in their groups, the Software Engineering course is open to a wide range of projects that each fit different appropriate approaches. Therefore, the observations are situational and based on TA judgment. Different TAs might judge certain situations of skills slightly differently. We cannot study these differences within the scope of this study.

5.7 Future work

This work focused on the course Software Engineering in 2023-2024. Future iterations of Software Engineering should be assessed to check whether the results are generalisable. This can also be extended to different courses. Our next study will include the creation of a replication package such that this work can also be applied to other courses.

One of the biggest limitations of this work is the lack of data. We only have 28 groups, made up of 118 students in total. Although this was enough to discover significant relationships in the data, we need more data to determine whether these variables are grade predictors. This could be done in a future study when data on new iterations of Software Engineering become available.

Additionally, the focus of this study was to explore the impact of various variables on the grades. With this information, we can create certain support measures to keep groups on track to success. Although we have already discussed some rough guidelines for support, in our next study, we wish to set up a customised support plan to support struggling groups and set up preventive measures such that struggles and conflicts can be detected and resolved in the early stages.

5.7.1 Self-assessment of soft skills

In the newest run of the course, Software Engineering 2024-2025, we let our students fill out a self-assessment of their soft skills. We could not fully include these results in this study as the course has not yet finished, so the data are incomplete. However, we performed a brief comparison of the self-assessed soft skills and TA-assessed block A soft skills. We visualised these results in a barplot, which can be seen in Figure 19.

We also performed a Wilcoxon signed-rank test to determine the significance of the difference between the soft skill self-assessment and TA-assessment. Additionally, we performed a correlation test using Spearman correlation to investigate the relationship between the two assessments further. The results of both these tests is available in Table 7. We report the p-value of both tests, plus the result of the Spearman correlation, as these values are essential to be able to draw conclusions.

From the comparison of self-assessed versus TA-assessed soft skills, the negative differences represent underestimated skills, and positive differences represent overestimated skills. We can see that most skills are often underestimated by students. This underestimation ranges from a few-tenths difference up to an entire point difference on a 5-point scale. Eight skills are often overestimated slightly, those are the following skills: ‘Continuous improvement’, ‘Growth mindset’, ‘Ability to benefit from constructive criticism’, ‘Communicating effectively’, ‘Problem solving’, ‘Analytical thinking’, ‘Managing conflict’, and ‘Ability to understand instructions’. These skills are overestimated by only a few tenths, up to 0.2.

From the Wilcoxon signed rank test, we can see that many differences between TA-assessment and self-assessment are statistically significant. Therefore, these initial results seem promising and could indicate that the self-assessed soft skills could be another grade predictor.

From the results of the Spearman correlation test, we observe very high p-values, indicating that there is no statistical significance. This means that, while the differences between the self-assessment and TA-assessment are significant, there is no consistent difference in ranking pattern between the two assessments.

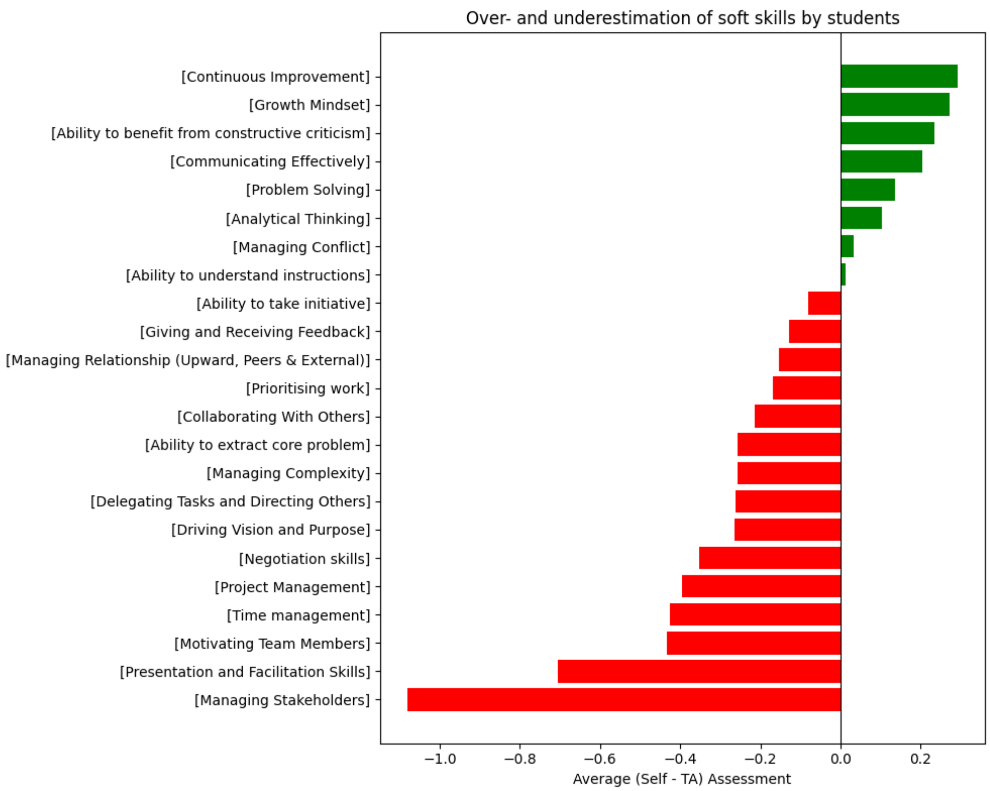

Figure 19: A visual comparison of the differences between the self-assessment of soft skills and the TA assessment of the same soft skills. The difference is calculated by subtracting the TA assessment from the self-assessment. A negative difference indicates that the student graded themselves lower than their TA did, a positive difference indicates that the student graded themselves higher than their TA did. TAs did not see the self-assessments of the students to remain as objective as possible.

Table 7: Results of the Wilcoxon signed-rank test and Spearman correlation on the differences between the self-assessment and TA-assessment of soft skills. Significant p-values are highlighted in green.

Soft skill attribute	Wilcoxon signed rank test	Spearman correlation	
	P-value	Result	P-value
Ability to benefit from constructive criticism	0.021	-0.03	0.734
Ability to understand instructions	0.753	-0.13	0.156
Collaborating With Others	0.002	-0.07	0.447
Communicating Effectively	0.169	-0.07	0.440
Giving and Receiving Feedback	0.074	-0.02	0.810
Managing Relationship (Upward, Peers & External)	0.026	0.02	0.857
Negotiation skills	0.004	-0.03	0.805
Presentation and Facilitation Skills	0.000	0.08	0.400
Managing Stakeholders	0.000	-0.09	0.335
Motivating Team Members	0.000	0.04	0.691
Ability to extract core problem	0.001	0.10	0.264
Analytical Thinking	0.299	-0.02	0.840
Continuous Improvement	0.004	-0.07	0.448
Growth Mindset	0.019	0.08	0.368
Problem Solving	0.271	-0.11	0.244
Ability to take initiative	0.155	0.06	0.507
Delegating Tasks and Directing Others	0.005	0.07	0.434
Driving Vision and Purpose	0.004	-0.09	0.351
Managing Complexity	0.005	-0.06	0.520
Managing Conflict	0.506	0.06	0.580
Prioritising work	0.025	0.02	0.857
Project Management	0.000	0.05	0.624
Time management	0.000	0.07	0.425

In other words, we could not find any pattern where high self-assessments correspond to high/low TA-assessments, or the other way around.

In the self-assessed communication skills from the technical questionnaire in Software Engineering 2023-2024, we noticed that a higher self-assessment yielded higher presentation grades, possibly because of increased confidence. This raises the question of whether more self-assessed skills are similarly related to certain grades.

To investigate these self-assessments and the relationship with grade variables further, the full Software Engineering 2024-2025 data needs to be investigated once the course is finalised.

6 Conclusion

In this study, our objective was to explore the impact of soft skills on the grades of groups in the Software Engineering course in 2023-2024. Besides the soft skills, we also explored the impact of technical skills, weekly statuses and indicated collaboration preferences.

We performed correlation tests, statistical tests and used visualisations to investigate the relationships in the data. From our experiments, we found that experience with programming is relevant in block A for setting up the requirements document and in presenting the intermediate results, with more experience leading to higher grades on these parts and the total block A grade. We also found that the self-assessed communication skills from the technical questionnaire have a positive and significant relationship with presentation grades.

From the TA-assessed soft skills, we found that the ‘Ability to extract the core problem’ is crucial for success in block A, and issues in communication could lead to a negative grade modifier in this block. For block B, we found that teamwork and communication are more important, with the most essential skills being the ‘Ability to benefit from constructive criticism’, ‘Collaborating with others’, ‘Communicating effectively’, and ‘Managing relationships’. The negative grade modifiers in block B are more complex, depending on more than half of the assessed soft skills.

From the weekly statuses, we found that week 5 in block A is most telling about the negative modifiers in block A. Week 3 in block B is a good indication for various grades, from both block A and block B, which indicates that the block A success of a group is reflected in this week’s status. Finally, week 5 in block B has a positive correlation with the grade for the testing document, making it a good indication of the status of product testing.

By constructing clusters using quadrants, we identified the nature of movements between the quadrants. Using these movements, we believe that the setbacks of some groups may have been prevented through additional support or intervention. We discussed some support measures for this, including support in developing conflict resolution skills, offering direct mediation, and discussing the expectations of the deliverables.

In addition, we compared the difference between TA-assessed and self-assessed block A soft skills in Software Engineering 2024-2025. We found that on average, students underestimate most of their skills. The most underestimated skill was ‘Managing stakeholders’. The underestimation of this, and possibly other skills too, could be due to the students’ inexperience with the task. We found that the difference between the self-assessment and TA-assessment was statistically significant for many of the soft skills, but we could not find a consistent ranking pattern.

While we could not use the Software Engineering 2023-2024 data to create a prediction model, we discovered and investigated crucial relationships that describe the impact of various variables on the grades. This study provides a good foundation for future studies that aim to investigate support measures for struggling groups, build prediction models, or investigate the differences between self-assessed skills and TA-assessed skills.

7 Acknowledgements

The authors extend their gratitude to Dr. Venustiano Soancatl Aguilar for his support and guidance in navigating our dataset and for his valuable insights into the interpretations of our results.

The authors acknowledge the use of Grammarly and Writefull, which use artificial intelligence technologies, to improve spelling, punctuation, and grammar.

References

[1] Ahmed Al-Sa'di, Parina Yamjal, Esraa Ahmad, Richa Panjabi, CAM Allott McPhee, and Olkan Guler. Assessing Educators' Soft Skills: Developing a Self-Assessment Instrument. *Administrative Sciences*, 13(9):208, sep 19 2023.

[2] A. Aranzabal, E. Epelde, and M. Artetxe. Team formation on the basis of Belbin's roles to enhance students' performance in project based learning. *Education for Chemical Engineers*, 38:22–37, 1 2022.

[3] Annabelle Beyer, Saskia Hohagen, Uta Wilkens, and Valentin Langholf. The impact of team competence on short- and long-term team performance. *TEAM PERFORMANCE MANAGEMENT*, 30(5/6):136–153, OCT 23 2024.

[4] Diego Buenaño-Fernández, David Gil, and Sergio Luján-Mora. Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study. *Sustainability*, 11(10):2833, may 17 2019.

[5] S. B. Channon, R. C. Davis, N. T. Goode, and S. A. May. What makes a 'good group'? exploring the characteristics and performance of undergraduate student groups. *ADVANCES IN HEALTH SCIENCES EDUCATION*, 22(1):17–41, MAR 2017.

[6] Amrita Dutt, Vijaishri Tewari, Shashikant Rai, and Shardul Shankar. Exploring the mediating role of student attitudes in the relationship between dysfunctional teams and student team performance. *ACTA PSYCHOLOGICA*, 252, FEB 2025.

[7] Álex Escolà-Gascón and Josep Gallifa. How to measure soft skills in the educational context: psychometric properties of the SKILLS-in-ONE questionnaire. *Studies in Educational Evaluation*, 74:101155, 9 2022.

[8] Fabian Fagerholm and Arto Vihavainen. Peer assessment in experiential learning assessing tacit and explicit skills in agile software engineering capstone projects. In *2013 IEEE Frontiers in Education Conference (FIE)*, pages 1723–1729, 2013.

[9] Ruti Gafni, Itzhak Aviv, Boris Kantsepolsky, Sofia Sherman, Havana Rika, Yariv Itzkovich, and Artem Barger. Objectivity by design: The impact of AI-driven approach on employees' soft skills evaluation. *Information and Software Technology*, 170:107430, 6 2024.

[10] Hina Gull, Madeeha Saqib, Sardar Zafar Iqbal, and Saqib Saeed. Improving learning experience of students by early prediction of student performance using machine learning. In *2020 IEEE International Conference for Innovation in Technology (INOCON)*, pages 1–4, 2020.

[11] Shelly Gupta and Jyoti Agarwal. Machine learning approaches for student performance prediction. In *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pages 1–6, 2022.

[12] Hariegwambe, Shakila Baskaran, and Prakash Marimuthu. Stacking based machine learning approach for student performance analysis. In *2024 Global Conference on Communications and Information Technologies (GCCIT)*, pages 1–5, 2024.

[13] Irmohizam Ibrahim and Norhasni Zainal Abiddin. The Critical Role of Soft Skills in Engineering: Enhancing Performance and Career Advancement. *Journal of Ecohumanism*, 3(7):691–703, oct 6 2024.

[14] R.K. Kavitha, W. Jaisingh, and S.K. Kanishka Devi. Applying learning analytics to study the influence of fundamental computer courses on project work and student performance prediction using machine learning techniques. In *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, pages 1–5, 2021.

[15] Kwok Hung Lau and Qian Jin. Chinese students' group work performance: does team personality composition matter? *EDUCATION AND TRAINING*, 61(3):290–309, 2019.

[16] Ruholla Jafari Marandi, Brian K Smith, Reuben F Burch, and Sara C Vick. Engineering soft skills vs. engineering entrepreneurial skills. *The International journal of engineering education*, 35(4):988–998, 2019.

[17] Fouzia Munir. More than technical experts: Engineering professionals' perspectives on the role of soft skills in their practice. *Industry and Higher Education*, 36(3):294–305, jul 26 2021.

[18] Sathiyapriya N, Rithika C, Sanjay S, and Subathra S. Utilizing machine learning to forecast a student's performance. In *2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA)*, pages 1–5, 2024.

[19] André Seixas de Novais, José Alexandre Matelli, and Messias Borges Silva. Fuzzy Soft Skills Assessment through Active Learning Sessions. *International Journal of Artificial Intelligence in Education*, 34(2):416–451, may 11 2023.

[20] Rick Parker, Shraddha Sangelkar, Matthew Swenson, and Julie Dyke Ford. Launching for success: A review of team formation for capstone design. *International Journal of Engineering Education*, 35(6):1926–1936, 2019.

[21] D. Petkovic, G. Thompson, R. Todtenhoefer, S. Huang, B. Levine, S. Parab, G. Singh, R. Soni, and S. Shrestha. Work in progress — e-tat: Online tool for teamwork and "soft skills" assessment in software engineering education. In *2010 IEEE Frontiers in Education Conference (FIE)*, pages S1G–1–S1G–3, 2010.

[22] Monika Pogatsnik, Éva Beke, and Rita Bodane Kendrovics. Essential teambuilding skills for project work in engineering studies. In *2024 IEEE 18th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 000121–000126, 2024.

[23] Juan Ruiz and Renee Wever. Design team formation using self-assessment and observer-assessment techniques: Mapping practices in a global network of universities. *Design Science*, 10:e7, 2024.

[24] Shayna A. Rusticus and Brandon J. Justus. Comparing student- and teacher-formed teams on group dynamics, satisfaction, and performance. *SMALL GROUP RESEARCH*, 50(4):443–457, AUG 2019.

[25] Bernd Schulz. The importance of soft skills: Education beyond academic knowledge. 2008.

[26] Shikha Sharma and Sanjeev Kumar Sharma. Team resilience: Scale development and validation. *Vision*, 20(1):37–53, 2016.

[27] Marlene A. Smith and Deborah L. Kellogg. Required Collaborative Work in Online Courses: A Predictive Modeling Approach. *Decision Sciences Journal of Innovative Education*, 13(4):563–581, 10 2015.

[28] Rahila Umer, Teo Susnjak, Anuradha Mathrani, and Lim Suriadi and. Current stance on predictive analytics in higher education: opportunities, challenges and future directions. *Interactive Learning Environments*, 31(6):3503–3528, 2023.

A Software Engineering 2023-2024 data

Table 8: Data types and possible values of all variables. Orange rows represent variables on student-level, green rows represent variables on group-level.

Variable	Datatype	Data category	Subcategory	Cols	Possible values
Group	String	Categorical	Nominal	1	G[01, 28]
Student ID	String	Categorical	Nominal	1	S[001, 118]
Group grade	Float	Numerical	Continuous	1	[0, 10]
Modifier block A	Float	Numerical	Continuous	1	[-10, 1]
Modifier block B	Float	Numerical	Continuous	1	[-10, 1]
Final grade	Float	Numerical	Continuous	1	[0, 10]
Rounded final grade	Float	Numerical	Continuous	1	[0, 10]
D1 grade in block A	Float	Numerical	Continuous	1	[0, 10]
D2 grade in block A	Float	Numerical	Continuous	1	[0, 10]
D3 grade in block A	Float	Numerical	Continuous	1	[0, 10]
Total grade block A	Float	Numerical	Continuous	1	[0, 10]
D1 grade in block B	Float	Numerical	Continuous	1	[0, 10]
D2 grade in block B	Float	Numerical	Continuous	1	[0, 10]
D3 grade in block B	Float	Numerical	Continuous	1	[0, 10]
D4 grade in block B	Float	Numerical	Continuous	1	[0, 10]
D5 grade in block B	Float	Numerical	Continuous	1	[0, 10]
Presentation grade in block B	Float	Numerical	Continuous	1	[0, 10]
Poster bonus block B	Float	Numerical	Continuous	1	[0, 1]
Total grade in block B	Float	Numerical	Continuous	1	[0, 10]
Weekly statuses block A	Integer	Categorical	Ordinal	8	[1,4]
Weekly statuses block B	Integer	Categorical	Ordinal	7	[1,4]
Soft skills assessment block A	Integer	Categorical	Ordinal	24	[1,5]
Soft skills assessment block B	Integer	Categorical	Ordinal	24	[1,5]
Years of programming experience	Integer	Numerical	Discrete	1	[0, ...>
Years of industry experience	Integer	Numerical	Discrete	1	[0, ...>
Specific programming experiences	Boolean	Categorical	Nominal	9	yes/no
Communication skills	Integer	Numerical	Ordinal	1	[1,5]
Collaboration preference	Boolean	Categorical	Nominal	1	yes/no

B Correlation heatmaps

B.1 Correlation heatmap of all variables

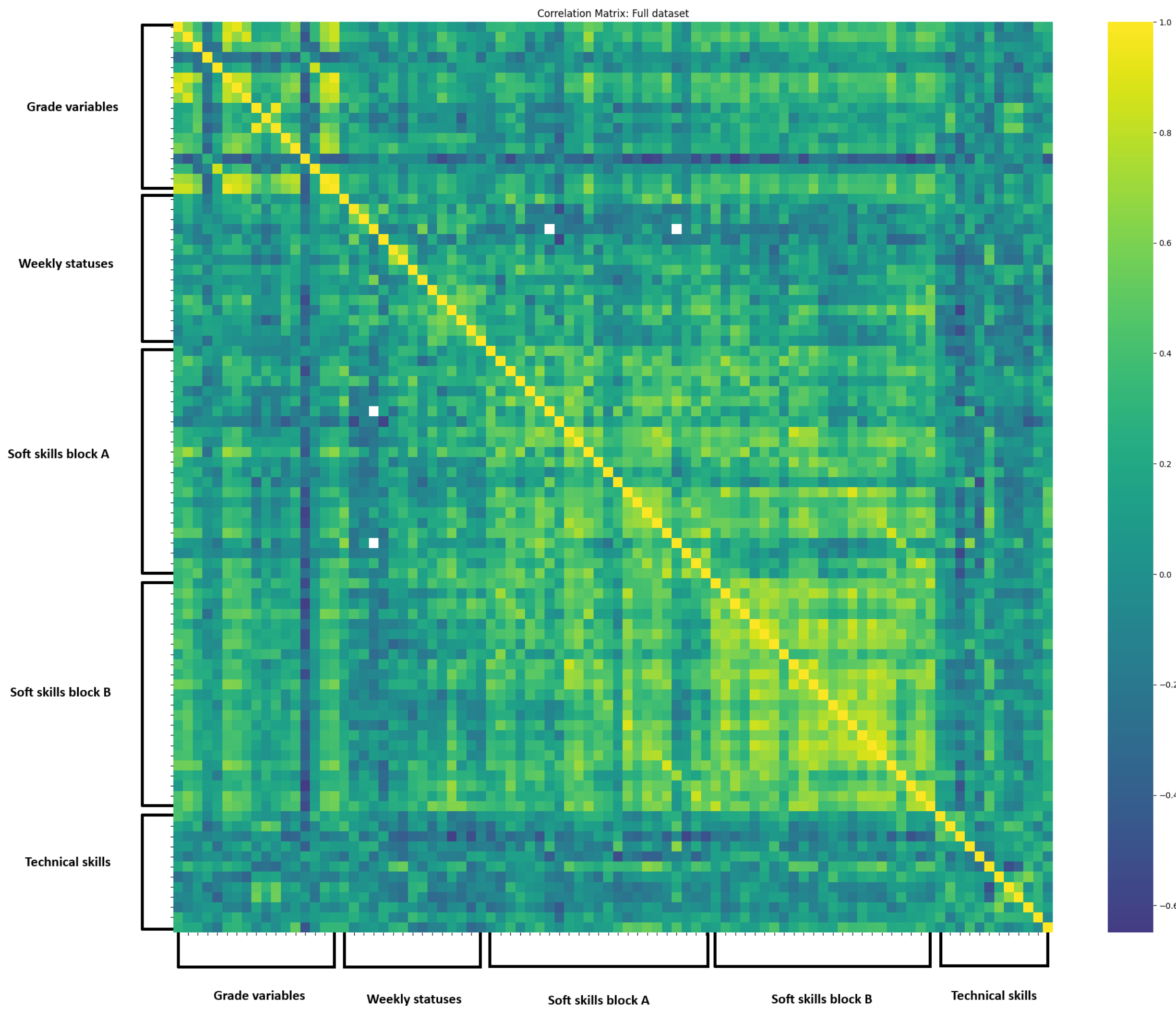

Figure 20: Heatmap of the correlation matrix of all variables in the dataset. To increase readability and interpretability, we listed the variable categories instead of individual variable names.

B.2 Correlation heatmap of soft skills versus grades

Figure 21: Heatmap of the correlation matrix of soft skills versus grades.

C Statistically significant differences in soft skills

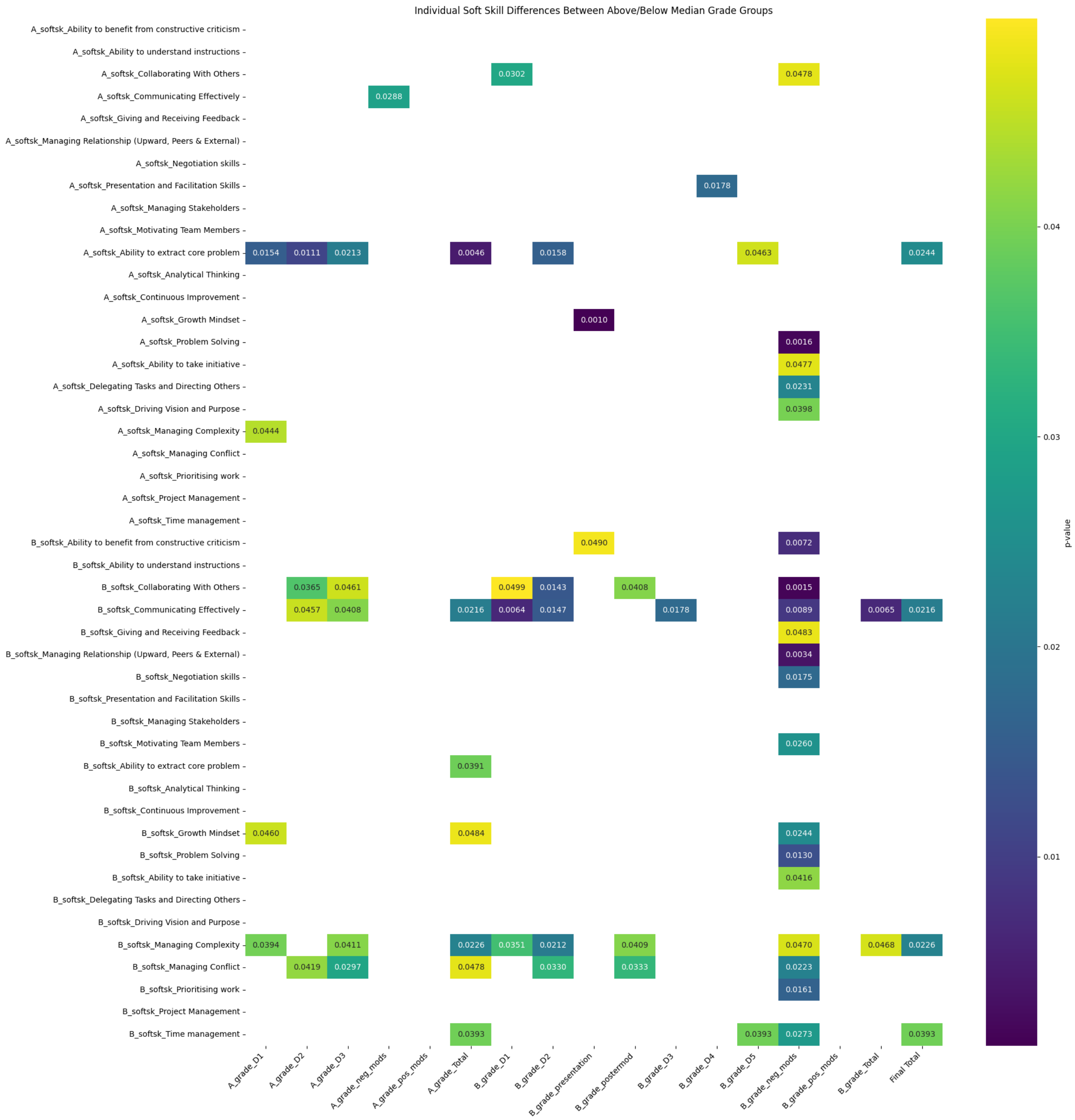

Figure 22: Heatmap of the p-values indicating the statistical significance between the below-median and above-median grade categories regarding the correlation between soft skill attributes and grades.