Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #47018

Exploring Collaborative Success of Undergraduate Students within a Cyberlearning Environment Using Gamification

Mr. Gabriel Eduardo Prieto, Florida International University

Gabriel E. Prieto received his BS in Computer Science from Florida International University (FIU) in 2022, and MS degree in Data Science & Artificial Intelligence from FIU in 2024. He is currently the lead iOS developer on an NIH-funded project and anticipates joining a startup company that will leverage the technology developed through this project.

Neila Bennamane, Florida International University

Neila has a Bachelors of Science in Mathematical Sciences and is now pursuing a PhD in Computer Science at Florida International University. Her research ambitions are strongly driven by a desire to bridge the accessibility gap through technological advancements.

Agoritsa Polyzou, Florida International University

Agoritsa Polyzou is an Assistant Professor at the Knight Foundation School of Computing and Information Sciences in Florida International University (FIU), Miami. Agoritsa received the bachelor's degree in computer engineering and informatics from the University of Patras, Greece, and her Ph.D. degree in computer science and engineering from the University of Minnesota. Next, she was a Postdoctoral Fritz Family Fellow with the Massive Data Institute of McCourt School of Public Policy at Georgetown University, Washington, DC. She is involved in projects in the intersection of education, data mining, machine learning, ethics, and fairness. Her research interests include data mining, recommender systems, predictive models within educational contexts, and the fairness concerns that arise from their use. Her goal is to help students succeed using data and machine learning models.

Dr. Peter J Clarke, Florida International University

Peter J. Clarke received his B.Sc. degree in Computer Science and Mathematics from the University of the West Indies (Cave Hill) in 1987, M.S. degree from SUNY Binghamton University in 1996 and Ph.D. in Computer Science from Clemson University in 2003. His research interests are in the areas of software engineering, software testing, model-driven software development, and computer science education. He is currently an associate professor in the Knight Foundation School of Computing and Information Sciences at Florida International University. He is a member of the ACM (SIGSOFT, SIGCSE, and SIGAPP); IEEE Computer Society; and a member of the Association for Software Testing (AST).

Exploring Collaborative Success of Undergraduate Students within a Cyberlearning Environment Using Gamification

Abstract

In higher education, particularly at the undergraduate level, active learning has become an increasingly important pedagogical approach. It encompasses approaches including collaborative learning and gamification aiming to enhance student engagement, promote collaboration, and improve learning outcomes. However, the effectiveness of these approaches can vary significantly across different courses and student populations. This variation is due to the manner in which the approaches are implemented, such as the specific class activities, the modality of the class, and the pedagogical tools used to support active learning.

In that direction, the Software Engineering and Programming CyberLearning Environment (SEP-CyLE) is a tool designed to provide students and instructors structured access to vetted learning content through learning objects and tutorials. It also employs three active learning approaches – a lightweight version of collaborative learning (team-based activities), gamification, and social interaction; and can be used in any class independent of the modality. Based on students' interaction with SEP-CyLE and the assigned activities, they are awarded virtual points that may be converted to points used towards their course grade. This tool has been used in software engineering, software testing, and programming classes at a large urban Hispanic-serving institution.

While the related work recognizes the need and benefits of collaborative learning and gamification, there is a gap in the analysis of its effects on student learning, especially while it is facilitated by a cyberlearning tool. This paper addresses the challenge of quantifying and analyzing student collaboration and engagement in the context of SEP-CyLE. Our research aims to identify distinct student engagement profiles and their correlation with academic performance. We utilize behaviorally based metrics collected through SEP-CyLE for 12 semesters, starting in the spring of 2017 until the fall of 2020, for a software testing class taught by the same instructor. The data collected for each student includes their SEP-CyLE interactions (the total virtual points awarded, time spent on learning objects, team performance, and number of comments posted), team assignments, and course grade data. The insights gained from this study may lead to more personalized learning experiences, enhancing student satisfaction and success rates in team-based learning environments.

1 Introduction

In recent years, there has been a significant focus on incorporating active learning strategies into higher education, in particular in science, technology, engineering, and mathematics (STEM) fields, and assessing their value and impact on student well-being [16]. The goal of active learning is to link the "activity" with "learning". It is a broader educational strategy that encompasses many learning tactics. Many of these involve the notion of teamwork; students are placed in teams to discuss a topic, solve a problem, explore alternative solutions, or complete an assignment. These strategies offer numerous benefits to students, such as improved academic performance and enhanced physical, emotional, and social life [6]. Additionally, active learning also empowers students with multi-competencies, as students develop essential collaborative skills for their professional future. Research shows that, when students work together, they experience enhanced learning [12], motivation, a sense of community [29], and overall educational satisfaction [12]. For students from underrepresented minority groups, these benefits are even greater [23, 14].

However, conventional approaches to team formation often fail to create optimal learning environments, frequently resulting in uneven participation and suboptimal team dynamics. Appropriate team composition is one of the key factors of a successful team [21]. This challenge is particularly evident in software engineering education, where effective collaboration is crucial for student success and professional preparation.

Modern educational environments generate substantial data on student interactions, performance metrics, and engagement patterns. This wealth of information presents an opportunity to take advantage of advanced data analysis techniques to optimize team composition and improve learning outcomes. Although research has explored various aspects of collaborative learning and gamification, there is room to further investigate how data-driven approaches can enhance our understanding of team dynamics and improve educational outcomes.

This study investigates **the relationship between teamwork dynamics and academic performance** by analyzing patterns of student collaboration across multiple semesters in a Software Testing course. Utilizing advanced data analysis and clustering techniques on educational data from the Software Engineering and Programming CyberLearning Environment (SEP-CyLE) database, we examine how collaborative interactions evolve and impact student achievement. Our research applies multiple clustering algorithms (K-means, Hierarchical, and DBSCAN) to explore how student behaviors manifest and transform over time. SEP-CyLE [3, 30] is a tool designed to provide students and instructors structured access to vetted learning content through learning objects and tutorials. It also employs three active learning approaches – a lightweight version of collaborative learning (team-based activities), gamification, and social interaction; and can be used in any class independent of the modality.

The primary objectives of this research are to: (1) examine how students' course grades relate to their team characteristics, and (2) investigate how student engagement profiles in SEP-CyLE correlate to course performance. By analyzing comprehensive data collected over 12 semesters, including metrics such as team assignments, SEP-CyLE participation data, and academic performance, our aim is to uncover insights into the relationship between teamwork dynamics and student achievement. Through various data analysis techniques, including clustering algorithms, we explore patterns in team behavior and collaboration that could inform our understanding of effective team-based learning in software engineering education.

Our findings contribute to the growing body of knowledge on collaborative learning by providing empirical evidence for the relationship between team composition and academic outcomes. The results of this study have practical implications for educational institutions seeking to enhance their team-based learning approaches and may inform the development of more sophisticated team formation methodologies in academic settings.

2 Related Work

Active Learning Approaches. Based on the work by Clarke et al. [4], LESs are a subset of active learning approaches that include collaborative learning, gamification, problem-based learning, and social interaction. *Collaborative learning* is where two or more people work in groups mutually searching for understanding, solutions, meanings, or creating a product [18]. *Gamification* uses game design elements and game mechanics to improve user experience and engagement with a system, which may be applied to an educational context [5]. Game design elements include leaderboards, a points system, and levels. *Problem-based learning (PBL)* is an approach to learning and instruction in which students tackle problems in small groups under the supervision of a tutor [17]. These problems aim to prepare students for real-world settings [28]. *Social interaction* is an approach that enhances knowledge acquisition through social activities, such as students establishing meaningful dialogue within student groups and with teachers [8, 11].

Study of Learning and Engagement Strategies (LESs). Clarke et al. [4] described LESs and compared them to the traditional lecture-style approach. The authors also performed a study that quantifies the increased use of LESs in face-to-face (F2F) class activities to determine student learning improvement between Exams 1 and 2 for a software testing class. In the control group, the use of LESs was minimized, while the treatment group had increased LESs. The exams for both groups were very similar or exactly the same. A statistical analysis of the results using the Mann-Whitney U test showed a statistically significant difference between the groups. The Exam 2 scores for the control (minimal LESs) group (Mdn = 72.4%) differed significantly from the treatment (increased LESs) group (Mdn = 77.7%), where $U = 2421$, $z = 2.875$, $p = 0.004$, $r = 0.26$. This work provides evidence that using LESs or other active learning approaches has consistently improved student learning outcomes, as reflected by the exam scores.

Benefits of Teamwork. Teamwork is essential for student development in terms of knowledge and understanding of key STEM concepts [27]. Teamwork and active learning have many learning benefits [6]. It enables students to learn with and from each other, creating a sense of belonging to a learning community [2]. In this way, students are exposed to diverse perspectives. Students feel they have a shared purpose, which can improve morale and increase motivation and student retention. When students work in groups, they often develop a deeper understanding of course material [2]. Additionally, teamwork has the potential to increase students' self-confidence and responsibility. This is a self-sustaining cycle; team activities teach students how to be better communicators, which in turn helps every member of the team feel valued and respected, which in turn, further encourages them to participate in team activities [13]. At the same time, it requires various skills to be successful, i.e., soft skills, communication, decision-making, problem-solving, critical thinking, and providing feedback [15, 24], all of which are considered the skills of the 21st century [7, 25].

Challenges Most research focused on evaluating the effectiveness of these instructional strategies by comparing them with traditional approaches and students' outcomes. However, there is a gap related to analyzing team characteristics and how they can affect student outcomes and benefits. A reason for that might be that teamwork skills are hard to assess. Towards that direction, Britton et al. developed a tool to evaluate individual teamwork skills [1]. In another example, Jaiswal et al. defined collective team orientation as the team's ability to efficiently use all teamwork skills to attain the desired objective [10]. The study found that teams with higher team orientation performed better. From the few works that delve deeper into teamwork characteristics, most of them are focusing a few characteristics. In our work, we will examine *multiple aspects of teamwork* in conjunction with *the use of the SEP-CyLE* tool, including the evolution of *student behaviors over time*, using a big dataset that includes teams formed over 12 semesters.

3 Background and Context

In this section, we describe the context of the learning environment in which the data for the study was collected. The data collected was during a software testing class, CEN4072 Fundamentals of Software Testing, where the students used SEP-CyLE to complete extra credit assignments.

3.1 Course Context

The *CEN4072 Fundamentals of Software Testing* course may be taken by students in their junior or senior years during their undergraduate degree program. Based on the course catalog description, the CEN4072 course covers test plan creation, test case generation, program inspections, specification-based and implementation-based testing, and testing tools. The course grade is based on three exams (two midterms, 25% each, and a final, 20%), a group project (25%), and attendance at class (5%). Students may also receive *extra credit* based on in-class activities, a maximum of 2.5% for each exam, and a maximum of 3% based on the assignments completed in SEP-CyLE [4].

The semester project, in-class learning activities, and the assignments in SEP-CyLE, in part, are completed in teams. At the beginning of each semester, students are placed into teams using a round-robin assignment based on the alphabetical order of names in the class list. The initial team assignments are the same throughout the semester. In special cases, teams may need to be reassigned when students drop the class. After the initial assignment of the teams, there is an attempt to ensure that each team has at least one female student. Each team self-organizes team members into three administrative roles as follows. *Team Leader* - manages the overall team activities, including scheduling team meetings, assigning tasks to team members, and contacting the instructor to schedule meetings. *Minute Taker* - records the activities of each team meeting using a predefined diary entry consisting of the meeting date, time, location, duration, members in attendance, agenda, a summary of the discussion, and assigned tasks to the team members. *Time Keeper* - keeps track of the time allocation for the agenda items.

The extra credit for SEP-CyLE assignments is based on completing learning objects (LOs). Smith defines an LO as *any grouping of materials that is structured in a meaningful way and is tied to an educational objective* [19]. In our context, we consider an LO to be a chunk of learning content that consists of a learning objective, content on the specific topic, practice assessment, recorded assessment, and a list of references. A learner should complete an LO within 5 to 15 minutes. The content on an LO may comprise documents, pictures, simulations, movies, sounds, animations, and so on. As students complete individual LOs, they are awarded virtual points on an individual and team basis. Below are the guidelines for awarding the virtual points per LO completed in the CEN4072 course.

- *Team Basis*: Each member of the team gets the following for completing the quiz with 70% or higher in each LO:
  - First place team: 7 virtual points
  - Second place team: 4 virtual points
  - Third place team: 2 virtual points
  - If all members of the team complete the LO with 70% or higher, the entire team gets 3 virtual points.
- *Individual Basis*:
  - Completed the LO quiz with $>=$ 70%: 5 virtual points
  - Wrote on the discussion board: 2 virtual points
  - Uploaded a picture: 2 virtual points

Additional details on how the LOs were used in the CEN4072 class may be found in [4].

3.2 Cyberlearning Tool

The Software Engineering and Programming CyberLearning Environment (SEP-CyLE) is a tool that provides students and instructors with vetted learning content in the form of learning objects (LOs) and tutorials. This learning content is presented in the context of the LES collaborative learning, gamification, and social interaction. Tutorials are similar to LOs except there is no assessment component, and the tutorials are mainly used to describe tools used in the class, e.g., testing tools such as JUnit [22] or Selenium [20]. Fig. 1 shows a block diagram of the key components of SEP-CyLE. The key components of SEP-CyLE are shown in Fig. 1 and described below.

- *User Management* - provides basic user operations, including logging in/out, changing passwords, creating users and assigning them roles, creating/updating user profiles, and anonymizes student personal data.
- *Course Management* - manages operations related to creating course templates and instances, assigning LOs and tutorials, generating grade reports, uploading class rolls, and assigning students to teams, among others.
- *LESs* - provides the activities needed to support collaborative learning, gamification, and social interaction. In SEP-CyLE, we use a lightweight version of collaborative learning where students benefit from being part of a team regarding how the bonus virtual points are awarded. The gamification LES included a leaderboard and a virtual points system. The social interaction LES allows users to post to the discussion board and comment on the LOs and tutorials being assigned.
- *LC Repository* - contains the learning content (LC) available to course instances in SEP-CyLE. LC includes 40 LOs and 9 tool tutorials [9]. The LOs are in the areas of introduction to programming (CS1), cybersecurity (CSY), software engineering (SWE), and software testing (SWT). The tool tutorials are for software testing (TT).
- *LC Creator* - is used to create the LC for the LOs [19] and tutorials.
- *Data Analytics* - compiles various data for the students and teams in the course instances. This data includes virtual points collected for completing various activities, e.g., completing

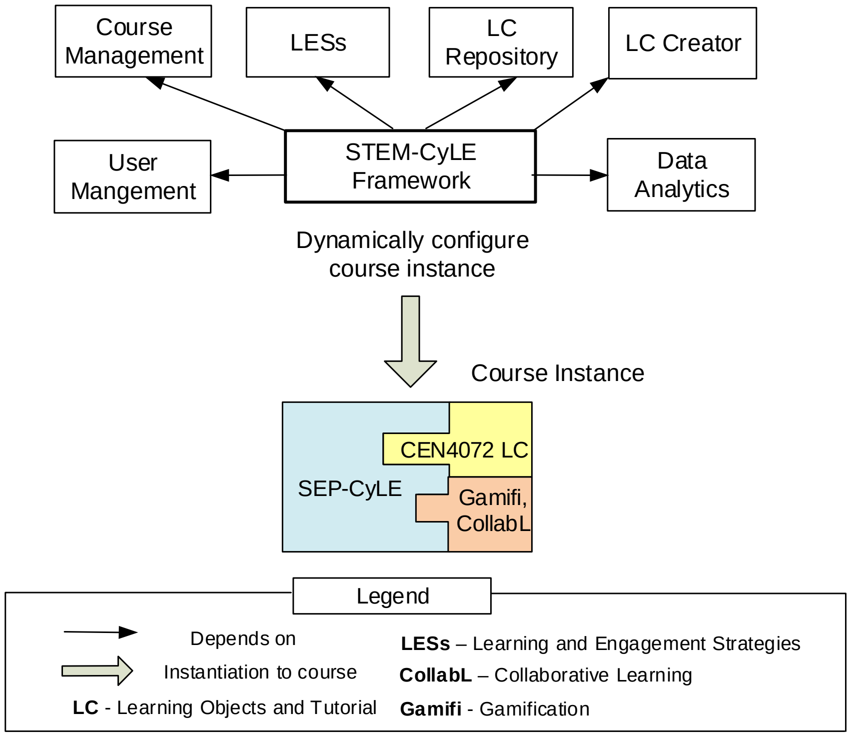


Figure 1: Block diagram of the Software Engineering and Programming CyberLearning Environment (SEP-CyLE), an instance of STEM-CyLE.

learning objects.

4 Research Questions

This paper presents a data-driven analysis of teamwork in the context of SEP-CyLE. Our study investigates the relationship between teamwork dynamics and academic performance by analyzing patterns of student collaboration across 12 semesters in a Software Testing course. We plan to answer the following research questions.

RQ1 How do students' course grades relate to their team characteristics?

RQ2 How does the student engagement profile in SEP-CyLE correlate to the students' performance in the course?

To answer these questions, we use various data analysis techniques to investigate the relationship between teamwork dynamics and academic performance, the collaborative interactions within a team. We also use clustering techniques to reveal any hidden patterns in team behavior. After our analysis, we expect to better understand the team collaboration mechanisms used in SEP-CyLE, identify improved team formation strategies, and potentially a pedagogical approach that leads to better student learning outcomes.

Table 1: Course data for CEN4072 from Spring 2017 to Fall 2020. Spr - Spring, Sum - Summer, Fa - Fall.

Year	Sem.	# Students	# Teams	Sem.	# Students	# Teams	Sem.	# Students	# Teams
2017	Spr	17	4	Sum	15	3	Fa	11	3
2018	Spr	19	4	Sum	14	3	Fa	20	4
2019	Spr	27	5	Sum	13	3	Fa	16	4
2020	Spr	10	2	Sum	21	4	Fa	33	6

5 Dataset

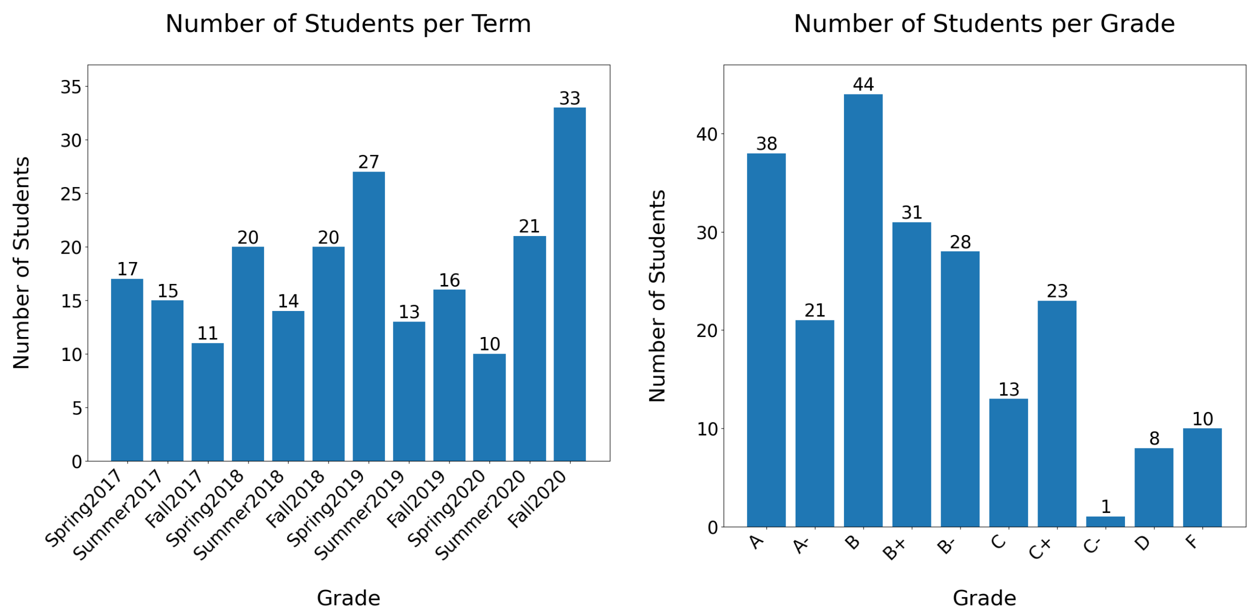
The data for the study was collected over 12 semesters of the CEN4072 course starting in Spring 2017 and ending in Fall 2020. Table 1 shows the course data related to the CEN4072 for the duration of the study. Each row of the table shows the semester (Spr - Spring, Sum - Summer, Fa - Fall), the number (#) of students, and the number (#) of teams for the years between 2017 and 2020. The total number of students in the study is 216, and the number of teams is 45.

In all the semesters shown in Table 1, students were required to complete two midterm exams. The students completed a course project consisting of two deliverables from Spring 2017 to Summer 2020. In the Fall of 2020, there was only one deliverable. The final exam was optional for the period of the study. The assignments in SEP-CyLE were available during all semesters of the study for a maximum of 3% extra credit course points. The assignment of virtual points in SEP-CyLE was described in Section 3.1. The extra credit points from SEP-CyLE were computed as a percentage ratio of the student(s) with the highest number of virtual points. The student(s) with the most virtual points received 3% extra credit course points.

The grades for each course project deliverable consist of four components: presentation (21%), demonstration (12.6%), documentation (50.4%), and peer evaluation (16%). The peer evaluation consists of the members of a team grading each other using a peer evaluation rubric provided by the instructor. The rubric includes four criteria: *Helping* - assistance provided by a team member to other team members, *Participating* - contribution and attendance by a team member at team meetings, *Questioning* - the level at which the team member interacts, discusses, or poses questions to other members, *Assigned Tasks* - the level of completion of the assigned tasks. Each criterion is rated on a scale of 1 to 4.

Data Preprocessing. The analysis utilizes a data subset that includes only students with both peer evaluation and final course grade data. We filter out any students with missing values, resulting in 217 total students and 45 teams across all terms. Fig. 2 shows the distributions of students over the different letter grades received (Fig. 2a) and over the semesters of the study (Fig. 2b). The majority of students achieve high to medium grades (A to B-), suggesting a generally positive outcome for the course. Course enrollment fluctuates through different terms, with peaks in Spring 2019 and Fall 2020.

Peer review metrics and final grades are converted to percentage values. Individual deliverable scores for Deliverable 1 and Deliverable 2 (D1 and D2) are maintained for each student. For team analysis, students' data are grouped by team ID to calculate average values for the desired feature



(a) Number of Students per Grade

(b) Number of Students per Term

Figure 2: Distribution of students per grade and per term

columns, representing team performance.

Visualizations. We use various types of plots to visualize the relationships between different aspects of the problem. In particular, when we use scatter plots, regression lines are calculated and added via least squares methods to indicate trends. These regression lines provide insights into the relationships within the data, with shaded areas around them representing the 95% confidence intervals. These intervals suggest the potential range of the true regression lines, given the data's variability. They also indicate the strength and direction of any observed correlations. Pearson's Correlation is used to calculate the correlation coefficient.

For example, consider Fig. 3. To better understand the nature of the data, this figure shows a scatter plot between the midterm 1 and 2 scores. Each dot corresponds to a student, and their midterm 1 and 2 grades are shown on the x-axis and y-axis, respectively. The correlation between Midterm 1 and Midterm 2 scores is moderate (r = 0.503), as illustrated in the graph. This correlation suggests that students' performance in these exams is somewhat consistent, yet there is room for variability which might be influenced by different exam formats, topics covered, or student preparation methods. The regression line shows a positive trend indicating that higher scores in Midterm 1 generally predict higher scores in Midterm 2, yet with some dispersion indicating other influencing factors.

6 Analysis and Results

In this section, we will explore different relationships of the data in multiple subsections to answer the two research questions in Section 4. Note that we will examine patterns at the data both at the team-level (per team), but also at the individual-level (per student).

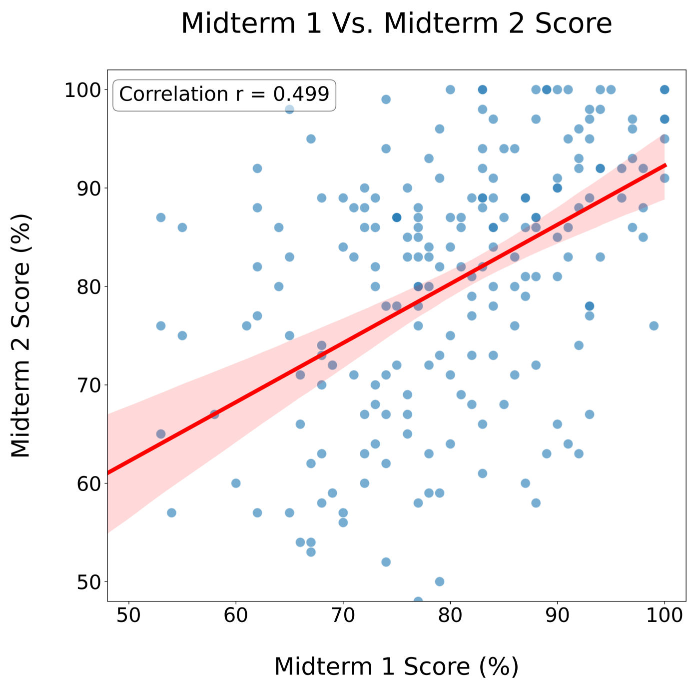


Figure 3: Midterm 1 vs. Midterm 2 Scores

6.1 Analysis - Team and Individual Performance (RQ1)

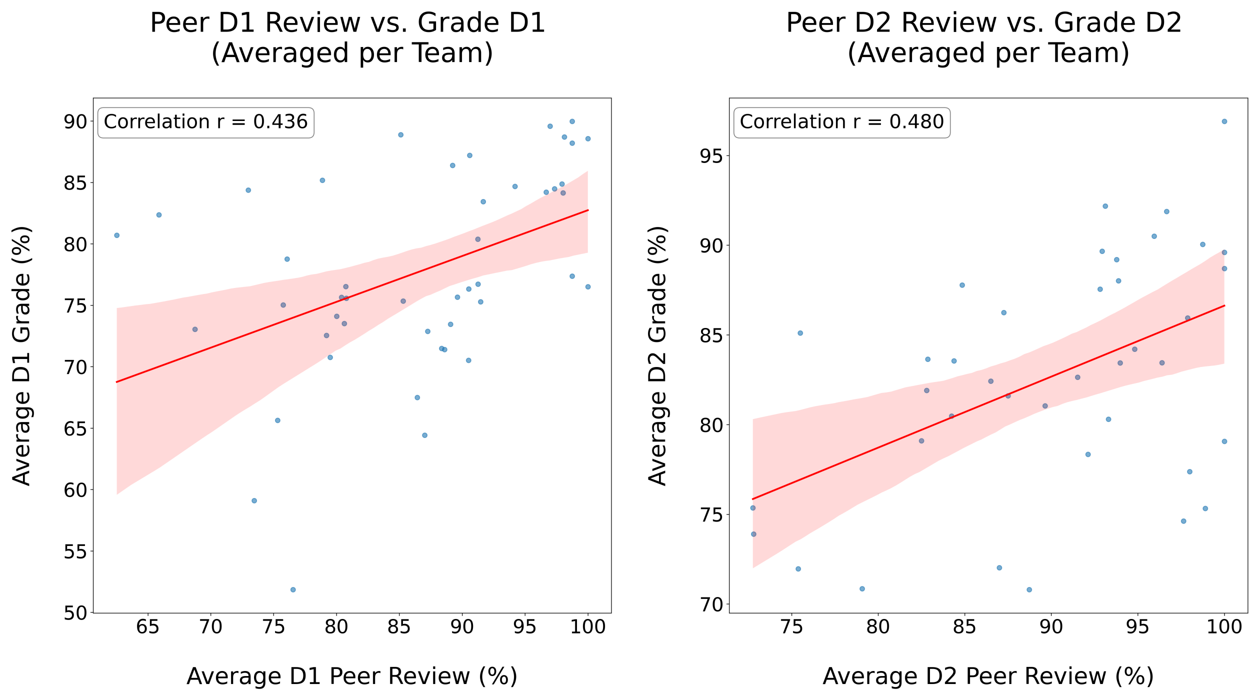
To answer RQ1, we analyze the data related to the students' teams and individual performances on project deliverables 1 and 2 and their exam grades.

6.1.1 Project performance and peer evaluations

We will start by exploring the relationship between the team performance (captured by the grades of their projects' deliverables) and students' contributions to the team (captured by the peer evaluation received from their teammates). Each student received peer evaluations from the other members of the team. We average them out so that every student has one peer review score.

To examine peer reviews at the team-level, team data was aggregated. A team's peer review grades correspond to the average grades received by its members. This summarization helps in understanding how well the team worked, on average. Well-functioning teams are expected to have high peer review scores, while teams that face challenges in their collaboration will have lower scores.

We plot that data against the grades received in each deliverable to visualize the correlation between peer assessments and project outcomes, in Fig. 4. For both deliverables in Figures 4a and 4b, the graphs show a moderate correlation, indicating that team peer reviews somewhat predict the grades received. The trend line and confidence interval suggest a positive relationship, albeit with some variance. *The better the peer reviews are (and the more satisfied the students are with their teammates), the more likely it is for teams to perform well on the corresponding deliverable.* Additionally, we notice that the range of average team peer review scores is narrower for the sec-



(a) Average Peer D1 Review vs Average Grade D1 (per Team)

(b) Average Peer D2 Review vs Average Grade D2 (per Team)

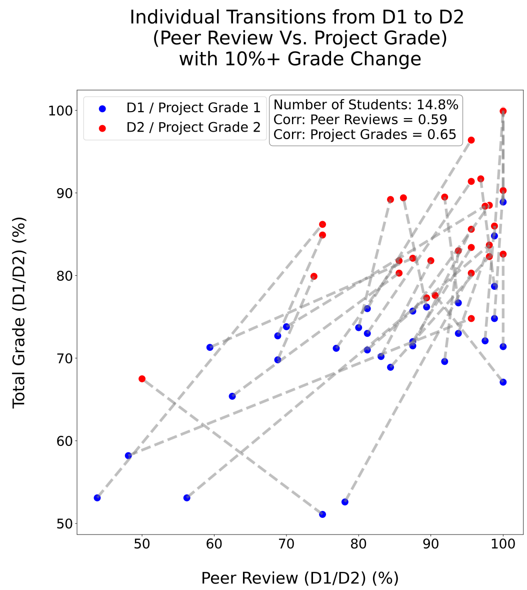
Figure 4: Comparison of Team Performance in Deliverable 1 (D1) and Deliverable 2 (D2)

ond deliverable, indicating that teams might manage to communicate and collaborate better after getting to know their teammates and the expectations of the course.

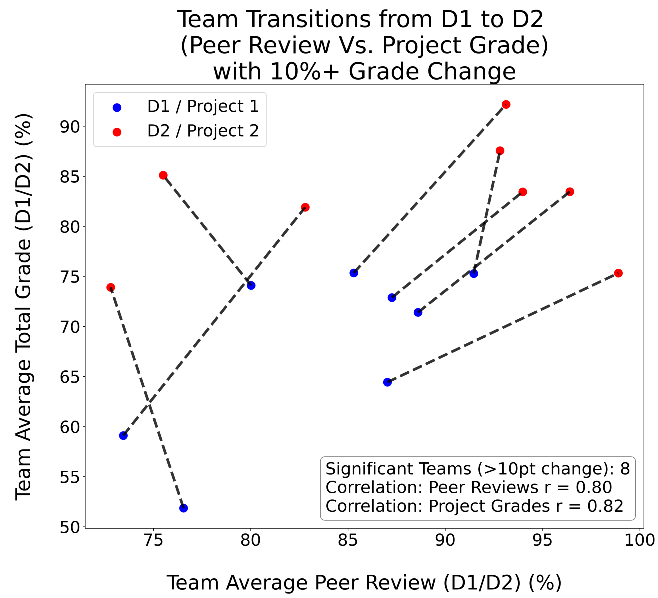
To focus on individual and team transitions from Deliverable 1 to Deliverable 2, we examine the cases with significant project grade changes ($\geq$ 10%). For this analysis, 'Grade Change' was calculated by finding the absolute difference between the grades of the two deliverables. Only instances with changes equal to or greater than 10% were included in the plots. Fig. 5 depicts the movement of each point from D1 to D2, illustrating not just the change in grades, but also the changes in peer review scores. Now we examine these changes both per team and per student. This dual-axis movement provides insight into how peer perceptions and actual grades evolve between two significant assessment points in the course.

Fig. 5a shows a variety of changes for students in grades and peer reviews. The blue dots correspond to a student's peer scores and project grade in Deliverable 1, while the red dot corresponds to Deliverable 2. We notice that *the majority of the students with a considerable change in the project grades across deliverables have a positive movement*, i.e., they receive better grades and the same or better peer reviews. As a result, the red dots are more prevalent at the top right of the plot. There are a couple of exceptions, though, which might be the result of other personal issues these particular students had to deal with during the semester. We observe a similar trend at the team level (Fig. 5b), with 6 out of the 8 teams with significant change to be teams that improved their performance and their collaboration.

For comparison, we include the graphs of all teams and all individuals (regardless of the percent



(a) Individual: Significant Transitions D1 and D2 Peer Reviews and Project Grades



(b) Teams: Significant Transitions from D1 to D2 Peer Reviews and Project Grades

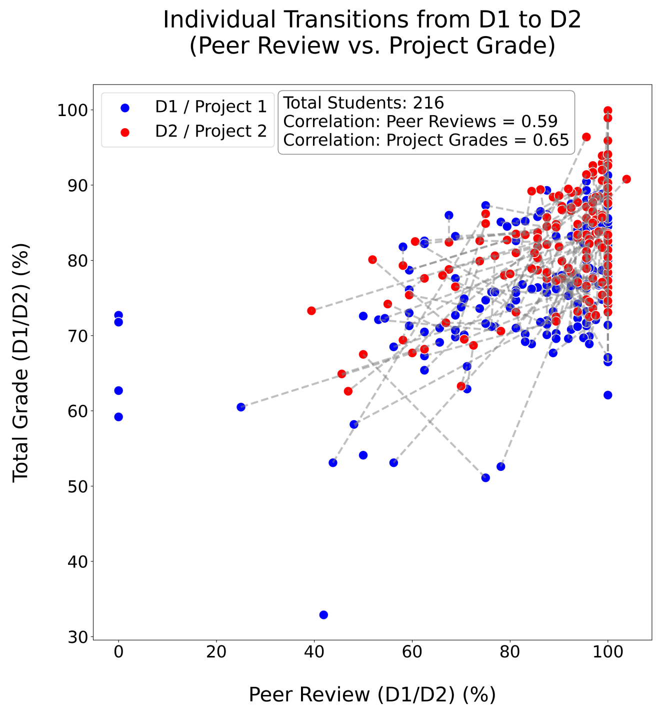
Figure 5: Comparison of Significant Peer Review and Project Grade by Teams and Individuals

change in their grade) in Fig. 6.

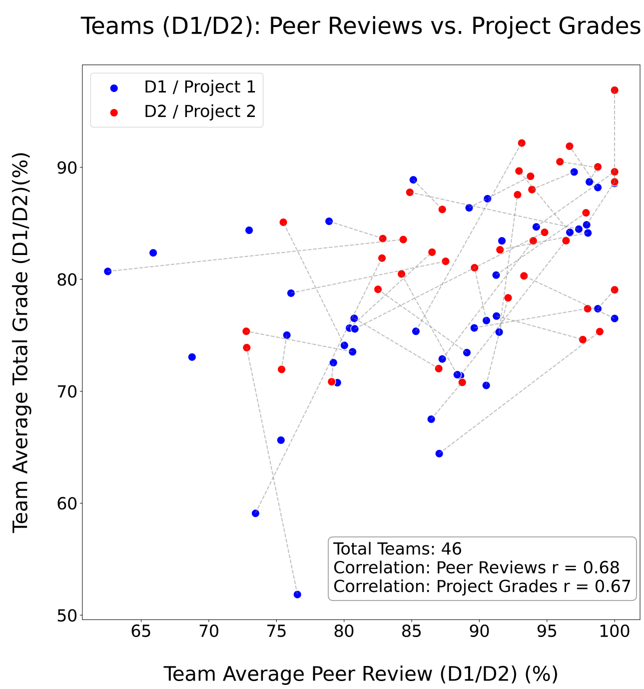
6.1.2 Exam performance and peer evaluations

The correlations observed in the previous section are expected to some degree since the Peer Reviews accounted for a small percentage (16%) of the final project grade. We will now explore the relationships between peer evaluations and the exam performance of the students. This will capture how well the student performs in the class in their own right, without the help of their teammates. Unless otherwise noted, exam performance is the average of their exam grades (midterm 1 and 2 and final exam). A team's average exam performance is the average exam performance of its members.

Fig. 7 displays the relationship between average peer review scores and exam performance. The two measures show a lower correlation (r = 0.284). This indicates that while there is a positive relationship, it is weaker than that between the project grades and peer evaluation. This suggests that peer reviews, which might reflect students' perceptions of each other's contributions and understanding, do not strongly predict exam outcomes. However, at the same time, we notice a lot of teams that have excellent average peer reviews, but a significant variance in the team's average exam grade. This can be the result of one of two phenomena, or their combination. (1) In these teams, students might not offer evaluations that reflect their true opinions; rather, they give everyone in their team a good evaluation. In that case, peer evaluations are of lower quality and they do not capture the collaboration within a team. (2) These teams might have students who are "free riders". They do not do much or gain much for the project, but they receive evaluations which are not reflective of their contributions. In both cases, the outcome is the same: peer evaluations



(a) All Individual: Transition D1 and D2 Peer Reviews and Exam Grades



(b) All Teams: Transitions from D1 to D2 Peer Reviews and Project Grades

Figure 6: Comparison of All Peer Review and Exam Grade by Teams and Individuals

are not a reliable measure to use in our evaluation. On the other hand, the lower correlation could be justified by the fact that different skills are assessed in exams versus peer evaluations. Over-

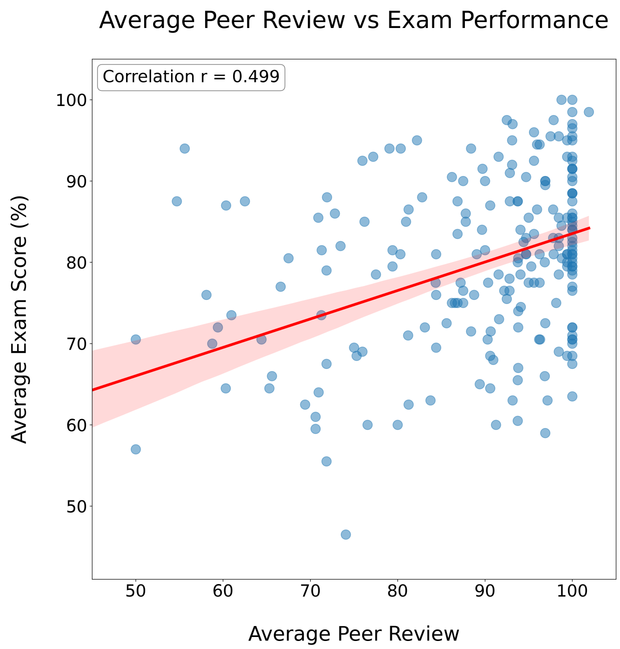


Figure 7: Average Team Peer Review vs Exam Performance

all, *it is trickier to analyze these peer evaluations and exam performance in isolation from other information about the students and the course.*

Summary of findings

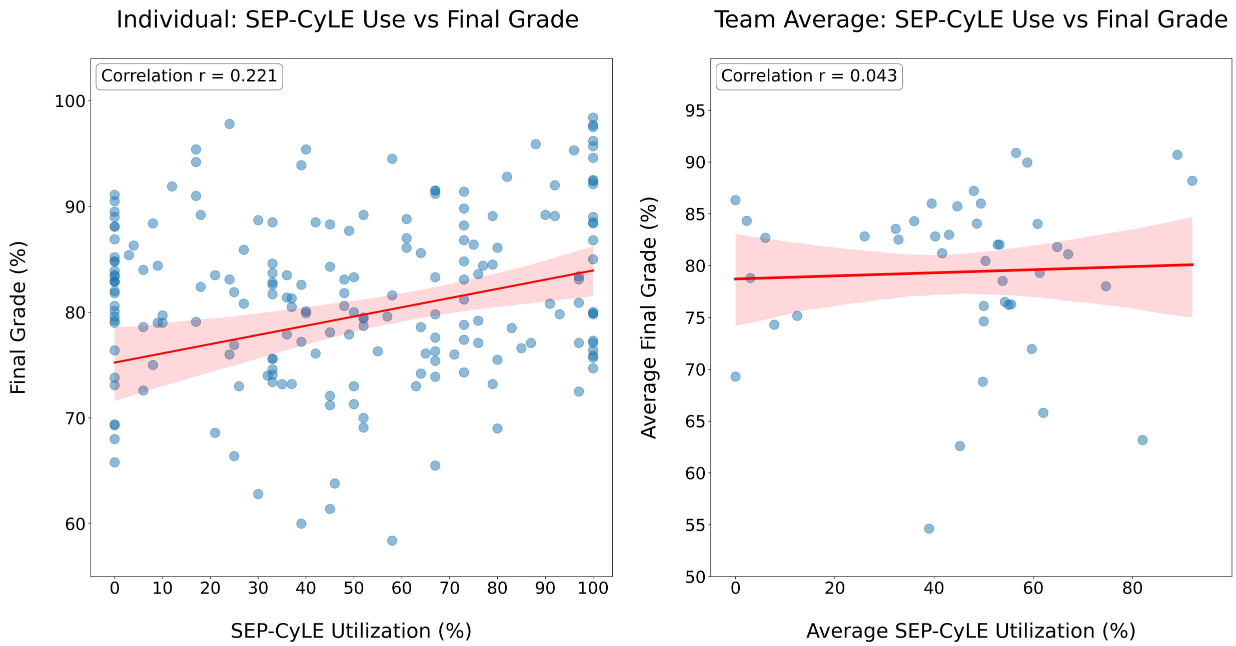
This comprehensive analysis has illuminated the complex relationships between peer reviews, project grades, and exam performance both at the individual and team levels. While peer reviews show a moderate correlation with project grades, suggesting their utility in assessing contributions and performance within team settings, their relationship with exam scores is more complex to understand. The progression from Deliverable 1 to Deliverable 2 indicates an improvement in the correlation of peer assessments with actual grades, reflecting either an increased accuracy in peer evaluations or a better alignment of project tasks with evaluation criteria as courses progress. However, peer reviews do not capture the consistent and variable correlation between midterm scores points to intrinsic student performance patterns. Ultimately, these findings underscore the complexity of educational assessment and the need for diverse evaluation strategies to accurately reflect student learning and performance dynamics.

6.2 Analysis - the effect of SEP-CyLE engagement (RQ2)

Now, we will take into consideration the SEP-CyLE tool that students use in the class, which also offers them extra credit as a team or as individuals.

6.2.1 SEP-CyLE utilization and final grades

Fig. 8 below analyzes the relationship between SEP-CyLE utilization and final grades, exploring both individual and team data to assess how the use of this resource impacts academic performance.



(a) Individual: SEP-CyLE Utilization vs Final Grade

(b) Team-Based: SEP-CyLE Utilization vs Final Grade

Figure 8: SEP-CyLE Utilization vs Final Grades for Individuals and Teams

The correlation coefficients reflect the relationships observed: a weak positive correlation of 0.186 for individuals and a negligible correlation of 0.007 for teams, indicating that SEP-CyLE utilization alone may not strongly predict final grades. One possible reason is its optional nature; students gain additional points by using it but they can still get full total points without it. As a result, some students might not use it consistently.

6.2.2 Clustering analysis for SEP-CyLE engagement

Our analysis now focuses on revealing insights into student behavior, team dynamics, and academic performance through advanced clustering techniques. We applied three distinct clustering algorithms to the SEP-CyLE educational database and systematically evaluated their effectiveness in capturing student collaboration patterns.

The first step in our analysis was to extract and analyze multiple features from the SEP-CyLE database that captured the nuanced aspects of student engagement and behavior. The features (with an example) used from the database included Course Name (CEN4072), Semester (Spring), Year (2018), SEP-CyLE user ID (840), Team ID (1), Grade (85%), Total Virtual Points (200), Submission Count (2), and Total Activities (30). Given these features, the following features were extracted *Content Completed* - elements in SEP-CyLE student completed, *Average Assessment Time* - time taken to complete the recorded assessment quiz, *Average Time to Submission* - the time the student spent working on the LO, *Average Points per Activity* - average points scored on the activities per assignment, *Average Pages per Day* - average number of pages visited in SEP-CyLE. Based on these features, we describe each student's profile (that describes their interaction with SEP-CyLE) using a vector, $\mathbf{p}_i$ for student i.

We performed the following steps to ensure compatibility across the diverse features obtained from SEP-CyLE.

- *Normalization* - used MinMaxScaler to transform all features to a value between 0 and 1. All other relationships between the data points were preserved.
- *Similarity Computation* - the similarity between two students i and j, $S(i, j)$, is computed as follows:
$$S(i,j) = 1 - \left(\frac{EuclideanDistance(\mathbf{p}_i, \mathbf{p}_j)}{MaxDistance} \right),$$
where the $MaxDistance$ is the maximum distance between all pairs of students.

We explored the following clustering algorithms to explore the different aspects of the data.

- *K-Means* - aims to minimize within-cluster variance.
- *Hierarchical* - captures nested relationships and multi-scale patterns
- *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* - identifies clusters of varying shapes and sizes. Effective at detecting outliers.

Three key evaluation methods were used to assess the quality and effectiveness of our clustering approaches.

1. Average Cluster Similarity:
$$Avg\ Similarity = \frac{\sum_i w_i * S_i}{N}$$
where S_i - average similarity for cluster i; w_i - size of cluster i; N - total number of students.
2. Average Distance to Centroid:
$$Avg\ Distance = \frac{\sum_i w_i * E_i}{N}$$
where E_i - average Euclidean distance of the data points in cluster i to its centroid; w_i - size of cluster i; N - total number of students.
3. Team Assignment Dissimilarity: We first need to have a binary $n \times n$ matrix that captures if students were assigned to the same team or not. Then, based on the clustering solution, we create a similar matrix that captures if two students belong to the same cluster. Fig. 9 shows an example with four students. Now, we can compute the Team Assignment Dissimilarity which captures how different the clustering solution is from the actual team assignments of the students.
$$n_{pairs} = \frac{n(n-1)}{2}$$
$$Team\ Assignment\ Dissimilarity = \frac{n_{\mathit{diffs}}}{2 * n_{pairs}}$$
where n - is the number of students in the group; n_{diffs} - number of different cell entries between the original and clustered matrices (like the ones in Fig. 9); $n_{pairs} = n^2$ is the total number of pairs.

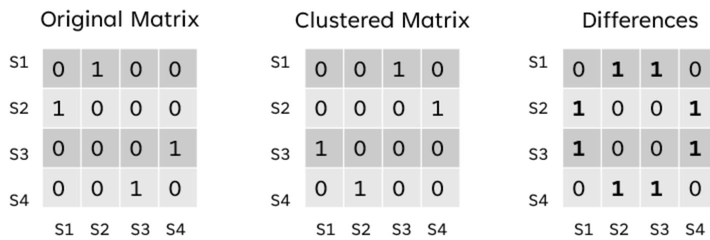


Figure 9: Matrices used to compute the Team Assignment Dissimilarity.

These evaluation methods are used to measure the differences between the original team assignment and the clustering solution.

6.2.3 Insights from the clustering analysis

Our analysis focuses on insights into student behavior, team dynamics, and academic performance through advanced clustering techniques. The three cluster evaluation metrics previously described were used to assess the clustering performance. Table 2 shows the scores obtained when applying the clustering algorithms to the student data in SEP-CyLE. The clustering method is shown in the first column of the table, the simple cluster score is shown in the second column, and Columns 3 - 5 show the scores for similarity, the distance to the centroid, and team dissimilarity.

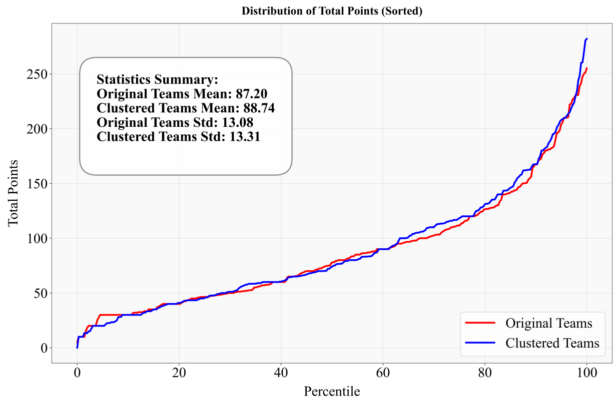
The key observations of the scores in Table 2 are as follows. The hierarchical clustering demonstrated the most cohesive and tightly knit clusters. DBSCAN tended to create fewer and larger clusters. *All clustering methods were able to create more homogeneous teams compared to the original team assignments*.

Patterns of Student Behavior. The initial findings demonstrated that randomly assigned teams exhibited low within-group coherence. In contrast, hierarchical clustering techniques effectively identified and grouped students with more consistent behavioral patterns, as shown in Table 2, showcasing the potential of data-driven team formation strategies to improve team performance. The clustering approach revealed nuanced student interaction patterns that were obscured in traditional team assignments. This behavioral pattern can be seen in Fig. 10, where the red line represents the original teams and the blue line represents the teams formed using clustering.

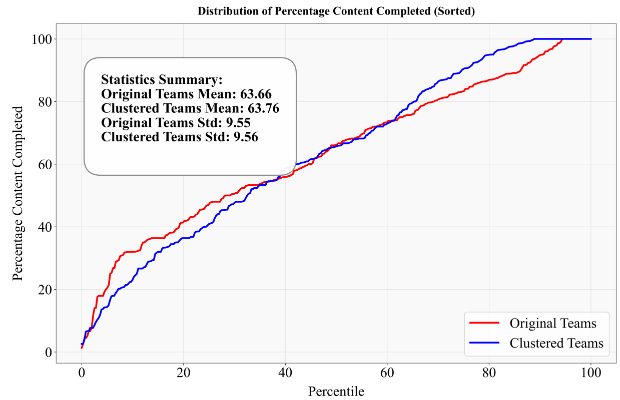
Team Tool Usage Patterns. Analysis of team consistency scores reveals that students within the same team tend to develop similar tool usage patterns over time. The data maintains a consistent mean consistency score of 0.66 across all periods, indicating moderate similarity in tool us-

Table 2: Cluster evaluation metrics for three clustering approaches using SEP-CyLE student data.

Method	Clusters	Similarity Score	Distance to Centroid	Team Dissimilarity
Original	8.813	0.549	0.651	-
K-Means	8.813	0.654	0.506	0.18
Hierarchical	8.813	0.674	0.472	0.18
DBSACN	2.438	0.658	0.552	0.06



(a)



(b)

Figure 10: Graphs showing distribution of (a) average similarity and (b) average distance to centroid.

Table 3: Cluster evaluation metrics for three clustering approaches using SEP-CyLE student data at three points in the semester.

Method	Clusters	Similarity Score	Distance to Centroid	Team Dissimilarity
Original	Beginning	0.45	0.72	-
	Middle	0.52	0.67	-
	End	0.55	0.65	-
Clustered	Beginning	0.43	0.70	0.18
	Middle	0.58	0.58	0.19
	End	0.67	0.47	0.18

age among team members. The distribution of consistency scores evolves notably throughout the semester, transitioning from an initial bimodal distribution (with peaks at 0.4 and 0.8) to a more uniform distribution in higher consistency ranges (0.6-0.9) by the semester's end. This evolution suggests that while team members may begin with diverse approaches to using the tool, they gradually align their usage patterns through shared experiences and established team workflows. The decrease in teams with very low consistency scores (below 0.4) from beginning to end further supports this convergence in behavior, though some individual variations persist even in well-coordinated teams.

Team Behavior Over Time. The SEP-CyLE student data for the CEN4072 assignments were captured at three points during the semester. Four-week intervals separated each assignment. Table 3 shows the data collected for the three assignments during the semester using clustering and the three evaluation metrics. The data in Table 3 is shown as a pair of bar charts in Fig. 12.

The data in Table 3 and Fig. 12 show that over time, the average similarity of the teams is better for the team formation using hierarchical clustering than for the original team structure. The average distance to the centroid decreases much more over time for the hierarchical clustering than the original team structure. Finally, the team dissimilarity remained consistent over time at approximately 18%.

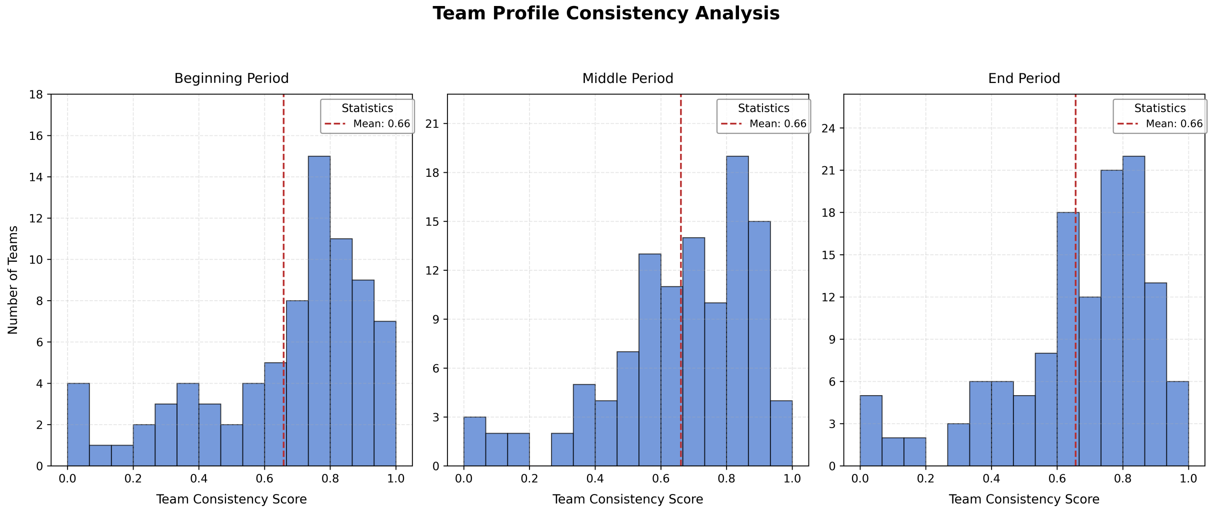
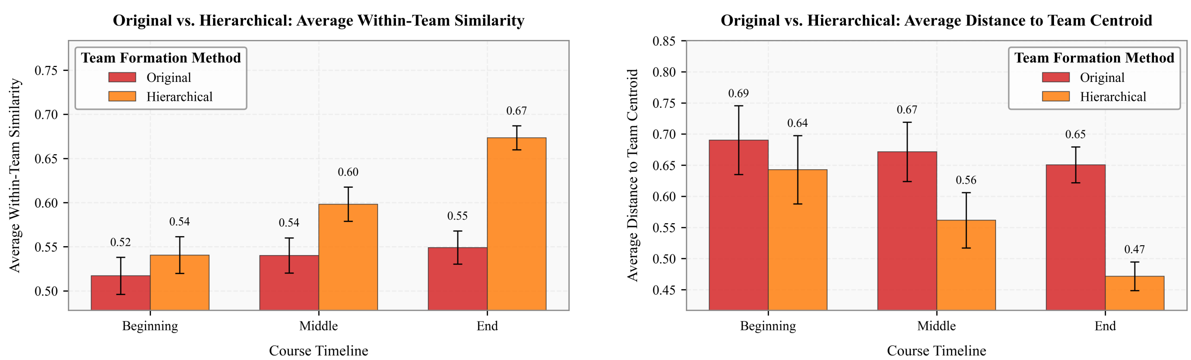


Figure 11: Team Profile Consistency Analysis showing the distribution of team consistency scores across semester periods. The analysis reveals the evolution of team behavior patterns, with a mean consistency score of 0.66 maintained throughout the semester. The distribution shifts from a bimodal pattern in the beginning period to a more concentrated distribution in higher consistency ranges (0.7-0.9) by the end period, indicating increasing alignment in tool usage patterns within teams.



(a) (b)

Figure 12: Bar charts showing original vs hierarchical clustering for (a) average similarity and (b) average distance to centroid.

Student Engagement Profiles. Through our analysis of engagement metrics, we identified three distinct student profiles that emerged from the clustering analysis. These profiles show clear differentiation in their engagement patterns and remain consistent across the semester. The high-engagement profile (Cluster 2.0) demonstrates consistently elevated metrics, with overall activity levels increasing from 12 to over 25 units throughout the semester, while maintaining content completion rates above 90%. The moderate-engagement profile (Cluster 1.0) shows intermediate values, with activity levels rising from 9 to 17 units and more variable content completion

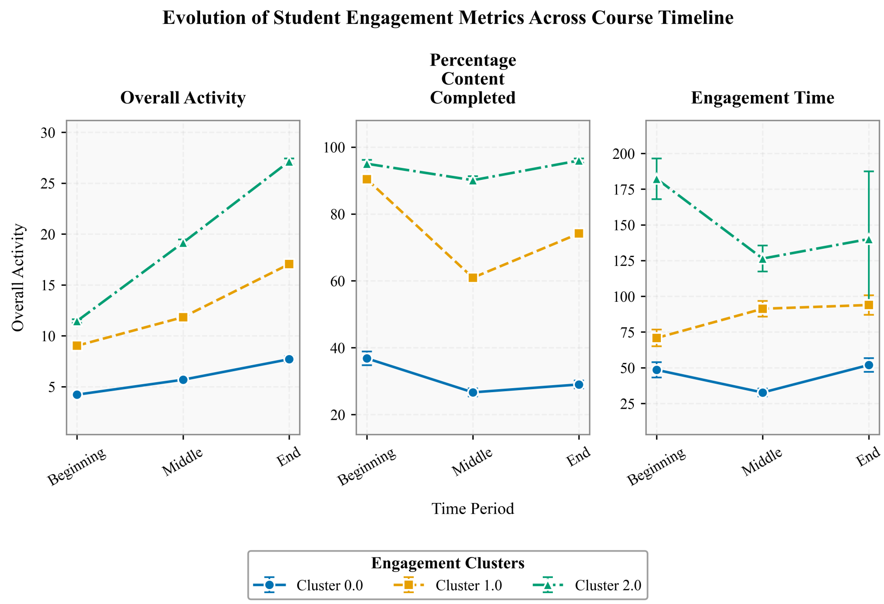


Figure 13: Evolution of Engagement Metrics by Cluster showing three distinct profiles across the semester. The metrics demonstrate clear separation between high (Cluster 2.0), moderate (Cluster 1.0), and low (Cluster 0.0) engagement patterns in overall activity, content completion, and engagement time. The divergence between clusters increases over time, particularly in overall activity measures.

rates, starting at 90% but declining to 60% before recovering to 75%. The low-engagement profile (Cluster 0.0) maintains minimal activity levels (5-8 units) and the lowest content completion rates (around 30%). The separation between these profiles tends to amplify rather than converge as the semester progresses, particularly in overall activity metrics, suggesting that initial engagement patterns have lasting effects on student behavior.

6.3 Limitations

While our work provides interesting insights, we should also be mindful of the following limitations:

- The use of the SEP-CyLE tool was optional, so it might be used by all students. At the same time and for the same reason, any signals and insights we observe are particularly strong, as they are driven by students themselves and not by the course syllabus.
- This analysis is only done for one course. In the future, we plan on analyzing additional courses in order to examine whether our insights hold across subjects.
- When students can gain points from different activities, there might be the phenomenon of "gaming the system". In these cases, students take advantage of the system's scoring mechanisms to gain points rather than properly thinking through the material [26]. That is a problematic situation that is tricky to handle, as the score received does not really capture students' efforts. While the instructor has taken steps against this phenomenon, it might still

be present in our dataset.

- Finally, we acknowledge that the data we use might not capture all the intricacies that are present in student learning and collaboration. When we believe this might be happening, we have mentioned it in the corresponding section.

7 Conclusion

Active learning activities that involve teamwork have become very important in higher education. In this paper, we explored a data-driven analysis of different aspects of teamwork, and also in conjunction with the use of a cyberlearning environment. We use 12 semesters' worth of data from the "Fundamentals of Software Testing" course, which has the same structure, instructor, and utilization of the SEP-CyLE through the years of the study. We analyze how team collaboration and students' performance correlate, at both the individual and the team level. We also explore students' SEP-CyLE engagement profiles, to understand the effect of SEP-CyLE and underlying patterns. Our findings include interesting patterns and indications of different student behaviors. In the future, we plan on exploring more data to verify whether our findings are consistent across different courses and instructions.

Acknowledgment

The authors would like to thank the reviewers for their thoughtful and helpful comments. The work was supported by the GAANN (Graduate Assistance in Areas of National Need) Fellowships from the Knight Foundation School of Computing and Information Sciences (KFSCIS) at Florida International University (FIU), funded by the U.S. Department of Education (award number P200A210087). Partial support for this work was provided by the National Science Foundation's (NSF) Improving Undergraduate STEM Education (IUSE) program under Award Numbers DUE-1525112 and CNS-2246004. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] E. Britton, N. Simper, A. Leger, and J. Stephenson. Assessing teamwork in undergraduate education: a measurement tool to evaluate individual teamwork skills. *Assessment & Evaluation in Higher Education*, 42(3):378–397, 2017.

[2] Center for Teaching Innovation. Active & collaborative learning | Center for Teaching Innovation. https://teaching.cornell.edu/teaching-resources/active-collaborative-learning. Accessed: January 2025.

[3] R. Chang-lau and P. J. Clarke. Software Engineering and Programming CyberLearning Environment (SEP-CyLE) Instances, July 2018. https://stem-cyle.cis.fiu.edu/instances.

[4] P. J. Clarke, D. L. Davis, I. A. Buckley, G. Potvin, M. Thirunarayanan, and E. L. Jones. Combining learning and engagement strategies in a software testing learning environment. *ACM Trans. Comput. Educ.*, 22(2), Nov. 2021.

[5] S. Deterding, D. Dixon, R. Khaled, and L. Nacke. From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '11, page 9–15, New York, NY, USA, 2011. Association for Computing Machinery.

[6] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences*, 111(23):8410–8415, 2014. Publisher: National Acad Sciences.

[7] O. Haatainen and M. Aksela. Project-Based Learning in Integrated Science Education: Active Teachers' Perceptions and Practices. *LUMAT: International Journal on Math, Science and Technology Education*, 9(1):149–173, 2021.

[8] B. Hurst, R. Wallace, and S. B. Nixon. The impact of social interaction on student learning. *Reading Horizons*, 52(4):375–398, 2013. `https://scholarworks.wmich.edu/reading_horizons/vol52/iss4/5` (Accessed August 2018.

[9] IUSE Project Team. Software Engineering and Programming CyberLearning Environment (SEP-CyLE) Resources, July 2024. https://stem-cyle.cis.fiu.edu/resources.

[10] A. Jaiswal, T. Karabiyik, P. Thomas, and A. J. Magana. Characterizing team orientations and academic performance in cooperative project-based learning environments. *Education Sciences*, 11(9):520, 2021.

[11] K. Jusoff and S. A. A. Samah. *Social Interaction Learning Styles*, pages 3101–3104. Springer US, Boston, MA, 2012.

[12] J. Lawlor, C. Conneely, E. Oldham, K. Marshall, and B. Tangney. Bridge21: teamwork, technology and learning. A pragmatic model for effective twenty-first-century team-based learning. *Technology, Pedagogy and Education*, 27(2):211–232, 2018. Publisher: Taylor & Francis.

[13] Marlborough. Four Benefits of Teamwork for Student Development & Success. https://www.marlborough.org/news/ board/stem/post/four-benefits-of-teamwork-for-student-development-success, 2019. Accessed: January 2025.

[14] National Center for Science and Engineering Statistics (NCSES). Diversity and stem: Women, minorities, and persons with disabilities: 2023. Special Report NSF 23-315. Alexandria, VA: National Science Foundation. Available at `https://ncses.nsf.gov/wmpd` (Accessed January 2025).

[15] T. L. Quick. *Successful Team Building*. AMACOM, American Management Association, 1992.

[16] E. Ribeiro-Silva, C. Amorim, J. L. Aparicio-Herguedas, and P. Batista. Trends of Active Learning in Higher Education and Students' Well-Being: A Literature Review. *Frontiers in Psychology*, 13, 2022.

[17] H. G. Schmidt. Foundations of problem-based learning: some explanatory notes. *Medical Education*, 27(5):422–432, 1993.

[18] B. L. Smith and J. T. MacGregor. What is Collaborative Learning? In A. Goodsell, M. Maher, and V. Tinto, editors, *Collaborative Learning: A Sourcebook for Higher Education.* National Center on Postsecondary Teaching, Learning, and Assessment, University Park, Pa., 1992.

[19] R. S. Smith. Guidelines for authors of learning objects. The New Media Consortium, 2004.

[20] Software Freedom Conservancy All. Selenium, 2025. `https://www.selenium.dev/` (Accessed Jan. 2025).

[21] P. Tarricone and J. Luca. Successful teamwork: A case study. In *Quality Conversations*, 2002.

[22] The JUnit Team. JUnit 5, 2025. `http://www.junit.org/` (Accessed Jan. 2025).

[23] E. J. Theobald, M. J. Hill, E. Tran, S. Agrawal, E. N. Arroyo, S. Behling, N. Chambwe, D. L. Cintrón, J. D. Cooper, G. Dunster, J. A. Grummer, K. Hennessey, J. Hsiao, N. Iranon, L. Jones, H. Jordt, M. Keller, M. E. Lacey, C. E. Littlefield, A. Lowe, S. Newman, V. Okolo, S. Olroyd, B. R. Peecook, S. B. Pickett, D. L. Slager, I. W. Caviedes-Solis, K. E. Stanchak, V. Sundaravardan, C. Valdebenito, C. R. Williams, K. Zinsli, and S. Freeman. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, 117(12):6476–6483, 2020. Publisher: Proceedings of the National Academy of Sciences.

[24] I. Topsakal, S. A. Yalçın, and Z. Çakır. The Effect of Problem-based STEM Education on the Students' Critical Thinking Tendencies and Their Perceptions for Problem Solving Skills. *Science Education International*, 33(2):136–145, 2022.

[25] E. van Laar, A. J. A. M. van Deursen, J. A. G. M. van Dijk, and J. de Haan. The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, 72:577–588, 2017.

[26] K. P. Vanacore, A. Gurung, A. Sales, N. Heffernan, et al. Effect of gamification on gamers: Evaluating interventions for students who game the system: Evaluating interventions for students who gaming the system. *Journal of Educational Data Mining*, 16(1):112–140, 2024.

[27] E. Viro and J. Joutsenlahti. Learning Mathematics by Project Work in Secondary School. *LUMAT: International Journal on Math, Science and Technology Education*, 8(1):107–132, 2020.

[28] A. Walker, H. Leary, C. Hmelo-Silver, and P. A. Ertmer, editors. *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows*. Purdue University Press, West Lafayette, Indiana, 2015.

[29] T. J. Yosso. Whose culture has capital? a critical race theory discussion of community cultural wealth. *Race ethnicity and education*, 8(1):69–91, 2005.

[30] L. Zahedi, J. Batten, M. Ross, G. Potvin, S. Damas, P. Clarke, and D. Davis. Gamification in education: A mixed-methods study of gender on computer science students' academic performance and identity development. *Journal of Computing in Higher Education*, 33:441–474, 2021.