POLITECNICO DI TORINO
Repository ISTITUZIONALE

The Impact of Team Diversity in Agile Development Education

Original
The Impact of Team Diversity in Agile Development Education / Torchiano, M., Coppola, R., Vetro', A., Musaj, X.. - (2025), pp. 1593-1601. (FSE Companion '25: 33rd ACM International Conference on the Foundations of Software Engineering Trondheim (NOR) June 23 - 28, 2025) [10.1145/3696630.3728696].

Availability:
This version is available at: 11583/3002894 since: 2025-09-09T15:42:30Z

Publisher:
Association for Computing Machinery

Published
DOI:10.1145/3696630.3728696

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

27 November 2026



The Impact of Team Diversity in Agile Development Education

Marco Torchiano
Riccardo Coppola
Antonio Vetrò
{first.last}@polito.it
Department of Control and Computer Engineering
Politecnico di Torino, Turin, Italy

Xhoi Musaj
xhoi.musaj@edu.unito.it
Department of Psychology
University of Turin, Turin, Italy

Abstract

Software Engineering is mostly a male-dominated sector, where gender diversity is a key feature for improving equality of opportunities, productivity, and innovation. Other diversity aspects, including but not limited to nationality and ethnicity, are often understudied.

In this work we aim to assess the impact of team diversity, focusing mainly on gender and nationality, in the context of an agile software development project-based course.

We analyzed 51 teams over three academic years, measuring three different Diversity indexes – regarding Gender, Nationality and their co-presence – to examine how different aspects of diversity impact the quality of team project outcomes.

Statistical analysis revealed a moderate, statistically significant correlation between gender diversity and project success, aligning with existing literature. Diversity in nationality showed a negative but negligible effect on project results, indicating that promoting these aspects does not harm students' performance. Analyzing their co-presence within a team, gender and nationality combined had a negative impact, likely due to increased communication barriers and differing cultural norms.

This study underscores the importance of considering multiple diversity dimensions and their interactions in educational settings. Our findings, overall, show that promoting diversity in teams does not negatively impact their performance and achievement of educational goals.

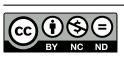
CCS Concepts

• **Software and its engineering** → *Agile software development*; **Programming teams**; • **Human-centered computing** → **Empirical studies in collaborative and social computing**.

Keywords

Agile, Team, Diversity, Software, Development, Education

ACM Reference Format:
Marco Torchiano, Riccardo Coppola, Antonio Vetrò, and Xhoi Musaj. 2025. The Impact of Team Diversity in Agile Development Education. In *33rd ACM International Conference on the Foundations of Software Engineering (FSE Companion '25), June 23–28, 2025, Trondheim, Norway*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3696630.3728696



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
FSE Companion '25, Trondheim, Norway
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1276-0/2025/06
https://doi.org/10.1145/3696630.3728696

1 Introduction

The lack of diversity in the tech industry is a well-known issue [4], even though several studies have recognized perceived diversity (i.e., the diversity factors that individuals are born with) as a high-value team property within Software Engineering (SE) teams, with companies willing to increase their effort to create more diverse work teams [19].

Regarding gender diversity, many studies have demonstrated that it brings benefits to the SE workplace, e.g. with increases in innovation [24], productivity [16], and with higher sentiment, although inequities are still evident: Guzmán et al. [8] reported that while both women and men hold leadership responsibilities in similar numbers, women are more likely to feel unsupported in their roles, have less authority, be dissatisfied with their compensation, and feel undervalued within their teams. Furthermore, their survey highlights that women face greater challenges related to team dynamics and biases, whereas men's primary barriers are technical and project-related issues.

Evidence about the increasing number of women (and of their responsibilities) collected from workplaces is however in contrast with the current trends in the educational context. In a snapshot presented in [2], related to Software Engineering education in a UK University, the female student population remains critically low, barely rising above 18%. Anh Nguyen-Duc and Letizia Jaccheri found that the active involvement of female students occurs in project management and requirement engineering, while the areas lacking active involvement are software architecture and Scrum methodology. [15].

Other aspects of diversity appear understudied in the related literature: Szlavi et al. [23] underline the necessity of addressing the systemic barriers in CS through intersectional lenses[1], and that an encouraging growing awareness can be found in the field. However, many of the provided results exploring these aspects are mixed, e.g. regarding nationality, age, race, disability, and sexual orientation. In the aforementioned analysis by Ortu et al. – for instance – it is indicated that diversity in nationality is linked with lower team politeness [16]; another study by Stahl et al. finds that cultural diversity has a negative impact on SE processes [22]. Sanchez-Gordon and Colomo-Palacios denounce a long-standing lack of diversity in Computer Science, that can create a toxic culture and

[1]The concept of *Intersectionality* is defined by Crenshaw as how race, class, gender, and other individual characteristics "intersect" with one another and overlap [5].

contribute to low rates of participation by underrepresented groups (URGs) [20].

The main objective of the present study is to perform an analysis of the current status in terms of DEI (Diversity, Equity, and Inclusion) in a Software Engineering capstone course held at Politecnico di Torino. To the best of our knowledge, this paper constitutes the first quantitative investigation about how DEI aspects influence team performance in Software Engineering courses. The course includes a mandatory team project, where the students are tasked with the development of a small-scale software project using the Agile and Scrum methodologies, thus fostering constant collaboration and opinion exchange between team members. All the project teams have been created by allocating the students to maximize the average diversity within each team. In this manuscript, we report our conclusion regarding the effects on course appreciation and team performance, in three consecutive editions of the course.

The remainder of this paper is structured as follows: in Section II, we provide a survey of prominent results about diversity and inclusion in Software Engineering; in Section III, we discuss the methodology used for the evaluation and the context where it took place; in Section IV, we present the results, that are discussed in Section V, along with the main threats and current limitations of the study; finally, Section VI concludes the paper and provides possible future research directions.

2 Background

In this section, we report findings from related works in the literature analyzing the effects of diversity within SE teams, and the metrics that can be used to quantitatively measure such diversity.

2.1 Effects of team diversity in SE

Many studies in the recent literature have analyzed, at different education levels and in different countries, how diversity and inclusion topics are perceived by SE teachers and students, and which measures are taken to approach the topic.

Happe et al. presented a study on the gender gap in SE education and careers in Germany, through a retrospective analysis of women's experiences at the university level. The author's findings reveal that significant barriers – such as the perception of software engineering as incompatible with personal interest, a male-dominated environment, and a lack of female role models – deter many women from pursuing the field. Persistent stereotypes and social challenges contribute to gender inequality, underscoring the need for supportive measures to encourage diversity in computing education [9].

Anderson et al. examined gender disparities in SE education within a UK university, focusing on both undergraduate and postgraduate levels. The study analyzes institutional data, national benchmarks and a survey of Master's students. The primary findings reveal a persistent gender imbalance, with women significantly underrepresented, particularly at higher academic levels. The authors point out that female students at the studied institution achieve comparable academic performance to their male counterparts, indicating no significant attainment gap, though female enrollment remains low. The authors underline how these results mirror broader UK trends and the need for interventions to generally support female participation in SE [2].

Cutrupi et al. presented the Draw-A-Software-Engineer Test (DASET), a research tool to study the perception of societal norms, gender biases and stereotypes. The authors asked nine university students to draw and describe their perception of software engineering. Of the artefacts, 78% of the drawings represented a man as a software engineer. Additionally, only 11% of the drawings included a software engineer with brown skin [6].

Hyrynsalmi studied the approaches of diversity and inclusion most commonly used by SE teachers in Finland, and how the importance of diversity and inclusion is perceived in their courses. The author found that whilst some of the respondents were hesitant to introduce such aspects, the importance of diversity and inclusion in SE education was recognized by most. It is underlined an existing familiarity with accessibility approaches, that are often used as an example of diversity and inclusion in teaching. Finally, the respondents recognized an increase in diversity and multiculturalism in the academic field, and the presence of trust issues between individuals that prioritize diversity and inclusion topics and those who do not [10]. Hyrynsalmi et al. also approached the topic of diversity in SE paths through an online survey sent to their students, showing that on average women apply later to SE studies than men, with statistically significant differences between genders [12].

Kovaleva et al. pointed out the necessity for the design of gender-neutral Software Engineering programs, as a tool to reach gender balance in the domain. The authors underlined the necessity of adopting a gender-neutral language in marketing and communication for SE programs, pointing out the existing pressure of stereotypes experienced by both genders [13].

Several studies have also analyzed diversity and inclusion for SE practitioners. Petrescu et al. conducted a study in Romania to investigate gender diversity in the Scrum Master role, within the context of agile software development, focusing on professionals in various companies. Primary findings suggest that women are well-suited to the Scrum Master role due to their strong communication, empathy, and conflict-management skills, valued more than technical expertise in this role. These findings at the same time may however underline a trend to underestimate women's hard skills in favor of soft skills [17].

Zhao and Young analyzed workplace discrimination in Software Engineering through a questionnaire administered to 97 respondents from all over the world. The findings highlighted that the main causes of workplace discrimination in SE are age, gender, and race/ethnicity/nationality. The intersectionality of the discrimination issue is evident when the authors point out that when analyzing other demographic factors than gender, female software professionals are more likely to be discriminated [26].

2.2 Diversity indexes

For our analysis of diversity within teams in SE, we considered all the three diversity indexes that were mentioned in the analyzed related work.

Richness: the number of different types present in the population.

Table 1: Diversity Indexes for exemplar team compositions (letters symbolize different types)

Data	Richness	Blau	Shannon
A b b b b b b b b b b b	2	0.153	0.287
A A A A A A b b b b b b	2	0.500	0.693
A A A b b b C C C C C C	3	0.625	1.040
A A A A b b b b C C C C	3	0.667	1.099
A b C C d d d d d d d d	4	0.514	0.983
A A A b b b C C C d d d	4	0.750	1.386

Shannon index [21]: originally proposed in information theory, the Shannon entropy quantifies the uncertainty in predicting the type of individual that is taken at random from the dataset. It is defined as:

$$H = \sum_{i=1}^{R} p_i \cdot ln(p_i)$$

where R is richness (the total number of types in the dataset), p_i is the proportional abundances the i-th type.

Blau diversity index [3]: The measure equals the probability that two entities taken at random from the dataset of interest represent the same type. It is defined as:

$$B = 1 - \sum_{i=1}^{R} p_i^2$$

In table 1, as an example, we show the corresponding diversity values for a population of 12 individuals with different random exemplar compositions.

3 Method

The general overview of the design of our study is illustrated in figure 1. This section details the context of our analyses, the research questions, the variables used to answer them, and the analysis method applied.

3.1 Context

The setting for the study reported in this paper is Master Software Engineering course. The main educational goal of the course is to provide the students with the essential skills and knowledge required to manage and execute software projects using agile methodologies. The focus is on instilling a comprehensive understanding of the agile development approach, emphasizing hands-on experience in carrying out team-based software projects efficiently and in an inclusive manner, i.e., enhancing the peculiarities of each person. All of the students involved have extensive experience in teamwork, acquired in previous courses; moreover, the instructors constantly monitor the behavior of the teams to capture possible communication issues and provide feedback for improvement.

Teams are formed after the first week, based on the information declared by the students in the preliminary questionnaire. We aim to achieve homogeneity between teams by balancing a variety of aspects that are assessed through an online form; the main aspects are gender, nationality, technical skills, and previously attended courses.

The actual composition of teams may vary due to some students dropping out of the course since they are not able to regularly attend the classroom practical activities and participate in the project work.

From the Gender Balance Report of our university [18], we know that overall female students are 30.5%, while in computer-related degrees the female students are 16.8% (in line with the findings reported in [2]). Focusing on nationality diversity, on average the university enrolls 20.6% foreign students.

The grading is performed by evaluating both team and individual performance. Team evaluation weights 80%, while individual evaluation weights 20%. The team evaluation is carried on according to four main criteria: team coordination and improvement, Scrum compliance, quality of meetings, and doneness.

3.2 Research questions

Since the novelty of this kind of studies especially in the SE domain, we defined a preliminary research question focused on methodology:

RQ0: *How are different diversity indexes related to each other?* Among the several diversity indexes proposed in the literature, we picked the three most common – presented in Sec. 2.2 above –. We are interested in understanding how they are related to each other and if there is a preferable one.

The main goal of our investigation is to understand how team diversity plays a role in agile software development projects and specifically how it can affect team performance. To this end, we formulate two main research questions:

RQ1: *What is the impact of diversity factors on team performance?* We consider different team diversity factors, measured according to diverse indexes, and observe the correlation with the team score.

RQ2: *What is the relative importance of different diversity factors?* We observe the magnitude of the impact of different diversity factors to understand which ones play a more meaningful role.

3.3 Variables

In order to address the main research questions, we consider a single dependent construct which is team performance. We operationalise it through one variable which is the final score assigned to the team. In particular, since the top grade differed in the three considered academic years, we adopt a normalized version so that 100 is the top grade attributable for the specific academic year.

As far as the factors whose influence we wish to study (i.e., the independent variables), they are:

- the **Gender diversity**, measured on the basis of the self-declared gender of the student, according to the definition[2] and guidelines provided by the American Psychological Association [1].
- the **Nationality diversity**, measured on the basis of the official nationality declared by the student at enrollment,

[2]Gender (n): the condition of being male, female, or neuter. In a human context, the distinction between gender and sex reflects the usage of these terms: Sex usually refers to the biological aspects of maleness or femaleness, whereas gender implies the psychological, behavioural, social, and cultural aspects of being male or female (i.e., masculinity or femininity)

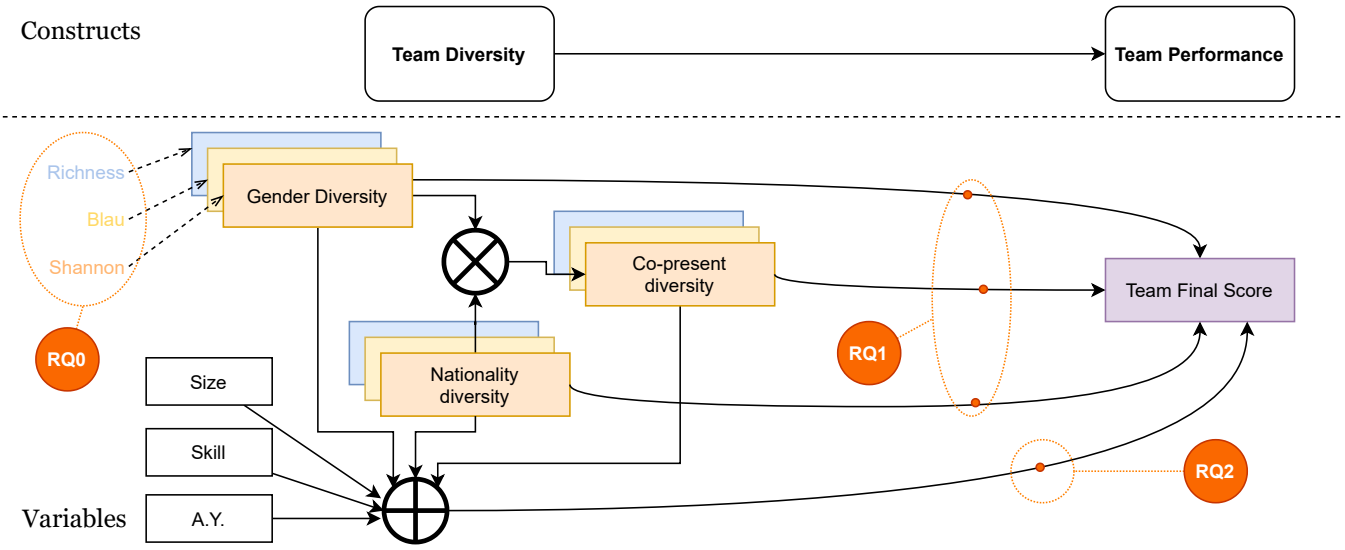


Figure 1: Overview of study design.

- the **Co-present (Gender and Nationality)** diversity, measured as the joint class of Gender and Nationality,
- the **Career diversity**, measured on the basis of the student being an Erasmus or International student, i.e. students who do not share a common background with the other students, since they typically come from foreign university for a single semester.

All diversity factors have been measured using the three diversity indexes presented in section Section 2.2 above. Therefore we have nine different combinations of factor and diversity index.

In addition, we also monitored a few co-factors that could affect the final team score:

- the **size** of the team,
- the **skill score**, measured as self reported skill in basic software development, web development, and software engineering (collected through a questionnaire administered at the beginning of the course),
- the **academic year** of attendance.

3.4 Analysis method

Concerning RQ0, we report the distribution of three different indexes for the considered diversity factors. In addition, we compute the overall correlations among the indexes.

Concerning RQ1, we compute the correlation coefficients as well as the relative statistical significance. Since we expect distributions far from normal we opted for the Spearman correlation coefficients which is suitable for non-normal distributed data.

As far as RQ2 is concerned, we compute a linear regression model of the final team score vs. all the diversity factors and the co-factors that we monitored.

Overall the goal of the study presented in this paper is exploratory, we do apply statistical tests as guidelines. In deciding whether to decide statistical significance we e consider the customary $\alpha = 0.01$ as the threshold.

Table 2: Summary of team compositions

		Size					Average
AY	Teams	Min	Median	Max	Women %	Foreign %	Score
2021/22	13	5	7	7	15.3%	14.1%	93
2022/23	20	5	6	7	16.7%	25.8%	91
2023/24	18	5	7	7	18.5%	27.7%	90

We aim to interpret the results and provide a general assessment of the effects of diversity on the team performance.

4 Results

In this section, we report information about the population of students evaluated in our study and the results of the three research questions described in the previous section.

4.1 Population

We consider the teams that participated in the latest three completed edition of the course, i.e. in academic years from 2021/22 to 2023/24.

Overall, we analyzed 51 teams over three academic years, which included a total of 324 members.

Table 2 reports the summary statistics of team compositions in the three academic years we considered in our analysis: for each year we report the number of teams, the distribution of team size, the average percentage of women in teams, the average percentage of foreign students in teams, and the average grade obtained by the team at the end of the academic year.

We observe that the size of teams is generally 6 or 7 members. We originally formed teams of the same size based on the online questionnaire that the prospective students filled in before the start of the course (see also section 3.1), but eventually few teams ended up with fewer components due to some students not showing up at the beginning of the course lectures or dropping out after a few lectures. The percentage of women increased from 15% to 18% in

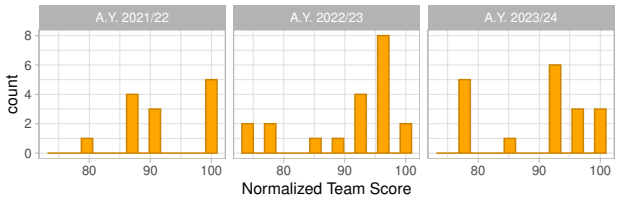


Figure 2: Team score distribution

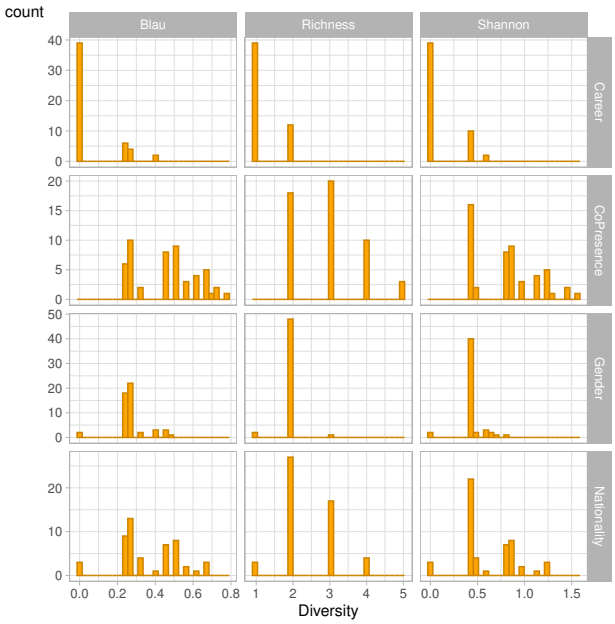


Figure 3: Overview of DI distributions

the three years, while the proportion of foreign students doubled (from 14% to 28%).

The average normalized team scores ranged from 93 to 90. Figure 2 reports the detailed distribution of team scores in the three considered academic years.

4.2 RQ0 - Relationship among indexes

The distribution of the values of the three Diversity Index applied to each of the four Diversity Factors is reported in Figure 3 that shows $3 \times 4 = 12$ histograms.

We observe that Richness (mid column) includes a few distinct values, as we would expect from its definition. While Blau and Shannon indexes (the two outer columns) are more spread over different values. When considering the indexes – i.e. looking by row – we can observe similar patterns.

A more holistic view is reported in figure 4, where for each index the values computed across all factors are considered together. In practice we summed together the histograms in the three columns of the previous figure, ignoring the distinction among the factors. In particular we report the correlation among the three different indexes.

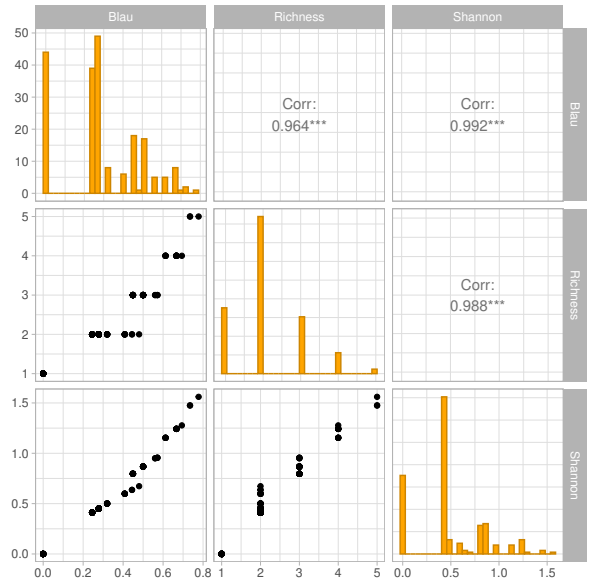


Figure 4: Cross correlations of the three Diversity Indexes

Table 3: Configuration causing index inconsistencies

Feature	Configuration	Blau	Shannon	Richness
Nationality	IR IT IT IT IT IT PT	0.449	0.796	3
Gender	F F M M M	0.480	0.673	2

There is a generally high correlation between the three pairs of Indexes, the highest being between Blau and Shannon. Since the Richness index is – by construction – discrete, it shows a slightly lower correlation with the other two indexes.

We also observe in the scatter plot of Blau vs. Shannon and Blau vs. Richness (two leftmost) that for a few values the two pairs of indexes are inconsistent. In particular, the ordering of two specific configurations is different w.r.t Blau vs. Shannon and Richness. The inconsistent pair corresponds to a configuration of Gender in a 5-member team and a configuration of Nationality in a 7-member team, whose details and index values are reported in table 3. Since such inconsistency involves two different features it does not affect the remainder of our analysis. However it must be noted if we aim to apply diversity indexes in larger data sets.

4.3 RQ1 - Impact of team diversity

Correlation coefficients of the different diversity measures vs. the final team score are reported in table 4. Figure 5 reports the scatter plots of the normalized final team score vs. the different combinations of diversity indexes and factors. The figure also reports the linear regression line together with the confidence interval and the value of the Pearson ρ coefficients.

We observe that almost all the correlations are not statistically significant. The only exceptions is the Gender Richness index (p=0.028) whose correlation index is $\rho = 0.307$, that can be considered a moderate strength correlation.

Table 4: Correlations of Individual Diversity Indexes values and Team Score

Feature	Index	Spearman ρ	p-value
Career	Blau	0.205	0.149
Career	Richness	0.191	0.179
Career	Shannon	0.205	0.149
CoPresence	Blau	-0.148	0.300
CoPresence	Richness	-0.081	0.573
CoPresence	Shannon	-0.145	0.309
Gender	Blau	0.002	0.989
Gender	Richness	0.307	**0.028***
Gender	Shannon	0.004	0.978
Nationality	Blau	-0.160	0.263
Nationality	Richness	-0.061	0.672
Nationality	Shannon	-0.164	0.251

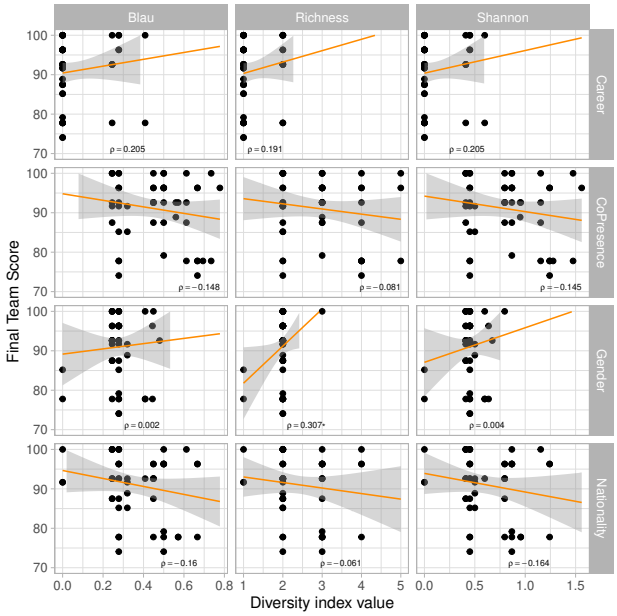


Figure 5: Correlations of Diversity and Team Score for different DI and factors

We also note that, although not significant, all correlations those involving Gender and Career are positive. While Nationality and Co-Presence are negative.

4.4 RQ2 - Combined Relative Importance

Finally, we evaluate the combined influence of the different diversity factors, as well as the additional co-factors that we monitored, using a linear model. We selected one index for each diversity feature based on the strengths of their individual correlations. Table 5 reports the estimated values of the coefficients for all the factors. In addition, it reports the results of an ANOVA test.

Table 5: Comprehensive linear model of diversity factors and additional co-factors.

Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	32.327	37.535	0.861	0.394
Skill	38.888	35.541	1.094	0.280
Gender Richness	11.185	5.248	2.131	**0.039**
Career Shannon	5.918	6.777	0.873	0.388
Nationality Shannon	2.910	7.071	0.412	0.683
CoPresence Shannon	-7.219	5.733	-1.259	0.215
size	1.614	2.282	0.707	0.483
AY2022_23	4.888	4.850	1.008	0.319
AY2023_24	-1.419	3.514	-0.404	0.688

From the overall model, we have the confirmation of only one factor for which we observed a significant correlation, that is the Gender measured with the Richness index. We use the model to understand how different diversity factors interplay with each other and with the co-factors.

We can roughly compare the different coefficients to each other since the different diversity indexes as well as the Skill have a similar range. Considering all the considered factors and co-factors we observe that the largest coefficient is linked to Skill which accounts for roughly 40% of the team grade (expressed in a 0-100 range). The second largest coefficient corresponds to Gender diversity, which positively affects team performance. Career diversity and Nationality diversity have a small positive effect on the team performance (5.92 and 2.91). The only factor having a negative absolute value for the estimate (-7.22) is CoPresence – i.e. the combined mix of Gender and Nationality within a team.

In addition, we observe that the contextual factor academic year influences the team grade, and that team size has a positive effect on the team performance, even though we cannot compare it directly to the other indexes since it has a value ranging from 5 to 7.

5 Discussion

The findings presented in the previous section allow us to address the three research questions.

5.1 Answers to Research Questions

We observed a general consistency and high correlation among the considered diversity indexes (RQ0). Nevertheless, we found a small inconsistency of order. While it can be considered practically irrelevant for our study, it represents a reminder that even the apparently simple question about what is more or less diverse does not have a unique answer. In general, we emphasize that critical adoption of numerical indexes carries the risk of hiding complexity behind the surface of presumed objectivity.

The main objective of this study was an investigation of the effects of different diversity aspects on the quality of the project works performed by teams in a Software Engineering course focused on agile software development (RQ1). Through statistical analysis, we found that gender diversity has a moderate-sized correlation

with the final outcome of the project (statistically significant). This outcome is in agreement with related studies in the literature.

The Nationality Diversity aspect had instead a negative but negligible (non-significant p-value, very far from the 0.05 selected threshold) effect on the results of the team projects. The non-significance of the results is deemed as an important result in demonstrating that promoting nationality diversity within groups has no relevant negative effects on the final results and grades obtained by students in a master's course. Thus, the diversity aspect shall not be sacrificed in favour of a tendency of students to form teams of a single nationality, which we still witness when teams are not defined by the teachers.

In addition to the effect of nationality on the project results, we have also evaluated the effects of the *Career* factor, regarding the fact that a foreign student was currently in an Erasmus programme (i.e., for a semester). Also in this case, we found a positive yet non-significant effect. Even though finer analysis would be required to investigate this phenomenon, we speculate that this aspect might be related to some pre-selection aspects inherent in the Erasmus programme selection procedures, i.e., there might be a dependency between the Career diversity and the Skill diversity.

Finally, the CoPresence that represents the intersection of Gender and Nationality has a negative, although not statistically significant effect.

When considering a combined model of the independent variables (RQ2) things change from considering the different factors in isolation. In this case, the effects (positive and negative respectively) of gender is confirmed. By contrast, in this case, the Nationality Diversity has a positive effect on final score. In the combined model, whilst both gender and nationality have a positive effect if considered individually, co-present diversity (the combination of gender and nationality within a team) had instead a negative effect on the group. This leads us to speculate that despite the isolated benefits of gender and nationality diversity, their co-presence within a team can lead to exacerbation of misunderstandings due to higher communication barriers and different cultural norms related to gender roles and interpersonal communication. All these findings are in line with the corpus of research in Global Software Engineering [25].

Overall, as expected, the most impacting factor on the final project results and scores was the skill level of the students, although not statistically significant. The lack of significance is probably due to an approximate measurement of skill, through a self-reporting questionnaire at the beginning of the course.

As a qualitative experience report from the studied three editions of the course, we witnessed the occurrence of a few issues between team members and scarce integration of foreign students within groups. Specifically, these issues happened for four groups in a.y. 2021-2022, three groups in 2022-2023, and four groups in 2023-2024. Two of these conflicts involved students in the Erasmus programme. All such criticalities were taken into account during meetings with the groups, which were encouraged to provide a more inclusive support to team members by keeping into account possible difficulties in communication or discrepancies in the respective skill sets. All the groups involved proved receptive to the suggestions.In the same period, we did not witness similar criticalities related to the other considered diversity aspects.

Overall we believe that the negative – albeit very limited – consequences of higher diversity mostly stem from the difficulty in collaborating with people coming from a different – although within the same discipline – background. We think that, from an educational perspective, exposing the student to different and challenging situations can have beneficial effects. Especially in preparing them to face circumstances that would likely happen in work settings, but in a relatively safe learning context.

5.2 Threats to Validity and Limitations

One potential threat to internal validity in this study is the selection of variables chosen to measure team diversity and its effects. This was controlled by using three different indexes: we found a general consistency and high correlation among the diversity indexes (see 4.2), with only minor order inconsistencies.

Other limitations impacting internal validity include: the short time span of the course (one semester, with the most relevant activities consisting of four two-weeks Scrum prints); the relatively small number of teams observed (51); and borderline statistical significance achieved.

Concerning construct validity, the primary threat arises from the data collection methods: evaluation of teams is partially based on self-reporting data on gender and skills by students.

As far as conclusion validity is concerned, we stress the fact that we adopted a fairly permissive value of α since our is an exploratory study. The low p-value for the correlation of Team Final Score and Gender Richness (0.028) is mainly related to the presence of a single person in the sample who preferred not to disclose their gender, when removing the relative team the resulting p-value (0,099) would still be significant.

Finally, a significant threat to external validity is the limited applicability of the findings to other contexts. The results of this study may not be generalizable beyond the specific course or educational setting in which the research was conducted. Factors such as the unique characteristics of the student population, the specific Agile methodologies used (Scrum), and the cultural context may limit the broader applicability of the findings.

An inherent limitation of the study in its current state is the lack of an analysis of other forms of diversity (e.g., age, sexual orientation, disabilities). For age and disability, the data at hand was incomplete. Regarding disability, we only had a single data point with such diversity which was not sufficient to perform an analysis on that aspect. Additionally, the analysis of the diversity in nationality (i.e., country of origin) can be considered as a proxy but not directly mappable to the ethnic diversity of the students. Future studies may involve the administration of questionnaires to capture the distribution of these diversities over the student sample, in order also to capture the concept of Intersectionality as established in the literature [5].

5.3 Comparison with related work

The work detailed in this paper is in line with several analyses performed in the literature, even though to the best of our knowledge it is the first investigation of DEI topics in advanced Software

Engineering courses revolving around the Agile philosophy of development, and the first research item quantitatively assessing the impact of intra-team diversity on team project results.

Hyrinsalmi et al. [10] investigated how Finnish universities address diversity and inclusion (D&I) in software engineering education through a survey of 30 faculty members. While the study acknowledges the rising importance of D&I in industry and academia, it finds that current teaching methods need improvement, particularly in raising awareness among educators. The focus is on identifying common practices, challenges, and opportunities for embedding D&I into curricula. The emphasis of the work lies on institutional strategies rather than practical, team-based educational activities.

A second work by Hyrinsalmi et al. [11] extends the previous study by presenting a framework for integrating D&I into software engineering courses. Based on the same survey data, the paper explores how educators perceive D&I, identifies implementation barriers, and offers curriculum-level solutions.

Gama et al. [7], describe a graduate-level course that addresses various diversity dimensions, including race, disability, and neurodivergence. The course uses both user-centered and team-centered perspectives, particularly in the context of software requirements and human aspects of software engineering. Teamwork is a fundamental component, with students collaborating remotely through digital platforms. The main results of the paper regard the students' awareness in D&I themes and the perceived safety during teamworks, with no negative results gathered.

Murphy et al. [14] explored how reading assignments on D&I topics were integrated into a technically focused undergraduate course. Spanning three semesters and involving nearly 600 students, the intervention aimed to increase awareness of diverse identities and the importance of inclusivity in technology development. The study focuses on individual reflection and class discussions rather than structured teamwork or agile practices.

6 Conclusions and future work

We analyzed 51 teams distributed over three academic years, measuring their diversity – Gender, Nationality and their co-presence, as well as career diversity – through three distinct indexes.

Through our preliminary analysis of the distinct diversity indexes frequently adopted in the literature, we observe a generally good correlation among the three considered team diversity indexes though a case of inconsistency emerged, thus this is an aspect that deserves further inquiry.

We observe a mild positive correlation between Gender diversity with team performance. All other individual diversity indexes have negligible correlations with the team performance.

Considering the combined effects of various diversity aspects we notice that team gender and country diversity may play a positive effect on performance while their co-presence in a team, as well as Career diversities, play a negative –but not statistically significant– effect.

Based on the statistical results and on the analysis of the outcome of the course, as actionable points for educators, we suggest:

- Encouraging the formation of teams with high diversity for different aspects, since inter-team diversity has no statistically significant negative effect on the outcomes of the team projects.
- If possible, allocating the teams to the group in a way finalized to maximize diversity, since an autonomous composition of the teams by the students themselves tends to minimize it.
- Maintaining an active line of support in case intra-group communication issues arise.

The study detailed in this paper was based solely on quantitative analyses of the results obtained by the teams at the end of the course. Starting from the 2024-2025 edition of the course, we plan to gather – through closed- or open-ended questionnaires administered to the students – qualitative and subjective opinions about their experience, and the overall safety and comfort perceived during their activities in their team.

References

[1] American Psychological Association and others. 2015. Guidelines for psychological practice with transgender and gender nonconforming people. *American psychologist* 70, 9 (2015), 832–864.

[2] Neil Anderson, Aidan McGowan, Leo Galway, Matthew Collins, and Philip Hanna. 2024. Learning to Improve Gender Equality: An Analysis of Software Engineering Education in a UK University. In *Proceedings of the 5th ACM/IEEE Workshop on Gender Equality, Diversity, and Inclusion in Software Engineering*. 37–44.

[3] Peter M. Blau. 1977. *Inequality and heterogeneity: A primitive theory of social structure*. Free Press.

[4] Bhaskar Chakravorti. 2020. To Increase Diversity, U.S. Tech Companies Need to Follow the Talent. *Harvard Business Review* (2020). https://hbr.org/2020/12/to-increase-diversity-u-s-tech-companies-need-to-follow-the-talent

[5] Kimberlé Crenshaw. 2013. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In *Feminist legal theories*. Routledge, 23–51.

[6] Claudia Maria Cutrupi, Irene Zanardi, and Letizia Jaccheri. 2024. Draw a Software Engineer Test-Preliminary Attempts to Investigate University Students' Perceptions of Software Engineering Professions. In *Proceedings of the 5th ACM/IEEE Workshop on Gender Equality, Diversity, and Inclusion in Software Engineering*. 45–46.

[7] Kiev Gama and Reydne Santos. 2024. It's not all about gender: A Multidimensional Course Perspective on Diversity and Inclusion in Software Engineering Education. In *Simpósio Brasileiro de Engenharia de Software (SBES)*. SBC, 487–498.

[8] Emitzá Guzmán, Ricarda Anna-Lena Fischer, and Janey Kok. 2023. Mind the gap: gender, micro-inequities and barriers in software development. *Empirical Software Engineering* 29, 1 (Dec. 2023). doi:10.1007/s10664-023-10379-8

[9] Lucia Happe, Kai Marquardt, Ricarda Trumpf, and Ingo Wagner. 2024. Decoding the Gap: A Retrospective Analysis of Women's Experiences in Software Engineering. In *Proceedings of the 5th ACM/IEEE Workshop on Gender Equality, Diversity, and Inclusion in Software Engineering*. 27–28.

[10] Sonja M Hyrynsalmi. 2023. How Diversity and Inclusion Are Approached in Software Engineering University-level Teaching. In *2023 IEEE/ACM 4th Workshop on Gender Equity, Diversity, and Inclusion in Software Engineering (GEICSE)*. IEEE, 17–24.

[11] Sonja M Hyrynsalmi. 2024. Challenges and opportunities: Implementing diversity and inclusion in software engineering university level education in Finland. *Journal of Systems and Software* (2024), 112239.

[12] Sonja M Hyrynsalmi, Ella Peltonen, Fanny Vainionpää, and Sami Hyrynsalmi. 2024. The Second Round: Diverse Paths Towards Software Engineering. In *Proceedings of the 5th ACM/IEEE Workshop on Gender Equality, Diversity, and Inclusion in Software Engineering*. 29–36.

[13] Yekaterina Kovaleva, Ari Happonen, and Eneli Kindsiko. 2022. Designing gender-neutral software engineering program. stereotypes, social pressure, and current attitudes based on recent studies. In *Proceedings of the Third Workshop on Gender Equality, Diversity, and Inclusion in Software Engineering*. 43–50.

[14] Christian Murphy, Anya Mushakevich, and Yunha Park. 2021. Incorporating readings on diversity and inclusion into a traditional software engineering course. In *2021 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*. IEEE, 1–5.

[15] Anh Nguyen-Duc and Letizia Jaccheri. 2023. Gender Equality in Software Engineering Education – A Study of Female Participation in Customer-Driven Projects. In *Sustainability in Software Engineering and Business Information Management*, Varun Gupta, Luis Rubalcaba, Chetna Gupta, and Thomas Hanne (Eds.). Springer International Publishing, Cham, 39–49.

[16] Marco Ortu, Giuseppe Destefanis, Steve Counsell, Stephen Swift, Roberto Tonelli, and Michele Marchesi. 2017. How diverse is your team? Investigating gender and nationality diversity in GitHub teams. *Journal of Software Engineering Research and Development* 5, 1 (2017), 9. doi:10.1186/s40411-017-0044-y

[17] Manuela Andreea Petrescu, Simona Motogna, and Liviu Berciu. 2023. Women in Scrum Master Role: Challenges and Opportunities. In *2023 IEEE/ACM 4th Workshop on Gender Equity, Diversity, and Inclusion in Software Engineering (GEICSE)*. IEEE, 49–55.

[18] Politecnico di Torino. 2024. *Diversità è Cambiamento - Bilancio di Genere 2023*. https://www.polito.it/sites/default/files/2024-01/BDG_2023_DEF.pdf

[19] Gema Rodríguez-Pérez, Reza Nadri, and Meiyappan Nagappan. 2021. Perceived diversity in software engineering: a systematic literature review. *Empirical Software Engineering* 26 (2021), 1–38.

[20] Mary Sánchez-Gordón and Ricardo Colomo-Palacios. 2024. On the Intersectionality of Software Practitioners and Role Models. In *Proceedings of the 5th ACM/IEEE Workshop on Gender Equality, Diversity, and Inclusion in Software Engineering* (Lisbon, Portugal) *(GE@ICSE '24)*. Association for Computing Machinery, New York, NY, USA, 22–26. doi:10.1145/3643785.3648483

[21] Ian F. Spellerberg and Peter J. Fedor. 2003. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. *Global Ecology and Biogeography* 12, 3 (2003), 177–179. doi:10.1046/j.1466-822X.2003.00015.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1466-822X.2003.00015.x

[22] Günter K Stahl, Martha L Maznevski, Andreas Voigt, and Karsten Jonsen. 2010. Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies* 41, 4 (2010), 690–709. doi:10.1057/jibs.2009.85

[23] Anna Szlavi, Marit Fredrikke Hansen, Sandra Helen Husnes, and Tayana Uchôa Conte. 2023. Intersectionality in Computer Science: A Systematic Literature Review. In *2023 IEEE/ACM 4th Workshop on Gender Equity, Diversity, and Inclusion in Software Engineering (GEICSE)*. 9–16. doi:10.1109/GEICSE59319.2023.00006

[24] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark GJ van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. 2015. Gender and tenure diversity in GitHub teams. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3789–3798.

[25] Shujian Wu. 2012. Overview of communication in global software development process. In *Proceedings of 2012 IEEE International Conference on Service Operations and Logistics, and Informatics*. IEEE, 474–478.

[26] Xin Zhao and Riley Young. 2023. Workplace Discrimination in Software Engineering: Where We Stand Today. In *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, 188–193.