ABSTRACT

AKINTUNDE, RUTH OLUWADAMILOLA. Learning Analytics for Data-Informed Interventions in Blended Classrooms. (Under the direction of Tiffany Barnes.)

Colleges and universities across the nation are making an immense effort to adjust to the effects of the COVID-19 outbreak on education and learning in general. Technological solutions are moving away from in-person to online modes of teaching and learning. Colleges that had rarely used online tools are now forced to embrace these tools at a greater speed than ever before. Consequently, now more than ever, it is imperative that we evaluate existing online learning tools to measure their effectiveness and expose their deficiencies in order to pave the way for a new generation of online tools that would adequately meet the needs of online learning. It is worth mentioning that most college courses make use of more than one online learning tool to support instruction, collaboration and help-seeking. Hence, it is increasingly necessary to know how students navigate these online tools, how these tools work together to help the learning process and if the tools used together in a class are sufficient to achieve learning goals. To investigate this, a project called Concert - Coordinating Educational Interactions for Student Engagement, was initiated.

Research on how students leverage, navigate and integrate their use of online tools is scarce. Concert aims to discover inadequacies in students' use of online learning platforms such as the Piazza forum, Jenkins Automated Grader, the GitHub submission system, MyDigitalHand, and Moodle to address the knowledge gap by developing an open platform to collect, integrate, and analyze data from students' use of multiple online tools.

This work applied both quantitative and qualitative approaches to achieve the goals of Concert. I performed four studies focused on tracking student online tool usage and made the following contributions to learning analytics and computer science education. First, this dissertation reveals a need for actionable learning analytics dashboards for help-seeking forums. I showed that online discussion forum analytics provides four major benefits to instructors. It provides instructors with: i.) an understanding of how students perceive programming assignments and signals when students are struggling ii.) reflection over previous assignment offerings for the improvement of future assignment specification design iii.) insights about the rate of social interactions going on while students work on assignments; and what higher or lower rates of interaction signal to instructors about an assignment and consequently class performance; and iv.) insights about the possible need for intervention before the end of an assignment's duration. Second, using machine learning and natural language processing techniques, I exposed the teamwork behaviors and roles of different genders while collaborating on programming assignments on Github, another online tool used by college students. Third, I designed, implemented and tested PEDI - Piazza Explorer Dashboard for Intervention, an analytics dashboard platform that achieves the above stated benefits. Lastly, this dissertation presents an experience report of seven instructors who explored the PEDI tool and how they intend to use the tool in their classes.

© Copyright 2021 by Ruth Oluwadamilola Akintunde

All Rights Reserved

Learning Analytics for Data-Informed Interventions in Blended Classrooms

by
Ruth Oluwadamilola Akintunde

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Computer Science

Raleigh, North Carolina

2021

APPROVED BY:

Collin Lynch	Sarah Heckman
Shiyan Jiang	Tiffany Barnes Chair of Advisory Committee

DEDICATION

I dedicate this dissertation to my Father in heaven, my creator and my God. I would have given up this quest if not for the strength, wisdom, inspiration, grace, favor and good health He gave me. Thank You Jesus!

My parents Tolu and Abimbola Okoilu who laid the foundation for my education by ensuring that I went to the best schools in Nigeria not minding the cost and in spite of their financial struggles. My Dad supports me spiritually and my Mum, a math teacher, always taught me advanced math concepts that are two grade-levels ahead of my grade. Most of my friends say my parents are the best in the world and I have to agree. I love them so much.

My husband, Tolu Akintunde who God placed with me to defend and support me with encouraging words. I'm grateful for his love, understanding and help with the kids during stressful periods.

My two little daughters Kikiopefoluwa (Kiki mama) and Kareoluwa (Bianca) who have definitely been affected by this pursuit but have always appreciated the time I could afford to spend with them. Thank you for your unconditional love. God bless you. Doing a PhD is like being on another planet. I'm so excited to be coming back to earth.

BIOGRAPHY

Ruth was born on March 4th 1991, in Lagos, a big city in Nigeria. Her father is a clergyman and her mother a math teacher. She is the first and only daughter of three children. Ruth grew up and attended elementary, high school and college in Nigeria. She graduated top of her class with a Bachelor of Computer Science degree from the University of Lagos, Nigeria. Afterwards, she got married to her best friend Tolu and gave birth to her first daughter Kikiope. Before heading off to graduate school, Ruth worked for about 4 years as a software developer and headed the mobile applications development team at Venture Garden Group, a leading provider of technology platforms in Nigeria.

In 2016, she got bored, missed doing research and studying for exams so she decided to pursue her PhD studies in Computer Science at NC State University. In 2019, she joined the Game2Learn lab at NC State where she met the best Professor she had ever worked with, Dr. Tiffany Barnes. In the same year, she gave birth to her second daughter Kareoluwa while writing the abstract of her first publication.

The author is a strong, determined and self-motivated individual who loves to motivate everyone around her. She enjoys fashion, singing and spending time with her family.

ACKNOWLEDGEMENTS

I acknowledge God Almighty, the maker of heaven and earth for making it possible for a Nigerian girl without a silver spoon in her mouth, to achieve great heights. Thank You Holy Spirit!

I am grateful to my advisor Dr. Tiffany Barnes. Dr. Barnes is not your typical PhD advisor; to me she has been a mother, friend and a very intelligent Professor. We learn a lot from our advisors, sometimes bad habits and sometimes good. However, with Dr. Barnes I have learnt only good habits. She taught me how to treat people well no matter their race or accent. She taught me how to care for the needs of others that go beyond the professional or academic world. She taught me to honestly praise others on every progress and give constructive criticism whenever it was required. She always says she loves and cares about members of her lab. I believe it because I have seen it in action.

I am also grateful to my committee members, Dr. Collin Lynch who gave great advice while working on the Concert project, Dr. Sarah Heckman, who always made out time to thoroughly review my papers and Dr. Shiyan Jiang who contributed her valuable time to be on my committee.

I am grateful for people in my lab, Dr. Veronica Catete who believed in me, Preya Shabrina, my first co-author, Dr. Yihuan Dong, Dr. Mehnak Manikala, Dr. Nick Lytle, Dr. Alex Milliken, Samiha Marwan, Rachel Harred, Lauren Alvarez, Dave Bell, Ally Limke, Isabella Gransbury and many. I appreciate you all.

TABLE OF CONTENTS

LIST OF TABLES . . . . . **viii**

LIST OF FIGURES . . . . . **x**

Chapter 1 INTRODUCTION **1**
1.0.1 Coordinating Educational Interactions for Student Engagement - Concert . . 2
1.0.2 Contributions Made by the Concert Project 4
1.0.3 Data and Population 5
1.0.4 Research Questions 5

Chapter 2 LITERATURE REVIEW **7**
2.0.1 Learning Analytics for Higher Education 7
2.0.2 Benefits of Learning Analytics 7
2.0.3 Learning Analytics in Smart Learning Environments 9
2.0.4 Learning Analytics Around Forum Usage and Help-seeking 10
2.0.5 Learning Analytics Around Teamwork Collaboration 12
2.0.6 How Forum Activity Analytics is Used in the Classroom 14
2.0.7 Learning Analytics Dashboard (LAD) 15
2.0.8 Summary 16

Chapter 3 Study 1: Exploring class level forum usage for instructor insights **17**
3.1 Introduction 18
3.2 Background 19
3.2.1 Class-level Performance Prediction 21
3.2.2 Forum Usage and Performance 22
3.2.3 Theory about the Effects of Milestones in Programming project 22
3.3 Methods 24
3.3.1 Data Collection 24
3.3.2 Exploratory Analysis 25
3.3.3 Correlation Analysis 28
3.3.4 Regression Analysis 29
3.4 Results 29
3.4.1 RQ1.1: What are the patterns of class-level student forum use for an assignment and what can these patterns indicate to instructors about that programming assignment? 30
3.4.2 Regression 31
3.5 Discussion 31
3.5.1 RQ1.2: What is the timing of class-level interactions in the class and consequently how can we encourage early forum use? 31
3.5.2 Instructor Feedback 32
3.6 Limitations 34
3.7 Conclusion 34

Chapter 4 Replicating Study One **35**
4.1 Introduction 35
4.2 Background 38

4.2.1 Notable Differences Between the Java Programming Course and Discrete Math Course Administration . . . 39
4.2.2 Zero Analysis . . . 40
4.3 Methods . . . 43
4.3.1 Data Collection . . . 43
4.3.2 Exploratory Analysis . . . 43
4.3.3 Correlation Analysis on class-level posts and performance . . . 44
4.3.4 Regression Analysis . . . 46
4.3.5 Random Forest Analysis . . . 47
4.4 Results . . . 47
4.4.1 Class-level Correlation Results . . . 48
4.4.2 Predictive Analysis . . . 48
4.5 Discussion . . . 49
4.6 Limitations . . . 49
4.7 Conclusion . . . 49

Chapter 5 Study 2: A Sneak Peek at a Collaborative CS Project: What part do female students play? . . . 51
5.1 Introduction . . . 51
5.2 Related Works . . . 54
5.2.1 Gender differences in Choosing Computer Science . . . 54
5.2.2 Gender differences in Collaborative Projects . . . 54
5.3 Method . . . 55
5.3.1 Pre-processing . . . 55
5.3.2 RQ2.1: How did each gender respond to the recommendation to tag their commit messages? . . . 56
5.3.3 Repo Processing . . . 57
5.3.4 RQ 2.2: In the class we studied, were tasks assigned to team members biased along gender lines? . . . 59
5.3.5 Visualizing Mean Number of Modified Files Committed in each category in Multi-Gender Teams . . . 60
5.3.6 Tag Prediction using Natural Language Processing (NLP) techniques . . . 60
5.3.7 Visualizing Mean Number of Modified Files Committed in each NLP category in Multi-Gender Teams . . . 62
5.3.8 RQ3: Can we predict the gender (M/F) of the authors using details of the commit messages as predictor variables and can gender predict the commit tag? . . . 63
5.4 Discussion . . . 66
5.5 Limitations . . . 66
5.6 Conclusion . . . 67

Chapter 6 Study 3: PEDI - Piazza Explorer Dashboard for Intervention . . . 68
6.1 Introduction . . . 68
6.1.1 Inspiration for PEDI . . . 69
6.2 Related Work . . . 70
6.2.1 Learning Analytics Dashboard (LAD) . . . 70
6.2.2 Functionality . . . 70
6.2.3 Ease of Use . . . 71

6.2.4 Design . . . 71
6.3 METHOD . . . 72
6.3.1 Design Principles . . . 72
6.4 DESIGN . . . 72
6.4.1 System Capabilities . . . 74
6.4.2 PEDI workflow . . . 76
6.5 Conclusions and Future Work . . . 77

Chapter 7 Study 4: What is PEDI to Instructors? . . . 78
7.1 Introduction . . . 78
7.1.1 Focus of the Study . . . 79
7.1.2 Hypotheses . . . 79
7.1.3 Research Questions . . . 80
7.2 Methods . . . 80
7.2.1 Recruitment . . . 80
7.2.2 Participant Procedure . . . 80
7.2.3 Planned analysis methods . . . 81
7.3 Case Studies . . . 82
7.3.1 Case Akin (P1) . . . 82
7.3.2 Case Dr. Bianca (P2) . . . 83
7.3.3 Case Caleb (P3) . . . 83
7.3.4 Case Dami (P4) . . . 83
7.3.5 Case Eden (P5) . . . 84
7.3.6 Case Fola (P6) . . . 84
7.3.7 Case Gad (P7) . . . 84
7.4 Themes . . . 85
7.4.1 Current instructor habits . . . 85
7.4.2 Instructors' perception of PEDI . . . 89
7.4.3 Future Recommendations - Future Work and PEDI Improvements . . . 96
7.4.4 Summary of Participant Survey . . . 99
7.5 Discussion . . . 102
7.6 Conclusion . . . 103

Chapter 8 CONCLUSION . . . 105
8.0.1 Summary . . . 106
8.0.2 Future Work . . . 107

BIBLIOGRAPHY . . . 108

APPENDIX . . . 120
Appendix A Figures, Tables and Interview Script . . . 121
A.1 Chapter 4: Figures and Tables . . . 121
A.2 Chapter 7: Pre-activity interview Questions: . . . 123
A.3 Exploration Task Zero: . . . 124
A.4 Exploration Task One: . . . 125
A.5 Exploration Task Two: . . . 127
A.6 Exploration Task Three: . . . 128
A.7 [10-15m] 4. Post-activity Interview Questions . . . 130

LIST OF TABLES

Table 3.1 Number of Students in each Semester 20
Table 3.2 Programming Projects in a CS2-OOP. Complexity (out of 5) was defined by the Course Instructor 20
Table 3.3 Duration of each assignment (Days) 20
Table 3.4 Average Class Performance on each assignment (Out of 100). Assignment with the lowest average performance in a semester is in italics, while that with the highest performance is in bold font 20
Table 3.5 Work Mode in Main Course Projects. Note that: All Guided Projects (GPs) were done individually 21
Table 3.6 Definition of Features of Piazza Posts 21
Table 3.7 Table comparing posts and performance 23
Table 3.8 Min-Max of features in Assignment used for normalization. Note: Average Post = Number of Post/N (students who attempted the assignment) 26
Table 3.9 Correlation ρ and p-values of Features. Statistically significant p-values based on Bonferroni Correction are in bold font 29

Table 4.1 Number of Students in each Semester 38
Table 4.2 Assignments in the Course 38
Table 4.3 Definition of Features of Piazza Posts 38
Table 4.4 Average Class Performance during Fall semesters 2014-2020 43
Table 4.5 Min-Max of features in Homework used for normalization. 44
Table 4.6 Min-Max of features in Exams/Tests used for normalization. 44
Table 4.7 Spearman correlation of class-level Average Performance with Features of Social Interaction (ρ). Statistically Significant p-values of features based on Bonferroni Correction are in bold font. The Feature Importance from the Random Forest Model is also displayed 47

Table 5.1 Statistics of Commit by Each Gender 56
Table 5.2 Commit Categories, Keyword and Definition 56
Table 5.3 Project Description 58
Table 5.4 Statistics of Total Commit Messages 59
Table 5.5 Results of NLP Analysis 61
Table 5.6 Percentage of commits messages in each category by gender O* - represents data from original tags. SVC - represents data from SVC - NLP tags. 63
Table 5.7 Performance of Models in Predicting Gender O* - represents prediction using original tags as a feature. SVC - represents prediction using SVC - NLP tags as a feature. Number in bold signify highest performance in each column. 65
Table 5.8 Accuracy of Models in predicting Commit Tags using Gender as a feature Highest values are in bold fonts 65
Table 5.9 Feature Importance of 6 features across top two performing models from column C and D - using permutation importance for each predictive analysis O* - represents prediction using original tags as a feature. SVC - represents prediction using SVC - NLP tags as a feature. Highest values are in bold fonts, commit tag features are in italics. 66

Table 6.1 Definition of Features of Forum Posts 70

Table 7.1 Each participant interview was tagged by two researchers as shown in the table. 82
Table 7.2 Table showing different types of analytics and what percentage of the seven participants are interested in them. 85
Table 7.3 Themes and Sub-themes 86
Table 7.4 Participants were shown periods of lower forum activity in task one. They were asked to select their preferred mode of intervention when unusually Lower Forum Activity is observed in the first half of an assignment. Participants could select all the options that applied. Therefore, these pies represent all the intervention choices, and the proportion of the choices that went to each intervention 100
Table 7.5 Reasons for high number of Posts. Participants were shown different posts with high likes as highlighted by the Treemap and asked to select all the options that applied. Participants could view the posts, so they may be reasoning based on actual post content. 101
Table 7.6 Reasons for high number of Likes 101
Table 7.7 Relative percentage of participants who preferred each reasons for high number of Replies 101
Table 7.8 Participants were shown periods of unusually higher forum activity close to the deadline (task two). They were asked to select their preferred mode of intervention in this case. Participants could select all the option that applied . 102
Table 7.9 PEDI Future Usage Frequency 104

Table A.1 Duration of each Assignment. Column D represents Duration in Days. Note that the assignments are displayed in the order used in each semester. This means that hw2 was Circuits in 2013 but hw2 was Counting 2 in 2020. See Zero Analysis Charts order of topics. 122
Table A.2 Duration of each Assignment. Column D represents Duration in Days. Note that the assignments are displayed in the order used in each semester. This means that hw2 was Circuits in 2013 but hw2 was Counting 2 in 2020. See Zero Analysis Charts order of topics. 122
Table A.3 Assignments and Percentage of the Class that Scored Zero in that Assignment 124
Table A.4 Assignments and Percentage of the Class that Scored Zero on that Assignment 125
Table A.5 Participants and their response to each question asked. 1 means Yes 0 means No 131

LIST OF FIGURES

Figure 3.1 The normalized number of forum posts and average performance for projects p0, p1 and p2 in Spring and Fall semesters of 2017-2019. 23
Figure 3.2 Projects in Spring 2017 and Fall 2018 showing forum use peaks a few days before the deadline 26
Figure 3.3 Projects in Spring 2019 and Fall 2019 showing the effect of intermediate milestones 27
Figure 3.4 Correlation analysis 29

Figure 4.1 Number of posts per day during Homework and Tests/Exams in 2015, selected to represent a typical semester 37
Figure 4.2 Percentage of class that had a zero score in Fall semesters of 2014-2017. . . . 41
Figure 4.3 Percentage of class that had a zero score in Fall semesters of 2018-2020. Cont'd 42
Figure 4.4 The normalized number of forum posts and average performance for homework in Fall semesters of 2014-2020. 45
Figure 4.5 The normalized number of forum posts and average performance for exams in Fall semesters of 2014-2020. 46
Figure 4.6 Predicted Average Performance from Random Forest Regression Analysis Vs Actual Average Performance 47

Figure 5.1 (a) Statistics of the class by Gender (b) Average commit message for each Gender 55
Figure 5.2 Mean commit messages per day for each gender 58
Figure 5.3 Each bin shows the mean number of modified files within each category as originally tagged by each gender. Focus Teams: (a) Team P2 and (b) Lab Teams with the original tags 61
Figure 5.4 Each bin shows the mean number of modified files by each gender within each NLP classified tag Focus Teams: (a) Team P2 and (b) Lab Teams with the NLP-SVC tags 62

Figure 6.1 (a) Sidebar Panel Control (b) Tree Map Control 73
Figure 6.2 (a) Interactive line graph showing average posts per day (b) Interactive histogram showing the number of posts per day. 73
Figure 6.3 (a) Pie chart showing number of students in the class who performed certain activity on the forum (b) Tree Map showing forum posts with the highest number of likes 73
Figure 6.4 Alert System displayed on the PEDI Dashboard 74
Figure 6.5 Tables on the PEDI Dashboard. (a) Average number of Questions asked (b)Fraction of the class that post (b) Assignment Deadlines 76

Figure 7.1 Relative proportions of interventions indicated by participants for periods with unusually (a) low forum activity or (b) high forum activity 100

Figure A.1 Number of posts per day during Homework in 2014 121
Figure A.2 Number of posts per day during Tests/Exams in 2018 123

CHAPTER

1

INTRODUCTION

The popularity of analytics in recent times is attributed to three major developments. The first is the increased availability, velocity, volume, and variety of data, an important pillar of analytics [Kir06; Hof17; She14]. The speed at which historical and real time data can be transferred and accessed, the immense amount of data produced by various technological systems and the diverse valuable sources of these data has made analytics not only beneficial but also serves as a motivation for skilled data analysts and scientists to be creative. Second, the emergence of robust artificial intelligence (AI) and Machine Learning (ML) techniques in recent time have given rise to quality analytics which are producing timely and increasingly accurate insights for effective decision making. Third, there is a growing demand from different public sectors including the educational, health and financial sectors for actionable insights. When analytics is focused on improving education, learning and teaching, it is referred to as Learning Analytics (LA).

Sclater et al. 2016, defines Learning Analytics as the measurement, collection, analysis and reporting of data about the progress of learners and the contexts in which learning takes place [Coo12]. Learning analytics leverages the availability of data such as digital footprints left by students' usage of learning environments and pedagogies, AI and ML techniques as well as the educational sector's demand for data-informed insights. Diverse learning environments and technologies have evolved over the years as a result of technological advancements. Learning is changing from what it used to be - the traditional teacher centered physical classroom environments to smart learning environments such as the flipped classroom, blended learning, hybrid learning, game based learning, gesture based learning [Kin16; Has10].

Blended learning environments in particular incorporate virtual learning space or online learning into the traditional offline learning environment [Wan17]. In blended learning environments,

online learning supplements traditional face to face (F2F) learning. Online learning systems provide several services. Piazza provides online discussion forums to enable students to interact and ask questions. Github enables collaborations with other students in order to complete group assignments. MyDigitalHand allows students to virtually meet with their instructor. WebAssign and Moodle are Learning Management Systems that host assignment resources and assignment submission boards. Other tools such as Zoom recordings and YouTube videos allow distance learning students to watch video lectures on their own schedules.

Incorporating online learning components with F2F learning enables students to observe instructors and vice versa, and helps students gain computational thinking skills as they interface with technological tools in the learning process [Lia17]. In the past year, the COVID-19 pandemic contributed to the sole reliance on the virtual learning space as physical gatherings of students are being discouraged and social distancing is being enforced. The challenge is that research also shows that total reliance on online learning and letting students learn on their own may not achieve the desired learning outcomes due to reasons such as low participation, lack of social presence, and delayed feedback [Wan17]. Research measuring the impact of increased reliance on fully virtual learning is still underway [Led20].

Another challenge is that major colleges use different combinations of these online learning platforms to achieve virtual learning and there is little research about the real time impact of these tools on learning outcomes and teaching. Instructors are tasked with the burden of keeping track of these different platforms, most of which do not provide adequate real time insights on how students are using these tools [Hoy20]. Instructors need to understand different signals from these various platforms about student's progress or failure. For example, what does lower forum activity on Piazza tell instructors about class progress on assignments? What does reduced collaboration on Github signal? Instructors need information about when to intervene based on signals from these platforms about specific students and groups, and need information on how to detect students who are working in isolation when collaboration is compulsory.

1.0.1 Coordinating Educational Interactions for Student Engagement - Concert

The Concert project aims to integrate data from these online learning tools using data mining techniques, detecting when students need help by training machine learning models on students' online footprints, providing a single student interface for notifications and tracking student progress in real time on all these platforms. The project will also bridge the knowledge gap and increase understanding of students' help-seeking and collaboration behaviors as well as social interactions that affect learning outcomes. Concert aims to explore the following online tools:

1.0.1.1 Piazza

Piazza, a Learning Management System (LMS) used by 2,000 Universities across 90 countries with over 2M students [Pia]. Piazza allows students to ask questions in a forum-type format. On this forum, discussion topics were classified into assignments. Each assignment has a folder containing

posts relating to that assignment on the forum. Students ask questions by creating a post on the forum and tagging the folder relevant to their question. Both instructors and students are free to respond to posts and for every question asked on the forum, Piazza records information such as time the question was asked, question content, replies, number of replies, number of views, number of likes, to mention but a few.

1.0.1.2 Github

GitHub has been embraced by the open source communities and is being employed as the major online learning tool for collaboration in programming related college courses. The platform is increasingly vital for managing software projects. GitHub provides decentralized version control for projects and allows different members of the project team to collaborate seamlessly by resolving changes to be merged and resolving resulting conflicts in an efficient manner. In addition Github provides a traceable project repository. Several research has shown the effectiveness of Github in transforming the learning experience [ZA15].

1.0.1.3 Moodle

Moodle is a popular LMS widely used by 60% of higher education institutions all over the world. Moodle's powerful pedagogical features provide both instructors and students with a robust online and offline learning experience. Moodle possesses the following features: allows instructors to set online exams using assessment tools, such as multiple choice quizzes, discussion forums, wikis, a store for course materials, polls and surveys for feedback, grading assignments etc [TT19]. Moodle is efficiently maintained and updated regularly. It also provides support and opportunity for open source development.

1.0.1.4 MyDigitalHand

MyDigitalHand (MDH) is an online tool for tracking instructor-student one-to-one meetings and interactions. During office hours, students can digitally request help from instructors. MDH allows instructors to manage help requests and close tickets [SA17].

1.0.1.5 WebAssign

WebAssign is a virtual homework application. It allows students to access and submit homeworks. It also allows instructors to grade assignments on the same platform [Web].

It is important to analyse data from each of this online learning platforms for the following reasons: Traditional classroom signals for intervention differ from online signals and different strategies might be required to detect them. Classroom behavior problems that disrupt learning objectives [par12], school bullying [MK08] which is tantamount to cyber bullying on online discussion forums are issues that might be easy to detect in a physical classroom but might not be obvious in virtual learning environments.

Virtual learning platforms introduce their unique challenges such as learner isolation [GS17], time management and lack of learner motivation. Technology such as online group forums, Facebook and slack groups are currently being applied to counter learner isolation. Learners in blended learning environments have to acquire time management skills to ensure that in addition to the F2F classes they have sufficient time to explore and leverage the online learning tools [Gla12]. Adequate analysis is required to detect poor time management and strategies such as course and assignments design that facilitate better time management should be created [Goo15]. Lastly, methods to detect lack of motivations and strategies to keep blended learners educated is of essence in recent times [Zav20].

Research about the impact of using these sets of online tools on teaching and learning is scarce. It is imperative to allow teachers to glean insights from these platforms that will help them measure the effectiveness of the platforms. The first step to achieving this is analysing individual impact of these tools and then moving up to analysing the combined impact of the tool on learning and instructions afterwards. This dissertation will introduce a LA tool to track students' activities on the Piazza platform.

1.0.2 Contributions Made by the Concert Project

Prior research aimed at achieving the goals of the Concert project have been successful in finding actionable insights from student activity across the different platforms. In a bid to understand how students manage their learning workflow across platforms, the concert project made the following findings: students who ask more questions on online discussion forums and receive more feedback tend to obtain higher grades Gitinabard et al. (2017) [Git17a]. Students usually work on single online tools rather than cross-platform [Git19b]. Higher performing students are more likely to go to Piazza after several submissions. Lower performing students tend to refer to class material on Moodle more [Git19a]. Students who work actively in teams and have active teammates often perform better in class and have better improvement.

Furthermore, students' online behavior and social connections can predict their performance in these classes before the first midterm exam [Git17a]. Students are always connected with similar performance peers through the entire class period [XY19]. Coding schema on MOOC discussion forum post [XY18]. Automatic classifiers design for imbalanced posts for students' discussion forums. Different types of posts have different impacts on different student cohorts. For example, logistic questions don't influence non-0 MOOC students' performance. Social interactions and students' online behaviors in MOOCS. Students' social metrics are better than the students' study behaviors in predicting student performance Gitinabard et al. (2019)[Git19a].

This current research, which is part of the Concert project, seeks to explore Piazza and Github learning tools to identify students' help-seeking and collaboration behaviors respectively based on students' activities on these learning platforms. This dissertation reports four additional novel studies performed under the Concert project.

1.0.3 Data and Population

For the first study, we analyzed class-level forum activities on Piazza, a popular online discussion forum used in a Java Programming course of a south-eastern research-intensive US university. We examined six (6) consecutive offerings of CS2-OOP [Por18b], a Java programming class in the Spring (January-May) and Fall (August-December) semesters of 2017 through 2019. The course offerings were identical, with similar schedules at the same university. We began with an exploratory analysis of how students use the Piazza forum on programming projects. Piazza data was collected as JSON files. Of the 1163 students (see Table 4.1) we studied, 74.3% of them posted at least once and 98.6% viewed questions posted on the forum. The second study explores gender roles in project task selection/assignment. We study how different genders in a CS2-OOP class collaborate on different projects. Statistics about this population are described in Figure 5.1. The third study designs and tests a forum analytics dashboard and investigates how instructors of two CS2 courses may wish to use forum trends and comparisons to intervene to improve class performance. Lastly, for study four, we worked with 5-10 professors and presented the PEDI dashboard having visual representations of the Piazza activities 3000 students from both classes.

1.0.4 Research Questions

This work has three overarching research questions listed below. Study 1 contributed to answering question 1. Study 2 focused on addressing question 2, and study 3 and 4 focused on addressing question 3.

- How can instructors get real time insights from online discussion tools?
  - Can Piazza signal useful learning information to instructors in real time?
  - What are the patterns of class-level student forum use for an assignment and what can these patterns indicate to instructors about that programming assignment?
  - What is the timing of class-level interactions in the class and consequently how can we encourage early forum use?
- How can instructors get insights about how gender affects students use of online collaboration tools
  - Are tasks assigned to team members biased along gender lines?
  - Can we detect biased cooperation and intervene for more equitable division of labor?
- Is the PEDI dashboard useful for instructors?
  - What are beneficial use cases of PEDI dashboard?
  - What particular information influences the instructor's decision to intervene while using the PEDI dashboard?

1.0.4.1 Filling the Knowledge Gap

There are two steps involved in filling these knowledge gap and support both instructors teaching and students' learning outcomes. The first step is to extract useful hidden students learning patterns that improve performance. RQ1 and RQ2 will focus on achieve this. The second step would then involve the implementation of Learning Analytics tools to detect these patterns in real time in order to impact students' performance. RQ3 will design, implement and test an analytics dashboard to achieve this.

CHAPTER

2

LITERATURE REVIEW

This literature review is organized as follows: the first two sections discusses the state-of-the-art contributions of Learning Analytics–the domain subject of my thesis, to higher education (HE) and college use of actionable learning - the backbone for innovation in college education. The third section describes related work on Learning Analytics in smart learning environments, the fourth presents existing literature on LA that improves forum use and help-seeking. The fifth section describes existing work on LA that improves students' collaboration.

2.0.1 Learning Analytics for Higher Education

The paradigm shift from teacher-centered to student-centered learning where students learn through teacher-guided activities such as talking, listening, writing, reading, and reflecting is being made possible by technological developments in education. Technology gave rise to software tools to facilitate these activities. The demand for insights and a need to model students while they learn through these series of activities on these tools in turn gave rise to Learning Analytics (LA). Learning Analytics is a natural result of rapid technological advancement in all areas of higher education [Alf19], [Man12a]. From flagging at-risk students to detecting hidden patterns of learners' behaviour and attitudes, LA has proven to be extremely valuable to education. The benefits of LA to higher education are discussed in more detail in the next section.

2.0.2 Benefits of Learning Analytics

The discovery of insights for actionable learning and data-driven decision making (DDDM) is one of the numerous contributions of LA to higher education [BG15]. DDDM has been used to inform

practises in HE and is one of the four pillars in the American Recovery and Reinvestment Act (2009) [Man12a]. Arne Duncan, the ex-Secretary of Education in an article stated that teachers need to know exactly what they need to do to teach and how to teach it and DDDM helps to achieve that [Man12a]. LA has been useful in exposing correlations between library services and learning outcomes. They found higher average GPAs among students who participate in the use of the Library's reference section, check out, retrieve, save, and download Library resources. They concluded that data about library usage should be integrated within institutions' learning analytics systems to ensure discovery of all-encompassing richer insights that take other aspects of students' learning into consideration [Oak17].

LA has engendered the design of tools for promoting student's integrity. In their work titled "Using Learning Analytics for preserving Academic Integrity," Amigud et al. implemented a machine-learning based framework which learns student content creation patterns and flags plagiarism with a mean accuracy of 93% which was 81% more accurate than instructors in the population they studied [Ami17].

LA has contributed to flagging low levels of student engagement in online learning. Rogers et al. described LA systems that associate low levels of students' engagement with poor performance. They argued that these systems only provide students with interventions such as study tips or more course-related resources to prevent students' failure without visiting both known (such as teaching or curriculum deficit) and unknown root causes of students' disengagement [Rog16]. Chapter 3 of this dissertation highlights how LA can be used to inform teaching and provide insights to instructors about how students perceive assignments based on different levels of student engagement with the discussion forum.

Furthermore, when LA is integrated in college education in real time, the result is personalized systems for flagging student's performance and progress when they fall below their peers. A 2011 study of 131 courses in University of Maryland Baltimore County (UMBC) showed that students who used the Campus' Online Content Management System (CMS) 39% less than their peers earned a D or an F. They highlighted the benefits of an LA tool that allows students to compare their activities on the CMS relative to their peers for students' self-regulation and to enable students to early detect their own inactivity in order to adjust and be more active on the CMS [Fri11]. The PEDI tool described in Chapter five provides a slightly similar tool but solely for instructors to compare current students' forum activities with those of previous semesters in order to detect late forum usage and provide necessary intervention that encourages early forum usage.

Another use case of LA is to measure the effectiveness of different intervention strategies for at-risk students. Jayaprakash et al. predicted at-risk students using students' learning management system (LMS) usage, demographic information and course signal. They concluded that interventions as simple as notifications and email provided earlier in the semester could highly improve students' performance, however it could also potentially trigger students to withdraw [Jay14].

LA has helped to understand how each member of a team would perform and teamwork habits [Git20]. Chapter 4 of this current dissertation explores how different genders in a team would con-

tribute to the progress of their assigned project. LA has also helped to expose the correlation between instructor's course preparation and dimensions of student engagement in activities. Ma et al. found a positive correlation between instructors' preparation and students viewing and understanding course materials [Ma2].

Before LA, most students could not detect their performance until usually after a major exam which in most cases is too late [Pis10]. Instructors on the other hand might not detect the effects of adjustments in instructions and assignment specification until after students' performance are reflected upon. LA systems such as the ones we've described are changing such narratives for both students and instructors. Interestingly, this current work includes aspects that provide instructors with an eagle eye and summary of forum activities on discussion boards, informing instructors on what section of course content to highlight during lectures based on the popular topics on the discussion forums. Other benefits of LA include early prediction of students' performance for effective interventions [Lu18], exposing factors affecting students's performance [Alt16], modelling students' learning habits [Mao20], exposing the impact of demography on performance [Han19] and engagement [Aki21], exposing the impact of different curriculum design and sequences on performance [Aki20].

All these benefits would not be possible without technology tools which serve as data warehouses to collect and collate data about teaching and learning. Technology tools include data collection systems, student and instructor assessment systems, monitoring tools, data dashboards, virtual learning environments, online learning, web-based tutoring tools, online discussion forums, interactive white-boards, computerized adaptive testing, teamwork tools. Mandinach et al. in chapter 3 - Technology to Support Data-Driven Decision Making of their book "Transforming Teaching and Learning Through Data-Driven Decision Making", explained these tools in detail [Man12b]. The process of passing data from data collection tools through qualitative analysis [Jia19], quantitative analysis [Che21], data mining [Des21] and/or machine learning pipelines [Mur21] is the core of LA which presents us with these benefits.

In subsequent chapters of this dissertation, I describe how studies I performed add to the existing benefits of LA. I performed learning analytics on data from two sets of tools - discussion forums i.e. Piazza platform and teamwork tool i.e. Github. Chapter 3 describes study one which exposes data-driven insights from Piazza while Chapter four describes study two, which presents insights from Github usage by different genders. In addition, I made recommendations such as the need for an additional data analytics dashboard layer on top of Piazza.

2.0.3 Learning Analytics in Smart Learning Environments

Smart learning environments include blended learning, flipped classroom, hybrid learning, game based learning, gesture based learning. **Blended learning** environments is one in which traditional F2F offline learning is supplemented by online learning. LA has been used to prove that this approach is beneficial for learning outcomes [Wan17]. Varshita et al. applied a machine learning technique called Data Time Warping (DTW) on LMS data logs in order to analyze the detect if correlation exists

between consistent LMS users and performance. They discovered that consistency in students' use of LMS systems does not always positively correlate with performance [She20]. In their paper titled "Predicting student success in a blended learning environment" Goidsenhoven et al. explored the blended learning environment on edX Edge and sought to create a robust course success prediction model using a total of 450 features. Using backward feature techniques they built their predictive model using the most important features. Their best performing classifier Random Forest achieved an accuracy of 0.84. Some of the important features for predicting success rates included "play video", "show answer of a problem", "submit problem" [VG20].

In a flipped classroom style students are introduced to the concepts at home and they apply the concepts in the classroom. Therefore classroom activities and home assignments are flipped. Harrak et al. in their work titled "Evaluating teachers' perceptions of students' questions organization" presented different dashboards intended to help teachers in the flipped classroom organize students' questions [Har20]. In a survey completed by 37 teachers in a flipped classroom, on teachers' dashboard preference, they discovered that teachers approved of the solution however teachers did not agree in one question organization dashboard design.

Hybrid learning is an area that is yet to be fully explored.Mandal et al introduced hybrid challenge based learning (CBL), an hybrid learning framework for cybersecurity courses. Although, the impact of this learning approach is yet to be evaluated among students, they discovered that students showed a strong motivation and high participation levels [Man21]. Chen et al. processed a low-cost scalable technology for hybrid learning and is currently implementing methods to achieve this [BC21].

Akintunde et al. and Shabrina et al. studied game play in a game based learning environment, they discovered patterns that could inform instructors on what math game students should play next [Aki20; Ped17], and preferable class formats (such as lab, free, and rotation seating) that improve learning outcomes [Sha20c].

Gesture based learning involves including gesture-based technologies in the learning process. Gestures such as eye tracking[Eme18], finger gestures [Zha21] and body movements [Den14] have been found to predict student performances.

2.0.4 Learning Analytics Around Forum Usage and Help-seeking

Forum usage can be defined as a collection of activities by students on online discussion forums. These activities could be help-seeking or help-giving activities [Bul01]. Online discussion forums used in colleges were solely designed for students to seek help and give help to their peers [Er15]. Instructors also respond to students' questions on the forum. Instructors in most colleges are increasingly replacing help-seeking via private emails with online discussion forums and prefer that students use the forum instead of sending non-private help-seeking emails. One reason for this is that other students could benefit from questions asked on the forum. There are numerous online forums used in today's colleges. Piazza [Pia], EchoLu [Er15], Discourse [Dis], InScribe [InS] and Yellowdig [Yel] are examples of discussion forums used by colleges.

2.0.4.1 Forum Use and Help-Seeking in Computer Science

Computing education researchers have become increasingly interested in the social aspect of learning and students' engagement in discussions as they work on programming assignments. Forum activities such as student questions asked and forum posts viewed are considered help-seeking activities, while instructor and student replies are referred to as help-giving. Together, help-seeking and help-giving constitute forum use or forum activities. In a study of help-seeking by novice programmers, Price et al. investigated factors that influence students and help-seeking behavior by analyzing 15 students' interviews. They found that participants trust help from a human tutor rather than a computer tutor. However, lower performing students are more likely to believe that asking for help is evidence of their academic incompetence, and see computer tutors as more accessible and less of a threat to their sense of independence.

Nelimarkka & Hellas studied sources of help, such as instructors, friends, and alumni, for participants of an introductory programming MOOC and how the behavior can be classified into help-seeking strategies. They found that students using social interaction strategies for help-seeking are less likely to drop the course. Students' engagement on online discussion forums have been found to correlate with students' performance.

Fritz [Fri11] and Gitinarbard et al. [Git19a] found that individual student engagement on on-line discussion forums positively correlated with individual student's performance. Romero et al. [Rom13] found that qualitative indicators such as messages, number of words or sentences in posts, reads, and time each student spent on the forum were useful in predicting their final individual performance in on-line discussion forums [Fri11].

Desai et al. developed discussion board networks for predicting final grade. They found a positive correlation between the number of posts made by a student and their academic performance in terms of the final grade[Des20]. Prior research on Piazza showed that the students tend to make communities on online forums where the students with similar performance connected to each other mostly and there were few links between the different communities [Git17c]. Gitinabard et al. showed that the students who asked and answered more questions and connected to more peers on the forum had better individual performance in class [Git17a]. In addition, students' social network properties such as degree, centrality, hub, and authority score as well as other online activity features could predict their individual performance in classes[Git19a].

Er et al. similarly explored online help-seeking behaviour in a large science classroom. They evaluated a web-based help-seeking tool, called EchoLu. They suggest four basic principles that positively influence students' decision to seek help. A discussion platform should (1) address privacy, e.g. allow students to post anonymously, (2) have a notification system for students to be aware of the availability of help, (3) ensure that all posts are visible to all students, and (4) promote social support by applying social interactions e.g. "likes" used by popular social networking sites. Piazza's current design adheres to these principles. The authors of Er et al. noticed that students using EchoLu used existing questions and answers to determine if they were actually in need of academic help, and what help they needed, or to review their understanding of the task. Piazza provides evidence of

this behavior through the number of unique views of posts on an assignment.

Chapter 3 of this dissertation describes class-level help-seeking and forum activity and explains a negative correlation between whole-class engagement of the forum and class average performance. The study also suggests what lower and higher levels of whole class participation in discussion forums might signal useful information to instructors about assignments presented to students.

2.0.5 Learning Analytics Around Teamwork Collaboration

2.0.5.1 Teamwork collaboration in Colleges

Teamwork collaboration is a necessary skill to learn for software engineering. Hamlyn-Harris et al [HH06] conducted a study to determine how to predict the satisfaction of group members on a team project. Data was collected from students rating agreeing or disagreeing with statements regarding their overall satisfaction and dissatisfaction of the group work. The survey also includes aspects of individuals such as previous experience with teamwork, perceived usefulness of team work training, familiarity with team members, and gender mix within the team. The first group to be surveyed was an introduction to databases course where students were working in groups of three to four students during a five week period. They chose to survey this class to analyse students' attitudes towards teamwork who may not have had previous experience with group work. The second group to be surveyed was a capstone software engineering course in which students worked on an eight month project. From the first group there was no correlation between teamwork satisfaction and usefulness of teamwork training, while a small correlation was found in the second group. There was no support for the hypothesis that working with friends would influence satisfaction. The final finding from the study was that working in mixed gender teams was more satisfying for students working in the software engineering course then students in the databases course [HH06]. This may have been from the time the students spent together, eight months versus five weeks, which allowed them to develop their communication skills with each other.

Maurice in 2018 explored data mining of students' GitHub VCS software repositories. The toolset they proposed was designed to enable course instructors who use GitHub in their classroom to gauge student success. The authors investigated the correlation between students' programming patterns on the projects and students' performance [Smi18].

2.0.5.2 Gender Differences in Collaboration

In a research study done by researchers from the University of Wisconsin, female students had less confidence in their ability to learn computer science concepts than male students in computer science. In fact, they felt less confident than the male students in non-computer science majors. The study also found that female students who have supportive friends and family members are more likely to major in computer science. The authors then concluded that they speculate the reassurance of one's competence by others may increase the probability that a woman musters the courage to enter into a maledominated major such as CS [**BeyerS2003**].

In their paper, Imtiaz et al demonstrated the capability of women in the open source community. The findings showed that women tend to have their pull requests accepted at a higher rate than men. In fact, women's pull requests are larger, indicating more lines edited. When accounting for the different programming languages, the findings showed women's pull requests written in almost all languages are more likely to be accepted. This can be explained in two categories. For insiders within an open source project, evidence showed little evidence of bias between women with gender-neutral profiles and women with gendered profiles, since both have similar acceptance rates. The researchers suspected that this is due to the fact that insiders likely know each other to some degree, since they all are authorized to make edits within the repositories, thus aware of each others' gender. However, this is not the case for outsiders within an open source project. Evidence showed a strong gender bias where women's acceptance rate dropped by 12.0% when their gender is identifiable. However, women continue to have high acceptance rates as they conduct pull requests on more projects than men [Imt19].

In their paper, Vasilescu et al studied the effect of gender and tenure on productivity in Github Open source software (OSS) development communities. They hypothesized that gender diversity has a positive effect on productivity and a negative effect on turnover while tenure diversity has a positive effect on both productivity and turnover. To prove their thesis, they conducted surveys and measured tem productivity by the number of commits by each team member. By using mixed effects analysis to measure significant predictive abilities of gender and tenure on productivity and turn out, they confirmed part of their thesis that gender and tenure diversity has a positive effect on productivity. However, they are unable to prove the negative effect of gender diversity on turnover and the positive effect of tenure diversity on productivity [Vas15].

In their paper, Judd et al. explored the differences in the amount and type of work completed by wikipedia contributors. Initially, the researchers believed that the gender gap may not be as large as previously thought. Antin et al explained that more recently, many of the gender differences can be explained by "occupational stereotypes" - preconceptions about the nature of specific jobs and the gender, ethnicity, or other descriptive characteristics of the people who do them. Therefore, a lack of gender equality in Wikipedia contributors may not necessarily be from active sexism from fellow Wikipedia contributors. To prove their hypothesis, they drew samples from 14,00 new users who had declared their genders on Wikipedia. Moreover, they found a gender distribution of 82% males and 18% women. These results proved that male Wikipedia contributors far outnumber female Wikipedia contributors. Furthermore, while they were examining the top 25% of Wikipedia contributors (based on citation and contents added and deleted) they found that the distribution is extremely skewed, where 27% of revisions are completed by women. Notably, two areas of work in which women made significantly larger revisions involved creative production, synthesis, and reorganization of text. This study suggests that the story of Wikipedia's "Gender Gap" is perhaps not as straightforward as initial reports have suggested [Ant11].

In their paper, Nafus et al. talked about how liberal notions of freedom have affected the F/LOSS (Free/libre/open source software) community. With only 1.5% of F/LOSS participants being female

compared to 28% of technical workers being female, it is particularly surprising for a community that is open to "everyone" [Naf12]. Nafus et al. conducted participant observation in three countries in Europe across multiple software projects, talked with contributors and leaders [Naf12]. In the study, they found that both men and women acknowledged there exists a problem of some kind in the open source community. However, men and a small number of women defend and protect this special form of "openness," despite the fact that it establishes an epistemological monoculture. From interviews of female participants in the community, they found that they were often sexualized, offended, and obliged to remind men to not stare at them. Some women took measures by hiding their gender and using gender-neutral names in these forums. While finding mentors, women were generally afraid to disclose their gender since most male mentors see this as a dating opportunity.

In chapter 4 of this dissertation we see what parts the female gender plays when they collaborate with males in mult-gender teams.

2.0.6 How Forum Activity Analytics is Used in the Classroom

Here, we will describe how forum analytics has been used in the classroom. Wise et al. in 2013 [Wis13], presented extracted and embedded analytics to students as a form of intervention while students used the discussion forum. They embedded real-time analysis, in the discussion interface and evaluated the impact of students' awareness of these analytics on help-seeking. Embedded analytics simply means that on the discussion forum student can see in real-time, how many posts they have participated in and how many more they are yet to participate in. Furthermore, they presented extracted analysis on the forum such as temporal distribution of participation, sessions, percent of sessions with posts, average session length, posts, average post length etc. During their semester-long study on a blended graduate seminar on educational technology, participants (instructors) appreciated the fact that the tool help them see that their students are making substantial non-visible even though they don't make visible posts. The analytics also detected dialogues where students do more talking but less listening [Wis13].

In addition, Moreno-Marcos et al. [MM19] introduced a methodology, named 3S (Social, Sentiments, Skills), for analyzing forum interactions and forum activity in MOOCs. The 3S methodology helps instructors to understand students' behaviors on the forum. Moreno-Marcos et al. discovered patterns such as the increase of activity near deadlines and period when instructor email reminders are sent out. Their work also found that not many learners use the forum Instructors. Lastly, they recommended that instructor should review assignment difficulty at certain points. The concept of adaptation should be introduced to ensure learners can complete the materials appropriately[MM19]. The PEDI tool discussed in chapter six of this dissertation already possesses features that could be instrumental in providing some analytics described by Moreno-Marcos et al. such as the line graph and histogram for tracking assignment difficulty.

2.0.7 Learning Analytics Dashboard (LAD)

LADs are basically designed for instructors to monitor students' progress and could be designed for the students to track their own progress. There are several approaches to administering LADs.

2.0.7.1 Process Mining

The process oriented approach considers the steps involved in the learning process as well as the frequency at which these processes are performed by students. Process mining is increasingly recognised as important for understanding and improving learning. LADs typically apply educational drill-down operations which allow instructors to explore the behaviour of sub-cohorts of students by filtering, an otherwise huge volume of information, using meaningful features and criteria relevant to the educational setting such as performance, Other drill downs include, drill downs by gender, demography, prior experience, background. Shabaninejad et al. describes these kinds of drill-downs as manual.

2.0.7.2 Automatic Recommendations

In their work on Recommending Insightful Drill-Downs Based on Learning Processes for Learning Analytics Dashboards, Shabaninejad et al. automated the drill-down process to eliminate the challenge of manually finding drill-downs that provide insights. Three aspects of students' learning behaviour should be displayed by LADs: 1. performed learning activities should be presented, 2. the frequency of each activity and 3. the order in which the activities are performed. It is advisable to implement LADs to solely track the learning process and seek to improve them rather than merely display frequency of engagement metrics. Shabaninejad et al. in their paper provided meaningful extensions to LADs with a functionality that provides recommendations of insightful drill-downs.

Common problem with LADs according to prior work [Sch17] [Bog18]], existing LADs have not been effective in allowing instructors to identify students' cohorts of interest. To tackle this problem, [Sha20a] recommends drill-downs based on the difference between cohorts' attribute values and while Shabaninejad et al. designed and implemented automatic recommendations based on the differences between cohorts learning processes.

Shabaninejad et al.'s algorithm evaluates all the possible drill-down actions to find the drill-downs that result in the students' group with the most anomaly in terms of learning processes. They defined anomalous cohorts as students cohorts with the highest distance from the rest when considering certain criteria such as performance and demographic information. The recommended filters by the system proposed in Shabaninejad et al. resulted in the display of sub-cohort whose learning processes are 72% different from the remaining students.

Other forms of LADs are animated visualizations. Sung et al. explored dynamic computer visualizations (DCVs) which are computer-based animated video clips of molecule interactions tp demonstrate the positive impact of short DCVs on students' performance on a poorly understood science concept - osmosis [Sun17].

In chapter 5, I designed and implemented PEDI, a LAD to monitor useful data indicators discovered in study one. Chapter 6 describes instructors' response to using the PEDI tool.

2.0.8 Summary

To summarize, prior literature has shown that:

- Technological advancements are the driving force of Learning Analytics in higher education today. Learning analytics brings about data-informed decisions and innovations in our colleges and universities.
- The combination of technology tools, application of data mining and machine learning methods on educational data make learning analytics possible.
- The importance of Learning Analytics cannot be overemphasized and will continue to be essential for improvement of student learning, performance and all round educational outcomes. When we know through the power of LA, how demographics, gender, background and diversity in general affects learning and students performance, we are able to make the necessary changes to pedagogical methods to accommodate these different backgrounds.
- Learning Analytics is important for instructor accountability and instructor assessment.
- Learning Analytics contributes to assessment of instructional methods and curriculum design.
- To achieve more robust data-driven decision making systems, integrating knowledge and analytics from several data collection systems and domains is inevitable; however, LA on individual systems are yet to be fully explored.

The following studies are organized as follows. Study 1 investigates class-level forum usage; it is also the first study aimed to understand how the whole class uses discussion forums. Study 2 applies machine learning and natural language processing techniques to understand how different genders collaborate in a multi-gender team. Study 3 reports the designs of an analytics dashboard layer on top of the Piazza platform for instructor insights and to spur interventions. Study 4 outlines the plan for a qualitative assessment on how instructors perceive the PEDI dashboard designed in study 3. These combined studies derive practical, actionable learning analytics and pedagogical insights from students' and instructors' data using both quantitative and qualitative approaches.

CHAPTER

3

STUDY 1: EXPLORING CLASS LEVEL FORUM USAGE FOR INSTRUCTOR INSIGHTS

abstract Discussing a problem is the first step in help-seeking, and online forums are one of the fastest ways students can communicate their concerns, doubts, and requests for help in a blended classroom environment. There are many studies aimed at detecting the relationship between a student's individual forum use and their individual performance on assignments. These studies found positive correlation between forum use and performance or final grades. In this study, we investigate *whole-class* Piazza forum use, and show a negative correlation between whole-class forum use and whole-class performance on programming projects in a second-year computer science course. On projects where the class performance was low, there was relatively more class forum activity and lower forum use on projects where the class performance was relatively high. We investigate the timing of forum use and show that activity trends generally peak closer to the assignment due date. In an interview, the course instructor, who has been teaching the course for over 10 years, confirmed that high Piazza forum activity correctly related to assignments where students faced difficulty, and that students are often deadline-driven, waiting until deadlines are near to begin work. When comparing class-level forum use across semesters, we show how the addition of intermediate deadlines/milestones for multi-week programming projects encourages earlier forum activity and positively correlates with improved project performance. We conclude with recommendations for how instructors may monitor and influence forum use to promote better

class performance.

3.1 Introduction

Prior research has shown that students learn better when their learning experience includes social interactions [Hur13]. Alvermann et al. (2012), for example, defines social interaction as a meaningful dialogue among learners and asserts that learning is a social activity which involves reading, writing, listening, speaking and thinking [Alv12]. Online discussion forums are the key piece of infrastructure that facilitate all of these activities in blended learning environments which have both in-person and online components. Questions asked on discussion forums are beneficial to the class as students can observe what kinds of help other students are seeking and what concepts or tasks they are struggling on [Er15].

This work is the first of many studies aimed at understanding class-level forum activity, monitoring forum traffic and what it can tell us about programming projects. We define whole-class, or class-level, forum activity as an aggregation of all forum activity by both students and instructors in the class. The purpose of this research is to learn about class-level forum usage and what this can tell instructors about assignments in their class. We do not take into account analysis about student's individual forum usage or when each student is working on a project. Instead, we ask the following question: what can we learn from class-level forum usage data alone? We will directly compare *class-level forum activities* on a programming project and *class-level performance* on that project. We define class-level performance as the class average performance, i.e. the total class performance divided by the number of students in the class who attempted the project (N). (Students who attempted the project and scored zero were included).

We are also interested in interventions such as prompting early forum use on assignments to ensure quality programming projects. Edwards et al. (2009) and Kazerouni et al. (2017) found that students who begin working on programming assignments early have more time to seek assistance when stuck and they submit better quality projects [Edw15b; Edw09b; Kaz17b]. Hence, prior work has shown that it is good to seek help and complete programming assignments early.

In this chapter, we analyze class-level forum activities on Piazza, a popular online discussion forum used in a Java Programming course of a south-eastern research-intensive US university. We examined six (6) consecutive offerings of CS2-OOP [Por18b], a Java programming class in the Spring (January-May) and Fall (August-December) semesters of 2017 through 2019. The course offerings were identical, with similar schedules at the same university. The course is described through enrollments (Table 4.1), assignments (Table 3.2), assignment timing (Table 3.3), average class performance (Table 3.4), and assignment completion mode (Table 3.5). In total we analyzed Piazza usage and course performance of 1163 undergraduate students.

Piazza is a popular wiki-based question and answer discussion forum used by 2,000 Universities across 90 countries with over 2M students [Pia]. Many instructors who use the platform, including the instructors of the classes we analyzed, direct students to ask questions on the forum rather

than via email, enabling teaching staff and students to share the load of answering questions while students with similar issues can also use the information exchanged in public posts. It is worth mentioning that in the classes we analyzed, there was no incentive for forum participation. Piazza allows instructors and students to categorize threads into "folders", which are non-exclusive tags that can represent assignments, exams, or other class topics. This feature can help students find information about a specific assignment easily. The instructor provides a folder for each of the major assignments and course logistics. Students ask questions by creating a post on the forum and are required to tag the folder relevant to their question. Both teaching staff and students are free to respond to public posts and for every question asked on the forum, Piazza records information such as the time the question was asked, question content, number and content of replies, number of views, and number of likes.

Piazza was designed to stimulate students to seek help from instructors and help each other. Bull et al. (2001) defined help-seeking in discussion forums to be of two kinds: visible help-seeking and lurking. Visible help-seekers are active forum posters who may also provide answers to fellow students' questions, whereas lurking refers to students who actively view posts in search of existing posts with answers to their questions [Bul01]. In this work, we focus on two types of help-seeking activities: asking questions and viewing prior posts. We refer to instructor and student answers to posts as help-giving. Together, help-seeking and help-giving cover all forum use or forum activities. Therefore, we define *class-level forum activity* for a class project to include both help-seeking (students' posts/views) and help-giving (answers from students and teaching staff).

In this chapter, we investigate the following research questions:

- RQ1.1: What are the patterns of class-level student forum use for an assignment and what can these patterns indicate to instructors about that programming assignment?
  - What does it mean when students ask many questions on an assignment or project?
  - What can class-level forum activities tell us about an assignment?
- RQ1.2: What is the timing of class-level interactions in the class and consequently how can we encourage early forum use?
  - Do students ask questions on the forum early in a programming project?
  - What strategies encourage early interactions and help seeking on the discussion forum?

To answer these questions, we visualized features of class-level forum activity as described in Table 3.6 and performed correlation and regression analyses to detect how these features relate with class average performance.

3.2 Background

Computing education researchers have become increasingly interested in the social aspects of learning as students work on programming assignments. Social interaction forums have recently

Table 3.1 Number of Students in each Semester

Semester	2017	2018	2019
Spring	147	150	172
Fall	203	247	248

Table 3.2 Programming Projects in a CS2-OOP. Complexity (out of 5) was defined by the Course Instructor

Project	Project Name	Topic	Complexity
gp1	Guide project 1	Composition and SE processes and practices	3
gp2	Guide project 2	Inheritance	3
gp3	Guide project 3	Testing	2
p0	Project 0	OO (Implementation and Test)	4
p1	Project 1	Finite State Machines (Implementation and Test)	4
p2	Project 2	Linear Data Structures (Implementation and Test)	4

Table 3.3 Duration of each assignment (Days)

Project	Spring 2017	Fall 2017	Spring 2018	Fall 2018	Spring 2019	Fall 2019
gp1	15	18	15	16	12	19
gp2	11	13	12	23	18	11
gp3	8	13	8	24	17	13
p0	23	23	23	16	-	-
p1	16	18	15	15	31	23
p2	17	30	18	16	47	23

Table 3.4 Average Class Performance on each assignment (Out of 100). Assignment with the lowest average performance in a semester is in italics, while that with the highest performance is in bold font

Project	Spring 2017	Fall 2017	Spring 2018	Fall 2018	Spring 2019	Fall 2019
gp1	79.98	90.00	91.64	87.88	91.95	89.36
gp2	**86.26**	**90.58**	**94.85**	**91.93**	**93.40**	**92.21**
gp3	85.24	89.90	93.17	89.30	91.46	91.21
p0	75.52	*64.12*	*73.42*	*72.18*	-	-
p1	76.29	84.04	88.00	82.74	84.39	*74.96*
p2	*73.02*	79.67	81.78	74.58	*68.37*	81.11

been employed for help-seeking and help-giving purposes [Nel18]. Er et al. (2015) explored online forum use behaviour in a large science classroom [Er15]. They evaluated a web-based help-seeking tool, called EchoLu and suggested four basic principles that an effective social help-seeking forum should possess. A social help-seeking forum should (1) address privacy, e.g. allow students to post anonymously, (2) have a notification system for students to be aware of the availability of help, (3) ensure that posts are visible to all students, and (4) promote social support by applying social

Table 3.5 Work Mode in Main Course Projects. Note that: All Guided Projects (GPs) were done individually

Project	**Spring 2017**	**Fall 2017**	**Spring 2018**	**Fall 2018**	**Spring 2019**	**Fall 2019**
p0	Individual	Individual	Individual	Individual	-	-
p1	Optional Pairs	Optional Pairs	Optional Pairs	Optional Pairs	Individual	Individual
p2	Required Pairs	Required Pairs	Required Pairs	Required Pairs	Optional Pairs	Optional Pairs

Table 3.6 Definition of Features of Piazza Posts

Features of Forum Activity	**Definition N= total number of students in the class who attempted the project**
Average # of Posts	(Total number of all posts related to an assignment)/ N
Average # of Likes	(Total number of likes on all posts related to an assignment)/ N
Fraction of the class that Posted	(Total number of students that posted about an assignment)/N
Fraction of the class that Viewed (Unique Views)	(Total number of students that viewed all posts of an assignment)/N
Average # of Questions	(Total number of questions asked by students in an assignment)/N
Average # of Replies	(Total number of replies by both students and instructors in an assignment)/N
% of Early Posts	(Percentage of total posts made half-way through the project)/N

interactions e.g. “likes” used by popular social networking sites. Piazza’s current design adheres to these principles. Er et al. (2015) noticed that students using EchoLu viewed existing questions posted by other students and corresponding answers to determine if they were actually in need of academic help, and what help they needed, or to review their understanding of the task [Er15]. Piazza provides evidence of this behavior through the number of post views on the platform.

3.2.1 Class-level Performance Prediction

One of the most important applications of educational data mining (EDM) is to predict students’ performance. Early prediction of poor performance is beneficial for effective intervention [Des20; Rab11]. Many EDM approaches have been used to predict students’ performance including a decision support tool based on students’ key demographic characteristics [Kot11], social network analysis, [Des20; Git19a] and discussion forum participation [Che11; Des20; Cor17; Git17b; Lóp12; Rom13; Vel17]. Class-level or class average performance on the other hand has rarely been explored. Class-level performance has been beneficial in evaluating the overall performance of student cohorts and comparing the effects of instructional methods [Win14; Rob14]. Hence class-level performance

on an assignment can be used by instructors to determine whether the assignment was deemed difficult by the class and evaluate the impacts of impacts of changes to assignment specifications. We are not aware of any prior work that has investigated the relationship between class-level forum usage and class-level performance. In this chapter, we describe how forum usage analysis can portend class average performance and signal information about a class and assignment to instructors.

3.2.2 Forum Usage and Performance

There are so many studies that found positive correlation between individual forum usage and performance [Che11; Des20; Cor17; Lóp12; Git17b; Vel17; Rom13]. Experiments by Cheng et al. (2011) showed a positive relationship between forum participation and course performance [Che11]. Desai et al.(2020) employed student interaction networks to measure students' engagements in discussion forums and found a positive correlation between a student's final grade and the number of posts made by the student [Des20]. Nelimarkka et al. (2018) explored social help-seeking strategies including sources of help for students in a programming MOOC, they discovered that students who seek help on social forums are less likely to drop the course [Nel18]. Gitinabard et al. (2019) explored social interactions and students' online behaviors in MOOCS. They found that students' social metrics are better than the students' study behaviors in predicting student performance [Git19a]. Romero et al. (2013) similarly found that qualitative indicators such as number of words in posts and time a student spent on the forum were useful in predicting their final individual performance [Rom13]. Lopez et al. (2012) found that students' participation in forums was a good predictor of their final performance [Lóp12]. Corrin et al. (2017) clustered learners based on forum use behavior and discovered that both passive and active engagement with discussion forums resulted in higher students' performances [Cor17]. Gitinabard et al. (2017) showed that students who ask more questions and receive more feedback tend to obtain higher grades [Git17b]. Vellukunnel et al. (2017) also studied posts content of 395 CS2 students across different Universities and discovered that some students' questions are useful in spurring discussions and that forum participation was positively associated with final grade[Vel17]. All these studies analyzed each student's forum usage and how it related with their performance individually. Our study is unique in that we find the correlation between levels of forum usage by the whole class while working on an assignment and what these levels indicate to instructors about that assignment. In this current study, we seek to understand what drives class-level forum activity and how it relates to *class average performance*. We monitor the forum use trend over six semesters and what it signals to the instructor.

3.2.3 Theory about the Effects of Milestones in Programming project

There are a number of theories that describe the effects of milestones on programming projects [Sha21]. Older literature [G75] introduced *Goal Theory*, a phenomenon where procrastination is highly likely to occur when goals are farther in the future. Ponton et al suggested breaking down large tasks into subtasks with deadlines [K01] whereas Bandura et al showed that an individual's self-efficacy and likelihood of success in smaller sub-tasks are higher than in larger all-encompassing

tasks, [Ban97]. Shaffer et al. in their 2021 study demonstrated how intermediate milestones inhibited procrastination and led to better student outcomes in programming projects [Sha21]. They found that milestones helped students to divide the projects into reasonable sub-tasks, helped students focus on one sub-task at a time and helped them keep track of their progress. They also showed in their survey that (75%) of the students participants in one semester found milestones to be helpful [Sha21].

Figure 3.1 The normalized number of forum posts and average performance for projects p0, p1 and p2 in Spring and Fall semesters of 2017-2019.

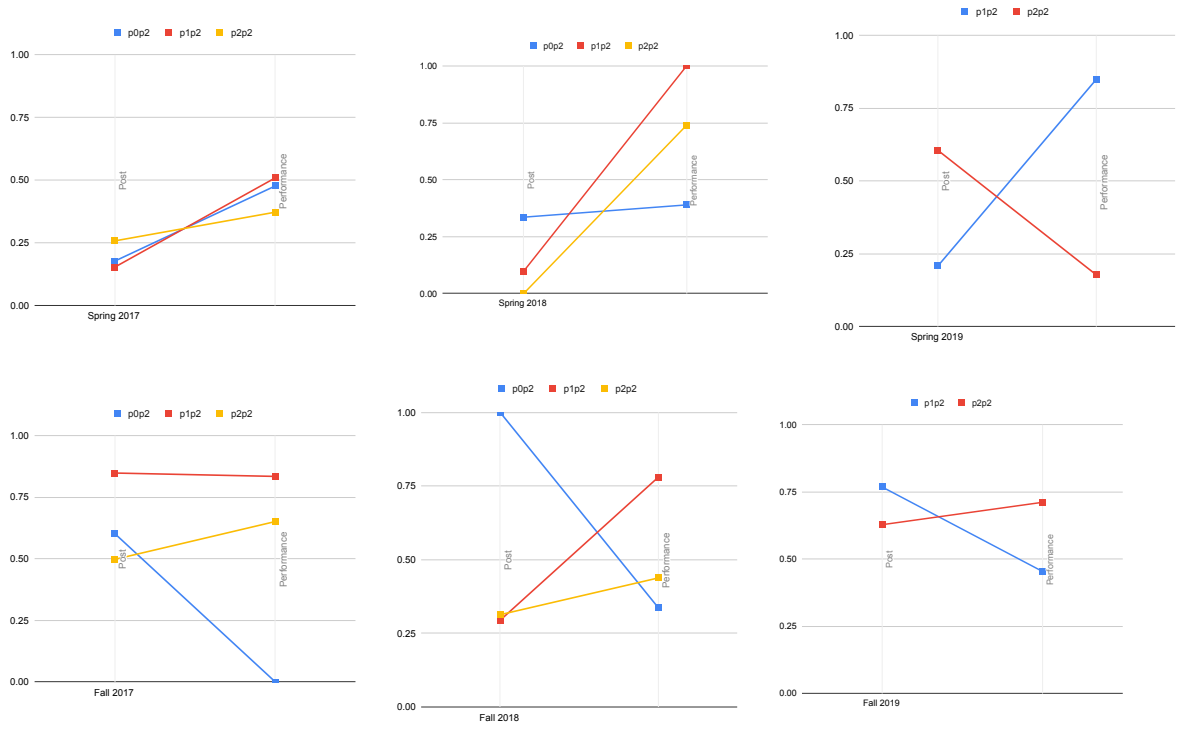


Table 3.7 Table comparing posts and performance

Project	Spring 2017	Fall 2017	Spring 2018	Fall 2018	Spring 2019	Fall 2019
gp1	79.98	90.00	91.64	87.88	91.95	89.36
gp2	**86.26**	**90.58**	**94.85**	**91.93**	**93.40**	**92.21**
gp3	85.24	89.90	93.17	89.30	91.46	91.21
p0	75.52	*64.12*	*73.42*	*72.18*	-	-
p1	76.29	84.04	88.00	82.74	84.39	*74.96*
p2	*73.02*	79.67	81.78	74.58	*68.37*	81.11

3.3 Methods

3.3.1 Data Collection

We began with an exploratory analysis of how students use the Piazza forum on programming projects. Piazza data was collected as JSON files. Of the 1163 students (see Table 4.1) we studied, 74.3% of them posted at least once and 98.6% viewed questions posted on the forum. These statistics are similar to the discovery in [Vel17] where data showed that 80% of the 560 students enrolled in classes posted to the discussion forum at least once during the course. The attributes of each post in our dataset are: folder name (e.g. gp1, p0), post content, created date, number of likes, post type (question, student answer, instructor answer, follow-up etc.), and number of unique views (i.e. number of students that viewed the post). We extracted ***Relevant posts*** for each project from the dataset using the student-assigned folder and the created date. For example, suppose guided project 1 (gp1)'s final due date is the 29th of August, then we grouped all posts with folder name "gp1", created before the due date 29th of August as *relevant posts* for gp1. We did not include posts in folder gp1 created after the 29th of August because these posts are about post-project information and would not affect performance in gp1. Furthermore, we manually checked *relevant posts* for each project to confirm that they are help-seeking and help-giving discussions related to that project. Analyzing this information at a lower-level (i.e. for each project/folder) rather than general forum use throughout a semester described in prior research [Che11; Des20; Cor17; Git17b; Lóp12; Rom13; Vel17] can provide insights into the class' forum use behaviors for a particular project. Piazza keeps a record of both the original and updated versions of posts, we ensured that only one version of the posts were included in the analysis by removing duplicate posts. The collected assignments include guided projects (gps) (these are projects with an interleaving of guided and independent tasks intended to introduce course technologies and review prerequisite material), and two to three main course projects (p0, p1, p2) see Table 3.2.

For the main course projects students write code to implement and test a given design. In this analysis we focus on work from the guided projects and the main course projects because we are interested in programming tasks. The word *"projects"* is used interchangeably with *"assignments"* throughout the paper. We explored all six (6) semesters and visualized the relationship between posts and performance for the main course projects of these six semesters in Figure 3.1.

3.3.1.1 Features of Forum Activity

We selected features of forum activity from Piazza data such as posts (which consist of questions, replies, post likes, and post views). Based on post type, we extracted two features from the post feature, namely questions and replies, to investigate what each of these can individually tell us about the assignment. Questions include initial questions posted by students while replies include instructor reply, student reply and follow-ups. Our features are defined in Table 3.6. Since each semester has varying numbers of students in each cohort, we took the average of the features of

forum activity and applied them to predict average class performance. Early posts are the number of posts made mid way through each assignment. We chose this feature to see if early interactions on an assignment could predict average performance on the assignment.

3.3.1.2 Complexity of a Project

The course instructor was asked to rate the relative complexity level for each project, ranging from 1 for the easiest to 5 for the most complex. The more complex a project is, the higher the rating the instructor gave. These ratings are found in Table 3.2. Our goal in obtaining these ratings is to assess how the instructor's estimated difficulty aligns with the class-level performance. For example, an instructor might rank a project as easy but might have included additional information in the assignment specification that may make the project more difficult as we would observe in subsection 3.5.2.

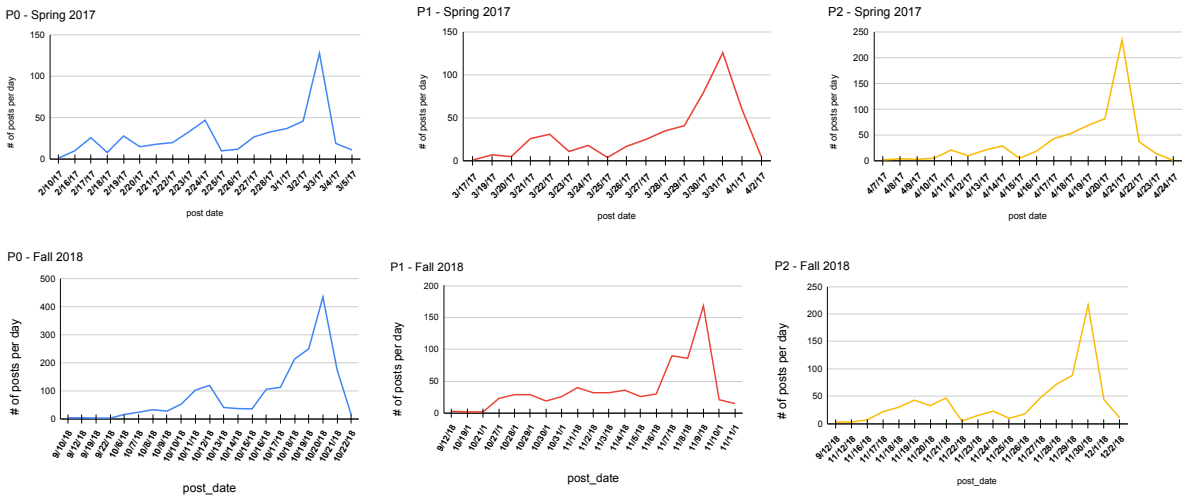
3.3.1.3 Duration of a Project

The duration of each project is displayed in Table 3.3. We define duration as the time students began seeking help on a specific project till the project due date. The relatively longer duration of some assignments compared to others in Table 3.3 shows that sometimes some students began working on assignments as soon as they were released (e.g. gp2 in Fall 2018 and p2 in Fall 2017) rather than after the deadline of the previous project. Instructors for this course often post projects early so students can inspect the project requirements, and for those who would like to work ahead (e.g. in some semesters, the instructor opened gp2 and gp3 before the gp1 deadline for students interested in working ahead). Another approach would be to assume that the start date is the day after the previous deadline, but with late windows, allowing for a prior assignment to be turned in after its original deadline, there is no definite start date. Instead, **our analysis reflects the time when students are actually working on the project as evidenced by forum posts.** Projects in Spring and Fall 2019 generally had longer durations, because students completed two main course projects (p1, p2) instead of three (p0, p1, p2).

3.3.2 Exploratory Analysis

We calculated the features defined in Table 3.6 and visually explored their relationship with project performance. We chose these features because research has shown them to be useful indicators that help-seeking and help-giving exist in the classroom [Er15]. In Spring 2019 and Fall 2019, the course included three projects instead of two. Instructors explained that p0 covered the same material as gp3, and having only two projects allowed for the creation of additional project milestones to encourage students to start (and possibly complete) their work earlier. We visually explored the relationship between class-level activity and class-level performance, the impact of an additional milestone and different work modes in course projects.

Figure 3.2 Projects in Spring 2017 and Fall 2018 showing forum use peaks a few days before the deadline



3.3.2.1 Class-level Forum Activity vs. Class-level Performance

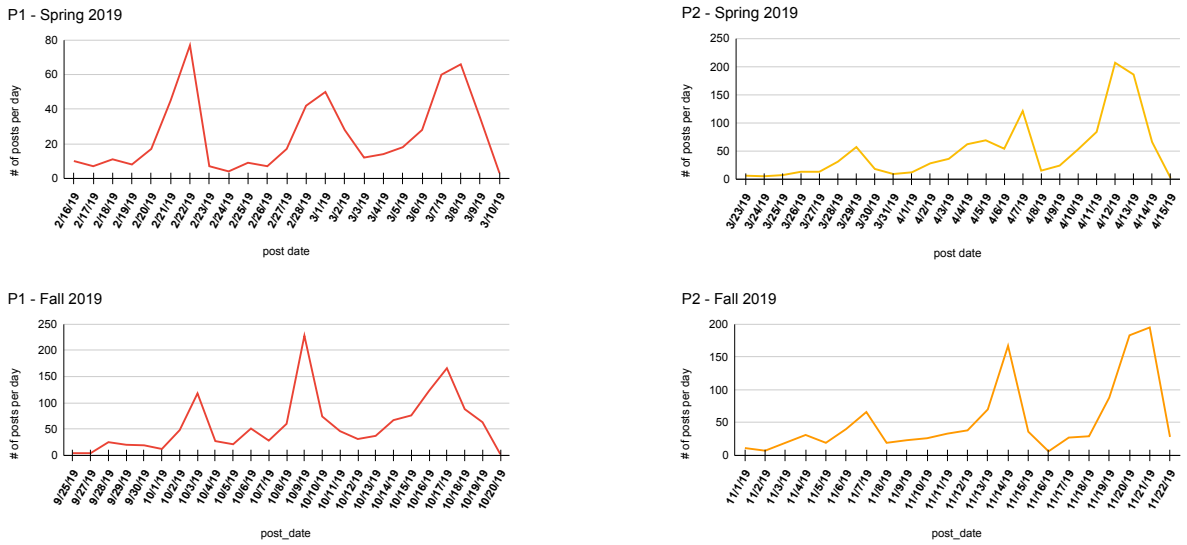
To investigate our hypothesis that the class-level forum activity are related to class average performance on a programming assignment, we plotted the interactions between class-level posts and performance in Figure 3.1. We normalized the number of forum posts in each assignment for easier comparison using min-max normalization (0-1). We also normalized the average performance in Table 3.4 in the same way for the line/interaction plots in Figure 3.1. The min-max values before normalization can be found in Table 3.8.

Table 3.8 Min-Max of features in Assignment used for normalization. Note: Average Post = Number of Post/N (students who attempted the assignment)

Feature	Minimum	Maximum
Average Posts	1.71	7.68
Number of Posts	256	1794
Average Performance	64.12	88

We hypothesize that the higher the posts, the lower the performance in each class. This can be seen in the interaction plots with lines for projects with lower posts crossing the lines for projects with higher posts. This trend is confirmed in Figure 3.1 for all projects in all semesters, except for p1 in Fall 2017, and spring 2018. In Fall 2017, p1 had both the highest number of posts and the highest class performance. In Spring 2018, p1 had a medium number of posts but the highest performance.

Figure 3.3 Projects in Spring 2019 and Fall 2019 showing the effect of intermediate milestones



3.3.2.2 Visually exploring the Impact of an Additional Milestone

Another important observation is presented in Figure 3.2 showing that most students wait till a few days before the due date before actively using the forum while working on their programming project. This trend was observed during the first 4 semesters, including the Spring and Fall semesters of 2017 and 2018. In these 4 semesters, students had to meet one milestone (compilation/ commenting) one week before the deadline. This milestone requires students to complete a skeleton of the project from a provided design that would compile against the teaching staff tests. Students were also expected to fully comment their skeleton to help with clarifying the purpose of each method in the implementation. Previous studies also suggested that students' major part of activities are cramped right before deadlines [She18]. In 2019, (see Figure 3.3) the compilation/commenting milestone was moved to two weeks before the project deadline and the instructor introduced an additional milestone to be completed a week before the due date of the course project. Therefore, in addition to compiling the project, students were required to finish roughly half the project a week before the final due date (the details of "half" varied by semester). We noticed that this extra milestone added in spring and Fall 2019 led to an increased rate of forum use weeks before the due date. This suggests that adding milestones to programming projects could be a strategy to encourage early interactions on the discussion forum.

3.3.2.3 Posts with a High Number of Likes

We also explored the number of likes on a post and observed that posts are rarely liked by students. The probability of liking a post is 0.03 and only 37 posts got more than five likes across all semesters. Most posts had zero likes and the maximum number of likes a post got was 26. A high number of likes on a post could be important to instructors for two reasons. The first reason is that the post

might be particularly helpful for an assignment. For example, 11 out of the 37 posts with more than five likes were about how to address challenges with using the automated testing tool employed for the course. These challenges and their solutions could be anticipated in future offerings of the course projects. For example, such posts could be refined, generalized and provided to students along with the assignment.

The second important reason a post may have a high number of likes, is that it may indicate problems for student cohorts that may lead to lower class performance, as illustrated in two cases here. In Spring 2019, on p2, a student made a post during the last week of p2 about a long response time for teaching staff replies (see Spring 2019 trend in Figure 3.3). The post had 16 likes, and the average class performance for p2 was the lowest in the Spring 2019 (68.37) (see Table 3.4). In Fall 2019, on p1, a student posted before the second milestone deadline (see Fall 2019 trend in Figure 3.3) with concerns about a slightly different pre-announced deadline (Wednesday rather than Thursday) due to Fall break. This post recorded the highest number of likes - 26 likes, while the class average for p1 was lowest in Fall 2019 (74.96). These cases are anecdotal, but are interesting exploratory findings which indicate that teaching staff should consider monitoring for posts with high numbers of likes, as they seem to indicate particular struggles shared by student cohorts, that lead to lower class performance.

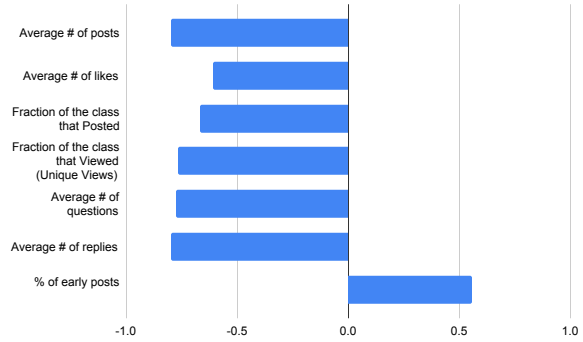
3.3.2.4 Work Modes in Main Course Projects

Table 3.5 shows students' work modes on the different project. Students were required to work in pairs when completing project p2 in Spring and Fall semesters of 2017 and 2018. We wanted to observe if students used the forum more when they worked individually than when they worked in pairs. By observing Figure 3.1, if this were the case p2 would have recorded lowest forum activity in spring and Fall semesters in 2017 and 2018. However, since we do not see significant differences in p2 posts, we cannot conclude that working in pairs leads to less forum activity.

3.3.3 Correlation Analysis

We performed direct Spearman's ranked correlation analysis comparing each of the selected features of forum activity with average performance for each project. We used Spearman's ranked correlation because we are comparing ratios and our data passes the assumptions that are required for Spearman's correlation[Spe]. For example, we wanted to know the relationship between average number of posts in a project (see definition in Table 3.6) and average class performance in that project. We took the average number of posts instead of the total number due to the different sizes of student cohorts. For example: Spring semesters generally have less students than the Fall semesters (see Table 4.1).

Figure 3.4 Correlation analysis



3.3.4 Regression Analysis

To confirm the patterns we observed in the exploratory analysis on class level forum activities and average class performance, we fit regression models to the data. We performed ordinary least squares (OLS) regression and Mixed Linear Model (MLM) regression analysis. MLM models are widely used for regression analysis to fit a linear model for the purposes of prediction [Pin]. These models extend OLS such that the linear predictor contains random effects in addition to the usual fixed effects. The fixed effects in our model represent the relationships between the features of each assignment and average performance on each assignment. The random effects are used to control for effects of groups within data [Bel19]. To account for the different student cohort in each semesters, we modeled semester_year as a random effect. We aim to achieve a model with good fit and observe the p-Value of each feature to confirm our exploratory analysis. We set our statistical significance threshold at $p < 0.05$. Our results are presented in the next section.

3.4 Results

Table 3.9 Correlation ρ and p-values of Features. Statistically significant p-values based on Bonferroni Correction are in bold font

Features	ρ	p-values(OLS)	p-values(MLM)
Average# of Posts	-0.800	**<0.001**	**<0.001**
Average # of likes	-0.609	0.022	0.032
Fraction of the class that Posted	-0.667	**<0.001**	**<0.001**
Fraction of the class that Viewed (Unique Views)	-0.768	**<0.001**	**<0.001**
Average # of questions	-0.776	**<0.001**	**<0.001**
Average # of Replies	-0.797	**<0.001**	**<0.001**
% of early posts	0.558	0.03	**<0.001**

3.4.1 RQ1.1: What are the patterns of class-level student forum use for an assignment and what can these patterns indicate to instructors about that programming assignment?

3.4.1.1 Correlation Results

We want to detect the patterns of students' forum use and what these patterns indicate to instructors about a programming assignment (RQ1.1). We also want to investigate strategies that encourage early interactions on the discussion forum (RQ1.2). We will now observe the correlations between all 7 features and our target, average performance on each project. From Figure 3.4 and Table 3.9, we can see that the average number of posts ($\rho = -0.800$), average number of likes ($\rho = -0.609$), the fraction of the class that posted ($\rho = -0.667$), viewed an assignment's posts ($\rho = -0.768$), average number of questions asked ($\rho = -0.776$) and replies ($\rho = -0.797$) are negatively correlated with average performance. The negative correlations between average posts, and the fraction of the class making posts with performance confirms our hypothesis that more posts by more students are made on assignments with a lower average class performance. The negative correlation between views and performance signals that relatively higher post views might be an indication that the class is struggling on an assignment. In the case of the post likes, this negative correlation indicates that high class likes relate to assignments where students are struggling (low class performance). Likewise, higher number of replies correlated with lower average performance on an assignment. One explanation for this could be that in a more complex assignment, teaching staff may need to post more replies to ask students to elaborate details that are needed for them to answer questions. This may not always be the case, therefore, analyzing the content of replies in depth could help determine if difficult assignments required more replies. On the other hand, the last feature we investigated, % of Early posts, is positively correlated ($\rho = 0.558$) with average performance. This means the higher the percentage of total posts made earlier (half-way through the project duration), the higher the performance on a programming assignment in the class. This result suggests that to an extent, the class average performance improves when the forum activities begin earlier in the project timeline.

3.4.1.2 Multicollinearity

Some of our features correlated with each other (e.g. average posts, questions, replies, and views). Average number of likes and questions also correlated with each other. To resolve this, for each regression analysis, we designed 5 sets of models in order to predict average performance using independent variables that were not multicollinear. For model 1 we predicted class-level performance using average replies and average number of likes as features. For model 2, we used average unique views and % of early posts, for model 3, we used average posts and % of early posts, for model 4 we used average number of questions and % of early posts and lastly, for model 5 we used fraction of the class that made a post and % of early posts as features to predict class average performance. All our models recorded **R-squared values that are greater than or equal to 0.6**.

3.4.2 Regression

Table 3.9 shows that the selected features have significant p-values indicating that they are significant predictors of class-level performance. After applying the Bonferroni Correction for multiple statistical tests, all features were found to be statistically significant predictors class-level performance except the average number of likes in both regression analysis and % of early posts in the OLS analysis.

3.5 Discussion

Our analysis indicates that difficulty or complexity of assignment related with increased forum activity. The average **number of questions** asked by the class is negatively correlated ($\rho = -0.776$) with class performance. This confirms the definition of help-seeking by Richardson et al. (2012) as the tendency to seek help when experiencing academic difficulties[RM12; Cro17]. The negative correlation ($\rho = -0.797$) between average replies and average performance might indicate that the higher the number of replies and follow ups (from other students who might be confused about the subject and need clarification), the lower the class average on that assignment. However, a deeper study on the content of replies is required to help us understand whether the initial answer was sufficient or led to more questions or confusion. In addition, the number of likes is an important feature because students actively pressed on the like button, meaning the post is important and addressed students' concerns or challenges. Some posts with the highest numbers of likes could be generalized and used as resources for the assignment in future semesters, but content analysis on the questions is needed to determine if that is appropriate.

We have identified patterns which indicate that complexity in assignments might lead to more forum activity which answers RQ1.1.

3.5.1 RQ1.2: What is the timing of class-level interactions in the class and consequently how can we encourage early forum use?

3.5.1.1 Early Forum Use Driven by the Introduction of Milestones

The motive for studying RQ1.2 is to determine the impact of encouraging early start or forum use on programming project performance. Related research [Edw15b; Edw09b] has analyzed the negative effects of procrastination on the quality of programming assignments and found a statistically significant relationship between the time when the assignment is completed and its quality. Early forum use is an indication that students have begun to work on their assignments early in the timeline. Our analysis showed that students' forum usage increased in the last few days before a deadline. The feature - % early posts (i.e. the percentage of posts made mid-way through the duration of assignment divided by N) is positively correlated ($\rho = 0.558$) with average performance (see Table 3.9). This indicates that more forum activity by the class, earlier in the project timeline, was observed for assignments with higher class average performance.

Our exploratory analysis revealed that most of the questions asked by the class in Fall and Spring semesters of 2017 and 2018, were asked at the last minute in 2017 and 2018. Figure 3.2 visualizes how forum use peaks a few days before the deadline. This is consistent with other work on general study habits which suggested that the majority of student work occurs right before deadlines [She18]. A data-informed solution to late help seeking can be observed in Figure 3.3, which shows that addition of a second milestone (i.e. deadline) increased early forum activity. Our exploratory analysis suggests that introducing milestones can improve class forum use and may encourage earlier student work. The positive correlation found between % of early posts and class-level performance and ability of % of early post feature to significantly predict class-level performance also suggest that early forum use relates with improved performance.

Also, some of the posts with high likes described in subsubsection 3.3.2.3 show that teaching staff might not have sufficient time to respond to questions asked right before deadlines. Providing students with resources such as project specific posts and replies relevant to previous offerings of the course could be helpful with this challenge. This could be achieved with an automated post recommendation system or a frequently asked questions (FAQ) database, but additional analysis on post content is needed to see if this is a viable option. An alternative could be to schedule deadlines during times when the teaching staff are most available. An instructor dashboard presenting the level of real time learning analytics in Figures 3.2 could help the instructor and teaching assistants to target additional support to meet students' needs or to identify persistent issues. A system for comparing current levels of forum activity with levels in similar previous offerings of the course could signal the instructor to encourage more forum use or introduce an intervention (e.g. a deadline) for the class to use the forum earlier on an assignment when the level of interaction is lower than expected.

To determine whether the exploratory and correlation results we observed align with instructor experience, we presented inferences from our exploratory, correlation and regression analyses to the main course instructor who has been teaching the course for over 10 years. The instructor observed these results and provided feedback on how the forum analysis related to classroom practice and experiences.

3.5.2 Instructor Feedback

The instructor confirmed the analysis results that **the most difficult assignments had higher rates of forum posts**. While, students ask more questions when assignments are difficult, as inferred by the strong negative correlation between average number of students' questions and average class performance, and confirmed by the instructor, we noticed a different pattern in Fall 2017. Project p1 recorded the highest number of posts but highest performance (see Fall 2017 in Figure 3.1). We asked the instructor to shed more light on this exception in Fall 2017. A senior teaching assistant, serving as a co-instructor, wrote p1 in Fall 2017. The project was challenging conceptually, but was straightforward to implement once the students understood the concepts. The project was also structured differently from other finite state machine projects used in other semesters of the course.

This suggests that although students found p1 conceptually complicated, they figured it out before the end, leading to better performance.

Reflecting on project p0 which had the the highest forum activity and lowest performance in Fall 2018, the instructor revealed that the p0 was a simulation project. There was some complexity in how the items in the simulation were generated and processed that was not fully explained in the project specification. Based upon our analysis and interviews with the course instructor, we also concluded that some aspects of performance were impacted by factors outside of the course design. For example, some of the lower performance by students on p1 in Fall 2019 (see Figure 3.1) relative to p1 in Spring 2019 can be attributed to a change in scope due to a change in Java versions. When the project moved to Java 11, a library previously used to support file I/O was no longer available. Students were then tasked with implementing their own file I/O classes rather than working with provided code. A difference was also noted on student performance on p2 between Spring 2019 and Fall 2019. In Spring 2019, there was a circular dependency in the project that made implementation and testing challenging. The circular dependency was removed in the Fall 2019 project p2. Class average performance in p2 was 68.37% in Spring 2019 and 81.11% when the dependency was removed in Fall 2019. Based on our analysis, the instructor reflected on the differences between Spring 2019 and Fall 2019 indicated an interest in continuing student implementation of file I/O as part of programming projects. Students have historically struggled with the file I/O task in gp1, so additional practice in the projects is needed. However, minimizing dependencies, particularly circular dependencies, is important so that students can focus on the project learning outcomes of finite state machines and linear data structures. Higher forum use therefore reflected this shared difficulty faced by the class.

Furthermore, we asked the instructor whether presenting generalized versions of posts with high likes could help future course offerings. The instructor showed interest, but also expressed that the summary may not lead to changes in the projects. The instructor explained that challenges may result from course goals (e.g. it may be important for students to overcome the challenge themselves). Or, the posts with many likes could be too semester-specific to provide value to future semesters. Additional analysis on post content will be needed. The instructor suggested that automatic nudges to TAs about unanswered questions could have prevented the issue of long response times discussed in subsubsection 3.3.2.3.

Overall, the instructor was excited that the addition of an extra milestone improved early forum use and positively correlated with performance. We conclude based on our analysis and the instructor interview, that the class forum analysis shows patterns in forum use and exposes some complexities in assignments, and some ways that real-time analysis of forum activity could be useful for instructors.

3.6 Limitations

The goal of this study is to explore class forum activity and observe what these alone can tell us about project assignments in a programming class. We also investigated the relationship between class-level forum activities on each programming assignment and *class-level performance*. We did not perform semantic text analysis to confirm the quality of help-seeking or help-giving for each post because we are more interested in forum traffic for a given project, however, we meticulously extracted posts relevant to each project in the manner discussed in subsection 4.3.1. One limitation of this study is that there might be other factors (not explored in this study) that might play a huge role in class-level performance such as out-of-forum help sources and resources, in-person office hours with teaching staff, different working habits of students and differences between student cohorts. Another limitation is that students receive notifications via emails about activities on the platform. When these are viewed via email, some students might not bother to view these posts on the Piazza platform, therefore the “number of post views" used in this analysis, is the minimum number of views of a post on the platform itself.

3.7 Conclusion

We performed an exploratory analysis to visualize forum use over 6 semesters of a data structures in Java programming course and compared them with average performance on their related assignments. We also performed correlation as well as regression analyses. Our results suggest that it is good practice to require students to meet intermediate milestones because early forum use was detected around milestones that were introduced in Spring and Fall 2019 offerings of the course. Early forum also positively correlated with class average performance. Related work presented in section 4.2 provides solid theoretical grounding for the impact of milestones in programming projects. We shared our findings with a course instructor who observed the positive impact of an additional milestone on early forum activity. Trends like higher forum activity which correlated with lower class performance might signal to instructors to check the forum to detect the reason for the high traffic. Another trend is that high early forum posts correlated with high class performance. Based on this result, we believe that instructors should find ways to encourage early forum use since it might improve class performance. In our future work, we plan to build and test a forum analytics dashboard and investigate how instructors may wish to use forum trends and comparisons to intervene to improve class performance.

CHAPTER

4

REPLICATING STUDY ONE

abstract The aim of this chapter is to replicate the analysis in study one (Chapter three) on another course, in order to improve the generalizability of our results. In this study, we first identify similarities and differences between the course we had previously investigated in terms of assignments and class logistics and then proceed to describe the analysis set up for the replication study. We perform some exploratory analysis of *whole-class* Piazza forum use during assignments in the math course and show a negative correlation between whole-class forum use and whole-class performance on assignments in the course. Our results confirm some of the findings in chapter 3 and we discuss why replicating some of the findings were not feasible in this particular math course.

4.1 Introduction

Prior research has shown that when students ask for help on the forum, they have usually begun working on that assignment, and may have been working on it for a while [B9]. We also found that students often use the Piazza forum near an assignment deadline rather than seeking help early. However, research shows that students who start working earlier have more time to seek assistance when they get stuck [Edw15a; Edw09a; Kaz17a].

In this chapter, we analyze class-level forum activities on Piazza, a popular online discussion forum used in a Discrete Math course of a south-eastern research-intensive US university. We examined seven (7) Fall offerings of the course in Fall semesters of 2014 through 2020. The course offerings were identical in each semester, with similar schedules at the same university. The course is described through enrollments (Table 4.1), assignments (Table 4.2), assignment duration (Table A.1 and A.2) and average class performance (Table 4.4). In total, we analyzed Piazza usage and

course performance of 2116 undergraduate students.

We will re-iterate the research questions from study one [B9]:

- RQ1.1: What are the patterns of class-level student forum use for an assignment and what can these patterns indicate to instructors about that assignment?
  - What does it mean when students ask many questions on an assignment or assignment?
  - What can class-level forum activities tell us about an assignment?
- RQ1.2: What is the timing of class-level interactions in the class and consequently how can we encourage early forum use?
  - Do students ask questions on the forum early in an assignment?
  - What strategies encourage early interactions and help seeking on the discussion forum?

In chapter 3, we answered RQ 1.1 by showing a negative correlation between the first five features of features of forum interactions on Table 4.3 and class performance, and a positive correlation between early forum interaction and class performance. In this chapter, we perform a similar analysis on the math course and report our findings.

In study one, RQ 1.2 was answered by showing patterns of help-seeking and forum use. We detected a common pattern among assignments and visually observed the impact of milestones on early forum use and class performance.

The course allows students to request automatic, individual, 3-day extensions on each assignment. Since students are aware of this benefit, it might influence the timing of early help-seeking. Students might not begin early if they know that they can get an extension.

Figure 4.1 shows the number of posts per day in homeworks 5 - 9 and all the tests in 2015 semester. The red line indicate deadlines. We will use 2014, 2015 and 2018 semesters as an example to represent the patterns found in other semesters. Figure A.1 (see appendix) shows the charts for 2014, Figure 4.1 shows the chart for 2015 and Figure A.2 (see appendix) shows the chart for 2018.

Based on findings in study one, higher rates of interaction are expected towards the deadline. This pattern is found in most assignments

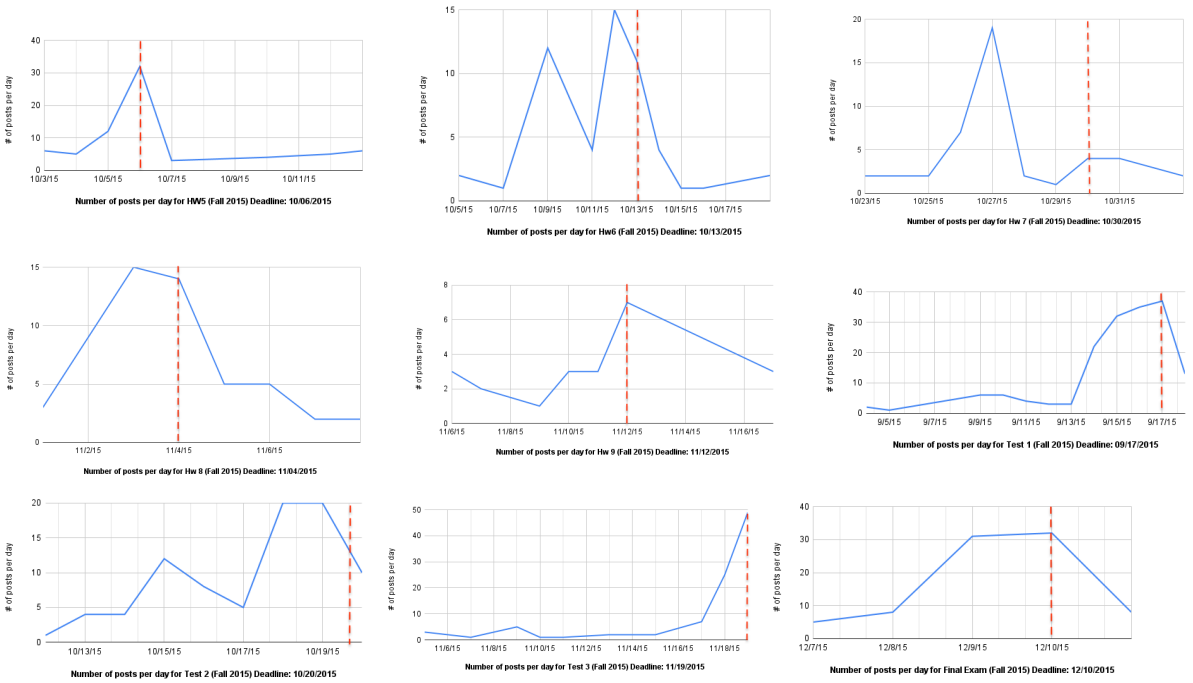
The 2nd chart in Figure 4.1, we see that in homework 6 some help-seeking activity still go on after the deadline, due to flexible deadlines. Based on the graphs, In test 1, the 6th chart in Figure 4.1, we see some patterns of late help-seeking, the exam date is the 17th of September and students started seeking help about 4 days to the exams. During this period, we can see that 30-40 post are made on the forum. That pattern matches the expected late help-seeking pattern in this homework. Figure 4.1 homework 6 shows that some help-seeking activity still go on after the deadline, due to flexible deadlines and extensions. Also the fall break might be a reason for the first spike as students tended to work more and seek help during that period.

Observe figure A.1 (see appendix) which shows homework 5 - 9 in 2014, the highest rates of help-seeking are on or close to deadlines. Figure A.2 (see appendix) shows the number of posts per

day in tests and Exams in Fall 2018 semester. Highest rate of help-seeking was also found towards the deadline. Late help-seeking was found in test 3.

To answer RQ1.2a, in study one, higher rates of interaction are expected towards the deadline, similar patterns are seen in Figure 4.1. There are no milestones in this assignment, so we could not assess the impacts of milestones This makes achieving RQ1.2b not feasible.

Figure 4.1 Number of posts per day during Homework and Tests/Exams in 2015, selected to represent a typical semester



In this chapter, we investigate the following hypothesis:

- We hypothesize that similar correlations results will be discovered between forum posts and class performance where features of forum activities which include average number of posts, average number of likes, fraction of unique views, average number of questions and average number of Replies, will be negatively correlated with class performance; and early forum activities will be positively correlated with class performance.

To confirm our hypothesis, we visualized features of class-level forum activity as described in Table 3.6 and performed correlation and predictive analyses to detect how these features relate with class average performance.

Table 4.1 Number of Students in each Semester

Semester	2014	2015	2016	2017	2018	2019	2020
Fall	285	273	254	274	331	359	340

Table 4.2 Assignments in the Course

Topics
Logic & Proofs
Circuits
Predicate Calculus
Sets & Predicate Calculus
Sets, Arithmetic Proofs, Induction
Induction, Recursion
Big-O, Induction, Recursion
Binary Relations
Counting 1 (Counting Introduction)
Counting 2 (Counting)
Graphs, Probability
Test 1 (midterm exam 1)
Test 2 (midterm exam 2)
Test 3 (midterm exam 3)
Exam (Final Exam)

Table 4.3 Definition of Features of Piazza Posts

Features of Forum Activity	Definition N= total number of students in the class who attempted the assignment
Average # of Posts	(Total number of all posts related to an assignment)/ N
Average # of Likes	(Total number of likes on all posts related to an assignment)/ N
Fraction of Unique Views	(Total number of students that viewed posts in an assignment)/N
Average # of Questions	(Total number of questions asked by students in an assignment)/N
Average # of Replies	(Total number of replies by both students and instructors in an assignment)/N
% of Early Posts	(Percentage of total posts made half-way through the assignment)/N

4.2 Background

Replication studies have been used to evaluate research and generalize the findings of prior pieces of research[Coi18]. We seek to perform the exploratory, correlation, and predictive analysis on a new course. One similarity between the math course (Discrete Math) and the previous (Java Programming) is that they are both 2nd-year Computer Science courses at the same university.

Also, they both require students to complete a number of assignments and both use a Piazza discussion forum for help-seeking. There are a number of differences between these courses. For example, the courses introduce different Computer Science concepts to students. In addition, the Java course is programming intensive whereas the Discrete math course is not, but involves proofs and formalizations. Below, we describe the differences between how the courses were administered:

4.2.1 Notable Differences Between the Java Programming Course and Discrete Math Course Administration

- **Assignments** In the math course, students were required to take a series of 10 assignments and 4 major tests. The assignments are shown in Table 4.2. These assignments were taken in a slightly different order each semester. For example, in Fall 2020, Counting was taken as the second homework, it was the taken as the eighth homework in 2016 and 2017 and in 2014, 2015, 2018 and 2019, it was taken as the ninth homework, One area of future work could be to evaluate the impact of these different homework sequences on class performance. In the programming course examined in study 1, students worked on 3 guided assignments and two to three course assignments, taken in the same order.

  While designing the replication study, we contacted the math course instructor who gave us an overview to guide us on replicating the study. She explained that assignments five through nine and the four tests were similar in nature to assignments due to the longer duration required for working on and preparing for them. We therefor focused on exploring those assignments.

  Students were given review questions before exams. We noted that a good number of posts were about solving these retake questions. This release dates of this review questions might control the timing of students' help-seeking[MM19]. Moreno-Marcos et al. discovered patterns such as the increase of activity near deadlines and period when instructor email reminders are sent out. Therefore, the release of review questions for exams might determine when students began seeking help.

- **Milestones** In the math course, there were no milestones and students were free to work on any portion of the assignments at any point in time, once they became available in the homework tool.

- **Students Attitude towards Assignments** A good percentage of the class did not attempt some of the assignments at some point in time as describe in the *Zero Analysis* in section 4.2.2. We would take this into consideration when interpreting the results. Furthermore, in some semesters, students that met a number of peer tutoring activities were exempted from taking the final exam. These students might have initially taken part in the help-seeking on the forum while solving review questions for the final exam. Their interaction on the forum for the final exam might skew the correlation between rate of forum activity and class performance since some students did not contribute to the class performance for the final exam.

- **Course Logistics**

  In two semesters (2017 and 2018), students were given incentives for participation and students proved their participation and peer tutoring by tagging a folder called “participation" whenever they made a post. Some posts that were about help-seeking in assignments and exams might have been tagged under participation for these two semesters, skewing the number help-seeking posts in homework and test folders. On the other hand, it might be the case that posts about peer tutoring were reserved for the participation folder, making numbers the posts in homework and test folders more related to help-seeking and help-giving. While extracting the posts in the homework and test folders, we skimmed through the posts to make sure that the posts in these folders were related to help-seeking for the relevant assignment.

- **Duration and Flexible assignment extensions** The math course allows for flexible assignment extensions which resulted in irregular duration. The duration documented in Tables A.1 and A.2 in the appendix section are the first and last date a student sought help on the forum for each assignment shown. Students could request an automatic 3-day extension on any homework, as needed, and were sometimes granted longer extensions upon request.

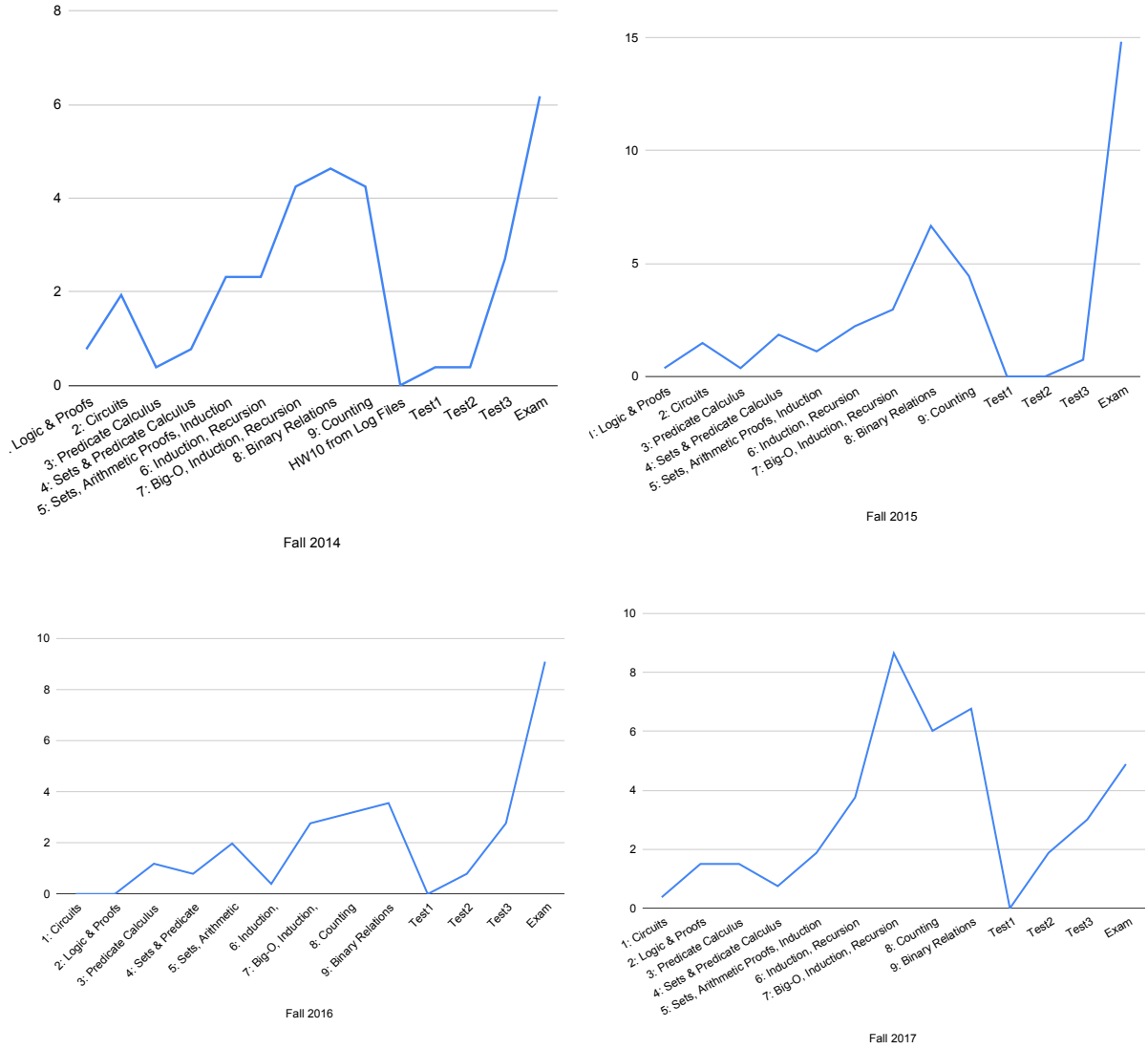
- **Average Performance Calculation** In the previous study, we calculated average performance as the total class performance divided by the number of students in the class who attempted the assignment (N). (Students who attempted the assignment and scored zero were included). In this study, average performance is the total class performance divided by the number of students in the class who attempted the assignment (N); however, in this study, we omitted all the students that had a score of zero because as at the time of the analysis, there was no data to differentiate students that attempted the assignment and scored zero from those who did not attempt the assignment. Furthermore, from the zero analysis we discovered that students enrolled in the course might choose not to attempt an assignment.

  In addition, the instructors allowed students to retake any one of tests one through three. While extracting posts relevant to tests, we were careful to omit posts about test retakes. We also calculated the class performances in tests using original scores only and ignored retake scores, to focus on relationships between the tests and relevant posts during the regular class schedule. Note that we included the scores of students that made late homework submissions when calculating the class average performance for homework.

4.2.2 Zero Analysis

This analysis was done to give more description of students’ behavior in the math course under analysis. Figures 4.2 and 4.3 shows the percentage of the class that had a score of zero in each assignment. The assignments were also arranged in the appropriate sequence for each semester. Note that we did not have enrollment data. Performance data was extracted from the WebAssign

Figure 4.2 Percentage of class that had a zero score in Fall semesters of 2014-2017.

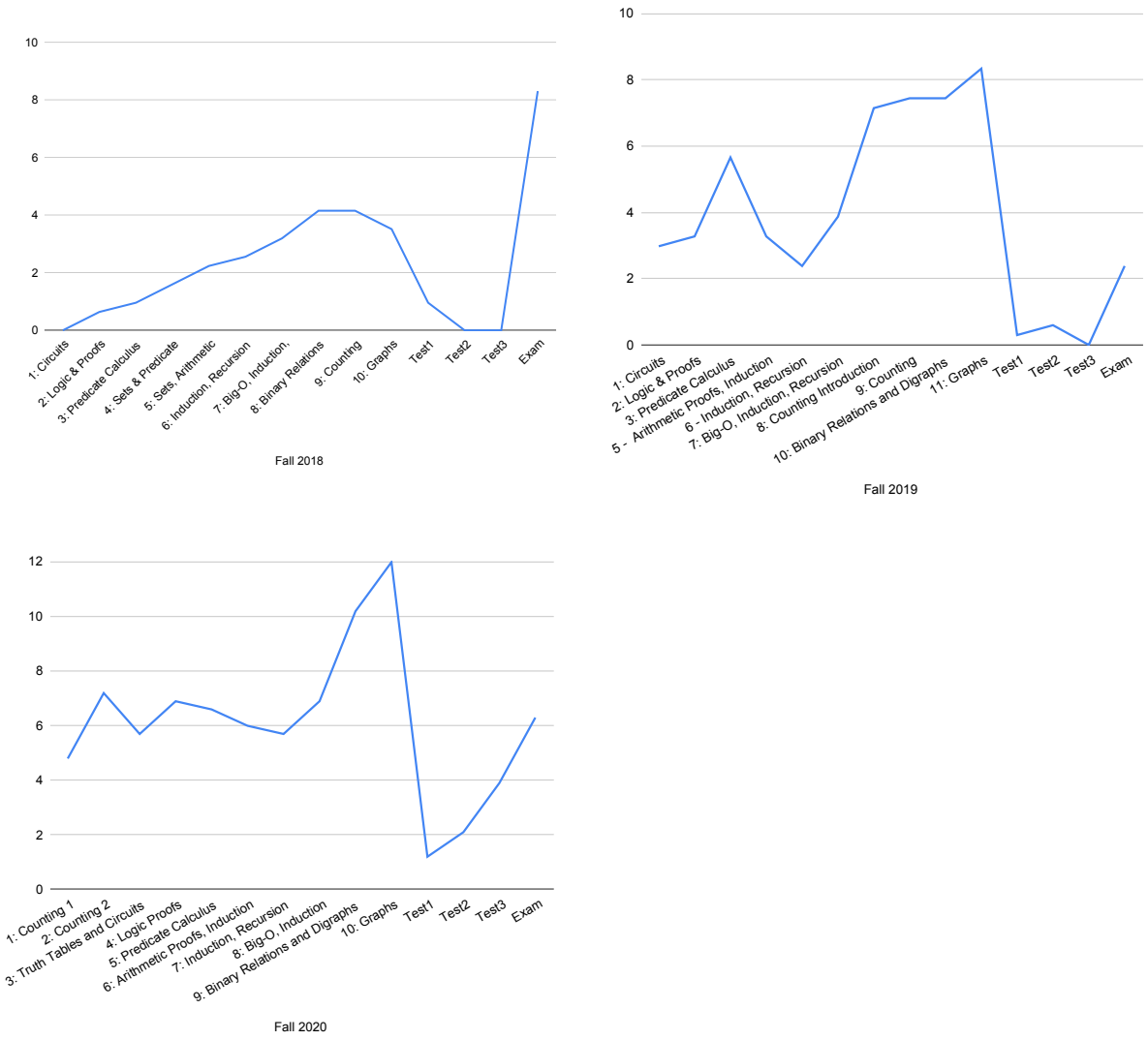


submission locker which did not indicate students who have dropped the course. We made the following assumptions for this zero analysis:

- students that had zero on all the four exams and had zero on the last six homework were assumed to have dropped the course and are not shown in this zero analysis and charts in Figures 4.2 and 4.3
- students that had zero on the last three exam and had zero on the last seven homework were assumed to have dropped the course and are not shown in this zero analysis and charts in Figures 4.2 and 4.3

This assumption was used for all semesters in the zero analysis alone. Students that met these

Figure 4.3 Percentage of class that had a zero score in Fall semesters of 2018-2020. Cont'd



assumptions were not included in the zero analysis. These assumption were chosen to provide **an estimate of the number of students that were enrolled but might not have attempted some assignments.** Students that submitted homework late were considered to have done the homework.

In 2016, (see Figure 4.2 2016) 9% of the class had a zero score in the final exam. The percentage was higher than the percentages of zero scores the homework. It might be the case that since most students have attempted a good number of assignments, they chose to skip the exam, or these students took part in the peer tutor program and were exempted from the final exams. In 2019 (see Figure 4.3 2019), more students tended to skip homework than tests. These behaviors might impact the relationship between forum activities and class performance, because some students that took part in the help-seeking process might end up not doing the homework.

In subsequent sections, we will describe methods for confirming our hypothesis. We seek to find

Table 4.4 Average Class Performance during Fall semesters 2014-2020

	2014	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Exam	81.50	80.79	81.34	78.98	83.24	79.42	90.82
Test1	92.37	80.37	96.16	93.64	83.11	84.34	74.39
Test2	80.67	83.37	91.35	89.77	80.48	83.10	81.49
Test3	82.30	74.33	75.59	86.88	80.48	70.67	79.20
Sets, Arithmetic Proofs, Induction	89.62	85.37	87.65	87.52	87.68	89.51	88.626
Induction, Recursion	93.51	89.65	92.57	93.15	92.62	94.94	93.64
Big-O, Induction, Recursion	87.43	84.10	74.85	83.12	80.71	86.33	82.47
Binary Relations	89.99	93.43	89.77	90.53	94.77	98.734	98.51
Counting	93.04	93.21	93.47	93.86	86.14	86.80	91.73

the relationship between levels of forum usage by the **whole class** while working on an assignment, and what these levels indicate to instructors about that assignment.

The table containing the assignments and the percentage of the class that had a score of zero in them are in Tables A.3 and A.4 (See appendix section).

4.3 Methods

4.3.1 Data Collection

We began with an exploratory analysis of how students use the Piazza forum. Piazza data was collected as JSON files in a similar way as in study one**??**.

4.3.1.1 Features of Forum Activity

We selected features of forum activity from Piazza data such as posts (which consist of questions, replies, post likes, and post views). Based on post type, we extracted two features from the post feature, namely questions and replies, to investigate what each of these can individually tell us about the assignment. Questions include initial questions posted by students while replies include instructor reply, student reply and follow-ups. Our features are defined in Table 3.6. Since each semester has varying numbers of students in each cohort, we took the average of the features of forum activity and applied them to predict average class performance. Early posts are the number of posts made mid way through each assignment. We chose this feature to see if early interactions on an assignment could predict average performance on the assignment.

4.3.2 Exploratory Analysis

We calculated the features defined in Table 3.6 and visually explored their relationship with performance.

4.3.2.1 Class-level Forum Activity vs. Class-level Performance

To investigate our hypothesis that the **class-level forum activity** are related to **class average performance** on an assignment, we plotted the interactions between class-level posts and performance in Figure 3.1. We normalized the number of forum posts in each assignment for easier comparison using min-max normalization (0-1). We also normalized the average performance in Table 4.4 in the same way for the line/interaction plots in Figure 3.1.

The min-max values before normalization can be found in Tables 4.5 for homework and 4.6 for exams. Notice that the lowest average performance recorded was 74.85% for homework and 70.67% for exams. The highest average performance recorded was 98.73% for homework and 96.16% for exams. Average performance across all assignments and exams are shown on Table 4.4.

Table 4.5 Min-Max of features in Homework used for normalization.

Feature	Minimum	Maximum
Number of Posts	5	170
Average Performance	74.85	98.73

Table 4.6 Min-Max of features in Exams/Tests used for normalization.

Feature	Minimum	Maximum
Number of Posts	36	440
Average Performance	70.67	96.16

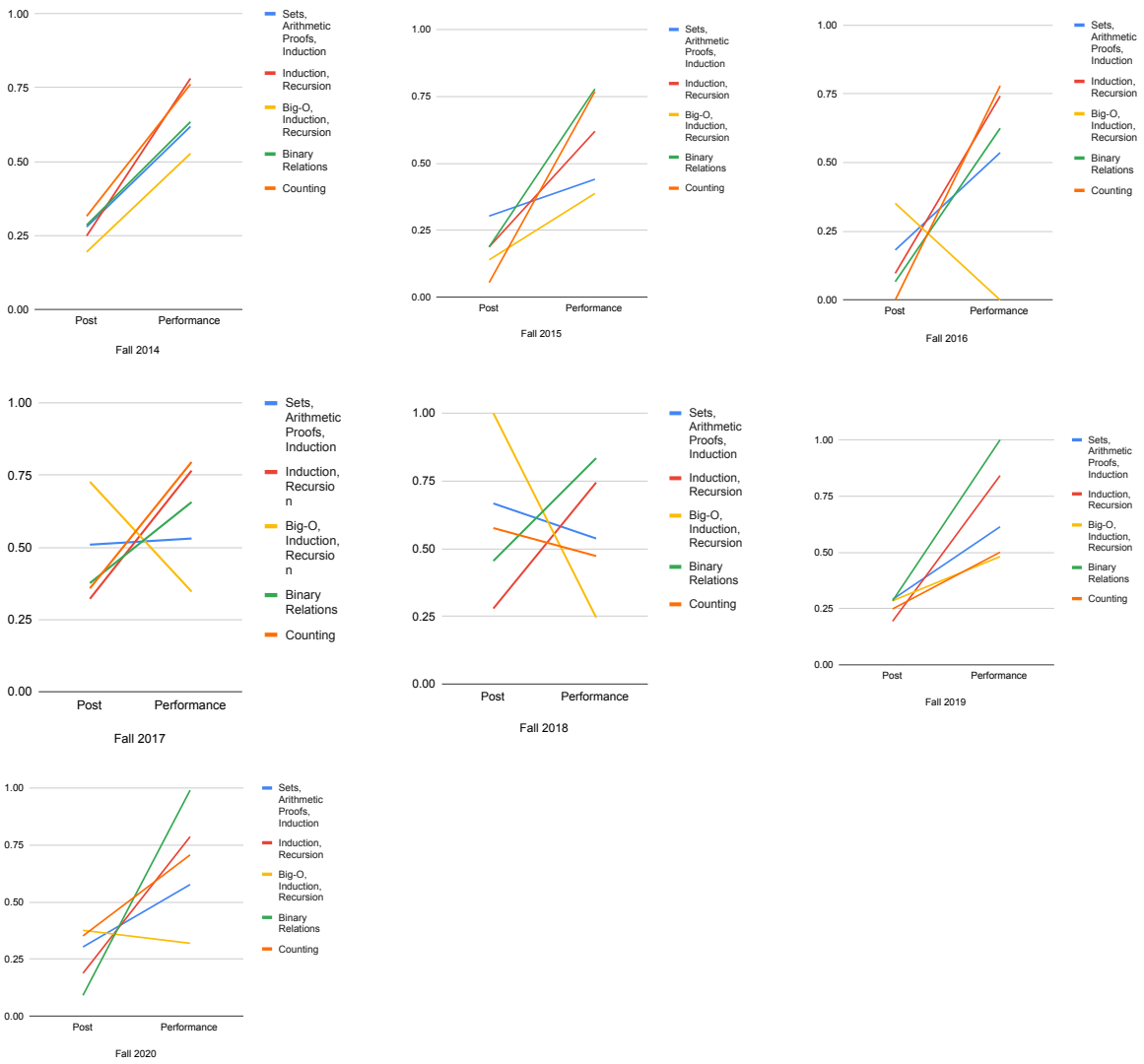
Hypothesis We hypothesize that the higher the **class-level** posts, the lower the **class-level** performance in each class. This can be seen in the interaction plots with lines for assignments with lower posts crossing the lines for assignments with higher posts. In 2017, this trend can be observed in all homeworks shown in Figure 4.4 (2017) except Induction, Recursion (Red line), and Counting (orange line) where the line did not cross. Like the homework chart for Fall 2017, other semesters' homework charts did not totally display this interaction.

4.3.3 Correlation Analysis on class-level posts and performance

We performed direct Spearman's ranked correlation analysis comparing each of the selected features of forum activity with average performance for each assignment. We used Spearman's ranked correlation because we are comparing ratios and our data passes the assumptions that are required for Spearman's correlation[Spe]. The Spearman's correlation assumptions are:

- Assumption 1: The two variables involved in the correlation should be measured on an ordinal, interval or ratio scale.

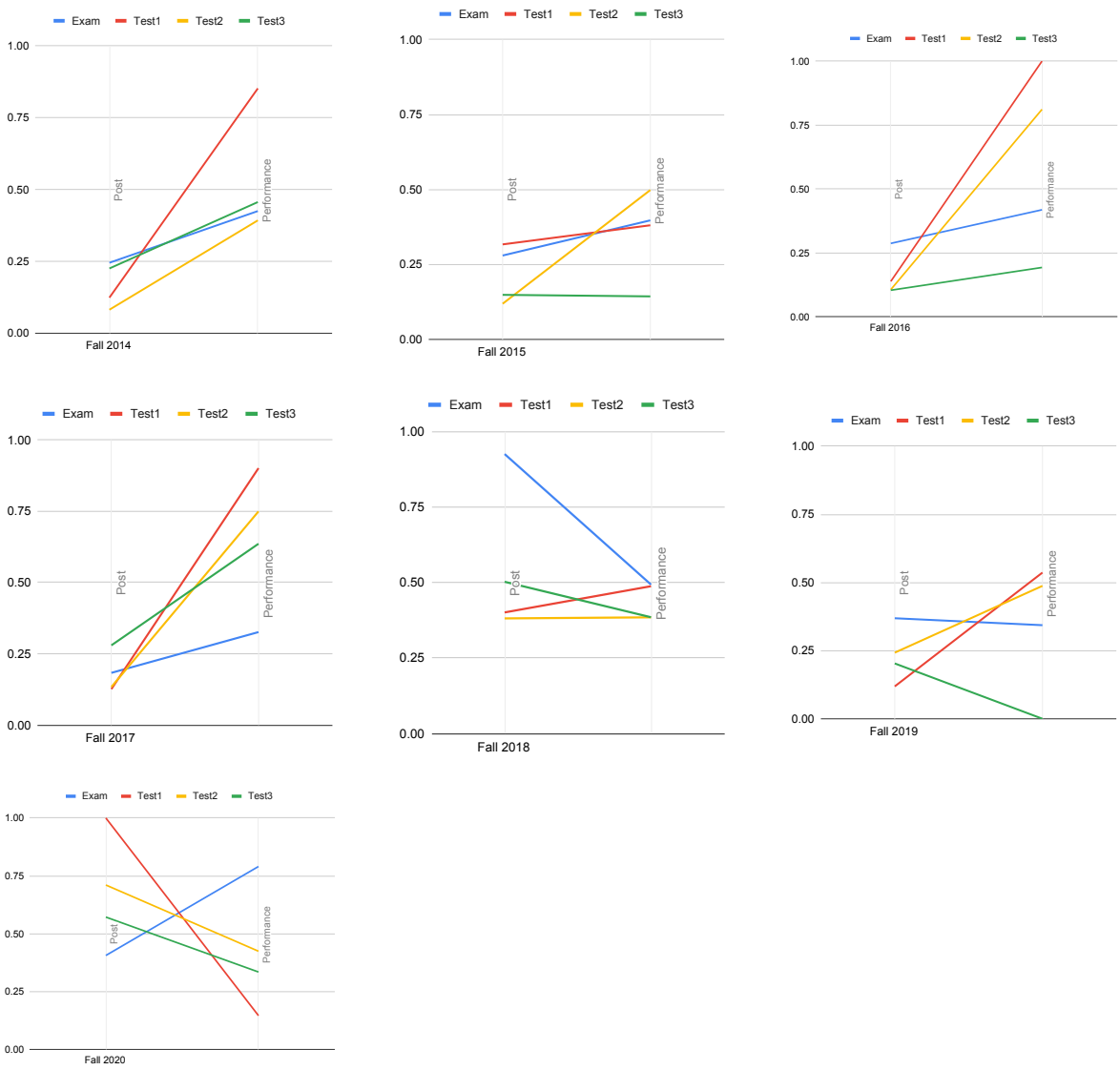
Figure 4.4 The normalized number of forum posts and average performance for homework in Fall semesters of 2014-2020.



- Assumption 2: The two variables must represent paired observations
- Assumption 3: There should be a monotonic relationship between the two variables

For example, we wanted to know the relationship between average number of posts in an assignment (see definition in Table 3.6) and average class performance in that assignment. We took the average number of posts instead of the total number due to the different sizes of student cohorts. For example: Spring semesters generally have less students than the Fall semesters (see Table 4.1).

Figure 4.5 The normalized number of forum posts and average performance for exams in Fall semesters of 2014-2020.



4.3.4 Regression Analysis

To confirm the patterns we observed in the exploratory analysis on class level forum activities and average class performance, we fit regression models to the data. We performed ordinary least squares (OLS) regression analysis. To account for the different student cohort in each semesters, we modeled semester_year as a random effect. We aim to achieve a model with good fit and observe the p-Value of each feature to confirm our exploratory analysis. We set our statistical significance threshold at $p < 0.05$ and applied Bonferroni Correction for multiple statistical tests. Our results are presented in the next section.

4.3.5 Random Forest Analysis

We set up a random forest analysis, since the fit for regression analysis is not ideal, as discussed in section 7.3. We trained a random forest model to predict class average performance based on some features of forum interactions. We use 90% of the data for training and the remaining 10% for testing. We left sufficient data for training and did not use cross validation because the out-of-bag (OOB) error calculated during model training is a good indicator of test set performance. Moreover, we are interested in the goodness of fit and the feature importance, rather than a generalizable model that is robust to different datasets. Here, we are interested in the features that significantly contributes to the model's accuracy. We report our result in Figure 4.6 as well as the feature importance in Table 4.7.

4.4 Results

Figure 4.6 Predicted Average Performance from Random Forest Regression Analysis Vs Actual Average Performance

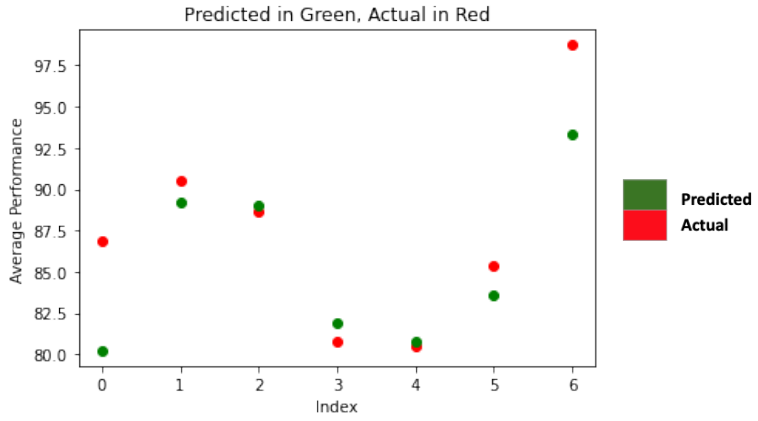


Table 4.7 Spearman correlation of class-level Average Performance with Features of Social Interaction (ρ). Statistically Significant p-values of features based on Bonferroni Correction are in bold font. The Feature Importance from the Random Forest Model is also displayed

Features	ρ	p-values(OLS) ($R^2 = 0.4$)	Feature Importance ($R^2 = 0.81$)
Average# of Posts	-0.61	**0.002**	**0.15**
Average # of likes	-0.57	0.45	**0.32**
Unique Views	-0.52	0.104	0.07
Average # of questions	-0.67	**0.001**	**0.25**
Average # of Replies	-0.56	**0.013**	0.068
% of early posts	0.19	0.16	0.14

4.4.1 Class-level Correlation Results

We seek to detect the patterns of students' forum use and what these patterns indicate to instructors about an assignment (RQ1.1). We will now observe the correlations between all features and our target, average performance on each assignment. From Table 4.7, we can see that the average number of posts ($\rho = -0.61$), average number of likes ($\rho = -0.57$), the fraction of the class that viewed an assignment's posts ($\rho = -0.52$), average number of questions asked ($\rho = -0.67$) and replies ($\rho = -0.56$) are negatively correlated with average performance. The negative correlations between average posts with performance **confirms our hypothesis** that more posts by more students are made on assignments with a lower average class performance. On the other hand, the last feature we investigated, % of Early posts, is positively correlated ($\rho = 0.19$) with average performance. This means the higher the percentage of total posts made earlier (half-way through the assignment duration), the higher the performance on an assignment in the class. This result suggests that to an extent, the class average performance improves when the forum activities begin earlier in the assignment timeline, as in Study 1, but with a lower correlation in this course.

Hypothesis Confirmed The correlation results confirm our hypothesis, although, the correlation values recorded were not as strong as in study one. This might be due to differences in the course administration discussed in section 4.2.1.

4.4.2 Predictive Analysis

All our models recorded **R-squared values of our models were approximately 0.4**. This was not a good fit. Table 4.7 shows that the selected features have significant p-values indicating that they are significant predictors of class-level performance. After applying the Bonferroni Correction for multiple statistical tests, the features: Average number of Posts, Average number of questions, and Average number of Replies like were found to be statistically significant predictors class-level performance and Average number of likes and % of early posts were not statistically significant. This confirms results in the original study. However, **unlike study one, unique views did not significantly predict class performance.**

Since, we were not able to achieve a good fit using the linear regression analysis, we decided to apply random forest analysis. The result of the random forest analysis can be measure using the goodness of fit (R squared). We got a better fit than the linear regression. Random forest model produced R squared of 0.81. We achieved a root mean square of 4.07 and visualized the actual versus the predicted test values in Figure 4.6. The feature importance of the features used are displayed on Table 4.7. The features with higher importance in our model are Average number of questions, Average number of Posts, and Average number of likes. Therefore, some of the results in study one were confirmed.

4.5 Discussion

Our analysis indicates that difficulty or complexity of assignment related with increased forum activity. The average **number of questions** asked by the class is negatively correlated ($\rho = -0.67$) with class performance. This confirms the definition of help-seeking by Richardson et al. (2012) as the tendency to seek help when experiencing academic difficulties[RM12; Cro17]. It also confirms our results in study one. The negative correlation ($\rho = -0.56$) between average replies and average performance might indicate that the higher the number of replies and follow ups (from other students who might be confused about the subject and need clarification), the lower the class average on that assignment. We observed that most of the results from the prior study were confirmed to an extent in this study despite the differences in course administration. We noted that assignments in this course were presented to students in different orders. Future work could perform detailed exploration on the impact of these different sequences on class performance in order to discover an optimal sequence for administering these topics.

4.6 Limitations

The goal of this study is to replicate the exploration of class forum activity in order to generalize our results. The replication was not totally successful due to a number of factors which we discussed in section 4.2.1, including student behavior in assignments, the opportunity for test retakes, class logistics, and the nature of the assignments (tests and assignment versus course projects).

Some students skipped some assignments, even though they might have sought help about the topic on the forum. Students were given review questions before exams. We noted that a good number of posts were about solving these review questions. This release dates of this review questions might control the timing of students' help-seeking[MM19]. Moreno-Marcos et al. discovered patterns such as the increase of activity near deadlines and period when instructor email reminders are sent out. Therefore, the release of review questions for exams might determine when students began seeking help. We should also note that some posts might not be properly categorized. Posts that pertained to help-seeking in one assignment might be tagged under other folders such as logistics for that assignment.

4.7 Conclusion

We performed an exploratory analysis to visualize forum use over 7 semesters of a Discrete Math course and compared them with average performance on their related assignments. We also performed correlation and regression analyses as well as random forest analysis. We found patterns of late help-seeking. We saw a medium negative correlation between whole-class forum use and class average performance. Although, these results are not as strong as the original study, instructors might find ways to encourage early forum use since it could potentially improve class performance.

We further found evidence of students choosing not to complete some assignments, which may inspire instructors to change course policies.

CHAPTER

5

STUDY 2: A SNEAK PEEK AT A COLLABORATIVE CS PROJECT: WHAT PART DO FEMALE STUDENTS PLAY?

abstract We present an exploratory analysis of how different genders collaborate while working on programming projects in a Java Programming Course of a south-eastern research-intensive US university. In the projects we studied, when students collaborated on Github, they declared what aspects of the project they worked on by including a tag in each commit message. Some aspects of the projects involved intensive coding whereas other did not. Our study analyzed these tags to determine what parts of the project each gender is prone to working on. We also performed predictive analysis to determine if elements of commit messages and tags are significant predictors of gender and vice versa. Our findings from the class we studied, show that gender is not a factor when it comes to what each student worked on. This could serve as a motivation for female high students who are considering choosing a career in Computer Science.

5.1 Introduction

Computing is a male dominated field and many studies have reported on the gender gap that exists in Computing related courses. Beyer et al. 2008 in their book titled: “Women in Computer Science or Management Information Systems Courses: A Comparative Analysis” asked why so few women major in computer science?’ They explained that gender imbalance is a complex issue and the

reason for the gender gap is yet to be understood despite numerous publications on the subject [Bey03; Syl08; Coh08; Ker14; Mil08]. Kamberi (2017) analyzed over 30 research studies investigating women in computing and produced the 4 Es model (expose, engage, encourage, empower) to entice women and spur their interest in computing [Kam17]. One challenge is that computer programming is negatively viewed by most female students and they consequently avoid programming related fields[Spi20]. Another challenge is the lack of confidence of women who are curious about going into computing related fields [Kam17]. Kamberi recommended that programming should be introduced to girls at an earlier age. They also suggested that educators should encourage high-school female students to pursue computing majors in college [Kam17]. Another solution involves introducing a gender-inclusive college curriculum which is a curriculum that acknowledges the different learning patterns and backgrounds of all genders [Mil08]. Lessons learnt from Kamberi's survey suggested that early exposure, engagement, encouragement of female students by educators and empowerment through mentoring would lead to an improved representation of women in computing related fields [Kam17].

In this current study, we seek to understand what goes on during programming assignments in a second-year programming class in College. We are specifically interested in exploring what part female students play while working on a programming project in multi-gender teams in this class. A programming project consists of several parts such as the implementation of the project requirements, bug fixing, unit and system testing, code cleanup and documentation. In this class, students work in teams to achieve a programming task. Teams collaborate over Github, a decentralized version control platform for projects that allows different members of the project team to collaborate seamlessly by resolving changes to be merged and resolving resulting conflicts in an efficient manner [Buf20]. GitHub has been embraced by the open source communities and is being employed as the major online learning tool for collaboration in programming related college courses [Zag15]. The platform is increasingly vital for managing software projects. Several research studies have shown the effectiveness of Github in transforming the learning experience [Zag15]. In addition, Github provides a system where every team member has a version of the project as well as a traceable project repository where each member can commit code changes to the project repository. Each commit is required to have a commit message which provides a basic description about the code being added to the repository. We aim to explore the contributions and portions of the code worked on by male and female students in mixed-gender teams based on their commit messages on Github.

The challenge is that there is little or no knowledge of how diversity affects online collaboration in team assignments. Imtiaz et al. in their paper: "Investigating the Effects of Gender Bias on GitHub" discovered that female contributions on Github tend to be accepted more often than contributions by their male counterparts. However, when a female's gender can be identified, they are rejected more often [Imt19]. Seeing that a bias exists in how gender efforts are perceived when gender is known, we want to know if this bias in turn affects females' choice of programming tasks while working in teams consisting of both genders in a blended learning environment.

This chapter reports tasks each gender chooses to work on while working in a multi-gender team. We investigate programming projects in one semester and seek to detect evidence for whether female students are prone to performing less programming intensive tasks, such as documentation, or styling, as opposed to the main implementation, debugging, resolving static analysis notifications or testing tasks. In this work, we investigate commit messages for team projects in the Fall 2020 semester of a CS 2-OOP class [Por18b]. Prior research on previous offerings of the class, showed that the commit messages sometimes included random texts that do not really explain the code committed [Thr]. Therefore, in the beginning of the Fall 2020 semester, we asked the class instructor to hint students on some good examples of meaningful commit messages. Hence, in addition to the commit messages the instructor included a tag between brackets. For example: if a commit message says: "committing a fix for linked list error", the students were recommended to commit the following instead: "[Debugging] committing a fix for linked list error". Another example is the commit message "adding code for listview" becomes "[Implementation] adding code for listview". We want to see if a particular gender is prone to performing one type of task relative to the others.

Prior research also discovered three teamwork habits among teams namely cooperative, collaborative and free-rider [Com14; Dui07; Thr]. Coman et al. described two forms of habits Cooperative, where each team member supports each other while they work on different tasks in a project and Collaborative, where the team members share the same tasks while completing a project [Com14]. There is a third teamwork habit described by van der Duim called free-rider where one team member does the majority of the tasks while the other members do little or nothing [Dui07].

We want to detect what habits are common among multi-gender teams. If Cooperation exists, we want to know what task each gender agreed to focus on. A positive result, which is collaboration on all tasks or cooperative habits where the programming tasks are not divided along gender lines, would serve as a great encouragement to female students with a negative view of programming and programming related fields. A negative result which is either cooperative habits with females working on less programming intensive tasks or a free-rider situation where a female team member does nothing might signal the need for an intervention, perhaps males should learn to share work more or more preparedness of female students before entering programming intensive fields.

It is important to note that academic performance and level of programming experience are other factors that affect a student's choice of task to work on but we want to see if there is a relationship between gender alone and the choice of tasks in the team project of the class we are investigating. This study is broken into two parts:

- Exploratory Data Analysis of commit messages
- Predicting Gender based on commit messages and vice versa

Our exploratory data analysis (EDA) answers the following research questions:

- RQ 2.1: How did each gender respond to the recommendation to tag their commit messages?
  - (a) How many students actually tagged their commit messages?

  - (b) Were females prone to following that recommendation or not?

- RQ2.2: In the class we studied, were tasks divided to team members biased along gender lines?
  - (a) Did female students do more documentation and less programming?
  - (b) Did one gender generally do more of one task than the other?

  For our prediction task, we ask:

- RQ3: Can we predict the gender (M/F) of the authors using details of the commit messages as predictor variables and can gender predict the commit tag?

5.2 Related Works

5.2.1 Gender differences in Choosing Computer Science

In a research study done by researchers from the University of Wisconsin, female students had less confidence in their ability to learn computer science (CS) concepts than male students in CS. In fact, they felt less confident than the male students in non-computer science majors. The study also found that female students who have supportive friends and family members are more likely to major in computer science. The authors then concluded that they speculate the reassurance of one's competence by others may increase the probability that a woman musters the courage to enter into a male dominated major such as CS [Bey03]. A study by Du et al. showed that males were exposed to programming at an earlier age than females. They also found that after a one hour coding tutorial, males were more likely to take a programming course than females even though females performed better than or equaled males in the programming quizzes after the hour of code [Du19]. This simply shows that although females are likely to do well in programming, there might be less likely to show interest in the field. This current work studies what parts females play in a collaborative CS college project. An insight into how females work on programming projects might embolden other female students to take up CS in college.

5.2.2 Gender differences in Collaborative Projects

In their paper, Imtiaz et al. demonstrated the capability of women in the open source community. The findings showed that women tend to have their pull requests accepted at a higher rate than men [Imt19]. In fact, women's pull requests are larger, indicating more lines edited. Vasilescu et al studied the effect of gender and tenure on productivity in Github Open source software (OSS) development communities. They hypothesized that gender diversity has a positive effect on productivity and a negative effect on turnover while tenure diversity has a positive effect on both productivity and turnover. To prove their thesis, they conducted surveys and measured productivity by the number of commits by each team member. By using mixed effects analysis to measure significant predictive abilities of gender on productivity and turn over, they confirmed part of their thesis that gender

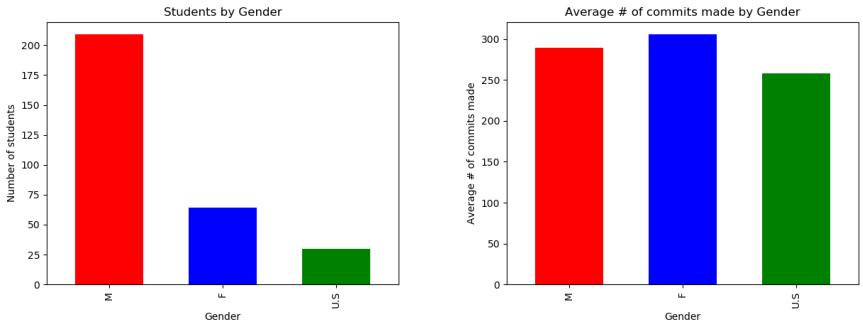
diversity has a positive effect on productivity. However, they are unable to prove the negative effect of gender diversity on turnover [Vas15; Dan13]. This current study applies an exploratory and predictive approach to exploring the productivity of female students while working in multi-gender teams.

5.3 Method

5.3.1 Pre-processing

This work explores the programming activities of students in the Fall 2020 semester of a CS 2-OOP class [Por18b]. Our dataset consists of two files, Gender file and Commit Messages file. The Gender file provided by the course instructor contains mappings of author_IDs to students' gender. Missing values in the gender column were labeled "Unspecified" (U.S.). As shown in Figure 1(a), there were 303 students in the class. 209 males and 64 females, and 30 student's gender were U.S.. While this makeup seems highly unbalanced at first glance, these metrics are representative of the current departmental makeup, which thereby assure us that our analysis population/sample was representative of the department as a whole.

Figure 5.1 (a) Statistics of the class by Gender (b) Average commit message for each Gender



(a) (b)

We collected the dataset containing commit messages directly from Github. This contains the author_ID, repo_name, commitID, commitTime, commitMessage, lines added, deleted, modified, copied, removed and typeChanged. Using the author_ID, we merged the two files into one dataframe. We dropped the columns typeChanged and commitID since they weren't relevant to our research question and sorted the dataframe by repo_name to make it easier to narrow our research on teams later on. There were a total of 1,942 unique repositories and 87,765 commits throughout the semester. 27 duplicates observations were removed resulting in 87,738 observations. Figure 5.1(b) shows the average number of commit made by gender. Female made more commits on the average than males in the class we studied.

5.3.2 RQ2.1: How did each gender respond to the recommendation to tag their commit messages?

Table 5.1 Statistics of Commit by Each Gender

Gender	Min	Max	Mean Lines Added	Mean Lines Modified	Mean Lines Deleted
Male	9	785	17,465	11,471	1,540
Female	10	693	18,741	12,479	1,647
U.S.	2	645	18,799	11,963	1,107

Table 5.2 Commit Categories, Keyword and Definition

New Tags	Original Tag Keywords	Definition
untagged		The commit message did not contain a tag
versioning	*git, version, merge*	The commit message is about versioning
documentation	*doc, deploy, comments, readme*	The commit message is about how the new update describes the code
style	*style, design, gui, uml*	The commit message is about front end tasks
implementation	*impl, impel, impe, optimization, integra, fsm, constructor, studentrecordio, tried, refactor*	The commit message is about main implementation tasks
debugging	*fix, bugs, revert, rollback, debug, exception, throw, error, deb, debug, degug, pmd, pls*	The commit fixes a bug, a.k.a bugfixes
testing	*test, random, hope, jenk, junit, unit, bb, box, black, coverage*	The commit tests the code update
skeleton	*skeleton, skel, libraries, initial, collection, setup, start*	This is usually the first commit or starter project
misc		The tag given by the student was not meaningful and did not fall into any category

5.3.2.1 Commit messages, cleaning & Tag Categorization

Of the 303 students in the class only 99% of both male and female students made commits to the course projects on Github and 90% of U.S. gender made commits on the platform. The minimum

commits within the class for each gender were roughly similar, 2 for U.S., 10 for females and 9 for males. Males had the maximum number of commits by a single student (785), maximum number of commits by a female student was 693 and U.S. gender, 645. On average, within the overall class female students made 5% more commits compared to males. As shown in Table 5.1, average male contributor added 17,465 lines of code, deleted 1,540 lines of code, modified 11,471 lines of code per male contributor, and removed 366 files.Females made more edits than males in every category. The average female contributor added 18,741 lines of code, deleted 1,647 lines of code, modified 12,479 lines of code and removed 772 files. The average U.S. gender contributor added 18,799 lines of code, deleted 1,107 lines of code, modified 11,963 lines of code and removed 348 files. The first gender to commit, commits the starter project provided by the instructor which has about 600 lines of code. The statistics above might be influenced by these, therefore we separated the commit messages into different tags/categories as described in Table 5.2. The category with of the starter project is called 'skeleton'.

In the Fall 2020 semester, students were expected to include 5 specific tags namely Implementation, Testing, Debugging, GUI, and Documentation, in their commit messages whenever they pushed to Github. We iterated over each commit and used a regex search to see if the student had tagged their commit. Only 51.6% of all commit messages were tagged. To answer RQ2.1a, 51.2% of all commit messages made by females were tagged and 50.6% of all commit messages made by males were tagged. 24 students did not tag any commit message at all and all students forgot to tag at some point. Of the 24 students, 17 males (8.2% of males), 4 females (6.3% of females) and 3 U.S. genders (11% of U.S.) did not tag at all. We transformed the commit message column to lowercase and found that students used 604 unique tags instead of the 5 tags we suggested. Some of the tags were as a result of typographical and spelling errors while others used synonyms of the 5 tags. We attempted to clean and categorize the commit tags into the 5 tags but ended up with 9 categories as described in Table 5.2. Using Table 5.2, we created a new column that identified whether or not a student had tagged their commit and what category (out of the 9) their original tag fell under. If the student hadn't tagged their commit, we tag the commit 'untagged', if their tag didn't fall into the categories in Table 5.2, we tagged the commit message as 'misc' which means miscellaneous.

5.3.3 Repo Processing

Students worked on 5 unique projects and several lab sessions during the semester. We created a new column to indicate the project being worked on in each commit. From the column "Repo_name', we identified whether the commit was for Guided Projects 1 through 3, Project1, Project2, or Lab. These projects are described in Table 5.3.

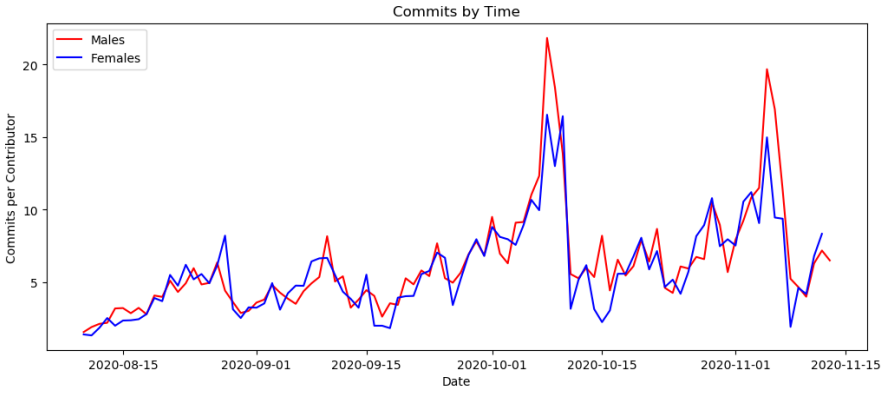
5.3.3.1 Commit Timeline

5.2 shows the average commits made per day by each gender. Interestingly, both genders had similar average commits per day. The earliest commit within the class was on August 10th. The last commit was on November 13th. The spike in Figure 5.2 shows the maximum number of commits. These

Table 5.3 Project Description

Project Name	Topic	Number of Repos
Guided Project 1 (gp1)	Composition and SE processes and practices	296
Guided Project 2 (gp2)	Inheritance	286
Guided Project 3 (gp3)	Testing	281
Project 1 (p1)	Finite State Machines (Implementation and Testing)	280
Project 2 (p2)	Linear Data Structures (Implementation and Test)	134 Individual repos, 71 Team repos, 13 multi-gender repos
Labs	Labs were on the following topics: InstallFest and Pair Programming, Project Creation, V&V Best Practices, Collections, Design, Debugging, FSMs, ArrayLists, Linked Lists, Stacks and Queues, Iterators, and Recursion and Lists	501 Individual repos and 93 multi-gender repos

Figure 5.2 Mean commit messages per day for each gender



were made on 10/08 (which was around the deadline for project 1). The second highest spike was on 11/05 (which was the deadline for project 2).The minimum number of commits was done on 9/17, 11/3 (around the start and end of the semester respectively).

5.3.3.2 Untagged Commit Messages

While preprocessing the students' commit messages, we noted that there were 43,557 untagged commits, which accounted for about half of all commits. The majority of these came from lab groups (18,553 commits) which also had the largest percentage of untagged commits (54%). The rest are

as follows, GP1: 2,795 commits (50%), GP2: 1,123 commits (35%), GP3: 1,654 commits (39%), P1: 10,087 commits (50%), Team P2: 4,841 commits (45%), P2: 4,504 commits (47%), Labs: 18,553 (54%).

Males and females within the class tended to leave messages untagged at a roughly similar rate, which was about 50% of all commits made by each gender or an average of about 150 untagged commits. In Project P2 and some labs, students were allowed to work in teams. Within the Lab team setting, the proportion of untagged messages by male and female was 57% and 56% respectively. However, within the P2 team setting, males had a significantly larger proportion of untagged commits (73%) as compared to females (56%). To answer RQ2.1b, both female and male students responded in about the same way (51.2% of all commits by females and 50.6% of all commits by males were tagged) to the recommendation to tag their commit messages. Except within the P2 team setting where males responded to the recommendation at a much lower rate than females in the team.

5.3.4 RQ 2.2: In the class we studied, were tasks assigned to team members biased along gender lines?

5.3.4.1 Mixed Teams

In order to answer the second research question on whether the tasks assignments were biased along gender lines, we need to analyze multi-gender teams. In Project 2 students had the choice of working in teams and in a few Labs instructors required students to work in teams. We created a subset of the data which included commit messages for project 2 and labs by teams with both male and female students. Within the mixed teams, for project 2, there were 71 teams, 9 teams having students with U.S. gender, 40 all male teams, 9 all female teams and 13 2-member teams, each consisting of 13 males and 13 females. Labs had 93 mixed 2-4 member teams consisting of 111 males and 52 female students in different ratios. The average number of commits made by gender shows an interesting difference between mixed teams and the overall class. On average, male students made 18% more commits than females in Project 2 with mixed teams having equal numbers of male and female team members. In Labs with mixed teams having an unequal number of male and female members, females made 80% more commits on the average than males whereas there were fewer females (52) within Lab teams than males (111). Within the whole class, female students made 5% more commits than males on the average.

Table 5.4 Statistics of Total Commit Messages

	# of Students			# of Commit Messages		
Gender	Mixed Team Project 2	Mixed Team Labs	Whole class (including mixed teams)	Mixed Team Project 2	Mixed Team Labs	Whole class
Male	13	111	208	1,050	5,360	60,466
Female	13	52	64	807	452	19,551

Team P2 with an equal number of male and female members and the Labs with multi-gender are our *focus teams*. Males and females had a roughly similar involvement rate proportional to their makeup within all team commits. Multi-gender team P2 made 1586 commits while working on P2. Males made 55% of all commits, females committed 45% of all commits. Male and female added and modified lines of code in the ratio 55:45 and 54:46 respectively. In summary, males in team P2 made slightly more commits, additions and modifications than females. Interestingly, males and females within mixed groups had varying commit patterns as related to the day of the week. Both made the majority of their commits on Thursday (Males (38.9%) | Females (44.6%)) which is usually the day of the deadline for project 2 and Wednesday (Males (18.7%) | Females (14%)) which is a day before the deadline for Project 2. Females tended to prefer working on the weekend a bit more than males.

The Lab teams were of different sizes, and there were 14 teams consisting of 1 male and 1 female student, who made 1,179 commits. Males in these teams made 49% of all commits, females committed 51% of all commits. Male and female added and modified lines of code in the ratio 44:56 and 54:46 respectively. This shows that although females made more commits, they added lesser lines of code but modified more lines of code. It also shows that both genders in the 14 teams contributed almost equally to the project. 55 teams had 2 males and 1 female with 5703 commits. 1 teams had 3 males and 1 female with 123 commits and 23 teams had 1 male and 2 females with 2261 commits.

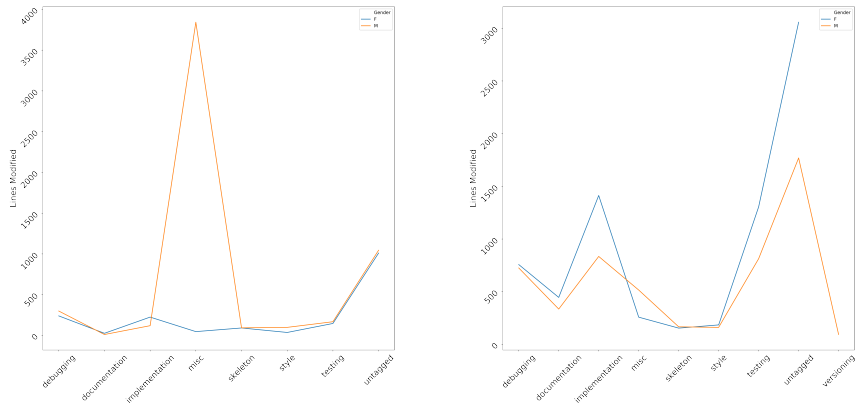
5.3.5 Visualizing Mean Number of Modified Files Committed in each category in Multi-Gender Teams

Each line chart in Figure 5.3 represents our focus teams and shows the mean of modified lines of code in each category as tagged originally by students. This contains untagged commit messages as well as misc tags. The red lines represent the male students and the blue represents the female students. The major inference from this figure is that due to the high number of untagged and misc messages, it is difficult to correctly conclude that one gender worked more on a certain category. In the first line chart, Figure 5.3(a) the female students in Team P2 performed slightly more implementation and documentation related tasks than the male student whereas the male students did more debugging, styling and testing. However, the majority of the commit messages by the male student had the miscellaneous tag. Similarly in Figure 5.3(b) the mean number of modified files by a female in the implementation category was higher than that of the male and since there were also a high number of untagged commits by both male and female, we cannot conclude on this result.

5.3.6 Tag Prediction using Natural Language Processing (NLP) techniques

By observing Figure 5.3, we cannot conclude what gender does more of one task simply because so many commit messages were not tagged. We used tf-idf NLP feature extraction techniques, three different classification techniques and an ensemble of the three techniques to predict the tags. The results are shown in Table 5.5. We applied One-vs-the-rest (OvR) multiclass strategy which

Figure 5.3 Each bin shows the mean number of modified files within each category as originally tagged by each gender. Focus Teams: (a) Team P2 and (b) Lab Teams with the original tags



(a) (b)

divides a multi-class classification into one binary classification problem per class [sci; Ovr]. To each observation OVR assigns a probability of falling into each class. The observation falls into the class with the highest probability. We tried to predict if commit messages fell into one of the 9 categories. We used the already tagged messages as ground truth. We used logistic regression (LR) coupled with the One-vs-the-rest (OvR) multiclass strategy to classify the tags which resulted in an f-score of 0.631. We also used support vector classifiers (SVC) and naive bayes (NB) with OvR which resulted in f-score of 0.682 and 0.611 respectively. We tried to see if an ensemble of the three predictions would produce a better model but it resulted in an f-score 0.679 which was not better than the best model, SVM. Finally, we worked with the predicted tags from the model highest f-score which is OvR using SVC.

Table 5.5 Results of NLP Analysis

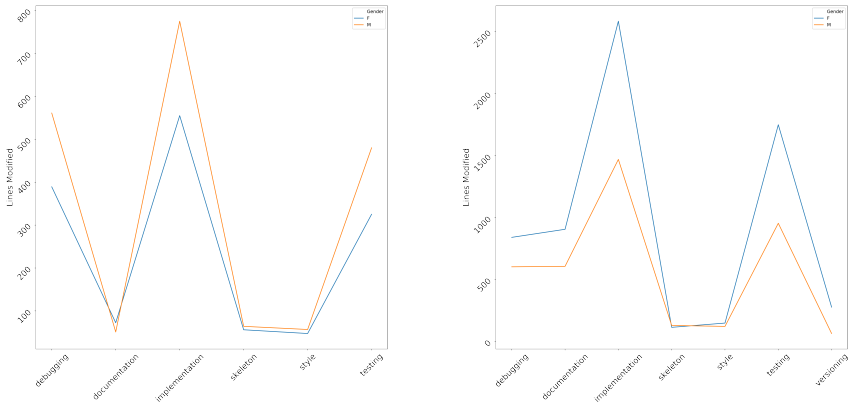
NLP Technique	**F-Score**	**Precision**	**Recall**
OvR using LR	0.631	0.631	0.631
OvR using SVM	**0.682**	**0.683**	**0.683**
OvR using NB	0.611	0.615	0.629
Ensemble of (LR, SVM, NB)	0.679	0.679	0.679

5.3.7 Visualizing Mean Number of Modified Files Committed in each NLP category in Multi-Gender Teams

Going by the NLP predicted tags, we can conclude based on Figure 5.4 that there was ample collaboration among genders. This means that each gender worked on a bit of each task and tasks assignment was not along gender lines in the projects we studied. Each line chart in Figure 5.4(a) represents the mean number of modified lines of code by each gender for the 13 P2 teams. We see that both genders worked the same tasks. On average, males modified slightly more lines of code than females in all tasks except documentation. Figure 5.4(b) represents the mean number of modified lines of code by each gender for the 93 Lab teams and for each category as predicted by our NLP SVC model.

Table 5.6 shows the percentage of commit message in each category by gender. Interestingly, although female students made slightly lower commits in most categories as shown in Table 5.6, on the average females modified more lines of code than male team members in all but the skeleton category while working on in Lab Teams. The lower average of modified lines under the skeleton category might be because more male teammates committed the starter project. Note that versioning tag was removed because some of the commit messages which include the keyword 'merge' were auto-generated by the platform. From the exploratory analysis, we can see that each gender worked on programming related tasks. To answer RQ2.2, by observing the line charts in Figure 5.4, female students collaborated with male students on each task.

Figure 5.4 Each bin shows the mean number of modified files by each gender within each NLP classified tag Focus Teams: (a) Team P2 and (b) Lab Teams with the NLP-SVC tags



(a) (b)

Table 5.6 Percentage of commits messages in each category by gender O* - represents data from original tags. SVC - represents data from SVC - NLP tags.

	Multi-Gender P2 Teams				**Multi-Gender Lab Teams**			
Category	Male % O* tags	Female % O* tags	Male % SVC tags	Female % SVC tag	Male % O* tags	Female % O* tags	Male % SVC tags	Female % SVC tags
debugging	40	60	59	41	62	38	**56**	44
documentation	36	**64**	43	**57**	**53**	47	46	**54**
implementation	25	**75**	**54**	46	48	**52**	**54**	46
misc	**89**	11	-	-	**78**	22	-	-
skeleton	45	**55**	**52**	48	51	**49**	**58**	42
style	**56**	44	**52**	48	**53**	47	**54**	46
testing	48	**52**	**56**	44	50	50	**54**	46
untagged	**63**	37	-	-	**55**	45	-	-

5.3.7.1 Percentage of commits messages in each Category by gender

Furthermore, we analyzed the percentage of commit messages by each gender. Table 5.6, shows that each gender in Team P2 contributed almost equally to the same types of tasks. [Thr] defined collaboration as a working scenario where both members contributed between 30%-70% to two or more common parts of the project [Thr]. Going by the **NLP-SVC provided tags**, the male members did slightly more of debugging (59%), implementation(54%), skeleton(52%), styling(52%) and testing (52%) and females did slightly more of documentation (57%) in mixed P2 Teams with equal number of male and female members. Females performed above 40% of each task, hence there was a healthy collaboration between multi-gender teams with equal numbers of male and female students in the class we studied. Similar results were observed in Multi-Gender Lab Teams columns shown on Table 5.6. To answer RQ2.2(a) and (b), based on the definition of collaboration in [Thr], females did not significantly do less programming tasks than males. One gender did not significantly do more task than the other. Both members contributed over 40% to all tasks.

5.3.8 RQ3: Can we predict the gender (M/F) of the authors using details of the commit messages as predictor variables and can gender predict the commit tag?

5.3.8.1 Predicting Gender

To answer RQ 3, we classify the gender (M/F) of a contributor on a project with mixed gender teams by analyzing the data of multi-gender Teams in P2 and Labs. We wanted to see what features are important in predicting gender contribution within a mixed team. Specifically, we want to see if commit tag (e.g. testing, implementation etc.) is an important predictor of gender. For example, to what extent can the commit tag tell the commit author's gender? Before we began training our models, we performed data pre-processing to transform some of the data to a format that the algorithms could use. To this end, we used LabelEncoder() in Python to binarize both the Original

and SVC commit tag values and gender values. We then split our data into 80% training and 20% testing for each of the models we used. We further split our training set into a training and validation set in order to optimize our hyperparameters. After achieving a satisfactory set of hyperparameters, we performed a final test on the testing set. When we created our models, we also included a 'class_-weight = balanced' parameter to ensure that there were an equal number of male/female commits that would be used to create each classifier. We applied Random Forest (RF), Decision tree (DT), Logistic Regression (LR), Naive Bayes (NB), KNN, Radius, and SVM and compared their performance in predicting gender.

5.3.8.2 Predicting Gender using only Commit Tags

We trained 7 models to predict gender using only the original commit tags. We included only commit messages tagged by students in this process. The results of this analysis is shown in column A of Table 5.7 where we show the balanced accuracy (BA) and f-score (F-S) of each of the 7 models. We trained 7 more model to predict gender using only the predicted tags from the NLP - SVC process and reported our results in column B of Table 5.7.

5.3.8.3 Predicting Gender using more features

We trained additional 7 models using 6 features namely 'added', 'deleted', 'modified', 'removed', 'original commit tag' and 'dayOfWeek' to predict the gender of the contributor behind a given commit. Lastly, we trained an additional 7 models using the 6 features but used the NLP-SVC classified commit tag instead of 'original commit tag'. We trained our model on Team P2 and Labs which has an equal number of male and female students in the team. We also removed the commit messages with the tag 'skeleton' because it contains the initial starter code provided by the instructor and not the code produced by the students.

We explored random forests and decision tree classifiers due their simple yet powerful techniques. For RF, we used 100 trees which produce the optimal validation accuracy. For DT, we used the criteria of entropy so that branches would be built only if they reduced the entropy of a classification. We then used Logistic Regression with min-max normalization and a C-value of .3 as well as Gaussian Naive Bayes. For the KNN and Radius classifiers, we assigned weights to the distance, used a power parameter of 3 to signify minkowski distance and tested k-values from 1 - 30. K=1 in columns C and D of Figure 5.6 provided the best results. We also tried the Radius neighbors classifier with R = 4 in column C and R =1 in D. Finally, we used SVM (LinearSVC) with a C-value of .005. The performances of these models are shown in Table 5.6. We determined the accuracy of a model by it's balanced accuracy (BA), which is the average recall obtained on both males and females. RF and KNN were our best classifiers for predicting gender using original tags with RF's accuracy of 83% and KNN's F-Score of 77%. Whereas RF and radius were best for predicting gender using NLP-SVC tags. Table 5.9 shows the feature importance of each feature in predicting gender. Each model gives the feature importance of each independent variable. We present feature importance for two top performing models RF and KNN. The lines of code added is the most important feature in predicting gender.

The next most important is the lines of code modified. Neither the original commit tags nor the NLP classified tags were highly important and as we observed in Table 5.6 columns A and B, these features alone are not good predictors of gender.

5.3.8.4 Predicting Commit Tags using Gender as a feature

Finally, we trained models using gender to predict the commit tag. We trained RF and KNN models which were high performing models from Table 7 but we got a very low performance as shown in Table 8, signifying that gender is not an important feature in predicting commit type.

Table 5.7 Performance of Models in Predicting Gender O* - represents prediction using original tags as a feature. SVC - represents prediction using SVC - NLP tags as a feature. Number in bold signify highest performance in each column.

	A (1 Feature - O* Tags)		B (1 Feature - SVC Tags)		C (6 Features - O* Tags)		D (6 Features - SVC Tags)	
Models	BA(%)	F-S(%)	BA (%)	F-S(%)	BA(%)	F-S(%)	BA (%)	F-S(%)
Majority Class	50	50	50	**71**	50	50	50	71
Random Forest	50	3	50	68	**83**	76	**76**	**77**
Decision Tree	56	42	52	54	82	75	73	75
Logistic Regression	50	0	50	0	55	56	50	7.6
Naive Bayes	50	0	50	70	60	46	51	42
KNN	50	30(K=1)	**53**	62	82	**77(K=1)**	72	73(K=1)
Radius	50	0(R=1)	51	68	78	73(R=4)	74	**79(R=1)**
SVM	**60**	**50**	52	62	63	58	55	47

Table 5.8 Accuracy of Models in predicting Commit Tags using Gender as a feature Highest values are in bold fonts

Models	**Original Commit Tags**	**NLP-SVC Tags**
Majority	**34**	**41**
Random Forest (Classify Commit)	23	20
KNN	24 (K=1)	20(K=4)

Table 5.9 Feature Importance of 6 features across top two performing models from column C and D - using permutation importance for each predictive analysis O* - represents prediction using original tags as a feature. SVC - represents prediction using SVC - NLP tags as a feature. Highest values are in bold fonts, commit tag features are in italics.

Model /Features	**Added (O*/SVC)**	**Deleted (O*/SVC)**	**Modified (O*/SVC)**	**Removed (O*/SVC)**	**Original Tags**	**Encoded NLP-SVC Tags**	**Day (O*/SVC)**
Random Forest	**0.33/0.32**	0.06/0.07	**0.30/0.31**	0.01/0.02	*0.12*	*0.13*	0.17/0.15
KNN	**0.38/0.32**	0.07/0.07	**0.38/0.36**	0/0.01	*0.15*	*0.11*	0.11/0.13

5.4 Discussion

Table 5.7 reports the performances of our 28 models in predicting gender. The column A in Table 5.7, is the result of using one feature which is Original 9 Tags to predict gender. We excluded commit messages with the misc tags and untagged messages. In column B, we applied NLP classified tags to predict gender. We observe that these results were poor and barely performed better than the majority classifier. In column C, 6 features including original clean tags were used. Whereas in column D we used the 6 features but replaced original tags with NLP classified tags. Table 5.8 shows the result of predicting tags using only the gender as a variable. The low performance in columns A, B, and Table 5.8 show that the commit tag doesn't accurately predict gender and vice versa. Columns C and D show that the performance of gender prediction improved when other 5 features were included. To answer RQ3, we cannot predict gender based on commit tag alone and vice versa which means in the class we studied the type of task worked on was not along gender lines. In the second two columns we used the 6 features to predict gender. Table 9 shows the feature importance of the 6 features in predicting gender. Elements of a commit such as the lines added and modified had higher feature importance than the commit tags. This shows that each gender worked on similar tasks therefore the tags alone could not be used to differentiate the genders.

5.5 Limitations

Other factors such as competence of each student programmer irrespective of gender might affect the choice of tasks undertaken by each gender. If teams were requested by students instead of assigned by instructor, this might affect the distribution of tasks as team members in requested teams might already have a plan on how to work together.

5.6 Conclusion

We explored the team work habits by each gender in a Java programming class. In the class we studied, students were advised to tag their commit messages to indicate what type of task they worked on. Each gender responded about the same way to the recommendation to tag commit messages. We found that both genders collaborated to complete the project. Future work will replicate this work for multiple semesters and years to confirm the validity of our results across a larger population. This study shows that although female students are less likely to show interest in the programming fields as seen from theory discussed in section 5.2.1, when females actually choose Computer Science, they are more likely to make significant contribution and healthy collaboration as defined by [Git20], comparable to their male counterparts when working with males students on programming projects. This study is beneficial for identifying teamwork habits among multi-gender teams. The knowledge that both genders have important roles to play in the success of a programming project is encouraging to females who aspire to take up a career in computing.

CHAPTER

6

STUDY 3: PEDI - PIAZZA EXPLORER DASHBOARD FOR INTERVENTION

abstract

There is little research on how professors could use learning analytics to gain insights into student forum usage or how these insights might impact courses. In this chapter, we introduce PEDI (Piazza Explorer Dashboard for Intervention) a software tool that presents useful visualizations of forum activity on Piazza, a learning management tool, to instructors. We outline the design principles and data-informed recommendations that led to the design of the PEDI tool. Our prior research has highlighted two critical periods in students' forum engagement. We've shown that early forum engagement while students work on assignments positively correlates with class average performance. On the other hand, extremely high engagement toward the final deadline predicted lower class average performance. PEDI takes these findings into consideration to detect and flag these critical periods. PEDI informs instructors using clear visualizations to promote data-informed instructional interventions. Such real time analytics could provide insights to instructors about their classes, consequently, improve class performance and pave the way for a new generation of online tools that would adequately meet the needs of online learning.

6.1 Introduction

Data analytics and dashboards are useful for providing awareness to stakeholders about visual features of data systems, processed information and overall system activity. Data captured from educational systems are data sources for learning analytics dashboards (LADs), these sources include

intelligent tutoring systems [Sab20], college programming and collaboration tools [AM20; Suh20; Isv20] and help-seeking tools [B9; Agh20; Mal20].

Due to the emergence of internet technology and the intense competition between numerous software applications for human attention, it is necessary that LADs should be designed to be memorable and effective in addition to being functional. It is imperative that LAD designers are intentional about the charts, graphs, plots, and animations they decide to include in their applications.

In this paper, we introduce PEDI, a learning analytics dashboard layer on top of the Piazza question and answer forum system. Piazza allows students to ask questions in a forum-type format. On this forum, discussion topics are grouped by assignments. Each assignment has a folder containing posts relating to that assignment on the forum. Students ask questions by creating a post on the forum and tagging the folder relevant to their question. Both instructors and students are free to respond to posts and for every question asked on the forum, Piazza records information such as time the question was asked, question content, replies, number of replies, number of views, and number of likes [Pia].

6.1.1 Inspiration for PEDI

PEDI was designed based on recommendations from our prior studies of students' activities on Piazza [B9]. When students ask for help on the forum, they have usually begun working on that assignment, and may have been working on it for a while [B9]. Students often use the Piazza forum near an assignment deadline rather than seeking help early. However, research shows that students who start working earlier have more time to seek assistance when they get stuck [Edw15a; Edw09a; Kaz17a].

Detecting Late Help-Seeking: PEDI provides filtering and ranking features for instructors to monitor current levels of forum activity in the first half of an assignment's duration. For each assignment, PEDI presents forum activities as visualizations and allows instructors to compare current levels of forum-usage and rate of help-seeking with those of previous semesters. This comparison can help instructors identify patterns of help-seeking for possible interventions.

Detecting Unusually High Rate of Help-Seeking: Our prior study also found that rates of forum activity towards the end of an assignment duration negatively correlated with class average performance. We identified the need to alert course instructors about unusually high levels of interactions on the forum compared to levels found in previous semesters [B9].

Piazza provides limited statistics and visualizations. It provides no filtering, ranking, or alerts. As found in our prior work, the relative timing of class forum posts to assignment deadlines, and the comparison with prior semesters are important. PEDI provides these features which present instructors with important insights which can lead to interventions that improve performance.

This is the link to [PEDI DEMO].

Table 6.1 Definition of Features of Forum Posts

Features of Forum Activity	**Definition N= total number of students in the class who attempted the assignment**
Average # of Posts	(Total number of all posts related to an assignment)/ N
# of Likes	Total number of likes on all posts related to an assignment
Fraction of the class that Posted	(Total number of students that posted about an assignment)/N
Fraction of the class that Viewed (Unique Views)	(Total number of students that viewed all posts of an assignment)/N
Average # of Questions	(Total number of questions asked by students in an assignment)/N
# of Replies	Total number of replies by both students and instructors in an assignment
% of Early Posts	(Percentage of total posts made half-way through the project)/N

6.2 Related Work

One major motivation for this research is that there is no real time knowledge or insights on Students' Usage of Online Tools. Hoyos et al, in their study highlighted that one major challenge of teaching and learning analytics is that times in which the learning activities occur and their analyses are very separate and not in real time [Hoy20]. This limits teachers from taking data-informed actions immediately to positively impact student's learning outcomes.

Specifically, real time insights from online discussion tools are scarce Kovanovic et al in 2017 highlighted the need to study students activity on discussion forums in order to leverage on the benefits of this kind of data. The challenge is that the analyses of online learning data are both time-consuming and labor-intensive, and require a highly advanced set of data science and analytics skills, often not available to instructors [V.17].

6.2.1 Learning Analytics Dashboard (LAD)

LADs are typically designed for instructors to monitor students' progress and for the students to track their own progress in self-regulated learning (SRL) environments. There are several aspects to developing LADs such as functionality, ease of use and design.

6.2.2 Functionality

A dashboard is functional if it provides features that supports its users' goals. One mistake in dashboard design is to present as much data as possible instead specific useful information. There are different approaches to building a functional LAD.

6.2.2.1 Process Mining Approach

The process oriented approach considers the steps involved in processes in a system as well as the frequency at which these processes are performed. Martinez-Millana et al. 2019 implemented and showed the effectiveness of a process mining dashboard in identifying in real time the different preferences and role each professional has in the operating room [A19]. In LADs based on these approach, the learning process is visualized.

6.2.2.2 Analytical Approach

This approach applies process mining but goes a step further to present data-driven recommendations. LADs allow users to filter, an otherwise huge volume of information, using meaningful features and criteria relevant to the educational setting such as performance, gender, demography, prior experience, and background. Shabaninejad et al. describes these kinds of user-driven drill-downs as manual [S.S20]. An analytical approach involves data-driven drill-downs. Shabaninejad et al. also automated the drill-down process to eliminate the challenge of manually finding drill-downs that provide insights and allow instructors to identify student cohorts of interest. The PEDI dashboard is based in the analytical approach, where we allow instructors (users) to manually choose from our analytical, data-informed drill-downs to find interesting patterns in forum activities.

6.2.3 Ease of Use

A dashboard's perceived ease of use is an important aspect of its acceptance [Jan13]. In choosing the right visualization, Janes et al. defined the pull and push approaches. In the pull approach, users specify what information they want to access, designers should provide visualizations that allow users to freely filter and explore the data. In the push scenarios, the dashboard is basically designed to push certain useful information to the users. The visualization components should be selected to highlight surprising or anomalous information that might be present in the data. An easy to use dashboard would employ data-driven strategies to choosing the right data, specific goals and apt visualization techniques [Jan13]. The PEDI tool pushes critical events such as low levels of early forum usage and higher levels of late forum usage.

6.2.4 Design

Human's attention span have decreased over time due to the onset of internet technology [McS15; Pan14]. The design techniques adapted by applications on mobile phones possess complex aesthetic interface to keep users continually drawn to their phones. Therefore in addition to a design being functional and easy to use, in order to retain a user's attention, the design should have good visual aesthetics.

6.2.4.1 Memorable Dashboards

In a 2013 study by Borkin et al., the authors collected a set of 2070 infographics and visualizations to discover what makes a visualization **memorable** [Bor13]. They argued that discovering factors that contribute to memorability of visualizations will helps us understand what makes visualizations effective. They concluded that complex, colorful and pictorial visualizations with familiar natural world objects were found to be memorable.

6.2.4.2 Effective Dashboards

Pandey et al. 2014 explored the effectiveness of visualizations [Pan14]. From their pilot studies of over 150 participants, they discovered that dashboards have more persuasive effects on neutral participants when they contain more visualizations than tables [Pan14].

The PEDI was designed to monitor useful data indicators discovered in prior studies[Chapter three]. We also applied design recommendations from the above studies we have discussed [Bor13; McS15; Pan14].

6.3 METHOD

6.3.0.1 Data and Population

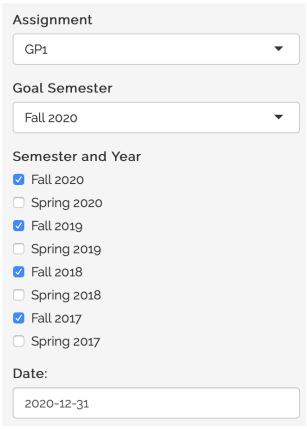
The goal of the PEDI tool is to provide instructors with useful forum trends before the end of each assignment. We analyze class-level forum activities on Piazza, used in one Computer Science courses of a southeastern research-intensive US university. We analyzed forum usage of 1615 students and about 10 course instructors from one CS2-OOP [Por18a] course from Fall and Spring semesters of 2017 through 2019; We omitted some semesters prior to 2017 because the data was either not available or forum posts were not categorized in the format we required. We plan to replicate our results on another course CSC 226 (a Discrete Math course) to improve generalizability of our results.

6.3.1 Design Principles

We carefully selected PEDI's design components to ensure a functional, easy to use, memorable and effective dashboard, with good visual aesthetics.

6.4 DESIGN

The goal of the PEDI tool is to provide instructors with useful forum trends before the end of each assignment.

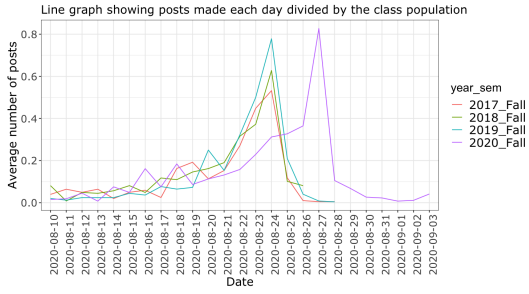


(a)

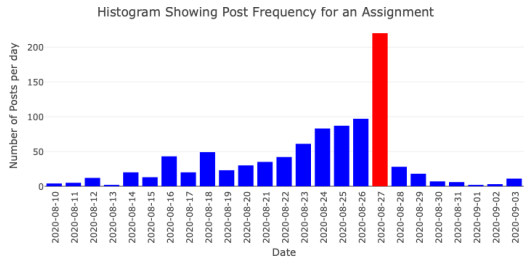
(b)

Figure 6.1 (a) Sidebar Panel Control (b) Tree Map Control

Figure 6.2 (a) Interactive line graph showing average posts per day (b) Interactive histogram showing the number of posts per day.

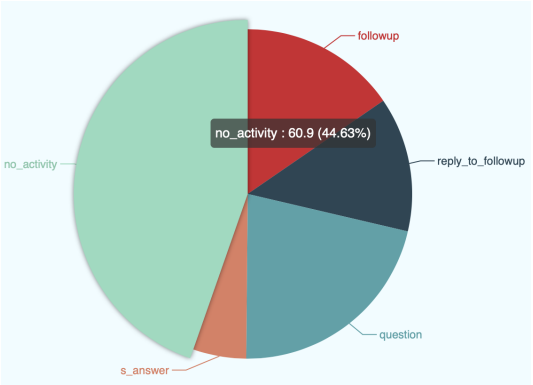


(a)

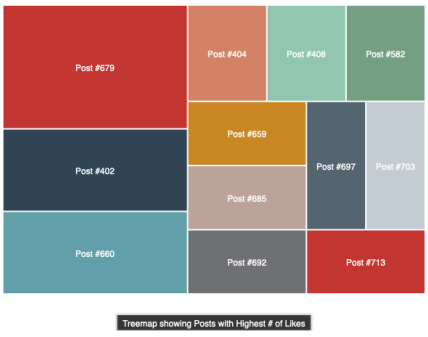


(b)

Figure 6.3 (a) Pie chart showing number of students in the class who performed certain activity on the forum (b) Tree Map showing forum posts with the highest number of likes

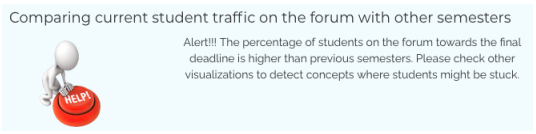


(a)

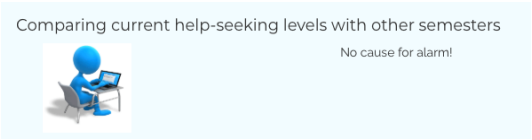


(b)

Figure 6.4 Alert System displayed on the PEDI Dashboard



(a)



(b)

6.4.1 System Capabilities

PEDI was built using Shiny, an open source R package. The main data analysis was done in Python[B9]. We used R's reticulate package [R18] to run Python scripts in R environment.

PEDI displays information to instructors using the following components: *Average Post Frequency* line graph, *Forum Activity* pie chart, *Post Frequency* histogram, *Post Likes and Replies* tree map and a few tables to provide context of students' forum activities.

6.4.1.1 Average Post Frequency

PEDI allows instructors to compare activities of the current semester with previous semesters using the sidebar panel control and *average post frequency* line graph in Figures 6.1a and 6.2a respectively (see Appendix A.2). The line graph shows the average posts per day (this is the number of posts per day divided by the class size) for an assignment during the goal semester. The goal semester is the current semester. The goal semester in Figure 6.2a is Fall 2020. Observe that the highest peak for the goal semester is slightly higher than those of previous semesters. This means that students are asking more questions than usual towards the end of an assignment duration. Instructors might wish to investigate the reason for this by exploring other charts.

6.4.1.2 Post Frequency

PEDI allows instructors to dig deeper into post traffic that catches their attention. For example, in Figure 6.2b the 08/27/2020 bar had the highest number of posts, when this bar is clicked, all the posts for that day are displayed as a table (not shown) below the histogram. This is important because questions asked on this day might be in the an area of shared difficulty by the class. Instructors might decide to intervene by sharing the posts and clarifying expectations or challenges during class time.

6.4.1.3 Forum Activity

The pie chart on the PEDI tool allows instructors to observe the number of students who have not been active on the forum and what activities are being performed on the forum (see Figure 6.3a). Instructors could decide to email students and motivate them to use the forum. They could also provide positive feedback to students providing answers to other students' questions or consider recruiting those students as future graders.

6.4.1.4 Post Likes and Replies

Our prior research found a negative correlation between the number of likes on posts on an assignment and the class average performance [B9]. The tree map in Figure 6.3b displays posts with the highest number of likes. The higher the number of likes, the bigger the rectangle. A high number of likes on a post could be important to instructors for two reasons. The first reason is that the post might be particularly helpful for an assignment and the post should be highlighted and visible for all students to benefit from it. The second reason is that posts with a high number of likes may indicate difficulty faced by majority of the student cohorts which may lead to lower class performance if not addressed properly. The tree map has a control in Figure 6.1b that allows users to also view posts with the highest number of replies. sers to also view posts with the highest number of replies.

6.4.1.5 Alert System

Based on our prior research [B9] the following scenarios should trigger an alert indicating a situation where an instructor may want to consider an intervention.

For an assignment:

- when the level of *no-activity* during the first half of an assignment's duration, is higher than the average across previous semesters
- when the percentage of students on the forum during the first half of an assignment's duration, is lower than previous semesters
- when student traffic (percentage of class actively posting on the forum) close to the final deadline is higher than previous offerings of that assignment
- when the average number of questions asked towards the final deadline is higher than other semesters

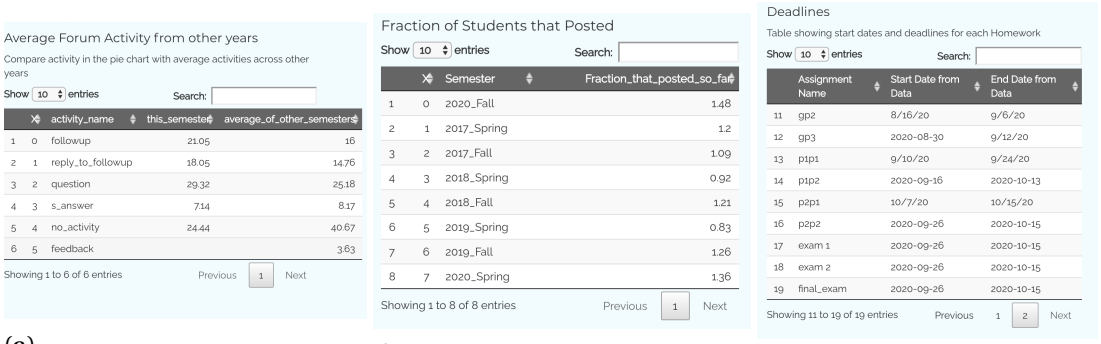
PEDI alerts reveal concerns that instructors may wish to address by promoting early help-seeking and resolving issues with assignments.

These components are all on one page to provide instructors with just the right amount of information to make a decision and act on it. PEDI's components provide insightful summaries to reduce the time instructors spend getting insights. Furthermore, PEDI follows the basic principles of the Visual Information Seeking Mantra: overview, zoom and filter, details-on-demand, and relate as described by Shneiderman [**b26**].

6.4.1.6 Tables for reference

We used more visualizations and less tables. We included one table at the bottom of the histogram to provide context of students forum activities to instructors. We also included a few tables (see Table 6.5) at the bottom of the dashboard for instructors to reference information such as assignment deadline and actual numbers of interaction that triggered the alerts.

Figure 6.5 Tables on the PEDI Dashboard. (a) Average number of Questions asked (b)Fraction of the class that post (b) Assignment Deadlines



Average Forum Activity from other years

Compare activity in the pie chart with average activities across other years

Show 10 entries Search:

		activity_name	this_semester	average_of_other_semesters
1	0	followup	21.05	16
2	1	reply_to_followup	18.05	14.76
3	2	question	29.32	25.18
4	3	s_answer	7.14	8.17
5	4	no_activity	24.44	40.67
6	5	feedback		3.63

Showing 1 to 6 of 6 entries Previous 1 Next

(a)

Fraction of Students that Posted

Show 10 entries Search:

		Semester	Fraction_that_posted_so_far
1	0	2020_Fall	1.48
2	1	2017_Spring	1.2
3	2	2017_Fall	1.09
4	3	2018_Spring	0.92
5	4	2018_Fall	1.21
6	5	2019_Spring	0.83
7	6	2019_Fall	1.26
8	7	2020_Spring	1.36

Showing 1 to 8 of 8 entries Previous 1 Next

(b)

Deadlines

Table showing start dates and deadlines for each Homework

Show 10 entries Search:

	Assignment Name	Start Date from Data	End Date from Data
11	gp2	8/16/20	9/6/20
12	gp3	2020-08-30	9/12/20
13	p1p1	9/10/20	9/24/20
14	p1p2	2020-09-16	2020-10-13
15	p2p1	10/7/20	10/15/20
16	p2p2	2020-09-26	2020-10-15
17	exam 1	2020-09-26	2020-10-15
18	exam 2	2020-09-26	2020-10-15
19	final_exam	2020-09-26	2020-10-15

Showing 11 to 19 of 19 entries Previous 1 2 Next

(c)

6.4.2 PEDI workflow

A walkthrough demo displaying PEDI's design and functionalities will be presented at the VL/HCC conference.

PEDI is meant to be used at least twice within the duration of an assignment (halfway and two-thirds of the way to an assignment deadline (a few days before deadline)).

6.4.2.1 The Early Forum Activity tab

helps instructors reflect on their student forum usage from the time the assignments are released until halfway to the deadline. The *Average Post Frequency* line graph which shows the average number of posts per day and the *Forum Activity* pie chart which shows the percentage of students and what activities they are performing on the forum, help the instructor get an overview of the current forum usage. To filter, zoom in, or get details on an assignment of interest, the instructor may select the assignment and adjust the due date using the date control in Figure 6.1a. They may also relate the current semester to a previous on the line graph in Figure 6.2a. This may help the instructors reflect on how effective their past interventions were or detect if their recent interventions have improved early help-seeking.

6.4.2.2 The Tracking Assignment Difficulty tab

is best for checking forum use within 2/3 of an assignment's duration. Heavy forum usage closer to the deadline may indicate higher assignment difficulty. This page provides a forum usage overview in the *Average Post Frequency* line graph, *Forum Activity* pie chart, *Post Frequency* histogram, and *Post Likes and Replies* tree map, each with its own filtering, zoom, and 'relate' options. The *Post Frequency* histogram provides access to more detail.

6.5 Conclusions and Future Work

In this chapter, we introduced PEDI - a LAD layer on top of the Piazza platform. We have discussed what makes an effective and memorable LAD. We have described how the design and implementation of the PEDI tool follows recommendations from prior work [B9], and expert design strategies [Bor13; McS15; Pan14; Por18a].

PEDI addresses the knowledge gap by analyzing and presenting visualizations of forum activities in a meaningful way to instructors while students work on assignments. PEDI enables instructors to gain important insights by comparing forum usage of one class to another, relative to assignment deadlines. As found in prior research, interventions based on these insights have the potential to improve class performance. In the next chapters, we present how we will apply qualitative methods to investigate how instructors use and respond to the PEDI tool.

CHAPTER

7

STUDY 4: WHAT IS PEDI TO INSTRUCTORS?

abstract

To gather data, we conducted interviews with 7 course instructors, as described in more detail in the methods section below. The interviews were transcribed and analyzed using thematic analysis. This data leads to the following 3 themes: 1) current instructor habits which describes current practices surrounding help-giving and forum use by instructors, 2) instructors' perceptions of PEDI which consist of instructors' opinion and review of PEDI, and 3) future recommendations to improve PEDI, Piazza, and help-seeking forums in general. These themes will be useful for understanding the optimal way to utilize the PEDI tool for instructional purposes.

7.1 Introduction

In Chapter 6, we introduced PEDI (Piazza Explorer Dashboard for Intervention). In this chapter, we report a qualitative analysis that investigates how the platform was used by instructors. We introduced PEDI to the potential users of the system - instructors, and obtained their opinions about the PEDI tool. This is important because our prior studies show that Piazza forum visualization and analysis exposes useful insights in a class to the instructor. We therefore designed and implemented the PEDI dashboard based on these findings and we wish to evaluate PEDI's usefulness with instructors' help. This research aims to provide insights to instructors, specifically to understand the state of their class better using the PEDI tool. The study was conducted with 7 instructors who have some years of teaching experience with Piazza in a Southeastern research-intensive University.

7.1.1 Focus of the Study

Our focus is to observe how instructors would use the learning analytics information about forum use in their classrooms. We formulated interview questions and activities to confirm the hypotheses as described in section 7.1.2. Participating instructors performed the four tasks listed in section 7.2.3 in order for instructors to:

1. Explore and familiarize themselves with the PEDI tool
2. Detect late forum usage half way through an assignment's timeline
3. Detect higher rates forum use during 2/3rd of an assignment timeline in order to determine if higher forum activity is as a result of assignment difficulty
4. Examine posts with the highest number of likes and replies to determine if these are useful to the class. We anticipate that instructors might wish to intervene based on these activities.

7.1.2 Hypotheses

For this study, these are our hypotheses:

1. Detecting Late Forum Use Scenario: We presented the following statement to participants: ***Our prior findings showed that the lower the rate of early forum use, the lower the class average performance***. In this scenario, the participants explored forum traffic in the first half of the assignment duration, in order to detect lower forum use in the first few days of an assignment's timeline. Hence, during the earlier points in the assignment's timeline, ***we hypothesized that since the teaching staff is informed of our findings stated above, they would show concern if they detected lower forum usage during the first half of the assignment timeline and seek to encourage students who have not begun the assignment to start***. Participants explored this scenario in Task one described in section A.2.
2. Detecting Higher Rates of Forum Use Scenario: We presented the following statement to participants: ***Our findings showed that higher forum usage towards the end of the assignment timeline is related to lower class average performance***. In this scenario, the participants explored forum traffic in the first 2/3rd of the assignment duration, in order to detect higher forum usage towards the end of an assignment's timeline and understand late forum activities. Hence, during the later points in the assignment's timeline, ***we hypothesized that if the teaching staff is aware of our findings stated above, they would show concern about higher forum usage towards the end and seek to intervene in different ways*** for example:
   (a) Ensure that students are clear about the requirements,
   (b) Encourage students who have not begun the assignment to start,
   (c) Encourage students who are inactive on the forum to be more active on the forum, and

(d) Give an extension or intervene in some other way. Participants explored this scenario in Tasks two and three described in section A.2.

We analyzed the interview data to answer the following research questions:

7.1.3 Research Questions

1. **RQ 3.1** Is the PEDI dashboard useful for instructors?
   (a) What are beneficial use cases of PEDI dashboard?
   (b) What particular information could influence instructors' decision to intervene if they were using the PEDI dashboard in real time?
   (c) What might an instructor choose to do to intervene given the information on PEDI?

7.2 Methods

In this section we describe the methods for this qualitative study with instructors.

7.2.1 Recruitment

We recruited 7 instructors for the study. These participants have had experience as instructor or teaching assistant in the Java programming class (CSC 216), which we examined in study one [B9]. After signing the informed consent form and completing the pre-survey about their teaching experiences, they participated in the PEDI interview session.

7.2.2 Participant Procedure

The participant procedure consisted of the following steps:

1. Completing a consent form and providing their availability,
2. Completing a 3 question pre-survey about their prior teaching experience,
3. Participating in a remote PEDI interview session (1 hour 50 minutes) on Zoom that was audio, video, and screen recorded where they :
   (a) Explored the PEDI tool using guided tasks,
   (b) Answered some questions while using the tool, and
   (c) Participated in interviews after exploring the tool.

7.2.3 Planned analysis methods

During the PEDI interview session, instructors were asked to perform the tasks highlighted in the A.2 section. As shown in section A.2, the pre-activity questions were necessary to assessing instructors' use of existing tools to better understanding their responses. The activity questions were related to RQ3.1a and RQ3.1b, to help us gauge PEDI's usefulness, when instructors would opt to intervene and what methods they would use. The post-activity questions was useful for collecting summaries of participant experiences with the PEDI tool.

We performed thematic analysis on data from the interview process. **Thematic analysis** is a data analysis method used in qualitative analyzes such as interviews and think aloud sessions [Vir21]. We examined the data to identify common themes – responses and patterns of responses that come up repeatedly. Thematic analysis was applied to analyze the audio, video and screen capture data. Three researcher raters tagged and created different codes from the transcripts, grouping them into categories, and then into data derived theories. We followed Braun & Clarke's 6-step framework [Bra06] which are:

1. **Becoming familiar with the data** We went through the interview responses to become familiar with the information provided by the participants.

2. **Generating Initial Codes** Our interview was driven by our research question and from step 1 above, we generated codes using open coding (i.e. the raters generated codes as we worked through the responses instead of pre-set codes). We distributed researchers such that each interview was reviewed by two researchers. Three researchers reviewed the data. Researchers 1 and 2 reviewed five transcripts each and Researcher three reviewed four transcripts. In total there where 14 reviews on the seven transcripts. While reviewing each interview, researchers got familiar with the data by reading the transcripts repeatedly, highlighting interesting phrases in each transcription, and gave descriptive codes for the highlight. After an initial run-through of the interviews, the researchers convened to merge similar codes, address overlaps, based on similar examples and resolve any ambiguities in the codes. A total of 334 codes were created. We then agreed upon on a concrete set of codes and all the interviews were reviewed a second time by a different researcher from the researcher that did the first pass. This is shown in Figure 7.1

   Table 7.1 shows that each participant's sessions of Zoom interview was coded by two researchers. The letter in the table represents the first letter of the participant's name and the number represents the order in which the researcher coded the interviews for that session. For example, Researcher 1 coded Eden's interview in the first pass, then Researcher 3 took a second pass on Eden's interviews.

3. **Themes** After the second review, researchers convened to categorize the codes under more general sub-themes. 26 sub-themes were formed from the process. Finally, three overarching

Table 7.1 Each participant interview was tagged by two researchers as shown in the table.

	Akin	Bianca	Caleb	Dami	Eden	Fola	Gad
Researcher 1		B2		D2	E1	F1	G1
Researcher 2	A1	B1	C2			F2	G2
Researcher 3	A2		C1	D1	E2		

themes were formed around patterns of responses and codes were organized by themes. These themes and sub-themes are found in Table 7.3.

4. **Review** We reviewed the themes again formed and made necessary modifications to resolve ambiguities and improve agreement between researchers. For example some sub-themes were initially grouped under participants' experiences with PEDI, however we realized that they fit better when grouped under the future work on PEDI theme because they are constructive criticism of the tool. We also determine via instructors' responses if PEDI would likely be adapted into instructors' help-giving routine.

5. **Define Themes** We defined all the highlighted themes in section 7.3 and created a thematic map in table 7.3 that maps the themes to sub-themes. We report the themes in detail in section 7.4.

6. **Write Up/report** Our results are presented as case studies in the following section along with results from overall participants' experiences.

7.3 Case Studies

In this section, we first provide initial results in case study form to enable the reader to develop an understanding of each participant's teaching background and experience with Piazza. Each case study presents a teaching staff's characteristics followed by their unique view of forum engagement on Piazza and interaction with PEDI. Participants consisted of were **six teaching assistants and one course instructor**. All participant names have been replaced with fictitious names.

7.3.1 Case Akin (P1)

Akin has been a teaching assistant (TA) for CSC 216 for 3 semesters and has used Piazza first hand every week during those semesters. She explained that she and other teaching assistants took turns in the week to review students' posts and provide answers to their questions in a timely manner. Akin shared some aspects in the PEDI tool that will potentially help the TA team to better how the students are fairing and find out what concepts students are having problems with and address that. Students ask a lot of questions, and deciding which question to answer first, is an issue for Akin. She shared that using the PEDI tool, she sees the post that got the highest likes or replies today, and tries to address the problem mentioned there first, that could be one solution to the problem.

7.3.2 Case Dr. Bianca (P2)

Dr. Bianca, the most experienced participant, having taught the course for over 10 years and has coordinated teaching assistants in using Piazza in her classroom to give help to her students, noted that she seeks to strike a balance between helping students in good time and allowing them to figure it out on their own. She stated that Piazza currently shows her what questions are yet to be answered and this triggers her to nudge the TAs to increase their speed in attending to students' questions. She added that most likely due to the volume of students' questions, Piazza's "Search" feature performs poorly for their large class and a system that mines students' questions to detect common patterns in posts would prevent students from asking questions already posted. She shared that it would be great to have a reflective tool that could show historical problems that they saw during every offering of the course in order to make necessary adjustments. Dr. Bianca liked to see forum activities in aggregate, "I really appreciate PEDI from a historical perspective, like being able to like think back about the classes I offered, and the changes I made, and then it's really cool to see some of the charts and have them showcase some of the changes, that was really cool". She shared that the next step would be to detect what type of intervention will help versus what type of intervention leads to behaviors that may not ultimately improve learning or performance.

7.3.3 Case Caleb (P3)

Caleb uses Piazza a few times a day in search of questions that students might have. He reads through each question in search of common overarching difficult concepts by noting the number of people asking the same kinds of questions. Although he was a teaching assistant for CSC216 for one semester, he shared that common questions highlight difficult areas most students are struggling with. While helping students out on Piazza, Caleb looks for questions that have not been answered but leaves some room for the students to help each other by not immediately responding to questions as soon as they are posted. Caleb noted that posts that had so much more likes than others are worth looking at. He shared that "if I didn't have this feedback from PEDI, because I'm not necessarily constantly stalking Piazza looking for these issues, I probably wouldn't have noticed lower levels of engagement compared to other semesters".

7.3.4 Case Dami (P4)

Dami is experienced, having been a teaching assistant for CSC 216 for 3 semesters, she started to note down things that could improve help-seeking and help-giving on discussion forums. She shared that a system to track posts history from a single person would be beneficial for student modelling. This would help instructors understand how to adequately help each student in a personalized way and reward students who provide good answers to their fellow students' questions. In addition, she indicated that it would be helpful if each post type (such as questions, answers, follow up) had a structure or template for students to follow, this would make it easier to know if a question has indeed been answered or not and make help-giving faster. Like Dr. Bianca, Dami would also like a

better way to search for questions on help-giving platforms. She currently checks Piazza before her office hours because most of the issues students would bring to office hours are usually the most pressing issues on Piazza. Dami stated that PEDI makes TAs more concerned about how the class as a whole is fairing as opposed to seeing TA duties or Piazza as a ticketing system where their goal is simply to answer the questions in a timely manner.

7.3.5 Case Eden (P5)

Eden would like to use a system that sorts out students by their level of engagement on help-seeking forums in order to detect students who might need some extra help in improving their performance. Eden, who was a teaching assistant for CSC 216 for 2 semesters, confidently noted that students who do not participate on the forum are either doing well on their own or really struggling but feel embarrassed to participate. He seeks for ways to encourage students to leverage help-seeking platforms to their full extent. Eden confidently stated that “PEDI provides cool data statistics I would never in a million years have been able to see on Piazza, because it gave me insights that I could never see on my Piazza feed, this delved into a much deeper level of data and I thought that was incredibly helpful”. Eden compared PEDI to a current assignment progress motivation strategy, “I have always wondered, is it wrong to show students what percentage of the class are finished with the assignment? It is almost like shaming those who haven’t started. I think this is really cool that instead, we could show students that analytically the overall class completion is far lower and engagement is far lower than in previous semesters".

7.3.6 Case Fola (P6)

Fola had experience helping students on Piazza while he was a teaching assistant for CSC 216 for a semester. He searches for questions that have not been answered and answers them from the easiest to the most difficult. Fola believes that a help-seeking platform should have a good user experience to encourage help-seeking and the help-giving process, he stated that he would like a less cramped user interface, that is intuitive and makes it easy for instructors to navigate through posts for different assignments. The major highlight of his interview was a need for a forum analytics system that provided sentiment analysis of posts in order for instructors to understand the major themes of students’ struggles for each assignment. Fola noted that “PEDI looks like Piazza, it has a similar vibe to Piazza. I guess it’s mainly because of the colors”. He wished Piazza could add PEDI to its statistics feature, “having used this tool, I know that Piazza doesn’t give me this information, I wish it would”.

7.3.7 Case Gad (P7)

Gad, a teaching assistant in one offering of CSC216, responded to students’ questions via Piazza. He noted that they received up to 200 posts per day towards project deadlines and this made it difficult for teaching assistants to provide timely responses. He suggested that a possible solution would

Table 7.2 Table showing different types of analytics and what percentage of the seven participants are interested in them.

S/N	Analytics	Interest (%)
Q1	The days with the highest rate of interaction	71.43
Q2	The number of likes on a comment	71.43
Q3	The number of replies on a comment	71.43
Q4	The percentage of the class that have not posted on the forum	28.57
Q5	The day that had the highest number of questions	85.71
Q6	The day had the lowest number of questions	28.57

be to have more teaching assistants during periods of high traffic and allocate questions to each teaching assistant as soon as the questions are posted. He also pointed out that having to answer repeated questions or multiple similar questions is time consuming. Gad thinks that posts with high likes are issues that are interesting and gaining attention among students.

7.3.7.1 Pre-Activity Questions

At the beginning of the interview session, participants answered some pre-activity questions where they selected what kind of insights from forum activities they were interesting in seeing. Table 7.2 shows each analytics and what percentage of the participants showed interest. Six out of the seven (85.71%) participants were interested in the day that had the highest number of questions (Q5) whereas, Q4 which represents the percentage of the class that have not posted on the forum and Q6 which represents the day had the lowest number of questions, were the least popular choices the as only two of the seven participants (28%) were interested in them.

7.4 Themes

Here, we discuss the themes and sub-themes the researchers observed from the thematic analysis process. We analyzed the interview survey responses.The themes and sub-themes are highlighted in table 7.3.

7.4.1 Current instructor habits

During the interviews, participants were specifically asked to share their experience with giving help to students and strategies they currently use in giving help to students in their classes.

7.4.1.1 Current Instructor Piazza Usage

Instructors use Piazza mainly as an asynchronous help seeking forum. Students typically post a question about an assignment to the forum and receive a response from another student, a TA, or instructor. During the interviews, many of the teaching assistants described a scheduling

Table 7.3 Themes and Sub-themes

Current Instructor Habits - Current Practice - Common Class Struggles	Instructors' Perception of PEDI	Future Work and PEDI Improvements
Current Instructor Piazza Usage	Importance of Post Likes Analysis	Visualization Improvements
Piazza Pain Points	Importance of High Replies Analysis	Future Work on PEDI
Meaning of Low Forum Traffic	Importance of Average Posts Frequency (Line Graph)	PEDI Future Usage Frequency
Meaning of High Forum Traffic	Importance of Posts Frequency (Histogram)	Interest in Grades
Current Knowledge of Student Piazza Usage	Importance of Forum Activity (Pie Chart)	
Common Class Struggles	PEDI Useful in Evaluating Instructional Methods and Interventions	
	Importance of Likes and Replies (Tree map)	
	PEDI Praises	
	Milestones	
	Surprises	

technique used by the instructor to help students get timely responses to their questions. Some instructors assigned TAs days of the week to respond to Piazza posts while others assigned Piazza shifts. Dr. Bianca scheduled 3 students to take Piazza passes each day: one in the morning, one in the afternoon, and one in the evening. Three of the participants (Dr. Bianca, Caleb and Fola) suggested that PEDI may be useful in scheduling TAs. Using the line graph to see spikes in Piazza activity from past semesters may help instructors predict when activity will be higher than normal; then, the instructors can allocate more TAs to that time period. Fola says that TAs expect highest activity before due dates; "there are times right before due dates" where the students "tend to question dump on Piazza".

These scheduled times are important as Dr. Bianca pointed out "I rely on my TAs to let me know if they see patterns" in the questions on Piazza because the TAs "do a majority of the answering on the course forum". Dr. Bianca said that while she doesn't use Piazza often, she tends to read the automated Piazza digest emails to monitor the forum.

There seemed to be some consensus between the participants on which Piazza questions they prioritize and answer first. Akin, Caleb, Dami, Eden and Fola thought that it's important to answer the oldest unanswered questions first. Akin admitted that although she tends to prioritize these questions, it is hard for her to know the right way to prioritize her time on the forum. When answering questions on Piazza, Fola adds that they tend to answer the questions that he is most comfortable with first.

Three of the participants mentioned that while they check Piazza, they intentionally look for patterns in student questions (Caleb, Dami, Fola). Fola commented that identifying common struggles among posts helps them be more prepared to help students during office hours.

7.4.1.2 Piazza Pain Points

While Piazza provides a useful space for students to receive asynchronous help on their assignments, the participants also found there to be many pain points that, if fixed, could help students receive better and more timely help. To start, Dami shared that when a student responds to a post, the post is marked as resolved even when the student may be responding to ask for clarification. This means that Dami spent the first part of their Piazza passes checking for answers marked as resolved that were actually still unresolved. Oftentimes, students who ask a question and figure out their answer before a TA responds will comment "resolved" on their post without sharing how they resolved the

problem, said Fola. Eden added that they appreciate when students post helpful solutions that they may have learned from office hours instead of simply indicating their problem was resolved as it is more helpful to other students to have a solution.

In addition to issues with resolving posts on Piazza, the participants also identified other struggles they faced regarding student interaction on the forum. Fola and Dr. Bianca shared that they regularly encounter posts that do not give them enough information to help the student. Fola explained that "having to do back and forth banter just to find out something really basic [from a student] is a pain". For example, Dr. Bianca said that "TAs ask followup questions to students if they don't post a link to their code". Wanting to change student behavior on Piazza goes beyond the content of the posts that students are writing. Eden explained that students who don't post often are "really struggling, but they don't enjoy participating in online forums or they feel embarrassed". Even though TAs like Eden wanted students who have questions to seek help, it is already a struggle for our participants to provide answers to all of the questions posted. A TA, Caleb, admitted that "when there's hundreds of posts being asked in a day, they're not going to respond to everything in a timely manner, no matter how hard they're trying." An instructor, Dr. Bianca, who delegates checking Piazza to TAs, reiterates the sentiment that there sometimes are too many posts to manage. Dr. Bianca would "love to, but [does not] have time" to read and find common themes in all of the Piazza posts. The large volume of posts might be confounded by the elementariness of Piazza's search feature. Dr. Bianca explained that "there are a lot of repetitive questions" because "Piazza search is so poor that students will just default to asking a question that's already been posted". Dami agreed saying that Piazza search "does not use natural language searching and is a strict all inclusive keyword search" which often leaves them asking "Why am I getting no search results". Dami explained another frustration about searching on Piazza. When answering questions, "there's not an easy way to track posts history from a single person" which would allow them to understand the student and their struggles better.

Even when there were large influxes of posts, it was important to the participants that students received timely responses from the TA staff. Dr. Bianca explained "When the question volume goes up, and students are kind of hitting panic mode, you know, we want the TA to be responsive to those questions." Dami suggested that Piazza could be improved by adding a filter to see "the time [it takes each post] to be resolved within each assignment would be useful. So I can see that as a metric of how helpful the teaching staff are per assignment."

In addition to the large volume of posts that the instructors need to answer, instructors also struggle with prioritizing posts and deciding which TA should answer a post. TAs have different orders they answer questions on Piazza and Akin admitted that it is a struggle "deciding which question to answer first". Gad suggested that it may make sense to have each post that a student makes on Piazza assigned to a specific TA to ensure that all the questions get answered.

7.4.1.3 Meaning of Low Forum Traffic

Lower forum traffic at the beginning of the semester should be alarming to instructors. The participants hypothesized a few reasons why some semesters had lower forum traffic. Akin and Caleb

suggested that a low number of posts on the forum may simply imply that students have not started the assignment. Caleb said because this is probably the case, they are not interested in seeing the days with the lowest number of posts. Fola had a few other ideas of why forum traffic may be low. "People lurk sometimes and they'll never post anything... I do that often" said Fola, suggesting that the population may just be "more introverted" than other semesters. According to Fola, another reason there may be low forum traffic is because the TAs were giving bad answers and students have stopped asking for help. Gad suggested that the assignment may just be easy and the students don't have any questions.

7.4.1.4 Meaning of High Forum Traffic

Many of the participants speculated as to why there might be high forum use during the semester. Akin suggested that students may not be clear about the assignment instructions and there may be a lot of questions about the requirements. Spikes may indicate, according to Eden, that students have started the "difficult part of the assignment". Caleb suggested that the spikes may just be a reflection of the deadline. High forum traffic may also just be a result of students not providing links to their repositories or not providing enough information in their posts. Fola and Dr. Bianca both mentioned that TA responses asking for clarification would significantly increase the number of posts. Finally, both Caleb and Akin mentioned that the semesters during the Covid pandemic may have high forum post traffic because there are less interactions with other students, the TAs, and the instructor.

7.4.1.5 Current Knowledge of Student Piazza Usage

Some of the ways participants talked about current student Piazza usage revealed that many of the students' posts and interactions on the forum are motivated by fear or urgency. Dr. Bianca explained that students often post the same question multiple times. They may start off by making a public and private post, but then also repost their question "if they don't get a response within 30 minutes". Unfortunately, when students were stressed, Dami and Dr. Bianca noticed that the quality of their posts decreased which also decreased the quality of help the students were able to receive. Instead of writing detailed questions, students start to write "more panicked questions" like "my code is not working, help me" (Dami, Dr. Bianca). This panic is also reflected in the types of questions that receive a high number of likes. Dami details an example from their own classroom experience, "Students were very upset about how their tests were going to be handled because of how the test was issued through Moodle. So, there were lots of posts with high [Piazza] like counts".

One reason that students may be interacting on the forum in an urgent manner is because students often write their posts close to the deadline. Dr. Bianca, Dami, and Eden shared that from their experience of using Piazza, they observed that students create posts close to assignment deadlines. Although it may be frustrating to see high late post activity, Dami commented that they would "rather people ask late than never". It may be especially discouraging to see questions posted

towards the deadline about the assignment instructions as Gad shared that "if students are still not clear about some specifics of the assignment, then that means they have just started".

According to our participants, follow up questions may be the result of a mistake made by a student or instructor. Dr. Bianca viewed student follow up questions as an indication that "the class didn't get enough information in the question or in the reply". Dami added that there are some "situations where you could have a long chain [of follow up posts]" that imply that a student is "particularly over reliant on teaching staff for getting feedback and supporting themselves".

7.4.1.6 Common Class Struggles

Participants Akin, Dr. Bianca, Eden, and Fola all mentioned that when they see multiple questions about the same topic on Piazza, they can infer that a large proportion has the same struggle. Akin and Dr. Bianca also mentioned that they try to notice those patterns to better address the class as a whole.

7.4.1.7 Summary of Theme One

These responses about instructors' current forum usage reveal that the help-seeking forum and PEDI could to be improved to enable better student help-seeking and instructor and TA help-giving. For example, the forum could be improved with mechanisms to indicate when posts are resolved properly. Since the class is so large, and there are so many posts that occur near deadlines, PEDI or the forum could benefit from a natural language processing based system that could categorize questions and help reduce the time TAs spend on individual questions so they could more efficiently help students. From participants' feedback, we can sense that the teaching staff we interviewed cared about helping students in their classes.

7.4.2 Instructors' perception of PEDI

Each participant was introduced to the PEDI tool and had 10-15 minutes to explore the tool on their own without any instructional guidance; this was to enable them familiarize with the components and workflow of the tool. The exploration session was followed by three detailed tasks where participants were asked to navigate the tool following specific instructions and answer some questions before and after they performed the instructions. We have summarized the participants' experiences with the PEDI tool with the different PEDI components explored under the following sub-themes.

7.4.2.1 Importance of Likes and Replies (Tree map)

Akin, Dami, and Gad think high numbers of likes and replies were interesting and helpful to look at. The larger the rectangle on the treemap, the higher the likes or replies. Dami was interested in the first few big rectangles of the treemap before looking at the medium sized ones as well. She suggested that "it would be really convenient to click on the block and be able to jump with a hyperlink straight

to the post on Piazza". Fola says posts with high likes means good post and noted that some posts with high number of replies could be as a result of low quality instructor response "I guess at the time, you probably don't even think about this, but now that you are gathering this information from PEDI, you will probably be more wary about giving such poor response".

Post Likes Analysis Participants reviewed PEDI's post likes and reply features by viewing the treemap which showcases the posts that have the highest number of likes. Dami shared that high likes means common questions and inferred from reading questions with the highest likes, that students liked such questions either because a good number of students observed that something was wrong with the project or the answer was not available, and they really wanted the attention of the teaching staff. Akin and Fola reviewed similar posts with high likes and concluded that such posts whether it be the question or the corresponding dialogue signify information that is useful to students or a viable solution to the problem.

High likes on a post is a way for students to express their feelings towards the issue being addressed in the posts. Akin, Caleb, Dami noted that feelings are important and posts that express strong sad feelings would attract sympathy from other students and teaching staff. Dr. Bianca shared that students might use the forum as a place to vent their frustrations about deadlines and assignment logistics and she is not really interested in likes on a post but rather she would pay more attention to likes on responses as the analysis of good responses could be useful for teaching assistant training research. Dami would like such good responses to be added to a reading list or broadcasted for other students to see. Gad noted that posts with high likes are issues that are interesting and gaining attention among students, and instructors need to be aware of it. It would therefore be helpful if the posts were flagged no matter the content of the post.

Akin and Dami also suggested that it is important to answer the questions with the highest number of likes before the other questions. Therefore this post like analysis, might help TAs prioritize what questions to attend to first.

Post Replies Analysis

We asked the participants to review the content of posts with the highest number of replies. Posts that have more than usual replies and follow ups might be because the initial question did not contain sufficient details to answer the question. Akin noted that posts with a high number of replies are usually less important than posts that don't have any reply since it could mean that the question in the former has already been answered. Notwithstanding, Akin and Eden emphasised that they review posts with high replies to ensure that the questions were answered correctly. Fola noted that most posts with a high number of replies didn't seem pin worthy, but recommends pinning posts with the most likes.

Dami noted that some posts with lots of replies might be because the solution to the question might be complex for both students and TAs to resolve via the forum and the system could recommend that such issues be discussed during teaching staff's office hours.

7.4.2.2 Importance of Average Posts Frequency (Line Graph)

All participants found the line graph useful in detecting when students are working on assignments. Eden shared that he would be worried if there are a lot of spikes too close to the deadline or too close to an examination. Spikes were usually seen close to a deadline. Caleb, Eden, and Fola were interested in the common issues in posts made during those spikes and indicated that the line graph sparked curiosity for more information. The line graph patterns were similar across the semesters for each assignment and Akin suggested that due to the similarity, it could be possible to predict when to expect a spike. Dami was on the lookout for early spikes and shared that earlier spikes compared to other semesters is a good indication that students have begun working on an assignment early compared to other semesters. However, she complained that because an assignment is of different length each semester, it was difficult to compare semesters using the line graph. In addition, she noted that spikes closer to the deadline reduce the efficiency of teaching assistants as they would hardly be able to respond to students in a timely manner. Dr. Bianca shared that she would not be concerned about seeing the spike before the deadline, because that means students are working and they can get help but rather would be concerned about seeing a spike of unanswered questions before the deadline. Posts during the highest spikes were about the teaching staff test cases and she wondered if giving a better hint to students would lead to fewer questions on the forum.

Commenting on the line graph, Dami explained that it was quite intuitive. Eden liked the nice control panel on the side bar that lets him select any assignment and view visualizations across different semesters as well as the calendar tool which lets him select different dates and observe the changes in engagement at different points in the assignment.

7.4.2.3 Importance of Posts Frequency (Histogram)

The Histogram shows the number of posts made each day for the goal assignment. Caleb and Fola indicated interest in the day with the most posts because it will expose the concepts that students are struggling with. Dr. Bianca came up with an hypothesis for high forum activity closer to deadlines she said, “I might hypothesize that the the more stress that students are under because of proximity to a deadline, the more potential they are to write poorly formed posts that require some of this back and forth, because the TA requires clarification”. She would like to investigate if there’s an increase in the number of follow ups, the closer we get to deadlines, just because of poorly formed posts? Dami could also see that it might be the case where students start later, and then they have less time to look for other answers to other questions that have already been asked. Leading to more duplicate questions, bloating the questions count. “By starting later, and students making lower quality, more panicked questions. But more people starting later, increased involvement is not a good trend and I would be curious about why that’s happening this semester compared to previous semesters”. Eden added that “towards the two thirds mark students were panicking, some of these posts have that panicky sort of tone to them”.

Reviewing the chart, Dr. Bianca thinks the histogram was interesting and to be able to click on a

bar and go to the post was cool. Eden indicated that the histograms drills down posts and showing all posts on the days with the highest rate of interaction is useful.

7.4.2.4 Importance of Forum Activity (Pie Chart)

The pie chart shows different forum activities and what percentage of the class are performing them. Akin, Dr. Bianca, Dami, and Eden read the pie chart correctly and found it intuitive. Recall that in study one [B9], we defined help-seeking in discussion forums to be of two kinds: visible forum activity and lurking. Visible help-seekers are active forum posters who may also provide answers to fellow students' questions, whereas lurking refers to students who actively view posts in search of existing posts with answers to their questions [Bul01]. Dami would prefer a pie chart without the "no_activity" slice showing only "visible" forum activities. Eden's comment about the pie chart was that "when I was saying in the beginning I would love to see additional data on student engagement, this is exactly the type of thing that I was thinking".

Dr. Bianca suggested a study on gender distribution of students performing no visible activity on the forum. Viewing posts fall under the no_activity slice of the pie chart, "there is an idea that women tend to be higher with viewing things on Piazza than they are with posting on Piazza" and so the "no_activity" slice might represent students with a different type of personality. Caleb and Fola shared that people lurk sometimes and they will never post anything but they will definitely go on there for answers, Fola stated that "I imagine it's always the top and the bottom people who tend to post, at least in my opinion".

7.4.2.5 PEDI, Useful in Evaluating Instructional Methods and Interventions

Six of the participants (Akin, Dr. Bianca, Caleb, Dami, Eden, Fola) noted that the average posts frequency analysis across semesters is useful in detecting the effects of instructional methods. The addition of project milestones or process points is an instructional strategy which was absent in 2017 but gradually introduced to the class in 2018. Dr. Bianca observed that some of the earlier milestones pushed students to ask questions earlier and try to meet these deadlines earlier, mitigating a much larger spike at the end just given the scale and the growth that she saw in semesters where milestones were introduced. Dr. Bianca shared that introducing milestones was a beneficial intervention.

Dami stated that one new thing she learned from exploring PEDI was that a higher early forum activity could be a positive indicator that a good change was happening that helped students and the line graph helped detect what that change might be. Eden shared that the data presented on PEDI allows teaching assistants to learn something about their classes and think of some creative ways to encourage more forum participation, "the effectiveness of these strategies can easily be measured by coming back to PEDI to see how the graph changed". Eden continued to explain that his class might have had terrible student engagement while he was a teaching assistant and wouldn't have known it, "PEDI allowed me to see which strategies were good practices and that is awesome". Fola shared that "with PEDI a question like "what would my actual intervention be?" would not

arise but I would go ahead and apply my intervention strategy and use the PEDI tool to measure the impact of my intervention on student engagement".

7.4.2.6 Milestones

Caleb and Gad think it is good to have an early milestone. Fola thinks early milestones set the tone for the rest of the semester "once you set that tone for students, they tend to follow that for the rest of the Semester and students expect things to get harder from as the semester progresses, so if you set an earlier milestone people kind of jumpstart and start working. Eden noted the effect of the milestones on forum activity is an example of trying something in the classroom and seeing how it responds. Dr. Bianca shared that students are starting the project earlier in assignments that have milestones is an indication that the process of adding that milestone was a success. Dr. Bianca stated that her reason for including process place milestones is because she recognizes that developing software is more than just hacking and cobbling some code together but developing software is a process. She added that people who might disagree with the milestones due to rigidness or reduction in autonomy of the students, should note that milestones could help students who struggle in the introductory course sequences. Dr. Bianca also said she would be concerned that no spike occurs, when a milestone is approaching, "the best way to get students working on the project is to have a deadline. Because as much as I would like, my students are not intrinsically motivated to program, you know, all the time every day to do their work. You know, the deadline has to be there to drive them as well. And so like, from an intervention perspective, I'm not necessarily surprised". Dami thinks it might be a good idea to add an early milestone to other assignments besides the main project: "I would struggle to think how you would do it, maybe grade some part of it early or some positive reinforcement outside of grades for working early for example on guided project one, especially, and two and three".

7.4.2.7 Real time reflection tool

Caleb saw PEDI as a good reflection tool that shows the big picture of the past and compares it with current semesters. Dr. Bianca was excited that she maximized a shorter semester (Fall 2020 during the COVID pandemic) and was able to deliver all required assignments to the students. In addition, she shared that despite her class growing between fall 18 and fall 20, data shows the same average number of posts per day. And so if she were seeing a higher average number of posts with an increased class size, that would not be good because it means potentially, that the students aren't getting the help they need. Dr. Bianca proceeded to reflect on the changes made to instructional methods such as the way that she did feedback. In the fall of 2019 and 2020, she implemented different feedback strategies so, she wondered if the feedback strategy in 2019 potentially suppressed the number of questions that were asked where the 2020 feedback strategy encouraged more of that question asking. If that was the case, she concluded that the later strategy would be seen as a beneficial thing. However, a confounding factor is COVID. Fall of 19 was in person, fall of 20 was online. And it might be that the online mode led them to do more online interactions rather than in person help seeking.

Moreover, she concluded that PEDI's insights agree with some outside data which showed that there was more help seeking in general during the COVID semester.

Dami stated that she found interesting information about the class "we think back over all the activities we did to perform a retrospective analysis on previous patterns to improve for later semesters".

7.4.2.8 PEDI Praises

Here particpants described what they love about the PEDI tool and how they would like to use the tool in their classes.

Participants want PEDI in real time

All participants indicated that they want to use the tool in real time for their classes. Akin stated that the line graph and tree map were most useful to her. Knowing those statistics live means TAs could probably get to posts receiving more attention on the same day and that way they will actually be helping students. Eden mentioned that PEDI is really beneficial because "you know you're not just talking about things that would have helped students in the past, you're talking about assisting students that need help now, and knowing that your response is going to help a high number of students who are facing such issues as shown in the statistics." He concluded that having PEDI live would be game changing.

"Even without any of the suggested improvements, PEDI is still helpful and it would be interesting to be able to look through some of these posts and look at them in this view compared to the normal Piazza view and just see what post was liked on that day. I would definitely get a lot of use out of PEDI", Dami said. She also emphasized that "I would definitely use this chart to see what percent of students who do not attend lectures but who are contributing on the forum".

PEDI Search

PEDI search allows users to search all posts made in a day. Users could filter by post type (questions, student answer, reply to follow up etc), post number, and post content. Dr. Bianca, Caleb, Dami, Fola indicated that the search feature was useful.

7.4.2.9 PEDI helps TA Recruitment, scheduling and Onboarding

Dr. Bianca noticed, on the pie chart, that 8.25% of one class (11 students) provided student answers and she sees this feature as a good recruiting pool for future class TAs. Three of the participants suggested that PEDI may be useful in scheduling TAs.

Fola indicated that providing the PEDI tool for first-time TAs would be great "it was also my first time teaching that class. I didn't really know what to expect". Akin stated that "knowing what happened in the past semester, for example, which were the most liked or most replied posts for one assignment and knowing them beforehand might be useful for us to be prepared in answering the questions".

7.4.2.10 Alerts

Alerts on PEDI flag unusually high or low help-seeking activity or forum engagement. Dr. Bianca thinks alerts are a really good indicator that there might be something wrong or interesting. She observed that alerts give a summary because it calculates if interactions are higher or lower than the average across previous semesters, she shared "I really don't want to do the math". Dr. Bianca's statement confirms Kovanovic's findings which highlighted that the challenge is that the analyses of online learning data are both time-consuming and labor-intensive, and require a highly advanced set of data science and analytics skills, often not available to instructors [V.17].

Eden shared that alerts draw attention to the very stark difference between current semesters and the average across previous semesters and stated that "if we were seeing lower levels than usual, we would see a cause for alarm here". He also noted that alerts help to interpret graphs, "when I initially saw the graphs I was ready to panic and say, oh no, we need to intervene but there weren't any alerts raised, meaning that the numbers were not too far off". Dami shared that alerts are a good cause of action for her to look into other parts of the PEDI tool and if she saw an alert, she would look for more detail in the charts, "having PEDI as an alternative view is good".

7.4.2.11 Help-seeking vs Student engagement

Eden always likes to see more student engagement. PEDI has two different sets of alerts, one for flagging unusual rates of help-seeking (questions only) and another for unusual rates of student engagement which include both questions, follow-ups, student answers, instructor answers, and feedback. Eden detected the importance of having both alerts. He stated that it is one thing for students to seek help and another for them to be less inclined to take the teamwork approach in helping each other on Piazza, "and that's super important because there might be a particular way the instructor is presenting the project that made student engagement way higher than all the averages. I always call it like a camaraderie amongst the students to ask follow up questions and help each other". Eden noticed that sometimes there are students that just go above and beyond, to answer questions and ask follow ups and really help each other out, but in some semesters that doesn't happen.

7.4.2.12 Tables on PEDI

PEDI presents three different tables to users, with one table to display the different deadlines for each assignment and two other tables that present the rates of engagement on the forum across all the semesters. Fola likes the deadline table on PEDI because "you need to know that right away and it's annoying to have to go back to the assignment and find it". Dami likes the table for comparing semesters because it gives a good full picture. Eden shared that "the tables are a nice way to compare and justify the alerts so when we fall below the averages, I can see how we would get an alert on that assignment".

7.4.2.13 Surprises

Sometimes when the participants were using the PEDI dashboard, they were surprised by what PEDI showed them about the semester that they instructed. For example, Dami was surprised that only about 40 to 50 percent of students participated on the forum during a given semester. Dami assumed that participation was closer to 60 or 70 percent. Similarly, Eden was surprised when activity was lower on the line graph compared to other semesters but no alert was raised meaning that the numeric difference was not significant. Akin and Caleb were surprised by how similar a lot of the semesters are in terms of the rise and fall in forum traffic.

7.4.2.14 Summary of Theme Two

The participants had a lot to say while exploring PEDI. Their feedback suggests that PEDI is an adequate forum analytics tool, and addresses a current gap and need in course administration. Their feedback suggests ways that PEDI would be used. PEDI could especially help TAs allocate time and help resources, and help instructors compare the impact of interventions on class-level forum behavior and performance.

7.4.3 Future Recommendations - Future Work and PEDI Improvements

7.4.3.1 Importance of Post Content Analysis

Akin would like to compare the concepts discussed on days with lower forum activity with those on days with high spikes, to detect the difference. Eden, Fola, and Gad also want to know what parts of the material are being taught during both periods of high and low spikes. Fola shared that "if all questions during a spike are related to one thing, it means students did not understand that concept". Fola suggested performing sentiment analysis on post replies to know what kind of post is getting positive attention versus negative attention. Dr. Bianca would like a system that categorises the posts and highlights the common set of problems that students are asking.

7.4.3.2 Visualization Improvements

Akin and Dami found the checkpoints controlling the pie charts to be unintuitive. Dami suggested that users should be able to select the time periods they want to see using the data calendar instead of the checkpoints. Akin and Fola stated that the treemap was unfamiliar and unintuitive. Akin, Eden, and Fola suggested that it would be better that the rectangles displayed the posts content and replies when clicked. Caleb believes tree map is only useful when combined with other charts.

7.4.3.3 No_Activity Ambiguous

Akin, Caleb, Fola, and Gad suggested that no_activity is ambiguous and should be broken down into lurkers, students that need help but are shy to ask, students that have not started working on the assignment, and students that do not need help. Caleb recommends that the percentages in the

pie chart should be explicitly explained on the tool. Akin asked a couple of questions in order to understand how alerts work, this means that alerts should be explained better on the PEDI tool. For example, it should be explained that the red color indicates an alert. The alerts are simply small gifs on the interface. Dr. Bianca suggested that alerts should be bigger on the interface.

7.4.3.4 Difference in Assignment Duration

Dr. Bianca and Dami think the line graph should account for differences in assignment duration across different semesters. One solution could be to normalize assignments' lengths into equal bins and aggregate posts in order to achieve equal assignment time lengths.

Dami noted that the definition of the line graph should be bolder at the top. Eden recommends that the colors of the line representing different semesters should be consistent while users navigate the system.

In summary, participants noted that the UI in general could be improved to be more intuitive. Fola would like PEDI to be modified and made available for non-CS courses.

7.4.3.5 Suggested Future Work on PEDI

Dr. Bianca suggests a question curation system based on patterns in the questions being asked. For example, she statedwhat are the three test cases that every student is failing or the common set of problems that we have? What are the historical problems that we saw that we can potentially mine and then use to update resources, like the guided projects, or the labs in the future? If it isn't written down or documented, it won't make it to my to do list". Caleb suggested that PEDI should help predict when there's going to be spikes and responses. Akin suggested a system to identify concepts that students struggled with. Dr. Bianca suggested that future work could apply NLP to provide that aggregate contextual content based refinement, so that TAs using the system can determine if some pattern needs to be brought to the instructors' attention or if a certain post should be included in her weekly emails to students. Eden would like to have charts that cluster the different types of students in the class, "what types of students are most engaged, what types of students, maybe could use some extra help"? Dr. Bianca shared that on Piazza, it is easier for the students to post a new post rather than collaborate on common errors that they might be seeing and wishes that PEDI could go beyond aggregate numbers and measure the quality of the questions and answers. One solution is to encourage students to post to ask better questions like giving students a questions template. Participants were asked to comment on the quality of questions and replies after viewing a number of posts on the forum. Dami noted that some of the students' posts she reviewed were well formed. However, Fola also noted that some questions were malformed and suggested a system that requires students to ask better questions. Dami suggested making a sample post as a hint for anytime a student wants to post something. Some context about previous semesters were given to TAs and Akin, Dr. Bianca Caleb, and Fola indicated that contexts were required to understand the reason for different patterns on the line graph.

7.4.3.6 Interest in Grades

Dr. Bianca wishes to see grades on PEDI for more context before making some specific intervention decisions. Gad would like to see which students have submitted homework assignments on PEDI.

7.4.3.7 Summary of Theme Three

Improving PEDI

Although PEDI is an effective analytics tool as attested by our participants, there is still room for improvement. For example, the line graph was effective in showing users overall trends across multiple semesters, but because assignments have different lengths across semesters, there is a need to normalize the assignment length to make semester comparisons easier. Also the tree map needs to be improved. Participants needed the tree map to link to the direct posts on Piazza when a rectangle is clicked. The Pie chart, which shows forum activity for an assignment, also needs to be improved, the **no_activity** slice could be replaced with the percentage of the class who has viewed the assignment so far. From what the participants have shared, we suggest that the following improvements and features are important areas of future work on PEDI:

- Sentiment analysis: this will be useful to gauge the tone of help-seeking in the class
- Natural language processing feature: for question grouping and summarizing
- User interface improvements of the line graph, pie chart and Tree map
- Grades: The course Professor with 10 years of experience in the course would like grades on PEDI for more context
- Submission Tracker: to view what percentage of the class have submitted homework assignments on PEDI and what percentage have not. A teaching assistant suggested that a submission tracking would provide additional information about the percentage of the class that have completed the assignment.

Improving Piazza

Participants suggested the following ways that Piazza could be improved:

- Question and reply template to ensure that posts and replies are properly formed
- Mechanism for resolving posts properly so that when posts are resolved they contain the correct solution to the question asked
- Make Piazza search a smarter search engine to help students find posts easily

7.4.4 Summary of Participant Survey

While exploring the PEDI tool, participants answered five sets of post-activity questions. Their responses have been documented in Tables A.5 (see the appendix section), 7.4,7.5, 7.6, 7.7 and 7.8. Table A.5 encompasses all the questions asked and the response of each participant to each question. Table 7.4 and 7.8 show the participants' preferred modes if intervention during unusually low and high forum activity respectively.

Participants were asked to perform task one and two (see the appendix section) and were shown periods of unusually lower forum activity during the first half of an assignment's timeline (described in Task one in the appendix section). They were also shown scenarios of higher forum activity close to the deadline and read posts during these periods (described in Task two in the appendix section). They were asked to select their preferred mode of intervention in each case. Participants could select all the option that applied.

Lastly, Tables 7.5, 7.6 and 7.7 show the reasons participants gave for high forum posts, high likes on Piazza posts and high replies on a Post respectively. They read posts with high like and replies highlighted by the treemap.

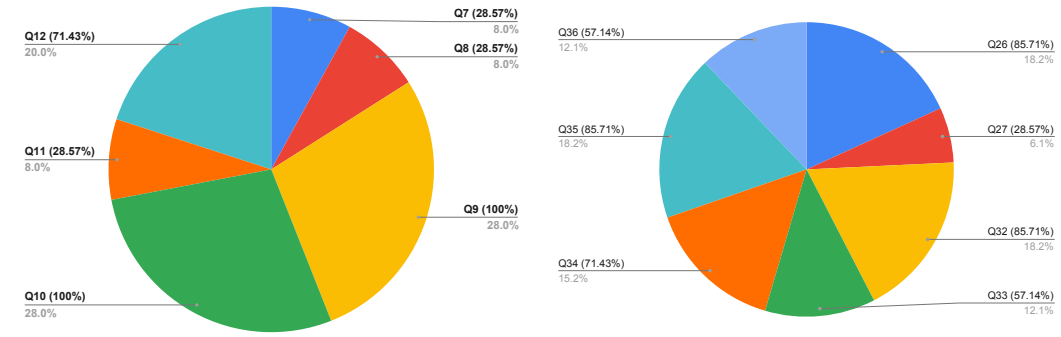
We asked participants to choose all the modes of intervention methods they might employ during periods of lower forum activity at the **halfway point of an assignment**, and the percentage of participants choosing each is shown in Table 7.4. They were also asked to select all the modes of intervention that they would apply during periods of unusually high forum traffic, from the options in Table 7.8.

The relative proportion of participants that would use each type of intervention is also depicted in the Pie charts in 7.1 for both low activity 7.1(a) and high activity 7.1(b). Participants could select all that applied. When seven out of seven participants select a mode of intervention, then it means 100% of all participants preferred that intervention strategy whereas if none selects it then the percentage is 0%. When lower than usual activity is observed, as shown in Figure 7.1a, seven out of the seven participants would give positive feedback to those who have made posts/replies/follow-ups (Q9) and encourage the whole class to ask questions and/or begin work (Q10), therefore 100% of the participants preferred these two modes (Q9 and Q10). Each of Q9 and Q10 constitutes 28% of the pie. Only five out of the seven participants (71.43%) went for option Q12 which is to *Set an additional early milestone* constituting 20% of the pie. All pie percentages sums up to 100%. However, since participants could choose all that applied, the percentage of participants' preferred mode of intervention does not sum of to 100%.

Lastly Two out of the seven participants would give incentives for meaningful participation on the discussion Forum (Q7), email students who have not posted/replied/follow-up on the forum to ask questions and/or begin work (Q8) and set new office hour sections for students who have not posted /replied/ follow-up (Q11); each of those modes represents 8% of the pie respectively.

Table 7.4 shows the mode of intervention when lower rates of early forum activity are observed. All participants would give positive feedback to students active on the forum and encourage the whole class to begin work (Q7). Dami would give extra credit for the amount of good comments students

Figure 7.1 Relative proportions of interventions indicated by participants for periods with unusually (a) low forum activity or (b) high forum activity



(a) (b)

Table 7.4 Participants were shown periods of lower forum activity in task one. They were asked to select their preferred mode of intervention when unusually Lower Forum Activity is observed in the first half of an assignment. Participants could select all the options that applied. Therefore, these pies represent all the intervention choices, and the proportion of the choices that went to each intervention

S/N	Mode of Intervention For Unusually Low Forum Activity (Halfway)	Preferred (%)
Q7	Give incentives for meaningful participation on the discussion Forum	28.57
Q8	Email students who have not posted/ replied/ follow-up on the forum to ask questions and/or begin work	28.57
Q9	Give positive feedback to those who have posted/replied/follow-up	100
Q10	Encourage the whole class to ask questions and/or begin work	100
Q11	Set new office hour sections for students who have not posted/replied/follow-up	28.57
Q12	Set an additional early milestone	71.43

receive on Piazza whereas Fola would give brownie points to encourage student engagement and make subtle comments in class to encourage students. Eden would ask students to apply for extra office hours because “maybe our office hours are not working for everybody, so if students who have not posted applied, we can help them get started, it can be more of a welcoming environment for those who haven’t done anything". Eden also thinks additional milestones and giving incentives to students to participate, is a really strong intervention formula but emailing students who have not posted is a counterproductive intervention strategy; “I have never been the type to want to call students out".

Caleb emphasized that a reason for lower forum activity might be that students tend not to want to ask questions super early except they really hit a wall or something was very confusing in the instructions because it looked like a lot of the early questions were about clarifying assignments.

The most popular intervention methods during periods of higher forum activity are Q26, Q32, and Q35 as shown in Figure 7.1b. During periods close to the deadline, if higher than normal engagement is observed, Table 7.8 shows that 85.71% of all participants would ensure that students

Table 7.5 Reasons for high number of Posts. Participants were shown different posts with high likes as highlighted by the Treemap and asked to select all the options that applied. Participants could view the posts, so they may be reasoning based on actual post content.

S/N	Reasons for high number of Posts towards the Deadline	Yes (%)
Q13	A concept in the assignment is not clear to students	71.43
Q14	Students have just begun work a few days to the deadline	51.14
Q15	Confusion about an item in the assignment	51.14
Q16	Request for assignment extension	0

Table 7.6 Reasons for high number of Likes

S/N	Reasons for high number of Likes	Yes (%)
Q17	The post was useful to students	57.14
Q18	The post highlighted difficulty faced by most students	100
Q19	Other reasons	14.29

are clear about assignment requirements (Q32), discuss posts with high likes and replies in class (35) and clarify instructions (Q26). Dr. Bianca's mode of intervention during high spikes would be to put more TAs to work during office hours and since she is not directly active on Piazza, she would rely on her TAs to inform her when such patterns are noticed. She shared that "I would put a peer teaching fellow (PTF) or a teaching assistant (TA) in charge of looking for these types of patterns; that's something I have not done before and that is an intervention that could be interesting. The TA in charge could let me know when there is a pattern of concern. Dr. Bianca would only pin a post if it is a major consistent problem or misconception and would address such in class as well. It allows me to really focus on the things that are going to be really important". Caleb would also give additional office hours. Eden continued, "I think we have the obligation for students to make progress, I think discussing some aspects of the assignments and class based on what was getting a lot of activity on Piazza is a good mode of intervention".

Dami is not really interested in all posts with a high number of likes and replies on a comment, only specific cases where the posts are really useful. Table 7.8 shows that only 42.86% of participants would pin posts with high likes and 28.56% might only pin posts that are useful.

The least popular mode of intervention is *Edit the assignment to: reduce workload (25)*. Due to the rigid structure of the class, instructors are wary of changing the assignment scope in order to achieve the pedagogical goals of the course.

Table 7.7 Relative percentage of participants who preferred each reasons for high number of Replies

S/N	Reasons for high number of Replies	Yes (%)
Q20	The initial question did not contain sufficient detail	57.14
Q21	The initial reply was not comprehensive enough	42.86
Q22	The initial question led to follow up questions	14.29

Table 7.8 Participants were shown periods of unusually higher forum activity close to the deadline (task two). They were asked to select their preferred mode of intervention in this case. Participants could select all the option that applied .

S/N	Mode of Intervention During Unusually High Forum Activity (Close to the Deadline)	Yes (%)
Q23	Discuss the assignment in class (Difficult concepts)	71.43
Q24	Ask students to work on the assignment in class and ask questions	42.86
Q25	Edit the assignment to: reduce workload	0
Q26	Clarify instruction	85.71
Q27	Give an extension	28.57
Q28	Would pin posts with high likes	42.86
Q29	Might pin posts with high likes when necessary	28.57
Q30	Would clarify questions with high number of replies in class	57.13
Q31	Might clarify questions with high number of replies in class when necessary	28.57
Q32	Ensure that students are clear about the requirements	85.71
Q33	Encourage students who have not begun the assignment or viewed posts to do so	57.14
Q34	Discuss some aspects of the assignment in class [during the lab session]	71.43
Q35	Discuss posts with high number of likes/replies in class	85.71
Q36	Tell students to work on the assignment in class and ask questions	57.14

7.5 Discussion

In this section, we address our research questions and inferences from the themes highlighted in participants' experience with the PEDI tool.

7.5.0.1 RQ 3.1a: What are beneficial use cases of PEDI dashboard?

In sections 7.4.2.9 and 7.4.2.11 and 7.4.2.7 participants indicated that would like to use the tool as a reflective tool, for TA Recruitment and On-boarding and to detect help-seeking versus student engagement. These show that participants were immediately able to identify how and when they would use PEDI in their classes by interacting with the tool for a few hours.

Table7.2 described the kind of analytics participants are most interested in. 71.43% of participants are more interested in the days with the highest rate of interactions, the number of likes and replies on a comment. On Tables 7.5,7.6 and 7.5, participants expressed what they feel is the reason for high forum activity. 71.43% believe that unclear concepts are the most popular reasons for posts made on the forum. 100% of participants believe when a post has high likes, it represents difficulty faced by students and if the initial question does not contain sufficient detail, 57.14% of the participants believe the question would attract more replies.

7.5.0.2 RQ 3.1b: What particular information could influence instructors' decision to intervene if they were using the PEDI dashboard in real time?

From the results in section 7.3, we see that most participants would like to intervene whenever they receive an alert or whenever they find unusually high or low activities. We observed that participants insisted on observing more than one chart before making their decision to intervene. Section 7.4.2.2 showed that participants were curious for more information after observing the line graph showing the average posts frequency.

7.5.0.3 RQ 3.1c: What might an instructor choose to do to intervene given information?

Tables 7.8, 7.8 and A.5(see the appendix section), show the percentage of participants and what mode of intervention they would like to employ at each point in time. Section 7.4.4 summarizes the most and least popular ways participants would like to intervene. Participants can select all modes of intervention that applies. When seven out of seven participants select a mode of intervention, then it means 100% of all participants preferred that intervention strategy whereas if none selects it then the percentage is 0%.

Participants' Opinions about our Hypothesis

Dami, Fola and Eden verbally agreed with our hypothesis that if the teaching staff is aware of our findings in prior research [B9], they would show concern about higher forum usage towards the end and seek to intervene in different ways. Dami stated that "so people are asking more questions later, which, like you said was an indicator of lower academic success. I appreciate it". Fola stated that "you can kind of gauge how well students are grasping information because sometimes you can tell when people really didn't understand something because right after class or right after an assignment was given out there would be an influx of Piazza questions". According to Eden, high forum activity either indicates that the instructions may not have been clear enough or it was not depicted while in class. The participants showed concern and decided that they would intervene.

7.5.0.4 PEDI Future Usage Frequency

Table 7.9 describes how our participants would like to use the PEDI tool in their classes. This information indicates that all participants would like to use the tool at some point for every assignment. PEDI is indeed a useful tool as highlighted by all participants.

7.6 Conclusion

This study allowed instructors to explore the length and breadth of the PEDI tool. We confirmed our first hypothesis which states that since the teaching staff is informed of our prior findings [B9], they would show concern if they detected lower forum usage during the first half of the assignment timeline and seek to encourage students who have not begun the assignment to start. All participants chose to encourage students who have not begun work halfway through the assignment's timeline.

Table 7.9 PEDI Future Usage Frequency

Participant	PEDI Usage Frequency
Akin	Every few days to compare current states with equivalent checkpoints in previous semesters
Dr. Bianca	Would have an experienced TA do a daily check
Caleb	If Caleb were an instructor, he would use PEDI very often. As a TA, not too often. Caleb doesn't believe it is TA's job to monitor PEDI
Dami	One to three times throughout an assignment. Maybe one early, one halfway through and probably one close to the deadline, like a day or two before
Eden	At least once a day
Fola	Before or after every Piazza usage
Gad	Three times during an assignment duration (25th, 50th, 75th checkpoints)

Our second hypothesis which states that if the teaching staff is aware of our prior findings in [B9], they would show concern about higher forum usage towards the end and seek to intervene in different ways, was also confirmed. Each participant showed concern and chose different ways to intervene. The most popular intervention method during this period was to clarify instruction, ensure that students are clear about the assignment and discuss posts with high likes and replies in class. The least popular intervention was to reduce assignment workload; no participant selected this intervention strategy.

Our results indicate that the professor/instructor and TAs would use the tool in different ways. The instructor would use PEDI for gathering the themes of common problems faced in the class and mine posts with high likes for TA training research. Teaching assistants, on the other hand, see PEDI as a tool that makes help-giving easier and less time consuming due to the aggregation and summaries PEDI provides. Furthermore, TAs would no longer be in the dark about common class struggles if they could use PEDI.

Interestingly, even though most TAs felt PEDI was useful, at least one felt that using PEDI was up to the instructor. Five out of the six teaching assistant would use PEDI for help-giving; the sixth teaching assistant suggested that PEDI would be more useful for instructors than TAs– but it was not entirely clear why. The instructor we interviewed would like to assign a dedicated TA to observe PEDI regularly and report important patterns for instructor decision making.

We anticipate that PEDI would be introduced for use in real time to provide TAs and instructors with insights about classes, and consequently pave the way for a new generation of online tools that would help the teaching staff support students.

CHAPTER

8

CONCLUSION

In this dissertation, I provide actionable learning analytics for intervention in blended classrooms by applying both quantitative and qualitative methods. My research has explored how students use online discussion forums, when they use it and what instructors can learn from the frequency and timing of students' online forum usage. I designed, implemented and tested a learning analytics dashboard (PEDI) to enable instructors to benefit from these insights. In addition, I performed a qualitative analysis to observe what PEDI means to instructors. I also explored the teamwork habits and task division in multi-gender teams in a programming project oriented class. In this chapter, I will summarize how the work in previous chapters contributed to answering my research questions.

8.0.0.1 RQ1: How can instructors get real time insights from online discussion tools?

We have seen that one major challenge of teaching and learning analytics is that times in which the learning activities occur and their analyses are very separate. This prevents teachers from taking data-informed actions immediately to positively impact student's learning outcomes. As we stated in section 1.0.4.1, the first step is detecting interesting patterns in the use of online tools. I have contributed to the first step by exploring and analyzing student traffic on an online discussion forum called Piazza and I showed that this analysis can signal useful information about assignments in the class to instructors. I also proposed that real time insights could be achieved by implementing learning analytics dashboards that present these insights in real time.

8.0.0.2 RQ2: How can instructors get insights about how gender affects students use of online collaboration tools

The second study also contributed to step one of the process for leveraging analytics to improve teaching. In the second study, I explored the team work habits among male and female students in multi-gender teams working on programming assignments and found a healthy collaboration between both genders. The types of analytic processes I devised could be used within programming projects to determine whether teams are sharing work equitably across gender and perhaps other types of students.

8.0.0.3 RQ3: Is the PEDI dashboard useful for instructors?

The second step needed for achieving real time analytics is to adapt the processes involved in the first step for real time use to ensure effective and timely interventions. I designed the PEDI tool to investigate how to adapt findings from the first study and how instructors may use them for real time interventions. Therefore, I implemented the learning analytics dashboard I proposed in study one. The results from our qualitative study were outstanding. All seven participants revealed their high level of interest in using PEDI in real time for their classes. In addition they gave important use cases for PEDI and revealed how frequent they would like to use the tool if it were introduced for giving real life instructions.

8.0.1 Summary

In summary, my studies have made the following contributions:

- Learning analytics to improve **whole-class** performance. There are numerous studies to improve individual student's performance through personalization. My studies, however, contribute to class-level and **whole-class** performance improvement.
- Learning analytics to improve class-level help-seeking and forum use
- Recommendation to include intermediate milestones in programming assignments because early forum use was detected around milestones that were introduced in Spring and Fall 2019 offerings of the course. Early forum usage is also positively correlated with class average performance [B9]. Related work presented in section 3.2.3 also provides solid theoretical grounding for the importance of milestones in programming projects.
- Research showing higher forum activities are related with lower class performance [B9]. This is important because based on this relationship, instructors might decide to check the forum to detect the reason for the high traffic on the forum and correct or address any problems that might be contributing to poor class performance.
- Research to detect teamwork habits in multi-gender teams. Although female students are less likely to show interest in the programming fields as seen from theory discussed in section

5.2.1, I found that when females actually choose Computer Science, they are more likely to make significant contributions and healthy collaboration as defined by [Git20], comparable to their male counterparts when working with males students on programming projects[B90].

- The design, implementation, and testing of PEDI, the analytics dashboard layer for Piazza contributes to forum analytics.
- Demonstrated how PEDI could be useful as a reflective Tool, TA recruitment, TA Scheduling, TA On-boarding, and to detect low student engagement
- The qualitative study provided instructors with:
  - An understanding of how students perceive assignments and signals when students are struggling
  - Reflection over previous assignments for the improvement of future assignments
  - Insights about the rate of social interactions going on while students work on assignments; and what higher or lower rates of interaction signal to instructors about an assignment and consequently class performance.
  - Insights about the possible need for intervention before the end of an assignment's duration.
  - The most and least preferred modes of intervention at different points during an assignment's timeline.

8.0.2 Future Work

The task of achieving real time analytics is an important one. We identified some potential research question and areas of improvement while working on our studies.

8.0.2.1 Future Work on PEDI

Participants have recommended ways to improve the user interface. Future work could apply participants' suggestions highlighted in section 7.4.3 as well as our recommendations in section 7.4.3.7 of chapter seven. This is important to improving the usability of the PEDI tool.

In particular, PEDI could be made better by grouping questions on Piazza using natural language processing and components such as a word cloud of major highlights of student struggles for each day. This would help TAs know what concepts students are struggling with. Also, a smart text classification system to group questions together, allowing TAs to provide summarized answers for multiple questions at once would help reduce the unnecessary time TAs spend on the forum. This would require a deeper integration with Piazza as questions asked on Piazza would be summarized by PEDI and TA responses to summarized questions would automatically be displayed on Piazza.

Adapting PEDI for use in other non-computer science courses would contribute to improving help-seeking and help-giving in those courses. However, a replication study similar to ours in

chapter four of this dissertation should be first performed on those courses to ensure that PEDI could potentially make a difference in improving the class's early forum activity and consequently class performance.

Some research questions that could be investigated include:

- How can sentiment analysis tell us the level and type of help students need? This question is important because instructors face challenge in understanding the overall class sentiment from the forum
- Potential research questions about visible and non-visible activities include:
  - What are the characteristics of lurkers in terms of performance (students having no visible forum activity)? Do students having no visible activity on the forum perform better or worse?
  - Is one gender or type of students prone to lurking on the forum than the other?

Currently, the *no_activity* slice on PEDI's pie-chart represents students with no visible activity. The contains both students who have never used the forum, and those that only viewed posts on the forum. Future work could break this down in order to investigate the above research questions.

After all the recommended PEDI improvements have been made, it would be interesting to evaluate PEDI's usage in a real life class in order to access the impact of interventions, spurred as a result of real time insights from PEDI, on forum usage.

8.0.2.2 Other Future Work

In chapter four, we described the different orderings of homework in a Discrete Math class. It is important to measure the impact of the different homework sequences on class average performance in order to answer research questions such as "what is the optimal sequence for presenting homework in assignments?"

I performed study 2 (chapter five) on the impact of gender in behavior in student collaborative teams using data from one semester. Future work could replicate this work for multiple semesters and years and in other courses to confirm the validity of our results across a larger and more diverse population and subjects. There is also a need for real time analysis to detect unhealthy students' collaboration patterns on Github. An adaptive support platform that provides instructors with similar insights from our analysis in real-time could signal the need for interventions to promote better teamwork.

The studies in this dissertation are unique and were implemented with the sole aim of improving learning and teaching for future generations.

BIBLIOGRAPHY

[A19] A, M.-M. et al. “Process Mining Dashboard in Operating Rooms: Analysis of Staff Expectations with Analytic Hierarchy Process”. *International Journal of Environmental Research and Public Health* (2019), pp. 2211–2220.

[AM20] A. Milliken V. Cateté, A. I. & Barnes, T. “Poster: Designing GradeSnap for Block-Based Code”. *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (2020), pp. 1–2.

[Agh20] Aghayi, E. et al. “Find Unique Usages: Helping Developers Understand Common Usages”. *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (2020), pp. 1–8.

[Aki20] Akintunde, R. O. et al. “Data-Informed Curriculum Sequences for a Curriculum-Integrated Game”. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK ’20. Frankfurt, Germany: Association for Computing Machinery, 2020, 635–644.

[Aki21] Akintunde, R. O. et al. “A Sneak Peek at a Collaborative CS Project: What part do female students play?” *Proceedings of THE 24TH ACM CONFERENCE ON COMPUTER-SUPPORTED COOPERATIVE WORK AND SOCIAL COMPUTING*. CSCW ’21 Submitted. Virtual: Association for Computing Machinery, 2021.

[Alf19] Alfy, S. E. et al. “Exploring the benefits and challenges of learning analytics in higher education institutions: a systematic literature review”. 2019.

[Alt16] Altujjar, Y. et al. “Predicting critical courses affecting students performance: a case study”. *Procedia Computer Science* **82** (2016), pp. 65–71.

[Alv12] Alvermann, D. et al. *Content Area Reading and Literacy: Succeeding in Today’s Diverse Classrooms*. Pearson Education, 2012.

[Ami17] Amigud, A. et al. “Using learning analytics for preserving academic integrity”. *The International Review of Research in Open and Distributed Learning* **18** (2017), pp. 192–210.

[Ant11] Antin, J. et al. “Gender Differences in Wikipedia Editing”. *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. WikiSym ’11. Mountain View, California: Association for Computing Machinery New York NY USA, 2011, 11–14.

[BC21] B. Chen, A. “Mission COVID: Low-Cost, Scalable Technology for Hybrid Learning”. *ACM SIGUCCS Annual Conference*. SIGUCCS ’21. Virtual Event, USA: Association for Computing Machinery, 2021, p. 9.

[Ban97] Bandura, A. “Self-Efficacy: The Exercise of Control”. 1997.

[Bel19] Bell, A. et al. “Fixed and random effects models: making an informed choice”. *Quality & Quantity* **53** (2019), pp. 1051–1074.

[Bey03] Beyer, S. et al. “Gender Differences in Computer Science Students”. *Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education*. SIGCSE ’03. Reno, Navada, USA: Association for Computing Machinery, 2003, 49–53.

[Bog18] Bogarín, A. et al. “A survey on educational process mining”. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **8** (2018).

[Bor13] Borkin, M. e. a. “What Makes a Visualization Memorable?” *IEEE Transactions on Visualization and Computer Graphics 19)* (2013), pp. 2306–2315.

[BG15] Bouwma-Gearhart, J. & Collins, J. *What We Know About Data-Driven Decision Making In Higher Education: Informing Educational Policy and Practice*. Proceedings of International Academic Conferences 2805154. International Institute of Social and Economic Sciences, 2015.

[Bra06] Braun, V. & Clarke, V. “Using thematic analysis in psychology”. *Qualitative Research in Psychology* **3**.2 (2006), pp. 77–101. eprint: `https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa`.

[Buf20] Buffardi, K. “Assessing Individual Contributions to Software Engineering Projects with Git Logs and User Stories”. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. SIGCSE ’20. Portland, OR, USA: Association for Computing Machinery, 2020, 650–656.

[Bul01] Bull, S. et al. “Help-Seeking in an Asynchronous Help Forum”. *In R. Luckin and B.* 2001.

[Che21] Chen, M. et al. “A Quantitative Analysis of Student Solutions to Graph Database Queries”. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. SIGCSE ’21. Virtual Event, USA: Association for Computing Machinery, 2021, p. 1385.

[Che11] Cheng, C. K. et al. “Assessing the effectiveness of a voluntary online discussion forum on improving students’ course performance”. *Computers & Education* **56**.1 (2011). Serious Games, pp. 253 –261.

[Coh08] Cohoon, J. & Aspray, W. “Women in Computer Science or Management Information Systems Courses: A Comparative Analysis”. *Women and Information Technology: Research on Underrepresentation*. 2008, pp. 323–349.

[Coi18] Coiera, E. W. et al. “Does health informatics have a replication crisis?” *JAMIA* **25**.8 (2018), pp. 963–968.

[Com14] Coman, I. D. et al. “Cooperation, collaboration and pair-programming: Field studies on backup behavior”. *Journal of Systems and Software* **91** (2014), pp. 124–134.

[Coo12] Cooper, A. *Vol 1, No 5. What is Analytics? Definition and Essential Characteristics*. 2012.

[Cor17] Corrin, L. et al. “Using Learning Analytics to Explore Help-Seeking Learner Profiles in MOOCs”. New York, NY, USA: Association for Computing Machinery, 2017.

[Cro17] Cross, S. et al. “Classifying Help Seeking Behaviour in Online Communities”. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. LAK ’17. Vancouver, British Columbia, Canada: Association for Computing Machinery, 2017, 419–423.

[Dan13] Daniel, S. et al. “The Effects of Diversity in Global, Distributed Collectives: A Study of Open Source Project Success”. *Information Systems Research* **24**.2 (2013), pp. 312–333.

[Den14] Deng, S. et al. “Multimodality with Eye tracking and Haptics: A New Horizon for Serious Games?” *Int. J. Serious Games* **1**.4 (2014).

[Des20] Desai, U. et al. *A Study on Student Performance Evaluation Using Discussion Board Networks*. Portland, OR, USA, 2020.

[Des21] Dessi, D. et al. “L2D 2021: First International Workshop on Enabling Data-Driven Decisions from Learning on the Web”. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. WSDM ’21. Virtual Event, Israel: Association for Computing Machinery, 2021, 1165–1166.

[Dis] Discourse.

[Du19] Du, J. & Wimmer, H. “Hour of Code: A Study of Gender Differences in Computing.” *Information Systems Education Journal* **17** (2019), pp. 91–100.

[Dui07] Duim, L. V. D. et al. “Good Practices for Educational Software Engineering Projects”. *29th International Conference on Software Engineering (ICSE’07)* (2007), pp. 698–707.

[Edw09a] Edwards, S. H. et al. “Comparing effective andineffective behaviors of student programmers”. *InPro-ceedings of the Fifth International Workshop on Com-puting Education Research Workshop, ICER ’09* (2009), 3–14.

[Edw15a] Edwards, S. H. et al. “Examin-ing classroom interventions to reduce procrastination”. *InProceedings of the 2015 ACM Conference on Inno-vation and Technology in Computer Science Education,ITiCSE* (2015), 254–259.

[Edw09b] Edwards, S. H. et al. “Comparing Effective and Ineffective Behaviors of Student Programmers”. *Proceedings of the Fifth International Workshop on Computing Education Research Workshop*. ICER ’09. Berkeley, CA, USA: Association for Computing Machinery, 2009, 3–14.

[Edw15b] Edwards, S. H. et al. “Examining Classroom Interventions to Reduce Procrastination”. *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*. ITiCSE ’15. Vilnius, Lithuania: Association for Computing Machinery, 2015, 254–259.

[Eme18] Emerson, A. et al. “Gaze-Enhanced Student Modeling for Game-based Learning”. *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, UMAP 2018, Singapore, July 08-11, 2018*. 2018, pp. 63–72.

[Er15] Er, E. et al. “Exploring college students’ online help-seeking behavior in a flipped classroom with a web-based help-seeking tool”. *Australasian Journal of Educational Technology* **31**.5 (2015).

[Ma2] “Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor”. *The Internet and Higher Education* **24** (2015), pp. 26–34.

[Fri11] Fritz, J. “Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers”. *The Internet and Higher Education* **14**.2 (2011). Web mining and higher education: Introduction to the special issue, pp. 89–97.

[G75] G, A. “Specious reward: a behavioral theory of impulsiveness and impulse control”. Psychological bulletin, 1975, 463–496.

[GS17] Gillett-Swan, J. “The Challenges of Online Learning Gillett-Swan Supporting and Engaging the Isolated Learner”. 2017.

[Git17a] Gitinabard, N. et al. “A Social Network Analysis on Blended Courses”. *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017(Workshops), Wuhan, China, June 25-28, 2017*. 2017.

[Git17b] Gitinabard, N. et al. *A Social Network Analysis on Blended Courses*. 2017. arXiv: 1709.10215 [cs.SI].

[Git17c] Gitinabard, N. et al. “Identifying Student Communities in Blended Courses.” *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017, Wuhan, China, June 25-28, 2017*. 2017, pp. 378–379.

[Git19a] Gitinabard, N. et al. “How widely can prediction models be generalized? performance prediction in blended courses”. *IEEE Transactions on Learning Technologies* **12**.2 (2019), pp. 184–197.

[Git19b] Gitinabard, N. et al. “What will you do next? A sequence analysis on the student transitions between online platforms in blended courses”. *CoRR* **abs/1905.00928** (2019). arXiv: 1905.00928.

[Git20] Gitinabard, N. et al. “Student Teamwork on Programming Projects. What can GitHub logs show us?” *Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference, July 10-13, 2020*. Ed. by Rafferty, A. N. et al. International Educational Data Mining Society, 2020.

[Gla12] Glava, C. C. & Glava, A. E. “Time Management on the On-Line Learning Platforms. A Case Study Regarding the BSCW Platform Used by Future Teacher Student”. *Procedia - Social and Behavioral Sciences* **46** (2012). 4th WORLD CONFERENCE ON EDUCATIONAL SCIENCES (WCES-2012) 02-05 February 2012 Barcelona, Spain, pp. 3551–3555.

[Goo15] Goodson, C. E. et al. “TIME MANAGEMENT SKILLS AND STUDENT PERFORMANCE IN ONLINE COURSES”. 2015.

[HH06] Hamlyn-Harris, J. H. et al. “Predictors of team work satisfaction”. *Journal of Information Technology Education: Research* **5**.1 (2006), pp. 299–315.

[Han19] Hanushek, E. A. et al. *The unwavering SES achievement gap: Trends in US student performance*. Tech. rep. National Bureau of Economic Research, 2019.

[Har20] Harrak, F. et al. “Evaluating Teachers’ Perceptions of Students’ Questions Organization”. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK ’20. Frankfurt, Germany: Association for Computing Machinery, 2020, 11–16.

[Has10] Hastie, M. et al. “A blended synchronous learning model for educational international collaboration”. *Innovations in Education and Teaching International* **47**.1 (2010), pp. 9–24. eprint: https://doi.org/10.1080/14703290903525812.

[Hof17] Hofmann, E. “Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect”. *International Journal of Production Research* **55**.17 (2017), pp. 5108–5126. eprint: https://doi.org/10.1080/00207543.2015.1061222.

[Hoy20] Hoyos, A. A. C. & Velasquez, J. D. “Teaching Analytics: Current Challenges and Future Development”. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje* **15**.1 (2020), pp. 1–9.

[Hur13] Hurst, B. et al. “The Impact of Social Interaction on Student Learning.” *Reading Horizons* **52** (2013), pp. 375–398.

[Imt19] Imtiaz, N. et al. “Investigating the Effects of Gender Bias on GitHub”. *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. 2019, pp. 700–711.

[InS] InScribe.

[Isv20] Isvik, A. et al. “Exploring Differences Between Student and Teacher Created Snap! Projects”. *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2020* (2020), pp. 1–5.

[Jan13] Janes et al. “Effective dashboard design”. *InProceedings ofthe 2017 ACM Conference on International Comput-ing Education Research ICER NewYork, NY, USA, 2017* (2013), pp. 17–24.

[Jay14] Jayaprakash, S. M. et al. “Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative”. *Journal of learning Analytics* **1** (2014), pp. 6–47.

[Jia19] Jiang, S. et al. “Examining how different modes mediate adolescents’ interactions during their collaborative multimodal composing processes”. *Interactive Learning Environments* **0**.0 (2019), pp. 1–14. eprint: https://doi.org/10.1080/10494820.2019.1612450.

[K01] K, P. M. et al. “Understanding the role of self-efficacy in engineering education”. *Journal of Engineering Education* **90**.2 (2001), pp. 247–251.

[Kam17] Kamberi, S. “Enticing women to computer science with es (Expose, engage, encourage, empower)”. *2017 IEEE Women in Engineering (WIE) Forum USA East*. 2017, pp. 1–5.

[Kaz17a] Kazerouni, A. M. et al. “Quantifying incremental development practices andtheir relationship to procrastination”. *InProceedings ofthe 2017 ACM Conference on International Comput-ing Education Research ICER NewYork, NY, USA, 2017* (2017), 191–199.

[Kaz17b] Kazerouni, A. M. et al. “Quantifying Incremental Development Practices and Their Relationship to Procrastination”. *Proceedings of the 2017 ACM Conference on International Computing Education Research*. ICER ’17. Tacoma, Washington, USA: Association for Computing Machinery, 2017, 191–199.

[Ker14] Kermarrec, A. “Computer Science: Too Young to Fall into the Gender Gap”. *IEEE Internet Computing* **18**.3 (2014), pp. 4–6.

[Kin16] Kinshuk et al. “Evolution Is not enough: Revolutionizing Current Learning Environments to Smart Learning Environments”. *International Journal of Artificial Intelligence in Education* **26** (2016), pp. 561–581.

[Kir06] Kirk, J. *’Analytics’ buzzword needs careful definition*. 2006.

[Kot11] Kotsiantis, S. “Use of machine learning techniques for educational proposes: a decision support system for forecasting students’ grades”. *Artificial Intelligence Review* **37** (2011), pp. 331–344.

[Led20] Lederman, D. *Inside Higher Ed*. 2020.

[Lia17] Liao, L. & Liang, J. “An Empirical Study on Blended Learning to Promote the Development of Computational Thinking Ability of College Students”. *2017 International Symposium on Educational Technology (ISET)*. Los Alamitos, CA, USA: IEEE Computer Society, 2017, pp. 256–260.

[Pin] “Linear Mixed-Effects Models: Basic Concepts and Examples”. *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer New York, 2000, pp. 3–56.

[Lóp12] López, M. et al. “Classification via clustering for predicting final marks starting from the student participation in Forums”. *EDM*. 2012.

[Lu18] Lu, O. H. T. et al. “Applying Learning Analytics for the Early Prediction of Students’ Academic Performance in Blended Learning”. *Journal of Educational Technology & Society* **21**.2 (2018), pp. 220–232.

[Mal20] Malysheva, Y. & Kelleher, C. “Using Bugs in Student Code to Predict Need for Help”. *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (2020), pp. 1–6.

[Man21] Mandal, S. & George, B. “Hybrid Learning Framework for Cybersecurity Courses”. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. SIGCSE ’21. Virtual Event, USA: Association for Computing Machinery, 2021, p. 1292.

[Man12a] Mandinach, E. B. “A Perfect Time for Data Use: Using Data-Driven Decision Making to Inform Practice”. *Educational Psychologist* **47**.2 (2012), pp. 71–85. eprint: https://doi.org/10.1080/00461520.2012.667064.

[Man12b] Mandinach, E. B. & Jackson, S. S. *Transforming teaching and learning through data-driven decision making*. Corwin Press, 2012, 59–88.

[Mao20] Mao, Y. & Marwan, S. “What Time is It? Student Modeling Needs to Know”. *In proceedings of the 13th International Conference on Educational Data Mining*. 2020.

[McS15] McSpadden, K. “Science: You now have a shorter attention span than a goldfish.” *IEEE Transactions on Visualization and Computer Graphics 19)* (2015).

[MK08] Merrell K Isaval D. M., G. B. A. e. a. “How Effective Are School Bullying Intervention Programs? A Meta-Analysis of Intervention Research”. 2008.

[Mil08] Mills, J. E. et al. “Perceptions and understanding of gender inclusive curriculum in engineering education”. *SEFI Annual Conference*. 2008.

[MM19] Moreno-Marcos, P. M. et al. “A Learning Analytics Methodology for Understanding Social Interactions in MOOCs”. *IEEE Transactions on Learning Technologies* **12**.4 (2019), pp. 442–455.

[Mur21] Muratova, A. “Interpretability and Effectiveness of Machine Learning Methods for Sequence Mining in Various Domains”. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. WSDM ’21. Virtual Event, Israel: Association for Computing Machinery, 2021, 1113–1114.

[Naf12] Nafus, D. “‘Patches don’t have gender’: What is not open in open source software”. *New Media & Society* **14**.4 (2012), pp. 669–683. eprint: https://doi.org/10.1177/1461444811422887.

[Nel18] Nelimarkka, M. & Hellas, A. “Social Help-Seeking Strategies in a Programming MOOC”. *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. SIGCSE ’18. Baltimore, Maryland, USA: Association for Computing Machinery, 2018, 116–121.

[Oak17] Oakleaf, M. et al. “Academic libraries institutional learning analytics: One path to integration”. *The Journal of Academic Librarianship* **43**.5 (2017), pp. 454–461.

[Ovr] *One Vs the Rest Strategy*. 2021.

[Yel] *Online Learning Platform for Student Engagement*.

[Pan14] Pandey, A. V. et al. “Science: You now have a shorter attention span than a goldfish.” *The Persuasive Power of Data Visualization," in IEEE Transactions on Visualization and Computer Graphics* (2014), pp. 2211–2220.

[par12] parsonson, barry s. “Evidence-based Classroom Behaviour Management Strategies”. 2012.

[Ped17] Peddycord-Liu, Z. et al. “Using Serious Game Analytics to Inform Digital Curricular Sequencing: What Math Objective Should Students Play Next?” *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY 2017, Amsterdam, The Netherlands, October 15-18, 2017*. 2017, pp. 195–204.

[Pia] Piazza. “FERPA Compliance”.

[Pis10] Pistilli, M. D. & Arnold, K. E. “Purdue Signals: Mining Real-Time Academic Data to Enhance Student Success”. *About Campus* **15**.3 (2010), pp. 22–24. eprint: `https://doi.org/10.1002/abc.20025`.

[Por18a] Porter, L. et al. “Developing course-level learning goalsfor basic data structures in cs2.” *nProceedings of the49th ACM Technical Symposium on Computer ScienceEducation, SIGCSE ’18, page , New York, NY,USA, 2018. Association for Computing Machinery.* (2018), 858–863.

[Por18b] Porter, L. et al. “Developing Course-Level Learning Goals for Basic Data Structures in CS2”. *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. SIGCSE ’18. Baltimore, Maryland, USA: Association for Computing Machinery, 2018, 858–863.

[Pri17] Price, T. W. et al. “Factors Influencing Students’ Help-Seeking Behavior While Programming with Human and Computer Tutors”. *Proceedings of the 2017 ACM Conference on International Computing Education Research*. ICER ’17. Tacoma, Washington, USA: Association for Computing Machinery, 2017, 127–135.

[R18] R. “Interface to python.” *nProceedings of the49th ACM Technical Symposium on Computer ScienceEducation, SIGCSE ’18, page , New York, NY,USA, 2018. Association for Computing Machinery.* (2018), 858–863.

[Rab11] Rabbany, R. et al. “Analyzing Participation of Students in Online Courses Using Social Network Analysis Techniques”. *EDM*. 2011.

[RM12] Richardson M Abraham C, B. R. “Psychological correlates of university students’ academic performance: a systematic review and meta-analysis.” *Psychol Bull.* **138**.2 (2012), pp. 353–87.

[Rob14] Robi, Y. S. “University Entrance Exam Result and Preparatory Class Average Score as Predictors of College Performance”. 2014.

[Rog16] Rogers, T. et al. “Learning Analytics and the Imperative for Theory Driven Research”. 2016.

[Rom13] Romero, C. et al. “Predicting students’ final performance from participation in on-line discussion forums”. *Computers & Education* **68** (2013), pp. 458 –472.

[Sab20] Sabab, S. A. et al. “An Automated Approach to Assessing an Application Tutorial’s Difficulty”. *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2020* (2020), pp. 1–10.

[Sch17] Schwendimann, B. A. et al. “Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research”. *IEEE Transactions on Learning Technologies* **10**.1 (2017), pp. 30–41.

[sci] scikit. *sklearn.multiclass.OneVsRestClassifier¶*.

[Sha20a] Shabaninejad, S. et al. “Automated Insightful Drill-down Recommendations for Learning Analytics Dashboards”. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK ’20. Frankfurt, Germany: Association for Computing Machinery, 2020, 41–46.

[Sha20b] Shabaninejad, S. et al. “Recommending Insightful Drill-Downs Based on Learning Processes for Learning Analytics Dashboards”. *Artificial Intelligence in Education*. Ed. by Bittencourt, I. I. et al. Cham: Springer International Publishing, 2020, pp. 486–499.

[Sha20c] Shabrina, P. et al. “Peeking through the classroom window: a detailed data-driven analysis on the usage of a curriculum integrated math game in authentic classrooms”. *LAK ’20: 10th International Conference on Learning Analytics and Knowledge, Frankfurt, Germany, March 23-27, 2020*. Ed. by Rensing, C. & Drachsler, H. ACM, 2020, pp. 625–634.

[Sha21] Shaffer, C. A. & Kazerouni, A. M. “The Impact of Programming Project Milestones on Procrastination, Project Outcomes, and Course Outcomes: A Quasi-Experimental Study in a Third-Year Data Structures Course”. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. SIGCSE ’21. Virtual Event, USA: Association for Computing Machinery, 2021, 907–913.

[She20] Sher, V. et al. “Analyzing the Consistency in Within-Activity Learning Patterns in Blended Learning”. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK ’20. Frankfurt, Germany: Association for Computing Machinery, 2020, 1–10.

[She18] Sheshadri, A. et al. “Predicting Student Performance Based on Online Study Habits: A Study of Blended Courses.” *Proceedings of the 11th International Conference on Educational Data Mining (EDM) 2018, Buffalo, US*. 2018, pp. 411–417.

[She14] Sheth, A. “Transforming Big Data into Smart Data: Deriving value via harnessing Volume, Variety, and Velocity using semantic techniques and technologies”. *2014 IEEE 30th International Conference on Data Engineering (ICDE)*. Los Alamitos, CA, USA: IEEE Computer Society, 2014, pp. 2–2.

[Smi18] Smith, C. M. “A Toolset for Mining GitHub Repositories in Educational Software Projects” (2018).

[SA17] Smith A. Boyer K., F. J. e. a. “My Digital Hand: A Tool for Scaling Up One-to-One Peer Teaching in Support of Computer Science Learning”. 2017.

[Spe] “Spearman’s Rank-Order Correlation using SPSS Statistics”.

[Spi20] Spieler, B. et al. “FEMALE STUDENTS IN COMPUTER SCIENCE EDUCATION: UNDERSTANDING STEREOTYPES, NEGATIVE IMPACTS, AND POSITIVE MOTIVATION”. *Journal of Women and Minorities in Science and Engineering* **26**.5 (2020), 473–510.

[S.S20] S.Shabaninejad et al. “Automated insightful drill-down recommendations for learning analytics dashboards”. *InProceedings of the Tenth InternationalConference on Learning Analytics & Knowledge, LAK ’20* (2020), 41–46.

[B9] “Study 1: Exploring class level forum usage for instructor insights”. *Chapter 3* ().

[B90] “Study 2: A SNEAK PEEK AT ACOLLABORATIVE CS PROJECT: WHATPART DO FEMALE STUDENTS PLAY?” *Chapter 4* ().

[Suh20] Suh, S. et al. “Coding Strip: A Pedagogical Tool for Teaching and Learning Programming Concepts through Comics”. *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2020* (2020), pp. 1–10.

[Sun17] Sung, S. H.-H. et al. “Comparing the effects of dynamic computer visualization on undergraduate students’ understanding of osmosis with randomized posttest-only control group design”. *Research and Practice in Technology Enhanced Learning* **12**.1 (2017), pp. 1–21.

[Syl08] Sylvia, B. & DeKeuster, M. “Women in Computer Science or Management Information Systems Courses: A Comparative Analysis”. *Women and Information Technology* (2008).

[TT19] Teo T. Zhou M., F. A. e. a. “Factors that influence university students’ intention to use Moodle: a study in Macau”. 2019.

[Thr] Three, C.

[V.17] V., K. et al. “Developing a MOOC experimentation platform: Insights from a user study”. *In Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (2017), pp. 1–5.

[VG20] Van Goidsenhoven, S. et al. “Predicting Student Success in a Blended Learning Environment”. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK ’20. Frankfurt, Germany: Association for Computing Machinery, 2020, 17–25.

[Vas15] Vasilescu, B. et al. “Gender and Tenure Diversity in GitHub Teams”. New York, NY, USA: Association for Computing Machinery, 2015.

[Vel17] Vellukunnel, M. et al. “Deconstructing the Discussion Forum: Student Questions and Computer Science Learning”. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. SIGCSE ’17. Seattle, Washington, USA: Association for Computing Machinery, 2017, 603–608.

[Vir21] Virginia, B. & Clarke, V. “Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches”. *Counselling and Psychotherapy Research 21.1* (2021), pp. 37–47.

[Wan17] Wang, Q. et al. “Designing and Improving a Blended Synchronous Learning Environment: An Educational Design Research”. *The International Review of Research in Open and Distributed Learning* **18** (2017), pp. 99–118.

[Web] WebAssign. “WebAssign Wikipedia”.

[Win14] Winquist, J. R. & Carlson, K. “Flipped Statistics Class Results: Better Performance Than Lecture Over One Year Later”. *Journal of Statistics Education* **22** (2014).

[Wis13] Wise, A. F. et al. “Learning analytics for online discussions: a pedagogical model for intervention with embedded and extracted analytics”. *Third Conference on Learning Analytics and Knowledge, LAK ’13, Leuven, Belgium, April 8-12, 2013*. Ed. by Suthers, D. & Verbert, K. ACM, 2013, pp. 48–56.

[XY18] Xu Y. Lynch CF., B. T. “How many friends can you make in a week?: evolving social relationships in MOOCs.” *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2018*. 2018.

[XY19] Xu Y. Gitinabard N., L. C. e. a. “What You Say is Relevant to How You Make Friends: Measuring the Effect of Content on Social Connection.” *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2018*. 2019.

[Zag15] Zagalsky, A. et al. “The Emergence of GitHub as a Collaborative Platform for Education”. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. CSCW ’15. Vancouver, BC, Canada: Association for Computing Machinery, 2015, 1906–1917.

[ZA15] Zagalsky A. Feliciano J., S. M. e. a. “The Emergence of GitHub as a Collaborative Platform for Education”. 2015.

[Zav20] Zavyalova, K. “Unlocking students’ motivation in the blended higher education classroom: Lecturers’ perspectives”. *E-Learning and Digital Media* **17**.5 (2020), pp. 425–441. eprint: `https://doi.org/10.1177/2042753020931774`.

[Zha21] Zhang, Y. et al. “A Feature Adaptive Learning Method for High-Density SEMG-Based Gesture Recognition”. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **5**.1 (2021).

APPENDIX

APPENDIX

A

FIGURES, TABLES AND INTERVIEW SCRIPT

A.1 Chapter 4: Figures and Tables

Figure A.1 Number of posts per day during Homework in 2014

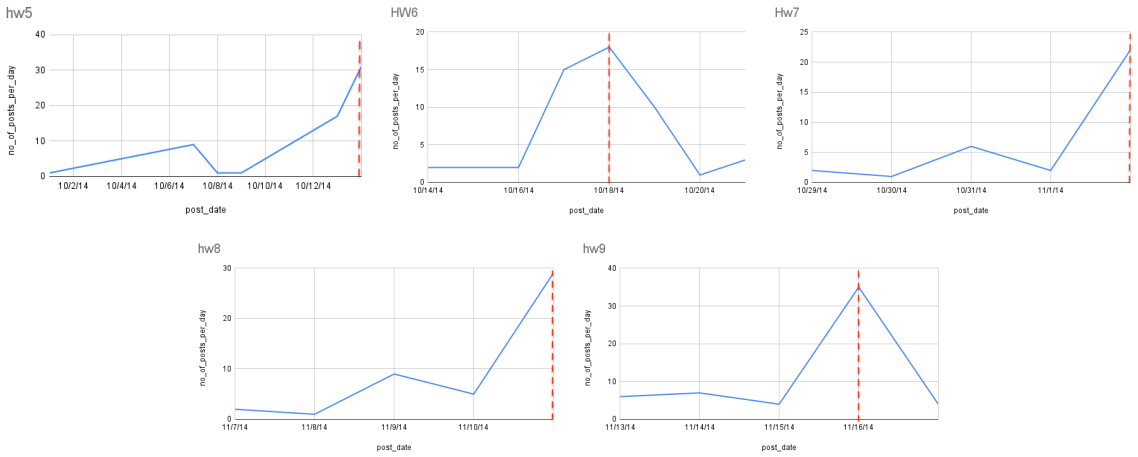


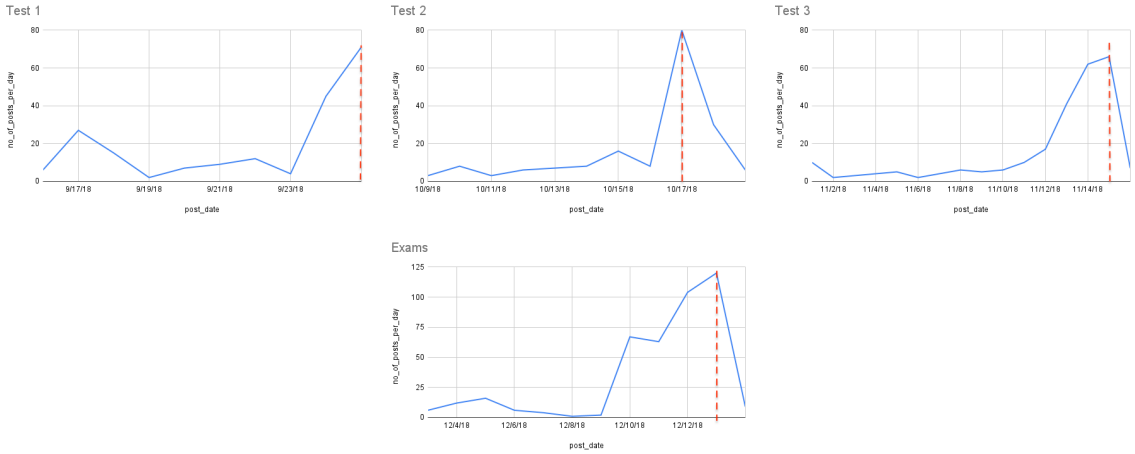
Table A.1 Duration of each Assignment. Column D represents Duration in Days. Note that the assignments are displayed in the order used in each semester. This means that hw2 was Circuits in 2013 but hw2 was Counting 2 in 2020. See Zero Analysis Charts order of topics.

2013		2014		2015		2016	
assignment	D	assignment	D	assignment	D	assignment	D
hw1	13	hw1	16	hw1	7	hw1	23
hw2	29	hw2	20	hw2	13	hw2	15
hw3	29	hw3	14	hw3	8	hw3	9
hw4	21	hw4	4	hw4	4	hw4	2
hw5	24	hw5	18	hw5	33	hw5	11
hw6	15	hw6	7	hw6	14	hw6	9
hw7	27	hw7	10	hw7	10	hw7	26
hw8	19	hw8	25	hw8	18	hw8	13
hw9	9	hw9	21	hw9	20	hw9	15
hw10	14	hw10	14	hw10	24	hw10	28
Exam	7	Exam	6	Exam	8	Exam	15
Test1	1	Test1	5	Test1	14	Test1	7
Test2	3	Test2	2	Test2	8	Test2	5
Test3	2	Test3	4	Test3	14	Test3	6

Table A.2 Duration of each Assignment. Column D represents Duration in Days. Note that the assignments are displayed in the order used in each semester. This means that hw2 was Circuits in 2013 but hw2 was Counting 2 in 2020. See Zero Analysis Charts order of topics.

2017		2018		2019		2020	
assignment	D	assignment	D	assignment	D	assignment	D
hw1	13	hw1	17	hw1	20	hw1	15
hw2	14	hw2	15	hw2	14	hw2	36
hw3	8	hw3	12	hw3	4	hw3	31
hw4	9	hw4	14	hw4	31	hw4	11
hw5	23	hw5	17	hw5	15	hw5	20
hw6	15	hw6	25	hw6	15	hw6	24
hw7	33	hw7	25	hw7	14	hw7	7
hw8	15	hw8	24	hw8	10	hw8	11
hw9	33	hw9	16	hw9	10	hw9	11
hw10	21	hw10	15	hw10	14	hw10	14
Exam	5	Exam	11	graph	28	Exam	7
Test1	20	Test1	9	Exam	10	Test1	10
Test2	29	Test2	12	Test1	7	Test2	9
Test3	9	Test3	15	Test2	6	Test3	7
				Test3	9		

Figure A.2 Number of posts per day during Tests/Exams in 2018



A.2 Chapter 7: Pre-activity interview Questions:

Please give brief answers to the following questions. For this task, you have up to 10 minutes.

1. How do you currently use Piazza monitor forum activities in your classes? [Insert Text.....]
2. The PEDI tool, which you are about to explore provides features such as:
   (a) Displays frequency of forum posts
   (b) Displays number of Posts per day for each assignment/project
   (c) Ranks posts by number of likes and replies
   (d) Displays posts with the highest number of likes and replies
   (e) Displays percentage of a class on the forum at any point in time
   (f) Displays alerts to flag late forum activity
   (g) Displays alert to flag unusually high forum activity

   Piazza currently does not provide any of the features listed above.
   (a) Do you think it is useful to know the number of posts per day and the days the highest number of posts? Please state the reason for your answer? [Insert....]
   (b) What does the number of likes on a post mean to you as an instructor? [Insert....]
   (c) What does the number of replies on a post mean to you as an instructor? [Insert....]
3. Do you think it would be useful to monitor the frequency of activities on Piazza E.g.Frequency of replies or likes on a comment; or the fraction of the class that posted? Why? [Insert....]
4. If Yes, which of the following would you pay attention to? (Select all that apply)

Table A.3 Assignments and Percentage of the Class that Scored Zero in that Assignment

Fall 2014		Fall 2015		Fall 2016		Fall 2017	
Assignment	Percent	Assignment	Percent	Assignment	Percent	Assignment	Percent
1: Logic & Proofs	0.78	1: Logic & Proofs	0.38	1: Circuits	0	1: Circuits	0.38
2: Circuits	1.94	2: Circuits	1.49	2: Logic & Proofs	0	2: Logic & Proofs	1.51
3: Predicate Calculus	0.39	3: Predicate Calculus	0.38	3: Predicate Calculus	1.19	3: Predicate Calculus	1.51
4: Sets & Predicate Calculus	0.78	4: Sets & Predicate Calculus	1.86	4: Sets & Predicate Calculus	0.8	4: Sets & Predicate Calculus	0.76
5: Sets, Arithmetic Proofs, Induction	2.32	5: Sets, Arithmetic Proofs, Induction	1.12	5: Sets, Arithmetic Proofs, Induction	1.98	5: Sets, Arithmetic Proofs, Induction	1.88
6: Induction, Recursion	2.32	6: Induction, Recursion	2.23	6: Induction, Recursion	0.4	6: Induction, Recursion	3.76
7: Big-O, Induction, Recursion	4.25	7: Big-O, Induction, Recursion	2.97	7: Big-O, Induction, Recursion	2.77	7: Big-O, Induction, Recursion	8.65
8: Binary Relations	4.64	8: Binary Relations	6.67	8: Counting	3.17	8: Counting	6.02
9: Counting	4.25	9: Counting	4.45	9: Binary Relations	3.56	9: Binary Relations	6.77
Test1	0.39	Test1	0	Test1	0	Test1	0
Test2	0.39	Test2	0	Test2	0.8	Test2	1.88
Test3	2.71	Test3	0.75	Test3	2.77	Test3	3.01
Exam	6.18	Exam	14.82	Exam	9.1	Exam	4.89

- ☐ The days with the highest rate of interaction
- ☐ The number of likes on a comment
- ☐ The number of replies on a comment
- ☐ The fraction of the class that posted on Piazza for an assignment
- ☐ None

A.3 Exploration Task Zero:

Go to the PEDI tool and explore the interface for the next 5 minutes.

Table A.4 Assignments and Percentage of the Class that Scored Zero on that Assignment

Fall 2018		Fall 2019		Fall 2020	
Assignment	Percent	Assignment	Percent	Assignment	Percent
1: Circuits	0	1: Circuits	2.98	Pre-Course Skills Test	7.19
2: Logic & Proofs	0.64	2: Logic & Proofs	3.28	1: Counting 1	4.8
3: Predicate Calculus	0.96	3: Predicate Calculus	5.66	2: Counting 2	7.19
4: Sets & Predicate Calculus	1.6	5 - Arithmetic Proofs, Induction	3.28	3: Truth Tables and Circuits	5.69
5: Sets, Arithmetic Proofs, Induction	2.24	6 - Induction, Recursion	2.39	4: Logic Proofs	6.89
6: Induction, Recursion	2.56	7: Big-O, Induction, Recursion	3.87	5: Predicate Calculus	6.59
7: Big-O, Induction, Recursion	3.2	8: Counting Introduction	7.15	6: Arithmetic Proofs, Induction	5.99
8: Binary Relations	4.16	9: Counting	7.45	7: Induction, Recursion	5.69
9: Counting	4.16	10: Binary Relations and Digraphs	7.45	8: Big-O, Induction	6.89
10: Graphs	3.52	11: Graphs	8.34	9: Binary Relations and Digraphs	10.18
Test1	0.96	Test1	0.3	10: Graphs	11.98
Test2	0	Test2	0.6	Test1	1.2
Test3	0	Test3	0	Test2	2.1
Exam	8.31	Exam	2.39	Test3	3.9

A.4 Exploration Task One:

Now we are starting with the first task. For this task, you have up to 20 minutes.

The first task aims to detect late forum use and encourage early forum use in the goal semester. Findings in 7.3 show that early forum use is positively correlated with class average performance. The higher the rate of early posts, the higher the class average performance

1. Go to the "Tracking Early Forum Activity" tab, select Fall 2020, select GP1, select date 8/22/2020 (Half way through GP1 timeline).

2. Observe the line graph. Compare this semester's Piazza activity to the following semesters

(Fall 2019, Fall 2018, Fall 2017). Look for differences between rate of forum use in this semester and previous semesters. (Remember to think aloud as you do).

3. Scroll down. Observe the Pie Chart and note the fraction of the class who have not performed any visible forum activity on the forum *(no_activity pie)*.

4. Scroll down. Observe if there are any alerts raised.

5. Scroll down. Observe the Table titled "Average Forum Activity from other years". Compare each activity on the table with other years. "How does the average activity on this assignment compare to that in other offerings?

   [Insert Text.....]

6. Scroll down. Observe the Table titled "Fraction of Students that Posted". Compare the table fraction of student that have posted on the forum with other years. How does the fraction of the class who posted on this assignment in this class compare to that in other offerings?

   [Insert Text.....]

Post-task Questions: Please answer the following questions:

1. Why do you think there were less posts than usual?

   - ☐ Most students have not began work
   - ☐ Change in the assignment scope
   - ☐ Other [Insert Text.]

2. Based on the number of posts and rate of forum use observed, if these assignments were live, would you have intervened using any of the following means? [Select all that apply]

   - ☐ Give incentives for meaningful participation on the discussion Forum
   - ☐ Email students who have not posted/replied/follow-up on the forum to ask questions and/or begin work?
   - ☐ Give positive feedback to those who have posted/replied/follow-up?
   - ☐ Encourage the whole class to ask questions and/or begin work?
   - ☐ Set new office hour sections for students who have not posted/replied/follow-up
   - ☐ Set an additional early milestone (At what point in the assignment would you set a new milestone and why?) [Insert Text..]
   - ☐ Other [Insert what mode of intervention that has worked in the past or how you would like to intervene in the future] [Insert Text...]

3. What was the reason for your choice in 1? [Insert Text...]

A.5 Exploration Task Two:

Now we are starting with the second task. For this task, you have up to 20 minutes.

The second task aims to detect higher rates forum use during the 2/3rd of the assignment timeline in order to detect assignment difficulty and ensure that the whole class benefits from answers on the forum. Findings in section 7.3 show that higher traffic towards the end of the assignment timeline is related to lower class average performance.

1. Go to the **"Tracking Assignment Difficulty"** tab, select Fall 2020, select GP1, select date 8/28/2020 (Two-Third of GP1 duration).

2. Observe the line graph. Compare this semester's Piazza activity to the following semesters (Fall 2019, Fall 2018, Fall 2017). Look for differences between rate of forum use in this semester and previous semesters. (Remember to think aloud as you do)

3. Observe the highest spike in the line graph. Are you interested in the forum conversations around this period?

   - ☐ Yes
   - ☐ No

   (a) Please give us a reason for your answer in question 3 above [Insert Text.....]

4. Scroll down. Observe the Histogram.

   (a) Based on the number of posts per day, what day are you interested in and why? Date: [Insert Text....]
   - ☐ The day had the highest number of questions
   - ☐ The day had the lowest number of questions
   - ☐ Other reasons [Insert Text..]

   (b) Click on the bar with the date you chose above to see the posts for that day

   (c) Observe the spike on day 08/28/2020, why are there so many posts on this day.
   - ☐ A concept in the assignment is not clear to students
   - ☐ Students have just begun work a few days to the deadline
   - ☐ Other reasons [Insert Text..]

   (d) What are the posts about?
   - ☐ Confusion about an item in the assignment
   - ☐ Request for assignment extension
   - ☐ Other [Insert Text..]

   (e) If you saw this during your class what will you do about it?

- ☐ Intervene
- ☐ Do nothing. Why [Insert Text...]
- ☐ Other [Insert Text...]

(f) If you would intervene, what form of intervention would you use?

- ☐ Discuss the assignment in class
- ☐ Ask students to work on the assignment in class and ask questions
- ☐ Edit the assignment to:
  - ☐ reduce workload
  - ☐ clarify instruction
  - ☐ Give an extension
- ☐ Other [Insert what mode of intervention that has worked in the past or how you would like to intervene in the future] [Insert Text..]

A.6 Exploration Task Three:

Now we are starting the third task. For this task, you have up to 20 minutes. We will explore the Tree Map displaying the number of likes and replies.

1. Still on **"Tracking Assignment Difficulty" t"** tab, select Fall 2020, GP1, select date 8/28/2020 (Two Third of GP1 duration).
2. Observe the Tree map
3. Observe the biggest rectangle, locate the post date and post number on the biggest rectangle
4. Click on this date on the Histogram and navigate to the post number to display the post. Please answer the following questions:

   (a) Students rarely like a post on Piazza. Only 3% of all the posts so far were liked by students. Why do you think the posts in the tree map received relatively high number of likes?

   - ☐ The post was useful to students because [insert text .....]
   - ☐ The post highlighted difficulty faced by most students
   - ☐ Other reasons [insert text ....]

   (b) Please comment on the content and the quality of the post?

   (c) Use the radio button on the navigation pane to toggle between likes and replies. Select replies, so that the tree map can show the posts with the highest number of replies. Why do you think the post had so many replies.

   - ☐ The initial question did not contain sufficient detail [Explain further .....]
   - ☐ The initial reply was not comprehensive enough [Explain further ....]

- ☐ The initial question led to follow up questions [Explain further ....]
- ☐ Other reasons [Insert text ....]

(d) Would you like to clarify the issues in any of these posts for the benefit of the whole class?

- ☐ Yes
- ☐ No Why/Why not? [Insert Text...]

(e) Would you pin this post in order for more students to view it?

(f) Would you like to have top posts with high likes/replies in real time for your current classes so that you could do interventions? (adaptability)

(g) Do you have any other comments about this tree map?

5. Scroll down. Observe the Pie Chart and note the fraction of the class who have not performed any visible forum activity on the forum *(no_activity pie)*.

6. Scroll down. Observe if there are any alert raised.

7. Scroll down. The following tables will clarify the reasons for the alert. Observe the Table titled "Average Forum Activity from other years". Compare each activity on the table with other years. How does the average activities of this year compare to that of other offerings of the current assignment?

   [Insert Text....]

8. Scroll down. Observe the Table titled "Fraction of Students that Posted". Compare the table fraction of student that have posted on the forum with other years. How doe the fraction of the class who have posted compare with that of other offerings of the current assignment?

   [Insert Text.....]

Post-task Questions: Please answer the following questions:

1. Why do you think there were more posts than usual?

   - ☐ Most students began working late
   - ☐ Change in the assignment scope
   - ☐ Students might be having difficulty with the assignment
   - ☐ Other [Insert Text.]

2. Based on the number of posts and rate of forum use observed, if these assignments were live, would you have intervened using any of the following means? [Select all that apply]

   - ☐ Ensure that students are clear about the requirements
   - ☐ Encourage students who have not begun the assignment or viewed posts to do so

- ☐ Discuss some aspects of the assignment in class
- ☐ Discuss posts with high number of likes/replies in class
- ☐ Tell students to work on the assignment in class and ask questions
- ☐ Edit the assignment to:
  - ☐ reduce workload
  - ☐ clarify instruction
  - ☐ Give an extension
- ☐ Other [Insert what mode of intervention that has worked in the past or how you would like to intervene in the future]
- ☐ Do nothing

3. What was the reason for your choice in question 8b above? [Insert Text…]

A.7 [10-15m] 4. Post-activity Interview Questions

Semi-structured interview questions [10-15m]

We would use these questions to guide and structure the interview, but allow the participant to take the conversation in unforeseen directions where appropriate. Approximate times are given for guidance, which the experimenter may adjust as necessary. An effort should be made to cover all questions.

1-3m First, the researcher should take time to clarify any remarks that were unclear during the think-aloud portion of the study.

2m What was your goal when exploring the PEDI tool today?

(a) Did your goal change while exploring the tool?

2m Did you find it useful to perform these activities on PEDI?

2m What is a pain point for you when you are trying to give help on Piazza?

(a) The current Piazza interface does not expose the features you have explored, did the PEDI tool adequately provide these features?

2m Did you find any new insight about your class?

(a) Were there situations where you were surprised?

2m Would you like the PEDI tool in real time for your classes?

(a) If you were to use the PEDI tool, what features or visualization would you expect to see in the tool?

2m Is there anything else you want to tell us?

Table A.5 Participants and their response to each question asked. 1 means Yes 0 means No

Participants/ S/N	**Akin**	**Dr. Bianca**	**Caleb**	**Dami**	**Eden**	**Fola**	**Gad**
Q1	0	1	0	1	1	1	1
Q2	1	0	1	1	0	1	1
Q3	0	0	1	1	1	1	1
Q4	0	1	0	0	1	0	0
Q5	1	0	1	1	1	1	1
Q6	1	0	0	0	1	0	0
Q7	0	0	0	1	1	0	0
Q8	1	0	0	0	0	0	1
Q9	1	1	1	1	1	1	1
Q10	1	1	1	1	1	1	1
Q11	0	0	0	1	1	0	0
Q12	0	1	1	0	1	1	1
Q13	0	1	1	0	1	1	1
Q14	1	0	0	1	0	1	1
Q15	1	0	0	1	1	0	1
Q16	0	0	0	0	0	0	0
Q17	1	0	0	1	1	1	0
Q18	1	1	1	1	1	1	1
Q19	0	0	0	1	0	0	0
Q20	0	1	1	0	1	1	0
Q21	1	0	0	0	1	0	1
Q22	0	0	0	0	1	0	0
Q23	1	0	1	1	1	0	1
Q24	0	0	1	1	0	1	0
Q25	0	0	0	0	0	0	0
Q26	1	1	1	1	1	0	1
Q27	0	0	0	1	0	0	1
Q28	1	0	0	0	1	0	1
Q29	1	1	0	0	0	0	0
Q30	1	0	0	1	0	1	1
Q31	0	1	1	0	0	0	0
Q32	0	1	1	1	1	1	1
Q33	0	1	0	1	0	1	1
Q34	1	1	1	1	1	0	0
Q35	1	0	1	1	1	1	1
Q36	0	1	1	1	0	1	0