

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Toward a Novel Tool for Continuous Peer Evaluation of Group Projects

Permalink

<https://escholarship.org/uc/item/1qr8q12d>

Author

Lu, Yu

Publication Date

2022

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Toward a Novel Tool for Continuous Peer Evaluation of Group Projects

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Software Engineering

by

Yu Lu

Thesis Committee:
Professor André van der Hoek, Chair
Professor Joshua Garcia
Professor James A. Jones

2022

DEDICATION

To my parents and my dear husband.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
ACKNOWLEDGMENTS	v
ABSTRACT OF THE THESIS	vi
1 Introduction	1
2 Background	5
2.1 Peer Evaluation	5
2.2 Formative and Summative Evaluation	6
2.3 Peer Evaluation Criteria	7
2.4 Existing Peer Evaluation Tools	7
3 Objectives	9
4 Daily Smirk	11
4.1 High-level Design Decisions	11
4.2 Main Interfaces	13
4.3 Primary Features	17
4.3.1 Evaluation Feature	17
4.3.2 Pre-defined Evaluation Criteria	18
4.3.3 Instructors Perspective	20
5 Evaluation Plan	22
5.1 Assessment Questions	22
5.2 Stages of Evaluation	24
5.2.1 Initial Deployment	24
5.2.2 Monitoring	25
5.2.3 Comparison	25
6 Conclusion	27
Bibliography	28

LIST OF FIGURES

	Page
4.1 Home page.	14
4.2 Guidance.	14
4.3 Smiley calendar.	15
4.4 Message center.	16
4.5 Team information.	16
4.6 Evaluation feature.	18
4.7 Peer evaluation criteria.	20
4.8 Dashboard for instructors.	21

ACKNOWLEDGMENTS

My deepest gratitude to my advisor, Professor André van der Hoek, for his guidance and support in my journey of research and the completion of this thesis. He always expresses unfaltering beliefs in me and inspires me to explore new research ideas. I sincerely appreciate his infinite patience in advising me and his constant encouragement to achieve my fullest potential.

I would like to thank my other committee members, Professor Joshua Garcia and Professor James A. Jones, for their support, challenging questions, and thoughtful comments that provide me with a wealth of ideas for future research on this thesis.

ABSTRACT OF THE THESIS

Toward a Novel Tool for Continuous Peer Evaluation of Group Projects

By

Yu Lu

Master of Science in Software Engineering

University of California, Irvine, 2022

Professor André van der Hoek, Chair

Many university courses are designed with group projects to provide students with “real-world” work experiences to better prepare them for their future careers. However, working in groups is not always a rewarding learning experience for students, as a variety of collaboration problems may arise when the students work together. Peer evaluation has been used as an approach to deal with these potential collaboration issues and has shown promise. At the same time, current peer evaluation approaches exhibit some serious shortcomings, including an increased workload for students and instructors, a tendency to give everyone high scores, and an inhibition to provide meaningful feedback from which group members can learn and improve how they engage. In this thesis, we introduce Daily Smirk, a novel peer evaluation tool that addresses these shortcomings. The tool supports a lightweight, high-frequency peer evaluation approach that aims to reduce students’ workload, provide useful insights for instructors to monitor ongoing teamwork, and reset expectations for using peer evaluation. We briefly review the state-of-the-art in peer evaluation, introduce key goals for Daily Smirk, and present its high-level design decisions and evaluation plan.

Chapter 1

Introduction

Group projects have become commonplace for undergraduate and graduate students who are majoring in STEM since there are many educational benefits to working and learning in a group. Many university courses are designed with group projects to provide students with realistic experiences in a “real-world” working environment that better prepares them for future professional careers [19]. Group projects are designed to promote student learning and teach students how to work effectively in settings mimicking students’ future careers. More specifically, group projects serve as an incentive for collaborative learning that allows students to learn to ask and share ideas, clarify differences, acquire problem-solving skills, and construct new understandings from each other [14]. It is acknowledged that being able to collaborate effectively in group projects is a vital and competitive skill that is considered when recruiting university graduates [18].

Working in groups is not always a rewarding learning experience for students. It can neglect students’ learning potential in group work. Students can struggle with dealing with some issues such as free-riders, social loafing, leadership, and conflicts between team members [21]. Free riding or social loafing is considered to be the most significant factor that can undermine

the effectiveness of group work affecting students' satisfaction with their team members' contributions [3]. "Last-minute group members" can also affect the progress and quality of the whole group project or assignment [7]. Additionally, some undesired collaboration issues such as poor communication and conflicts often emerge in group projects and they cannot be addressed promptly due to the lack of visibility into the ongoing teamwork [36].

Peer evaluation is widely used in group projects as an effective approach to help develop students' teamwork skills [31] and address the issues in group work [22]. By giving and receiving feedback through peer evaluations, students are provided with opportunities to develop their collaboration skills such as communication, problem-solving, and teamwork skills [10]. Peer evaluations can also increase students' self-awareness because they allow students to know how their team members perceive them [27]. Students might have a better understanding of their strengths and weaknesses by receiving feedback from their teammates, so they can be aware of how they can perform better in later project activities. Moreover, peer evaluations can also reduce the incidence of some issues like social loafing in group work [30].

While peer evaluation has proven to be beneficial for the student learning process, some drawbacks also need to be taken into account when incorporating peer evaluation into group projects. Some weaknesses include the accuracy and effectiveness of feedback from peers, students' insecurity about their team members' evaluations, and the difficulty of awarding a mark [29]. It is common that the students may not provide feedback sincerely or give high marks to their peers if the peers are their friends, so peer evaluation does not reflect individual performance accurately. In most situations, the peer evaluation process takes place at the end of group projects, so it cannot provide useful and timely feedback for students during the project or course. Moreover, peer evaluation is considered a time-consuming task for students when they are not familiar with the process [23].

To overcome these problems, we propose a simplified continuous peer-review approach for

students to evaluate their peers in group projects. We designed a novel peer evaluation tool, Daily Smirk, that aims to address the shortcomings of current peer evaluation tools. The concept of Daily Smirk is to enable high-frequency and lightweight peer evaluation that provides useful insights for instructors to monitor the students' teamwork. It keeps the peer evaluation process simple by providing five faces (satisfaction levels) for students to choose from for all their team members, indicating whether they are satisfied with their team members' work or not, and asking students for providing more information if necessary. The primary features of Daily Smirk are: (1) an evaluation feature that allows students to choose whether they are satisfied with their teammates, (2) a list of pre-defined collaboration behaviors with checkboxes for students to select, and (3) a dashboard for instructors to inspect all evaluation results.

Our approach aims to provide continuous feedback and to reduce both students' and the instructors' workload compared to the existing peer evaluation approaches. It allows students to receive ongoing and timely feedback from their peers throughout group projects that help them improve their performance in later group tasks. It focuses on reducing the students' workload in doing peer evaluation because it does not require students to answer too many questions in the form of essays or rate other teammates by reading guidance with many questions and answers. Instead, it only needs students to choose satisfaction levels, and select checkboxes for pre-defined collaboration behaviors if required. It is also devoted to providing useful insights and feedback for instructors by offering a visualization dashboard for them to inspect all evaluation results. Overall, this novel continuous peer evaluation tool aims to help students improve collaboration within groups and develop their teamwork skills by giving and receiving continuous feedback.

The remainder of this thesis is organized as follows. Chapter 2 introduces relevant background material in peer evaluation, explains the roles of formative evaluation and summative evaluation, presents related work in designing peer evaluation criteria, and analyzes existing

peer evaluation tools. We establish four objectives in Chapter 3 and explain each of them. Chapter 4 presents Daily Smirk, our proposed peer evaluation tool that aims to address the shortcomings of current peer evaluation approaches by introducing its high-level design decisions, four main interfaces, and primary features that are designed for achieving our objectives. In Chapter 5, we propose four guiding assessment questions and design a long-term evaluation plan with three stages to validate the effectiveness of Daily Smirk. Chapter 6 summarizes our primary contribution in this thesis.

Chapter 2

Background

This chapter discusses relevant background material regarding peer evaluation in group projects, formative and summative peer evaluation, peer evaluation criteria with categories of questions to ask, and presents existing peer evaluation tools.

2.1 Peer Evaluation

Peer evaluation plays a significant role in group projects for several reasons, including facilitating critical reflection on individual performance, developing students' teamwork skills, and improving students' behaviors by providing and receiving feedback [10]. First, it has been shown that peer evaluation leads to the development of self-motivated, responsible, reflective individuals [15], thus enabling students to monitor their performance. Second, involving students in the evaluation process provides them with opportunities to develop their collaboration skills for future professional development, such as communication, problem-solving, and teamwork skills [10]. Third, peer evaluation makes students feel more motivated and involved in the group, and motivates them to put more effort into group activities than usual

[29]. Fourth, giving and receiving feedback is beneficial for students' learning because it improves the development of skills that are required for professional responsibility, judgment and autonomy, and it also emphasizes the students' responsibility in their learning and evaluation process [24]. In general, peer evaluation helps students reflect on their own learning and facilitate team-based collaborative learning [12].

2.2 Formative and Summative Evaluation

Peer evaluation can serve two purposes: (1) formative evaluation, which is conducted throughout a course or a project to provide timely feedback on students' performance for improvements, and (2) summative evaluation which takes place at the end of a course or a project and is used for evaluative purposes [33][34]. Summative evaluation focuses on whether learning results are correct or not [24]. It is used for measuring student learning outcomes and student satisfaction, but it does not define why students are satisfied or what changes can lead to the most important learning improvements [28]. It can undermine collaboration between students [24] because it only takes place at the end of a course. Although it helps the instructor understand students' achievements and performance more easily, it does not accurately reflect students' performance.

Formative evaluation concentrates on the in-depth qualitative evaluation of students' various learning outcomes and enhances learning [24]. It is interactive, taking place during the course rather than only at the end [35], so it provides timely feedback [11] to develop students' collaborative skills and provide critical reflection on individual contributions. It can provide students with valuable diagnoses that they might not be aware of [28], allowing them to improve their performance by receiving continuous feedback. However, formative evaluation is a time-consuming task and difficult to manage, which is a common concern among instructors who lack experience in conducting peer evaluation [4].

Since this study aims to explore whether peer evaluation can promote collaboration in group projects, meaning that it focuses on students' learning process instead of learning outcomes, we adopt a formative evaluation approach. To be more precise, we plan to implement a continuous peer evaluation approach that requires students to provide feedback after each day they have group activities.

2.3 Peer Evaluation Criteria

There are various categories that should be taken into account when designing a peer evaluation process. Farrell et al. divided peer evaluation into two categories for students to assess their group activities: the degree to which each team member collaborates with other team members and the value that each team member contributes to the project's progress. They proposed eleven subcategories associated with these two broad categories [16]. Beatty et al. listed the following categories based on actual group project experiences and tested these criteria through multiple regression analysis, including dependability and availability, input and work quality, peer group equity and interaction, and overall evaluation [6]. Brutus and Donia designed a centralized peer evaluation system that categorized peer evaluation into four criteria: cooperation, conceptual contribution, practical contribution, and work ethic, with associated expected behaviors for each of them [9]. Overall, there are many variations on these criteria, though most cover roughly similar kinds of concerns.

2.4 Existing Peer Evaluation Tools

Peer evaluation tools have been an active research direction due to their potential to ease peer evaluation in team-based learning environments. Luxton-Reilly provided us with a systemic review of available peer evaluation tools by introducing and comparing their features [26].

Some recent web-based peer evaluation tools such as Peerceptiv [2] and CATME [1] have been designed for classes that have group projects. Peerceptiv is a peer evaluation tool that allows the instructor to scale evaluation criteria and allows students to provide and receive feedback on course assignments and how well their peers contributed to the group projects. The double-blind peer evaluation feature of Peerceptiv encourages students to deliver more critical feedback [13]. In addition, the Comprehensive Assessment of Team Member Effectiveness (CATME) peer evaluation tool has been shown as an effective tool to evaluate students' teamwork learning [8]. CATME is a behaviorally anchored point rating scale that collects and summarizes data in five dimensions, and each dimension is associated with a list of the description of the behaviors for high, medium, and low performance of team contribution [25]. The data collected from peer evaluations are anonymous to students but it is visible to the instructor. It allows students to provide feedback to their peers and the instructor about the ongoing teamwork as well as provides students with a learning opportunity to think deeply about teamwork and to improve self-reflection on their contributions [25]. While Daily Smirk got inspiration from these peer evaluation tools, it differs from them in objectives, assessment criteria, and functionality.

Chapter 3

Objectives

The goal of this thesis is to develop a novel peer evaluation tool that can improve students' collaborative learning experiences and provide insights for instructors to monitor students' teamwork. To identify the major requirements of our proposed peer evaluation tool, Daily Smirk, we studied why peer evaluation exists in group projects in Section 2.1, and the advantages and disadvantages of the existing peer evaluation tools in Section 2.4. As a result, we established the following objectives:

- *To conduct high-frequency peer evaluations.* Since this study focuses on improving students' collaborative learning process, we utilize high-frequency (continuous) peer evaluations to assess performance. The primary benefit of high-frequency peer evaluation is continuous and timely feedback, so to provide students with ongoing feedback that gradually develops a clear understanding of their work expectations [17]. The intended result of using continuous evaluations is to provide enough time for them to adjust and improve their performance to meet the team's expectations by receiving timely feedback regarding their contribution.

- *To reduce students' workload.* While we ask students to provide high-frequency peer evaluations, we aim to reduce their workload in doing peer evaluations because time and effort spent are the main concerns students have about peer evaluation [5]. Many instructors are also concerned that peer evaluation is a time-consuming additional task, that is not appreciated [20]. In this study, providing lightweight peer evaluation work for students is one of our main objectives, and there is little literature focusing on this aspect of peer evaluation.
- *To provide useful insights for instructors.* Some students provide very general feedback or very positive feedback to their peers, which is not helpful for instructors to monitor the ongoing teamwork. This is due to students' lack of experience in doing peer evaluations, or students' fear of being harsh to their peers resulting in penalizing their peers' grades [32]. One of the primary goals that instructors want to achieve by implementing peer evaluation in class is to identify potential dysfunction and conflicts [32], but sometimes they cannot get useful or helpful feedback from students. Therefore, our peer evaluation approach should be designed to guide students toward some important aspects that instructors want to see in evaluating their peers. By receiving useful feedback from students, instructors are able to identify and intervene early.
- *To reset expectations.* We aim to reset students' expectations to "normal" by using Daily Smirk because students often exaggerate or underestimate the performance of their team members. It is common for some students to give all of their team members the same scores such as 5/5, but this cannot reflect the students' real performances. In order to reset their expectations to normal, our peer evaluation should ask students for providing more information and feedback if they assess the performance of their team members as above expectations or below the expectations.

Chapter 4

Daily Smirk

In this thesis, we introduce Daily Smirk, a novel peer evaluation tool that enables high-frequency evaluations and addresses some of the shortcomings of current peer evaluation approaches. This chapter presents its high-level design decisions, show its four main interfaces, and presents how its primary features are designed for achieving our objectives.

4.1 High-level Design Decisions

- *Provide daily smiley-based assessment of teammates.* This decision addresses our first objective of enabling high-frequency peer evaluations. It is acknowledged that peer evaluation is a valuable learning tool that provides students with the experience of evaluating their peers and receiving feedback about their own performance [9]. However, poorly conducted peer evaluations may undermine team performance [25]. Our primary design decision for Daily Smirk is to make it useful for improving students' learning experience, including developing their collaboration skills, critique abilities, and self-awareness. Therefore, we adopt a formative (continuous) peer evaluation ap-

proach to assess students' performance in group projects in order to facilitate their learning experiences. We came up with a daily smiley-based assessment of teammates that provides smileys for students to choose from to indicate whether they are satisfied with their team members' work almost every day. It is similar to a check-in application that monitors how student teams perform over time. As a result, it allows students to receive timely feedback on their ongoing performance so that they can invest more effort in achieving the group's expectations, and gradually nurture their critical thinking skills by continuously responding to feedback.

- *Ask for more information if students signal something going well or not well in teams.* The second decision aims to achieve the objective of providing useful insights to instructors. More often than not, things are actually going well within a team, so we do not want to assign students additional tasks unless something is not quite right. Once students signal there is something going pretty well or not well in teams, Daily Smirk asks students to provide more information for instructors. Collaboration issues often emerge in teams, for example, if a team member never attends a group meeting or contributes little to teamwork; some other students in the team may want to signal instructors about this kind of problem. In addition, we hope this decision can help students reset their expectations in doing peer evaluations because giving everyone a high score has become commonplace in peer evaluations. So if a student gives their team members very high scores or very low scores, the system will ask for more information, moving the expected normal assessment to neutral.
- *Design evaluation criteria to assist students in evaluating their teammates.* The third design decision is the answer for reducing students' workload in doing peer evaluation and providing meaningful feedback for instructors as well. Some people may ask how this can reduce students' workload if this approach requires students to evaluate their peers almost every day and the measurement itself is a heavyweight task. Therefore,

we do not adopt some common forms of peer evaluation, such as reading long-form criteria or writing many short answers to evaluate team members. Instead, we scale the evaluation criteria with lists of associated behaviors in the form of checkboxes that allows students to provide feedback easily if they signal the system. So students simply check whether the listed behaviors match the performance of their group members instead of spending too much time reading pages of criteria or inputting a lot of text. Such pre-defined evaluation criteria are able to guide students to evaluate their peers in terms of what instructors want to see. As a result, instructors have the ability to intervene early because many students lack experience in doing peer evaluations or often provide very general feedback, all of which can result in the instructors often not receiving helpful feedback.

- *To be user-friendly.* We design Daily Smirk as a mobile application because one of the advantages of the mobile application is easy to access as we require students to evaluate team members almost every day. Since it is to be a mobile application, we need Daily Smirk to be user-friendly. Therefore, we need to make every effort to provide users with a great user experience by designing intuitive interfaces that are easy to use.

4.2 Main Interfaces

Daily Smirk has four main interfaces, including the home page, calendar interface, message center, and team information. It should be noted that Daily Smirk is intended to be a mobile application for students to stay engaged.

Figure 4.1 is the home page of Daily Smirk, which displays how many smileys this student has received from their peers by using bar graphs, and there is a “Rate Your Peers ” button on the bottom. Once students click this button, the system will bring up guidance (Figure 4.2) for them before they start to evaluate, and it explains the degree of satisfaction corresponding

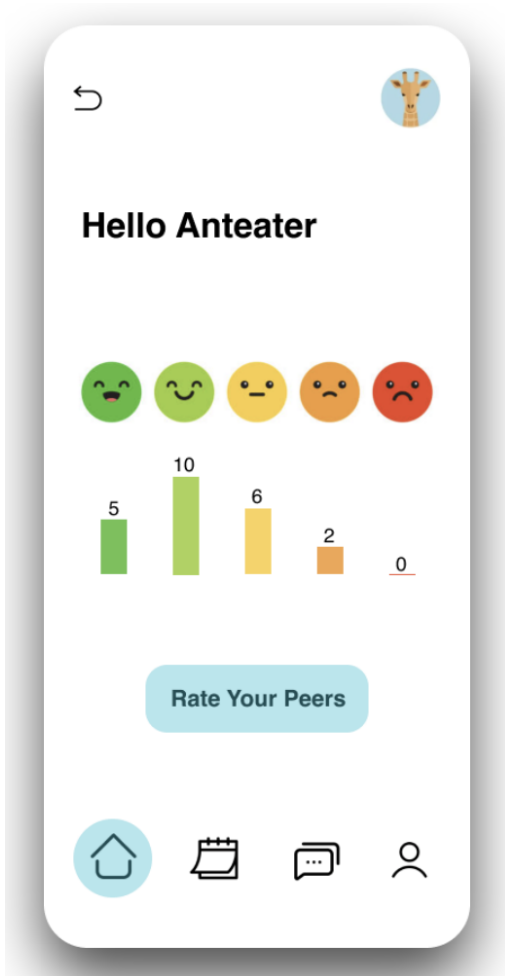


Figure 4.1: Home page.

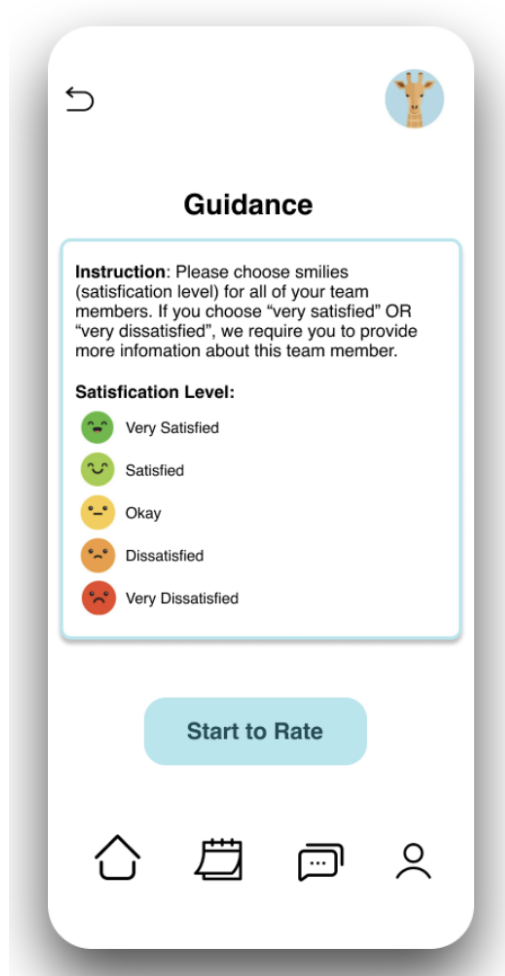


Figure 4.2: Guidance.

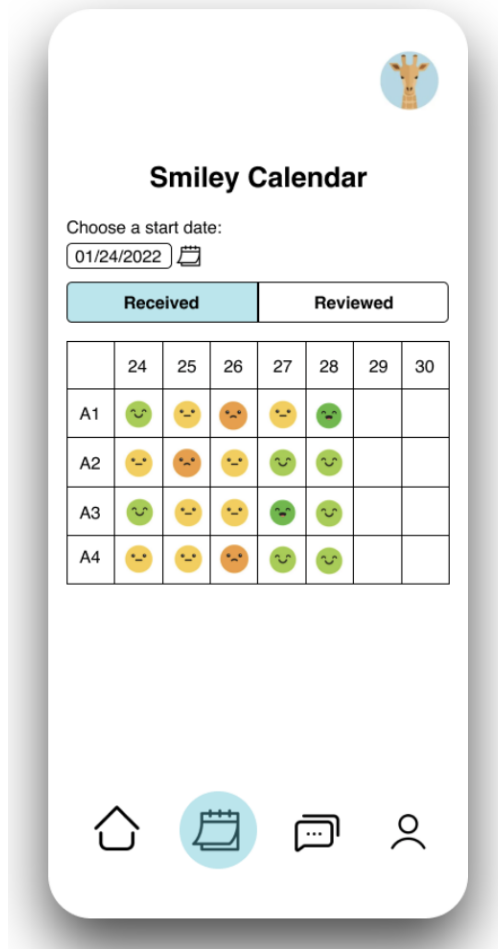


Figure 4.3: Smiley calendar.

to each face. This interface provides students with an intuitive visualization of whether their peers are satisfied with their performance or contribution. For example, if students receive mostly dark green faces or green faces, it is intended to mean that their group members are content to collaborate with them. Conversely, if students receive mostly red faces or orange faces, it is intended to indicate their team members are not content to work with them.

The second main interface of Daily Smirk is a smiley calendar, as shown in Figure 4.3. It provides a detailed visualization of how they evaluated their peers and how they got reviewed. The students can choose the start date and see the evaluation results for a week following this date. The results of how their team members evaluate them will be anonymous because students may be afraid of being too hard on evaluating their peers as we discussed in Chapter

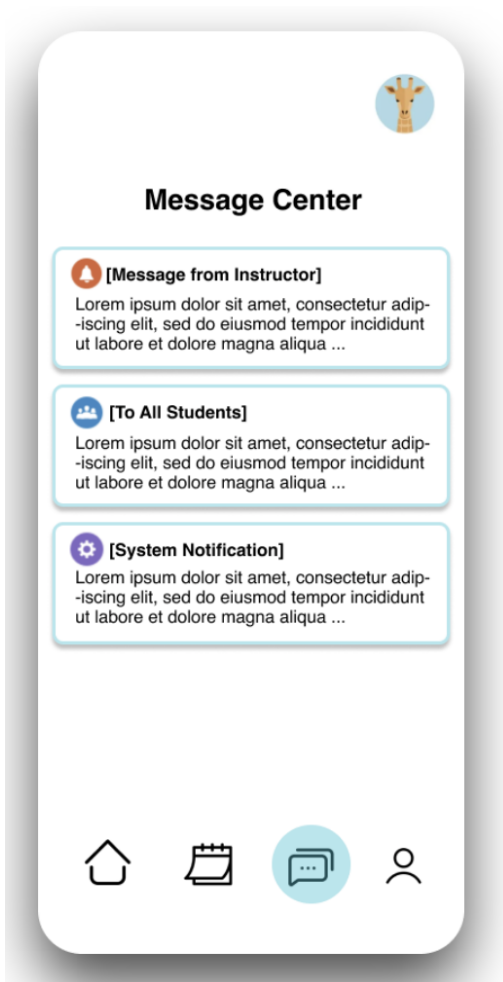


Figure 4.4: Message center.

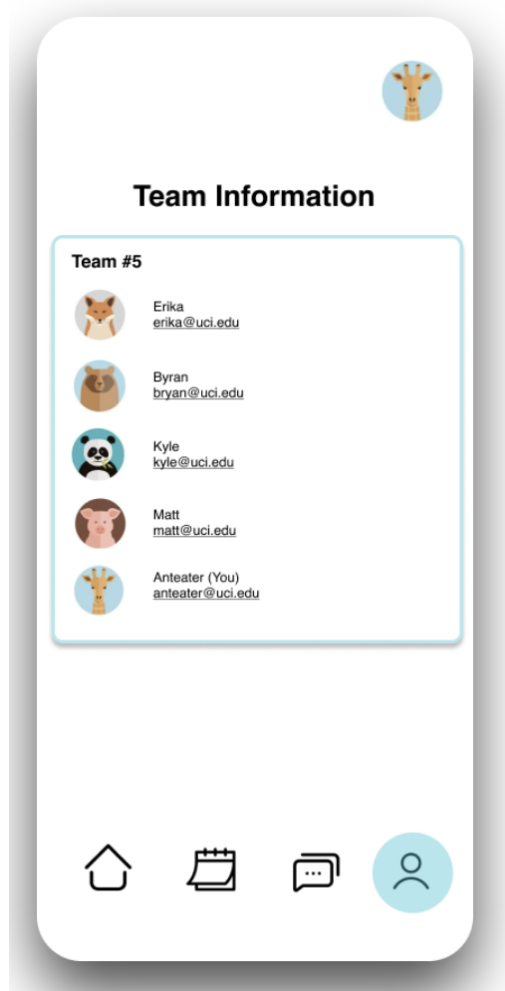


Figure 4.5: Team information.

3. Students are able to evaluate their peers' teamwork more objectively without too much pressure by providing anonymous feedback.

Furthermore, the message center (Figure 4.4) allows instructors to release announcements or remind students of important deadlines to all students. It also allows students to receive messages from their instructors so that instructors can provide timely feedback for students and solve the potential issues for student teams. Finally, Figure 4.5 displays the related information about the team, including the team name, and team members' names with their email addresses. It allows students to quickly contact their team members by clicking on the email addresses.

4.3 Primary Features

In this section, we present three primary features of Daily Smirk in detail and discuss how these novel features are designed for achieving the objectives we discuss in Chapter 3.

4.3.1 Evaluation Feature

The first primary feature of Daily Smirk is the evaluation feature, which is the solution to enable high-frequency peer evaluations and reduce students' workload. As shown in Figure 4.6, the evaluation feature provides students with five faces to choose from for all their peers. Each face is corresponding to a different satisfaction level (see satisfaction level in Figure 4.2). By moving the slider along the bar to correspond with the faces above the bar, students can easily rate their team members. In particular, all team members will be evaluated in a single interface, so students do not need to go through many screens.

We keep the evaluation process simple unless students signal that there is something wrong, in which case the system requires them to provide more information. In other words, if a student selects a red face or an orange for a team member, the system will ask them for more information about this student based on our predefined evaluation criteria (see Section 4.3.2).

Furthermore, if a student selects a dark green face or a green face for a team member which means this student is very satisfied with this team member's work, the system will also ask them for providing detailed feedback. It allows students to learn from their peers by evaluating their peers' outstanding performance and to self-reflect on their own performance.

Some current peer evaluation approaches require students to spend a certain amount of time evaluating their team members, for example, writing short answers to describe and

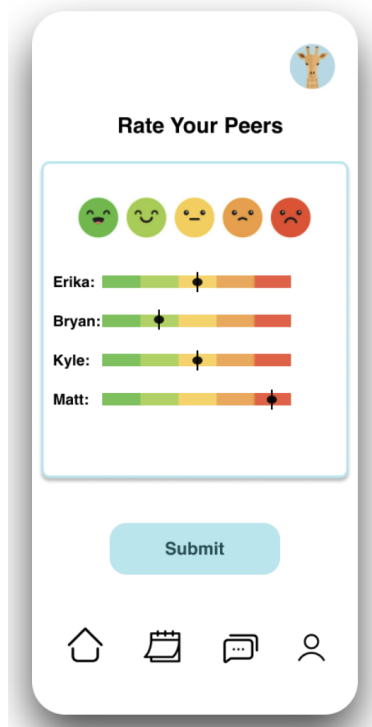


Figure 4.6: Evaluation feature.

assess their peers' performance or giving student scores by reading pages of instructions or lists of assessment criteria, which adds another burden to the students' already heavy workload. Our evaluation feature reduces the amount of time and workload for students to do peer evaluations. This simplified peer evaluation is intended to motivate students to keep providing continuous feedback. As a result, it may help students and instructors to keep track of each team member's perceptions of other team members to help them discover the collaboration issues in teams promptly.

4.3.2 Pre-defined Evaluation Criteria

Pre-defined evaluation criteria aim to achieve the objective of providing useful feedback for instructors. Once a student signals there is something quite wrong or good about a team, the system provides them with more information to evaluate their team members.

As shown in Figure 4.7, we categorize peer evaluation into three aspects by considering students' behaviors in the group projects and referencing the literature we discussed in Section 2.3, including collaboration, participation, and contribution. Each criterion has a list of associated behaviors in the form of checkboxes that are provided to students for their reference to team members' behaviors. Under each criterion, there is a comment box that can be used to provide additional detail. Students can use comment boxes to tell instructors about their concerns and suggestions. Only one criterion is displayed at a time, and students do not have to work through all three screens. They can skip some screens, but the system does require them to choose checkboxes, provide feedback or leave comments at least for one criterion. Figure 4.7 shows the criteria for students who select dark green or green faces. Daily Smirk has another criterion for students who select red or orange faces, asking them why they are not satisfied with their team members' work. The only difference between these two criteria is the description of associated behaviors. For example, in Figure 4.7, the first behavior of collaboration is "Be responsive to others", and the description will change to "Not be responsive to others" in another criterion for students who choose red or orange faces.

We considered two broad categories in designing our evaluation criteria: the extent to which this student collaborates with others and the contributions of this student to teamwork, and categorize these into three specific aspects. The first criterion is collaboration as we will examine whether Daily Smirk can improve collaboration among student teams in the future. It includes the following associated behaviors: be responsive to others' messages, easy to work with, try to resolve conflicts, assist group members when needed, and behave cooperatively and respectful to others. Furthermore, we choose participation as our second criterion because of the consideration of the free riders and some students who work individually. Participation consists of attending group meetings/discussions regularly, volunteering for group tasks, and building and listening to others' ideas. Finally, we design the third criterion, contribution, in terms of both quantity and quality. It includes suggesting or creative

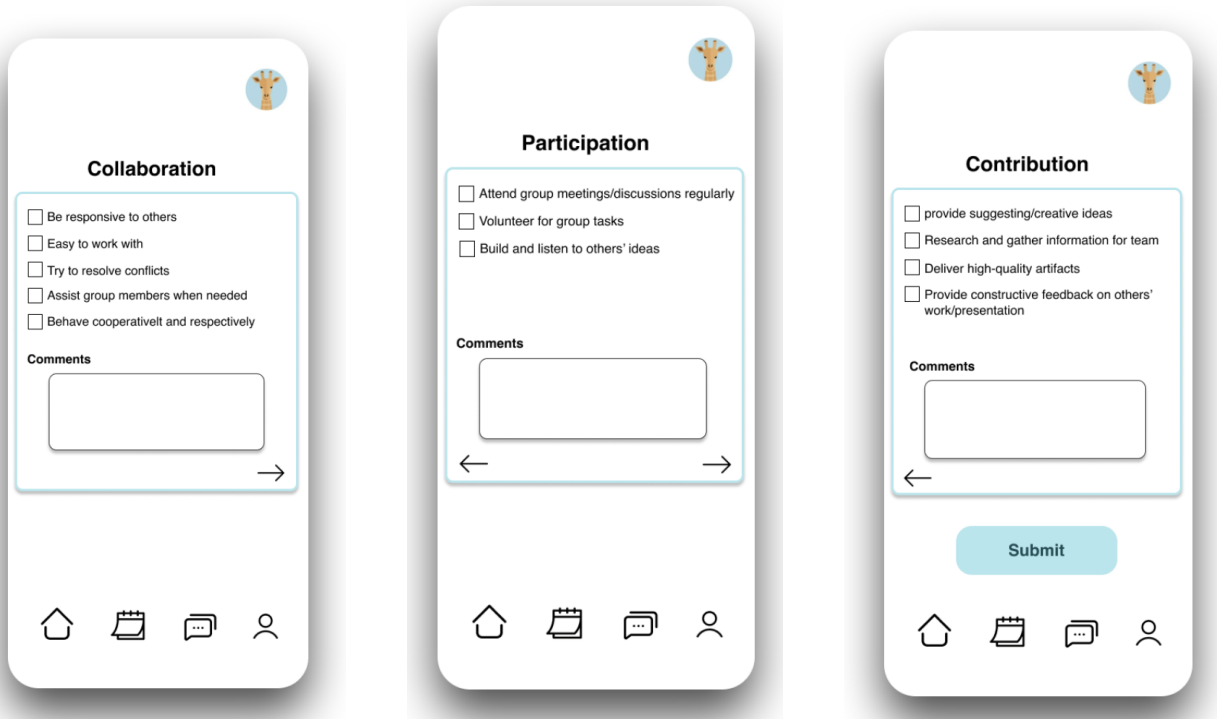


Figure 4.7: Peer evaluation criteria.

ideas, researching and gathering information for teams, providing constructive feedback on others' work/presentation, and delivering high-quality artifacts.

4.3.3 Instructors Perspective

The third feature of Daily Smirk is the visualization for instructors to inspect all evaluation results because most peer evaluation tools are only visible to the students. As shown in Figure 4.8, this feature is designed for addressing the objective of providing helpful insights for instructors. Commonly, peer evaluation results are displayed in pivot tables or spreadsheets, so it is difficult for instructors to figure out a student's performance and to monitor the ongoing teamwork in different groups. Daily Smirk provides a dashboard for instructors to display all students' details. By clicking on a student's name, instructors can view all related information about this student, including this student's team information, a calendar

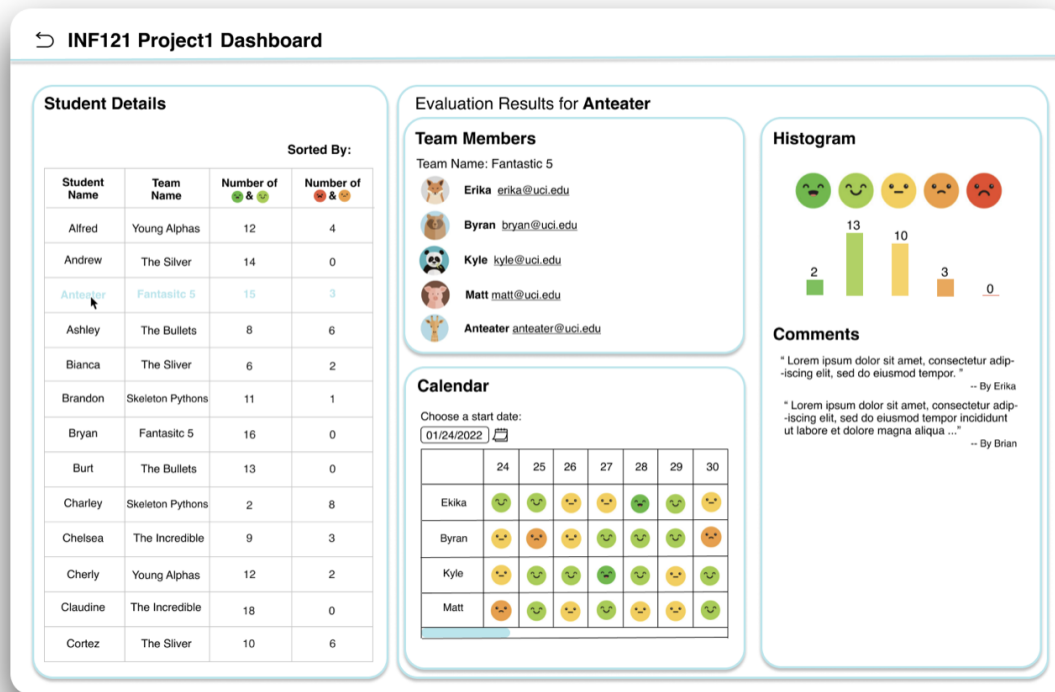


Figure 4.8: Dashboard for instructors.

to display how this student got reviewed, histograms to show how many of each type of smileys this student has received, and comments. It should be noted that all evaluators' names will not be anonymous to instructors. The intended result of using this feature is to allow instructors visually see how each student is performing over time and to assist them in identifying the potential collaboration issues early.

Chapter 5

Evaluation Plan

In this chapter, we propose a long-term evaluation plan to assess the effectiveness of using Daily Smirk to evaluate team members' performance and monitor the teamwork. The rest of this chapter is structured as follows. Section 5.1 proposes four guiding assessment questions and Section 5.2 introduces three stages of our evaluation plan by illustrating the methods by which we plan to collect the data and analyze it.

5.1 Assessment Questions

Since this thesis seeks to investigate whether our continuous peer evaluation approach can address some of the issues with current peer evaluation approaches, including reducing students' workload, providing useful insights for instructors, and resetting students' expectations in doing peer evaluation by designing and developing Daily Smirk, we propose four guiding assessment questions:

1. *Which features effectively support students in doing peer evaluations?* Two of the three primary features (see Sections 4.3.1 and 4.3.2) are designed to help students with doing peer

evaluations. We want to explore how well these features assist students to evaluate their team members, and whether our mechanism of asking for more information if a student selects a face other than a yellow face would be feasible and effective for peer evaluation. We also wish to evaluate our pre-defined evaluation criteria that can arm students with evaluation skills to rate peers in a more professional way instead of providing very general feedback.

2. How does continuous and timely feedback change students' behaviors in group projects?

Our peer evaluation approach is intended to enable students to provide and receive continuous feedback. We aim to study whether the continuous and timely feedback that students receive motivates them to adjust their performance or contribution and increases their self-awareness. In particular, we want to understand whether the continuous feedback encourages them to improve how they contribute to group projects or causes fewer negative effects.

3. How does Daily Smirk assist instructors in monitoring students' ongoing teamwork?

This question will be studied through two features of Daily Smirk: the visualization of evaluation results and pre-defined criteria. Since one of Daily Smirk's contributions to peer evaluation is to provide a visualization of evaluation results for instructors to view, we want to assess the effectiveness of this primary feature. More specifically, we will examine whether this feature allows instructors to easily understand how each student is performing over time and whether this feature allows them to discover problems in student teams and intervene early. Furthermore, the pre-defined evaluation criteria of Daily Smirk will be assessed in terms of whether they allow instructors to receive more useful feedback from students.

4. What is the impact of using Daily Smirk on team collaboration?

The ultimate goal of our research is to investigate whether Daily Smirk can facilitate collaboration among students. The most important ability we want the students to acquire is to reflect on all the feedback they provided and received to improve their teamwork by using Daily Smirk. Specifically, we want to assess whether Daily Smirk helps students develop teamwork skills that actively participate in group projects and communicate with other team members, identify problems

promptly if they arise, and proactively attempt to solve them in order to create a more productive and harmonious team environment.

5.2 Stages of Evaluation

In order to validate the effectiveness of Daily Smirk along with these four assessment questions, we propose a long-term evaluation with three stages, including an initial deployment to collect feedback on the use of Daily Smirk, monitoring how students use it, and their thoughts about Daily Smirk, and comparing data to investigate the differences between Daily Smirk and other peer evaluation approaches. We will deploy Daily Smirk in several team-based courses such as software design courses and capstone courses at UC Irvine.

5.2.1 Initial Deployment

The first stage is a formative assessment process that focuses on examining usability and collecting early feedback from students. This stage aims to address the first assessment question of which features effectively support students in doing peer evaluation, and test whether Daily Smirk is user-friendly. We will deploy Daily Smirk in a team-based course for a few weeks, and survey students' opinions on using Daily Smirk together with some interviews to understand what students think of the tool. What we look for in the data we obtain is to get a sense of the user experience and how they feel about the primary features we designed. This would provide us with a first look at what features are useful and what features might need improvement or to be added altogether to better assist students in evaluating their team members.

5.2.2 Monitoring

The second stage will also utilize a formative assessment strategy, concentrating on longer-term use in a team-based course to learn students' thoughts about Daily Smirk and study what students do with the tool. The goal of this step is to address the second assessment question about how continuous feedback changes students' behaviors in doing peer evaluations and how it changes their performance in group activities. It also strives to address the third assessment question on whether Daily Smirk provides useful insights for instructors to monitor the ongoing teamwork. We will deploy Daily Smirk in a class for a whole quarter and design before and after surveys to learn students' thoughts about Daily Smirk. A pre-survey will be used for understanding students' previous experience in team work and peer evaluations, such as what kinds of tools they have used and the strengths and weaknesses of these tools, what they think about the aspects that their team members should be evaluated on, and so forth. For the post-survey, we will ask students about the usability of Daily Smirk, whether continuous feedback helps them to self-reflect on their own performance and learn from others, whether this tool helps them evaluate their team members, how it can be improved, and so forth. We will also investigate what they do with the tool by analyzing the usage logs and the grades they assigned to others. Furthermore, we will conduct interviews with instructors on the visualization of evaluation results to investigate if this feature assists them in monitoring the whole team's performance as well as individual performances and if it empowers them to intervene early on collaborative issues and provide timely feedback for students.

5.2.3 Comparison

We will make the improvements to Daily Smirk based on the data collected from the previous two stages, after which the third stage of our assessment will use a summative evaluation

strategy to actually compare the differences between Daily Smirk and other peer evaluation approaches or tools. This stage targets whether we address the issues that current peer evaluation tools have, and the strengths and the weaknesses of Daily Smirk compared to other tools. We plan to do this by deploying Daily Smirk in half of a team-based class, and the other half will use another peer evaluation tool. Surveys and interviews will be conducted to ask students questions such as the difficulty level of using the evaluation feature, whether the peer evaluation criteria are helpful for them, the level of workload to evaluate their team members, and whether the students would be willing to use this peer evaluation tool in a future course. We will also design quantitative surveys to ask students questions such as how frequently they use the tools, how long the students take to complete their evaluations, how many collaboration issues have been reported, and so forth. In addition, we will interview instructors to ask them if Daily Smirk allows them to better figure out how students are performing in their teams and find problems more promptly compared to the peer evaluation tool they used before.

Chapter 6

Conclusion

Many university courses provide students with opportunities to work as a team, and they utilize peer evaluation as an effective approach to address the collaborative issues in student teams and to develop students' teamwork skills. This thesis strives to tackle the shortcomings of current peer evaluation approaches by studying the literature and describing the design of a novel peer evaluation tool, Daily Smirk, to address these weaknesses. The concept of our approach is to conduct high-frequency and lightweight peer evaluations that enable students to provide and receive continuous and timely feedback. In particular, the features of Daily Smirk are designed to achieve our stated objectives of reducing students' workload in evaluating their team members, resetting expectations for normal level of peer performance, and providing helpful insights for instructors. We identified four guiding assessment questions to study in our future work, and we developed a three-phase evaluation plan to assess the effectiveness of Daily Smirk that seeks to deploy it in team-based courses at UC Irvine. Finally, we hope that the anticipated educational benefits of Daily Smirk will be used to promote team collaboration and develop students' lifelong skills that help them to be better prepared for their future careers.

Bibliography

- [1] CATME - Smarter Teamwork [Computer Software]. <https://info.catme.org/>.
- [2] Peerceptiv [Computer Software]. <https://peerceptiv.com/>.
- [3] P. Aggarwal and C. L. O'Brien. Social Loafing on Group Projects: Structural Antecedents and Effect on Student Satisfaction. *Journal of Marketing Education*, 30(3), Dec. 2008.
- [4] G. Akom. Using Formative Assessment Despite the Constraints of High Stakes Testing and Limited Resources: A Case Study of Chemistry Teachers in Anglophone Cameroon. *ProQuest LLC*, 2010.
- [5] C. Aoun. Peer-assessment and learning outcomes: product deficiency or process defectiveness? *Proceedings of the 34th International Association for Educational Assessment (IAEA) Conference*, 2008.
- [6] J. R. Beatty, R. W. Haas, and D. Sciglimpaglia. Using Peer Evaluations to Assess Individual Performances in Group Class Projects. *Journal of Marketing Education*, 18(2), Aug. 1996.
- [7] T. Beaubouef and J. Mason. Why the high attrition rate for computer science students: some thoughts and observations. *ACM SIGCSE Bulletin*, 37(2):103–106, June 2005.
- [8] B. Beigpourian, D. Ferguson, F. Berry, M. Ohland, and S. Wei. Using CATME to Document and Improve the Effectiveness of Teamwork in Capstone Courses. June 2019.
- [9] S. Brutus and M. B. L. Donia. Improving the Effectiveness of Students in Groups With a Centralized Peer Evaluation System. *Academy of Management Learning & Education*, 9(4), Dec. 2010.
- [10] A. Burgess, C. Roberts, A. S. Lane, I. Haq, T. Clark, E. Kalman, N. Pappalardo, and J. Bleasel. Peer review in team-based learning: influencing feedback literacy. *BMC medical education*, 21(1):426, Aug. 2021.
- [11] J. C. Cassady and B. E. Gridley. The Effects of Online Formative and Summative Assessment on Test Anxiety and Performance. *The Journal of Technology, Learning and Assessment*, 4(1), Oct. 2005.

- [12] C.-h. Chen. The implementation and evaluation of a mobile self- and peer-assessment system. *Computers & Education*, 55(1):229–236, Aug. 2010.
- [13] S. Chen. Student Perspectives on Peer Learning. *WGU Labs — Advancing Innovation in Education*, Nov. 2021.
- [14] E. Chiriac. Group work as an incentive for learning – students’ experiences of group work. *Frontiers in psychology*, 5:558, June 2014.
- [15] R. English, S. T. Brookes, K. Avery, J. M. Blazeby, and Y. Ben-Shlomo. The effectiveness and reliability of peer-marking in first-year medical students. *Medical Education*, 40(10):965–972, Oct. 2006.
- [16] V. Farrell, G. Ravalli, G. Farrell, P. Kindler, and D. Hall. Capstone project: fair, just and accountable assessment. In *ITiCSE ’12*, 2012.
- [17] W. A. Friess and A. J. Goupee. Using Continuous Peer Evaluation in Team-Based Engineering Capstone Projects: A Case Study. *IEEE Transactions on Education*, 63(2), May 2020.
- [18] M. Hamilton, A. Carbone, C. Gonsalvez, and M. Jollands. Breakfast with ICT employers: What do they want to see in our graduates? In *Proceedings of the 17th Australasian Computing Education Conference (ACE 2015)*, pages 29–36. Australian Computer Society Inc, 2015.
- [19] J. Hayes, T. Lethbridge, and D. Port. Evaluating individual contribution toward group software engineering projects. In *25th International Conference on Software Engineering, 2003. Proceedings.*, pages 622–627, 2003.
- [20] E. Isaac and A. Olutimilehin. Influence of Peer-Assessment on Students Academic Performance in Home Management in Public Senior Secondary Schools in Akwa Ibom State. *International Journal of Innovative Education Research* 9(4):112-125, page 14, 2021.
- [21] S. Isaac and R. Tormey. Undergraduate group projects: Challenges and learning experiences. *QScience Proceedings (Engineering Leaders Conference 2014)*, Aug. 2015.
- [22] D. Kelley and M. Sadowski. Peer Evaluation within a Team Design Project. *Journal of Engineering Technology*, 32, Mar. 2015.
- [23] S. Kunwongse. Peer Feedback, Benefits and Drawbacks. *Thammasat Review*, 2013.
- [24] S. Lindblom-ylänne, H. Pihlajamäki, and T. Kotkas. Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education*, 7(1), Mar. 2006.
- [25] M. L. Loughry, M. W. Ohland, and D. J. Woehr. Assessing Teamwork Skills for Assurance of Learning Using CATME Team Tools. *Journal of Marketing Education*, 36(1), Apr. 2014.

- [26] A. Luxton-Reilly. A systematic review of tools that support peer assessment. *Computer Science Education*, 19(4), Dec. 2009.
- [27] M. Mayo, M. Kakarika, J. C. Pastor, and S. Brutus. Aligning or Inflating Your Leadership Self-Image? A Longitudinal Study of Responses to Peer Feedback in MBA Teams. *Academy of Management Learning & Education*, 11(4), 2012.
- [28] E. B. Nuhfer. The Place of Formative Evaluations in Assessment and Ways to Reap Their Benefits. *Journal of Geoscience Education*, 44(4), Sept. 1996.
- [29] A. Planas Lladó, L. F. Soley, R. M. Fraguell Sansbelló, G. A. Pujolras, J. P. Planella, N. Roura-Pascual, J. J. Suñol Martínez, and L. M. Moreno. Student Perceptions of Peer Assessment: An Interdisciplinary Study. *Assessment & Evaluation in Higher Education*, 39(5), 2014.
- [30] A. Poddar. Continuous Additive Peer Review: A New System to Control Social Loafing in Group Projects. *Journal for Advancement of Marketing Education*, 17, 2010.
- [31] K. Pond, D. Coates, and O. A. Palermo. Student experiences of peer review marking of team projects. *International journal of management education*, 6(2), pp. 30-43, Jan. 2007.
- [32] W. W. Shi, A. Jagannadharao, J. Lee, and B. P. Bailey. Challenges and Opportunities for Data-Centric Peer Evaluation Tools for Teamwork. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):432:1–432:20, Oct. 2021.
- [33] B. Sridharan, J. Tai, and D. Boud. Does the Use of Summative Peer Assessment in Collaborative Group Work Inhibit Good Judgement? *Higher Education: The International Journal of Higher Education Research*, 77(5), May 2019.
- [34] T. Staubitz, D. Petrick, M. Bauer, J. Renz, and C. Meinel. Improving the Peer Assessment Experience on MOOC Platforms. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale, L@S '16*, pages 389–398, New York, NY, USA, Apr. 2016. Association for Computing Machinery.
- [35] J.-W. Strijbos, T. Ochoa, D. Sluijsmans, M. R. Segers, and H. Tillema. Fostering Interactivity through Formative Peer Assessment in (Web-Based) Collaborative Learning Environments. *Cognitive and Emotional Processes in Web-Based Education: Integrating Human Factors and Personalization*, Jan. 2009.
- [36] I. Čavrak, M. Orlić, and I. Crnković. Collaboration patterns in distributed software development projects. In *2012 34th International Conference on Software Engineering (ICSE)*, June 2012.