2023 IEEE 35th International Conference on Software Engineering Education and Training (CSEE&T) | 979-8-3503-2202-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/CSEET58097.2023.00014

The Daily Smirk: A Preliminary Prototype for Continuous Peer Assessment of Team-based Projects

Yu Lu
Department of Informatics
University of California, Irvine
Irvine, CA, USA
ylu31@uci.edu

André van der Hoek
Department of Informatics
University of California, Irvine
Irvine, CA, USA
andre@uci.edu

***Abstract*—Team-based projects are common in many software engineering classes, but come with a host of well-known problems such as social loafing, free-riding, and undesirable team dynamics. To counter these issues, many instructors use peer assessment. Unfortunately, current peer assessment approaches exhibit several limitations, including a typically time-consuming evaluation process, feedback being too infrequent or too late to make a difference for the project, and student ratings being unreliable. In this paper, we introduce the Daily Smirk, a preliminary prototype of a new peer assessment tool that is based on three primary design decisions: (1) peer assessment is continuous throughout the project, (2) assessment is through a lightweight smiley-based rating system, and (3) assessment is normalized around a neutral state so that additional feedback is only needed for outlier ratings. We elaborate on the design of the Daily Smirk and its primary functionality, and report on a preliminary evaluation in a software design course with three team-based projects.**

***Index Terms*—Software Engineering Education, Team-based Projects, Peer Assessment**

I. Introduction

Team-based projects have become commonplace for software engineering and other computing courses, often to provide students with a realistic experience in a "real-world" working environment that better prepares them for future professional careers [1]. From an educational perspective, they serve as an incentive for collaborative learning, which enables students to develop problem-solving skills and build new understandings with each other [2]. While team-based projects offer many educational benefits, there are considerable challenges in accurately identifying individual contributions for assessment [1]. Moreover, students struggle with collaboration issues within teams, including free-riders, social loafers, conflicts, poor communication, and so on [3].

Peer assessment has been employed as an approach advocated by many educational researchers to integrate into team-based projects, as it is considered to deliver feedback efficiently to learners [4]. There is strong evidence in the literature that peer assessments facilitate learners' learning and develop their key competencies [5] while reducing the burden on instructors [6]. It benefits learners by promoting critical reflection, developing communication skills, and improving teamwork behaviors by providing and receiving feedback [7]. In addition to improving the efficiency of grading [8], instructors also leverage peer assessment to instill a sense of responsibility for individual contributions in order to reduce the problem of free-riding or social loafing [9].

While peer assessment has been proven to be beneficial for collaborative learning, some challenges also need to be taken into account when incorporating peer assessment into team-based projects. These include the accuracy and effectiveness of feedback from peers, students' insecurity when assessing their peers, the difficulty of awarding a mark, and the tendency to under-mark or over-mark [10]. Potential dishonesty in doing peer assessment includes bias from interpersonal relationships, students' fear of being harsh to their peers resulting in penalizing peers' grades, and lack of true anonymity in small teams [11]. In addition, other challenges in conducting peer assessment include the class time consumed by the assessment process and difficulties in capturing the learning outcomes [6].

The specific problems with peer assessment that we address in the paper are the heavyweight workload to complete a single round of peer assessment, providing and receiving useless feedback, and inflation of the rating results. To address these problems, we introduce the preliminary prototype of a novel peer assessment platform, Daily Smirk, which aims to address some of the limitations of existing peer assessment platforms and provide insights for both learners and instructors. It is designed for learners to deliver continuous feedback to promote collaborative learning, as well as for instructors to avoid heavyweight assessment mechanisms yet be able to effectively monitor ongoing teamwork. A preliminary evaluation of the Daily Smirk was conducted in a software design course with three team projects using a pre- and post-survey. The purpose of the surveys was to examine the Daily Smirk's effectiveness, to study students' reactions to continuous feedback, and to seek opportunities for improving the tool.

We found that the Daily Smirk reduces heavyweight workload of completing peer assessments, changes learners' behaviors in collaborative learning environments to some extent,

979-8-3503-2202-6/23/$31.00 ©2023 IEEE
DOI 10.1109/CSEET58097.2023.00014

and provides instructors with useful insights when monitoring teamwork.

The remainder of this paper is organized as follows. Section II provides an overview of related work on peer assessment in teamwork, and overviews existing peer assessment platforms and their benefits and limitations. Section III introduces our proposed peer assessment platform, Daily Smirk, including its main design decisions, the initial prototype and its main features for the students and instructors, and implementation details. In Section IV, we present the preliminary evaluation of the Daily Smirk, including the detailed setting, the pre-survey and post-survey, and a discussion of the results. The last section outlines our conclusions and discusses future work.

II. Background

Team-based projects are a significant component of higher education as they provide many educational benefits. Such projects provide learners with a realistic experience in terms of collaboration, leadership, decision making, and communication, while also improving each team member's discipline-specific knowledge [12]. However, working as a team is not always a rewarding learning experience for students, as undesired collaboration issues may emerge within teams [13]. Free riding or social loafing is considered the most significant factor that can undermine the effectiveness of teamwork, affecting students' satisfaction with their team members' contributions [14]. Students also struggle with dealing with other issues such as communication difficulties, problems of leadership, and conflicts among team members [15].

Peer assessment was introduced in 1731 by the Royal Society of Edinburgh, which published a collection of peer-reviewed medical articles [16]. Peer assessment has since been applied to a variety of contexts, including writing, portfolios, oral presentations, test performances, and other skilled behaviors [17]. In recent years, there has been a renewed interest in peer assessment, particularly in the area of formative assessment [17]. Formative assessment is interactive and takes place during the course or a project, rather than at the end of it [18]. It takes advantage of one of the major benefits of computer-assisted assessment, which is providing timely feedback [19]. In addition, formative peer assessment focuses on the in-depth qualitative evaluation of various student learning outcomes and reinforces learning [20].

Nowadays, peer assessment is widely used as an approach advocated by many educational researchers to integrate into team-based projects as its use improves the student learning experience and its positive effects on student achievement and attitudes [21]. Involving students in the peer assessment process provides them with opportunities to develop their soft skills for future professional development, such as communication, problem-solving, and teamwork skills [7]. Furthermore, providing and receiving feedback is beneficial for fostering students' cognitive development, including constructive reflection, a greater sense of accountability and responsibility, and so on [22]. Peer assessment also motivates students to involve in more group activities and put more effort into teamwork than normal [10].

Peer assessment can be conducted manually (e.g., on paper) or through the use of online peer assessment tools. Many tools have shown promise in team-based learning environments. SPARK [23] is one of the earliest peer assessment platforms for student team-based projects and was designed for reducing the limitations of paper-based peer assessment systems, improving the fairness of team assessment, and enhancing the learning experience for students. It enables students to evaluate their own or their peers' contributions toward team tasks and team maintenance. SPARK allows students to practice using the assessment criteria and provides explicit help for them to get a better understanding of the assessment process, and submit their ratings confidentially. Furthermore, SPARK has a separate system for instructors that allows them to create teams, set assignment deadlines and assessment criteria, and calculate final grades. The system supports either multiple criteria or holistic rating schemes and offers a small library of predefined assessment criteria for instructors to design the self and peer assessment.

Comprehensive Assessment of Team Member Effectiveness (CATME) [24] is another web-based assessment system that can be used for demonstrating the achievement of learning goals related to teamwork. It consists of three tools: (1) Team-Maker allows instructors to assign students to teams based on collected demographic information; (2) CATME Peer Evaluation is used for self-assessment and peer assessment according to team member contributions; and (3) Rater Calibration allows students to practice rating fictitious team members in order to ensure students are familiar with the rating criteria. The second tool, CATME Peer Evaluation, collects self and peer ratings on five dimensions of team members' contributions, namely contributing to the team's work, interacting with teammates, keeping the team on track, expected quality, and having relevant KSAs (knowledge, skills, and abilities). It leverages a behaviorally anchored point scale to describe the behaviors of high, medium, and low performance for each dimension. The platform is capable of collecting multiple rounds of assessment data and distributing feedback to students repeatedly over a long term. In addition, the data collected from peer assessments is anonymous to students but visible to instructors to ensure that ratings are confidential and more honest. The CATME system has been shown to support teamwork in higher education, promote better experiences for instructors to monitor ongoing teamwork, and provide a better learning experience for students to collaborate as a team.

Beyond SPARK and CATME, several peer assessment platforms for team-based learning have been proposed, each with varying features. Most of which have also been used in one or more courses. Collectively, while making a difference, some concerns have arisen as well when used. Particularly, issues have been observed concerning the accuracy of feedback, insecurity to rate peers, and a protracted assessment process. These factors need to be considered when employing a peer assessment tool in team-based projects, as poorly conducted

peer assessment can undermine team performance [25]. The tool described in this paper, Daily Smirk, is designed to address these concerns. While it draws inspiration from existing peer assessment platforms, it differs from them in goals, design decisions, and overall functionality.

III. Daily Smirk

A. Main Design Decisions

In this paper, we propose a high-frequency and lightweight peer assessment system that aims to provide instructors with useful insights about teamwork and assist students in assessing their peers. The Daily Smirk is specifically designed to address some of the challenges that current peer assessment systems exhibit. The first step in designing the Daily Smirk was to decide on the main design decisions that govern how we envisioned it to work. We list and motivate these design decisions in the below.

- **High-frequency assessments.** The primary design decision is to focus on a high-frequent formative peer assessment, as it allows students to receive timely and continuous feedback regarding their ongoing performance in teamwork, which can provide them with valuable diagnoses that they might not be aware of [26]. Specifically, students are required to rate their peers every day on which they perform team activities. The expected outcome of using formative peer assessments is to improve students' self-reflection skills, gradually develop their critical thinking skills, and provide them with sufficient time to adjust their performance to meet the teams' expectations. By giving students feedback on a daily basis, it becomes possible to quickly discover potential issues and allows students to adjust their behavior, rather than only learning about teammates' attitudes towards their own work at the end of a project. Continuous assessment can also identify problems that arise later in the project or intermittent issues.
- **Lightweight workload.** Since our approach aims to conduct high-frequency peer assessments, it is important that each assessment can be conducted quickly and does not impose a serious burden on the students. We propose a smiley-based peer assessment approach that provides five different faces for students to choose from, indicating their level of satisfaction with their peers' performance or contribution in group activities on a given day. Students simply choose one of five faces for each of their peers to complete the evaluation, ranging from (1) did something extraordinary, (2) went above what was expected, (3) as expected, (4) below expectations, and (5) seriously problematic.
- **Neutral expectation of everything is going well.** When things are going well within a team, especially we do not want to have students do more than what is needed to signal that all is as expected. To do this, a neutral (yellow face) is the default on each day, meaning that with a simple click, students can signal to one another that from their perspective, all is going as expected that day.

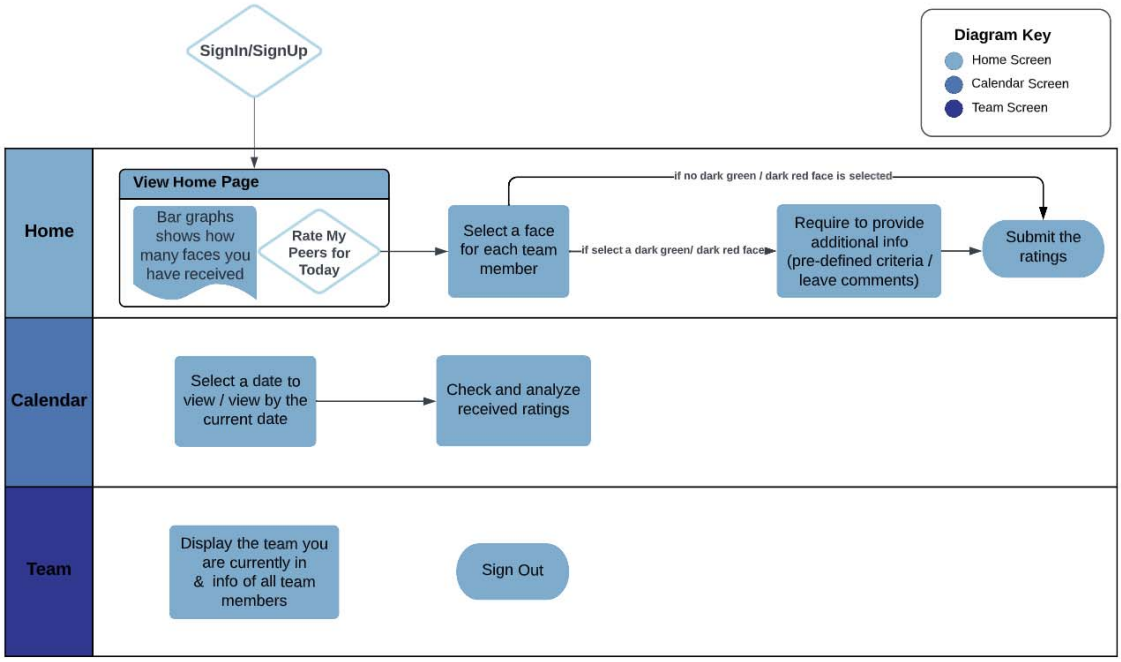


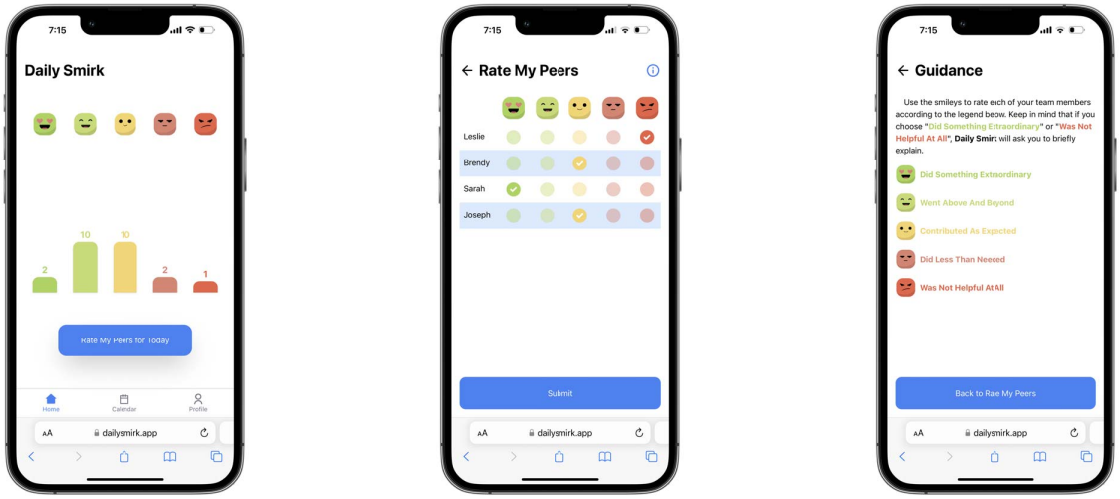
Fig. 1: A diagram showing the workflow of the Daily Smirk student side.

- **Elaborate on ratings when a peer's contributions on a given day are egregiously bad or very good.** The Daily Smirk will ask students to provide more information if they want to signal that someone did something extraordinary or was seriously problematic. If students report that one of their team members is doing nothing or not helping at all, they need to identify their specific issues affecting teamwork. This provides an opportunity for that student to correct their behavior and keeps the instructor informed of collaboration issues that arise within the team. Also, if students want to commend someone on their team for doing something extraordinary, they are asked to provide specific reasons.
- **Provide and receive useful feedback.** Manually providing written feedback can lead to feedback that is poorly written and not informative due to a lack of experience in how to give useful feedback [27]. Therefore, the Daily Smirk assists students in rating their peers by providing pre-defined assessment feedback. The criteria are designed based on the literature on student behaviors in team-based projects (see next subsection). In the Daily Smirk, students can simply checkmark applicable feedback, which makes it easy for students to provide useful feedback to each other. Nonetheless, the Daily Smirk also provides a freeform field, should a student have feedback different from the standardly available phrases, or feels like elaborating on those phrases to provide more context.
- **Easy to use.** Since the vision of the Daily Smirk is frequent use, it is important that the ability to provide feedback is quick and painless. The prior design decisions already set the stage for that being the case, but it is important that the remainder of the Daily Smirk's functionality also does not impose too much burden on the user. As one example, it should not be necessary for a student to work through multiple levels of menus to understand how their peers are rating them.
- **Dashboard for instructor.** Instructors have a need to identify potential dysfunction and conflicts [11]. Therefore, the Daily Smirk provides instructors with insights to monitor ongoing teamwork through a dashboard that is designed to display all team information and assessment results so as to be able to intervene as needed.

B. Initial Prototype

The initial prototype of the Daily Smirk is designed in accordance with the primary design decisions we just described. It consists of two parts: the student side and the instructor side. The student side is shown in Figure 1, which highlights how each of the three main screens of the Daily Smirk student side works, including home, calendar, and team. The home screen summarizes the assessment results and enables students to start the peer assessment process; the calendar screen allows students to view their assessment results over time in a calendar visualization; and the team screen displays a user's current team and information about his or her all team members, including their names and contact information. In the following, we present the Daily Smirk's primary features, demonstrate how they work, and how they address the main design decisions in detail.

1) Assessment: Figure 2(a) presents the home screen of the Daily Smirk, which consists of two parts: a summary of rating results and an evaluation button. It uses a bar graph to



(a) home screen

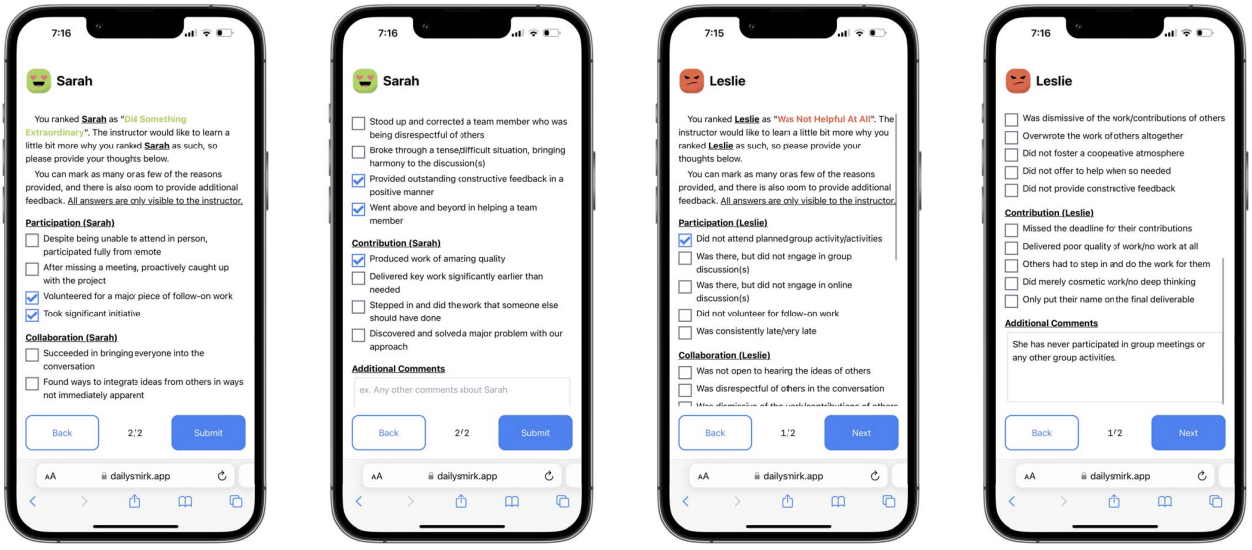
(b) evaluation screen

(c) guidance for explaining satisfaction levels

Fig. 2: Screen collage showing how the evaluation feature works.

display the number of faces students received from their peers from the start date of the group project to the current date, with each bar corresponding to the face above it. The graph provides an intuitive visualization of whether their peers are satisfied with their performance based on the number of green, yellow, and red faces they received. In addition, there is a "Rate my Peers for Today" button at the bottom. Once a user clicks this button, the assessment process starts by providing the student with five faces to select for each of their peers (Figure 2(b)). Each face corresponds to a different satisfaction level (Figure 2(c)) that represents students' perceptions of their peers. It allows students to easily rate their peers by simply clicking on the radio button corresponding to each face, and all peers will usually be evaluated on a single screen instead of going through many screens. The evaluation process is kept simple unless students signal to the system that there is something not going well or going very well in the team, in which case the system requires the students to provide additional information. For example, a student has a very good team member (Sarah) who contributed the most to the team and always took initiative to help other team members, and a not-so-great team member (Leslie) who did not attend planned group activities at all. Once that student selects a dark green face for Sarah and a dark red face for Leslie (see Figure 2(b)) and hits the "Submit" button below, the system will ask that student to provide detailed feedback about Sarah and Leslie based on our pre-defined evaluation criteria (see Figure 3) or leave comments. This process by default these being yellow faces and choices of dark red and dark green only requires additional information. It is designed to reduce the time and workload of students doing peer assessments and to motivate them to keep providing feedback.

2) Pre-defined Criteria: Appropriate assessment criteria can better identify individual efforts within a team [28]. Therefore, we designed pre-defined assessment criteria that aim to assist students to evaluate their peers' work, while also providing useful feedback to instructors. There are various categories that should be taken into account when designing peer assessment criteria. Farrell summarized peer assessment of team-based projects into two broad categories: the extent to which students collaborate in team activities with their peers and the value each team member brought to the progress of the project, and they introduced eleven subcategories associated with these two broad categories [29]. Many other assessment criteria were proposed as well. Brutus and Donia categorized peer assessment into four criteria, including cooperation, conceptual contribution, practical contribution, and work ethic [30]. Beatty also proposed criteria based on actual group project experiences and tested these through multiple regression analysis [28]. Borrowing from this body of the literature on peer assessment criteria, and taking into account student behaviors in team-based projects, we categorized our peer assessment criteria into three specific areas: participation, collaboration, and contribution. The first criterion, participation, was chosen to address students actively engaging in team activities. Collaboration is used to assess whether the student helps the team operate together in a smooth manner. The last criterion, contribution, addresses the overall quality of the student's contribution. Each criterion has a list of associated



(a) criteria for dark green faces (part 1)

(b) criteria for dark green faces (part 2)

(c) criteria for dark red faces (part 1)

(d) criteria for dark red faces (part 2)

Fig. 3: Screen collage showing the pre-defined criteria.

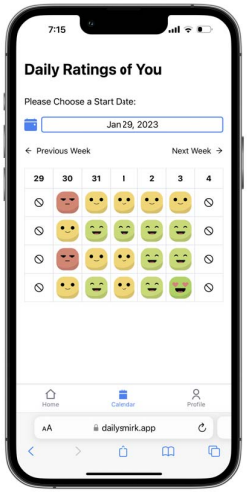


Fig. 5: Calendar screen visualizing the ratings that a student received for a particular week.

behaviors in the form of checkboxes that are used as references to evaluate their peers' performance towards the team. In addition, a comment box at the bottom can be used by students to provide additional details. The system requires the students to select at least one check box or leave a comment in order to successfully submit an evaluation that involves a dark green or dark red face. Figure 3(a)(b) shows the criteria for the student who selected a dark green face for Sarah and marked several check boxes to indicate why that student felt Sarah did something extraordinary that day. Conversely, Figure 3(c)(d) shows the criteria the student selected to explain the dark red face for Leslie. That student also left a comment further detailing the reasons for giving Leslie a bad rating. The only difference between these two criteria is the description of associated behaviors.

3) Calendar: Figure 5 presents the smiley calendar that provides an intuitive visualization of how a student was reviewed by their peers. It allows students to choose the start date and see the assessment results for a week following this date. Each column indicates the face students received on that day, and each row indicates the faces they received from the same group member. The assessment results are anonymous to all team members when students are able to evaluate their peers more objectively without too much pressure by providing anonymous feedback [31]. To prevent students from guessing which team member gave that rating, the position of each row is randomly changed once this page is refreshed. This calendar visualization intends to provide an opportunity for students to improve self-reflection on their contributions over time.

4) Instructor Dashboard: The Daily Smirk also offers instructors a dashboard (Figure 4) to inspect all assessment results. Typically, peer assessment results are displayed in pivot tables or spreadsheets, so it is difficult for instructors to monitor the ongoing teamwork in different student teams and to check a student's performance over time. Therefore, the dashboard was designed to work on a technology team basis. When the instructor clicks on one of the teams, it will display all team members' information including the name,

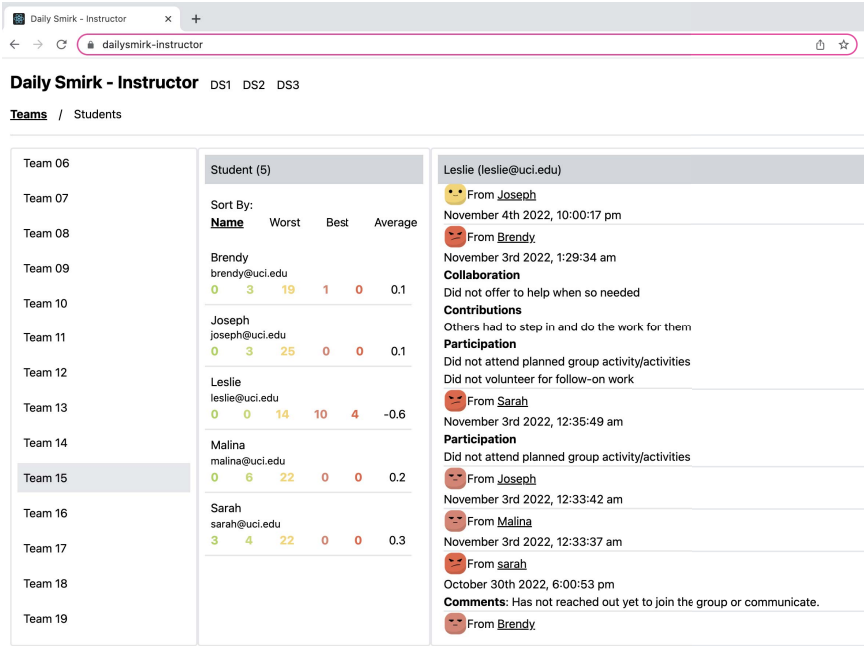


Fig. 4: Screenshot showing the instructor dashboard.

email address, and the number of each of the five faces they received. Clicking on a student's name can view who gave that rating at what time and possibly some comments if they got a dark red or a dark green face. This way an instructor can get a holistic sense of a team and how they view each other's performance. To more directly identify top performers or worst performers in a class, the instructors can also view the results by student names which can be sorted by alphabetical order of names, average scores, or who got the largest number of dark green faces or red faces. Doing so brings up a list of students from which the instructor can navigate to a team's evaluations as desired.

C. Implementation

The Daily Smirk is implemented by using React and Firebase. As a front-end library, React enables the creation of reusable components and allows developers to iterate on the development quickly. Firebase is a Backend-as-a-Service (Baas) that offers many services such as Authentication and Cloud Firestore. Cloud Firebase is an extensible and scalable database that helps to store and synchronize the data for mobile and web development.

IV. Preliminary Evaluation

A. Setting and Goals

As a preliminary evaluation of the Daily Smirk, we deployed it in a software design course during the Fall quarter of 2022. This course is an upper-division course for students who are majoring in Software Engineering, Computer Science, Game Design & Interactive Media, and Informatics at the University of California, Irvine. The course centers on teaching students expert practices in designing software applications and consists of three design group projects, each of them lasting three weeks. The group projects vary in focus, with one project focused on identifying requirements and creating mockups, another project on designing data structures and writing algorithms, and a third project combining all the tasks of the other two projects.

The Daily Smirk was incorporated into all three design group projects for conducting peer assessment activities. There were 144 students enrolled in this course, with each team consisting of four or five students. Students were randomly assigned to different teams for each project so that each student would be more likely to collaborate with different team members for each project. At the beginning of the first group project, students were simply asked to use the Daily Smirk throughout to evaluate their team members. The goal was for students to become familiar with it first, and it was also a good opportunity to test the Daily Smirk on a large scale. For the second group project, we did not force students to use the Daily Smirk, but regular emails were sent out to remind them to do peer assessment if they performed team activities on that day. As a backup, students were also introduced that peer assessment could influence their final grade, especially based on the number of dark red or dark green faces received. In addition to gathering the data from the platform, we also conducted a pre-survey to understand students' previous experience in doing peer assessments, and a post-survey to examine the usability of the Daily Smirk and to collect users' feedback that could help us improve the Daily Smirk in preparation for the next version.

B. Pre-Survey

To understand students' previous experiences with group work and peer assessments, we conducted a pre-survey before students started to use Daily Smirk. The survey consisted of multiple choice and open-ended questions with the questions centering on their past involvement in doing peer assessments. Forty-nine (34%) of the students filled out this survey before the first project started. According to the survey results, 91.66% of these students had experience with working on group projects, and 81.82% of these students encountered one or more collaboration issues in working as a team. Major collaboration issues that repeated include free-riding, lack of communication, unresponsiveness, time conflicts, unbalanced contribution, low-quality work, and so on (as described in Section II). These match the problems we found in the literature about working in groups. Of those students who had participated in team-based projects, 73.81% had experience with using peer assessments to rate their team members.

In terms of the form of peer assessment, the survey results showed that 70.97% of these students who had encountered peer review had only participated in summative peer assessments (which only took place at the end of the course) across various courses, while the rest (29.03%) had previously participated in both formative (took place multiple times during course) and summative peer assessments. Among these students who had experience with doing peer assessments, 50% of them stated that peer assessments were not helpful in avoiding various problems they encountered during the group projects. This was typically due to the summative assessment being only conducted after the completion of the project so they did not have time to make improvements as peers received the feedback too late. Other reasons include the assessment results only going to the instructor, unhelpful feedback, and so on. All of these resulted in the problems not being addressed.

The survey also asked students to what extent their past assessments reflect their true feelings about the team members in the peer assessments. Only 29.73% of them considered their reviews as mostly accurate concerning their feelings; the majority (67.57%) assessed more positively than they truly felt about their team members; and 2.7% of them assessed more negatively than they felt. The results indicate that students tended to give more positive ratings than their peers' real performance.

The pre-survey also explored the peer assessment approaches or tools that students have used, such as Google Forms and Canvas Quizzes as well as the strengths and weaknesses of these tools. Providing feedback in an anonymous way was one of the students' favorite features, and they also liked providing comments on team members' performance, which was more informative and helpful than just giving a

rating. Some weaknesses of the tools include difficulty on mobile devices, lack of individual feedback, and too many words to read. It should be noted that some students considered peer assessment as a tedious process if nothing was going wrong within teams.

C. Post-Survey

After the completion of all three group projects, we conducted a post-survey with 36 questions to evaluate the effectiveness of the Daily Smirk, to understand how students react to the assessment results, and seek opportunities for improvements. Out of 144 students who participated in doing peer assessments with the Daily Smirk, fifty-five (38%) students filled out this form at the end of the quarter. We summarize the survey results of the effectiveness of the Daily Smirk into the following aspects:

1) Average time to complete the peer assessment: To assess whether the Daily Smirk is easy and quick to use, the survey asked about the average time for students to complete a single round of peer assessment. The results (Table I) show that the majority of students (61.36%) complete rating their team members using the Daily Smirk in less than one minute, 34.09% of them took from one to three minutes, 4.55% of them took from three to five minutes, and none of them spent more than five minutes evaluating their team members.

Average time	Percentage of responses
Less than 1 minute	61.36%
1 - 3 minutes	34.09%
3 - 5 minutes	4.55%
More than 5 minutes	0.00%

TABLE I: Results of average time to complete a single round of peer assessment using the Daily Smirk.

2) Thoughts on providing continuous feedback: Since high-frequency formative peer assessment was adopted in the design of the Daily Smirk, we wanted to understand students' perceptions of providing continuous feedback to each other. Most students felt that giving feedback frequently from the beginning is better than giving a single summative assessment at the end as "it gives people a chance to adjust the performance". Some other students illustrated that providing and receiving continuous feedback can increase productivity because immediate feedback allows them to know each team member's contribution and identify the problems in a timely member that "keep up their efforts from getting positive feedback and put in more effort from getting negative feedback". However, a few students preferred to combine simple daily ratings followed by a summative assessment to provide detailed overall feedback at the end. In addition, the survey asked about the burden of using the Daily Smirk each day on a scale of 1 (I feel extremely burdened) to 10 (not a problem at all). As shown in Table II, the majority of students (17.95% gave a 10, 12.82% gave a 9, 23.08% gave an 8) did not feel burdened to use the Daily Smirk on a daily basis.

Scale	Percentage of Responses
1 (I feel extremly burdened)	0.00%
2	2.56%
3	5.13%
4	7.69%
5	15.38%
6	7.69%
7	7.69%
8	23.08%
9	12.82%
10 (Not a problem at all)	17.95 %

TABLE II: Results of the burden of using the Daily Smirk each day on a scale of 1 to 10.

3) Students' true feelings about assessing team members: Approximately half of the students (49.06%) stated that their reviews were mostly accurate with respect to their feelings, and 45.28% felt that they evaluated more positively than their team members' real performance. Students explained that they stayed true to the ratings because the assessment results were anonymous and they wanted the team to know if they were working well together. However, other students tended to evaluate more positively, e.g.,"feel good giving green smileys to members that are enthusiastic and friendly", and "I consider myself a pretty positive and forgiving person so it may take more for me to actually give someone a red knowing that it might affect their grade negatively". Also, they were afraid that negative ratings would prevent their peers from contributing more. One interesting finding was that for giving a green or yellow face, students tended to be more honest, but for red, they had trouble giving the real rankings they felt, and inflated their rankings at times (from dark red to light red, from light red to yellow).

4) Comparison to other peer assessment tools/approaches: 63.89% of students preferred the Daily Smirk over other peer assessment tools or mechanisms they had used in class before, 25% felt it was indifferent to others, and 11.11% preferred other tools or mechanisms. Students who enjoyed using the Daily Smirk commented that "the interface was clearer and more streamlined to give quick feedback in fewer steps", "giving simple feedback frequently keeps people accountable to do their part", "in review peer assessment forces you to pick the side of you liked them or not, with the Daily Smirk you can have a problem with someone on one day and change the next", "I prefer the Daily Smirk to have to think through and write up meaningful evaluations of peers in an environment where I know most don't really care once the course is done", and so on. Other students explained that the Daily Smirk was quicker and easier to use than Canvas or Google Forms with writing too many words to rate their peers, and it also assisted them in providing meaningful evaluations. However, a few students indicated that although it is a good tool, not everyone uses it the same because the ratings are biased and based on moods.

5) Overall experience: In terms of the overall effectiveness of the Daily Smirk, 43.25% of students felt that it made team cohesion and collaboration better, 51.35% felt that it

did not really make a big difference, and only 5.41% felt negative about using the Daily Smirk. Some students pointed out that the Daily Smirk encouraged others to be involved in order to get a good rating and they saw bad ratings made to peers leading to the peers putting more effort into teamwork. However, some students felt that their team members did not improve their behaviors by receiving red faces or they did not care about receiving bad ratings. We also asked students' reactions to the use of the Daily Smirk in future team-based courses and 74% of them were willing to use it in the future. Moreover, in order to the usability of the Daily Smirk interface, the survey offered a 10 Likert-scale, ranging from 10 (fantastic) to 1 (terrible). The results are shown in Table III: the majority of students are satisfied with its usability.

Scale	Percentage of Responses
10 (fantastic)	17.95%
9	10.26%
8	28.21%
7	20.51%
6	10.26%
5	2.56%
4	5.13%
3	5.13%
2	0.00%
1 (terrible)	0.00%

TABLE III: Results of the usability of the Daily Smirk interface on a scale of 10 to 1.

The post-survey studied the students' reactions to the assessment results as well. Specifically, it asked students about their feelings towards receiving green faces or red faces, and whether they changed any of their behaviors moving forward with the project. According to the survey results, receiving dark or light green faces made them feel happy and grateful, and 77.5% of them changed their behaviors to some extent after receiving good ratings. Students took the initiative to participate and contribute more to the teamwork, and some of them wanted to get more green faces so they worked harder. The others did not change their behaviors because they collaborated well with team members and contributed consistently without the threat of undesirable behaviors.

Conversely, students started to self-reflect or felt angry after receiving dark red or light red faces. 73.33% altered their behaviors to some extent, such as showing up at meetings, communicating more, and starting to contribute, but the remaining students never changed or improved after receiving bad ratings. It is interesting that 9.53% of students never gave their peers green faces, and 42.86% never selected red faces for their peers, so students were still apprehensive about giving bad ratings.

Finally, the post-survey also sought to identify opportunities for future improvements. Currently, the Daily Smirk provides five faces, ranging from dark green red to dark red, for students to choose from, and we wanted to explore alternatives in terms of more or fewer choices. 69.23% of students preferred to keep the faces as they are now. As discussed in the section of the initial prototype, the system will only ask for further information on the selections of a dark green or dark red face, so a question was asked if students want to leave comments for any of the other three faces (light green, yellow, and light red). 39.66% of them wanted to keep it as it is now, but more students preferred to be given the optional comments for the other three faces. Dark green faces or dark red faces were used for relatively extreme cases, but sometimes students wanted to send a kind of warning to their peers. They want to explain why they are slightly satisfied or dissatisfied with their peers' work and allow their peers to adjust their performance accordingly. In addition, the survey also asked students which features of the Daily Smirk need to be improved and if there is any other feature they want to add. Students want to add ratings from the past days if they forgot to assess their peers, as well as edit ratings if they want to adjust their ratings or comments. Students also want to customize their profiles, such as adding preferred names and pronouns and providing more contact information. All these results help us in preparation for the second version of the Daily Smirk.

V. Conclusion

Many software engineering and other computing courses provide students with opportunities to work and learn as a team in order to better prepare them for their future careers. Many educational researchers have advocated the use of peer assessment in team-based projects to address collaborative issues that arise within student teams as well as to develop students' lifelong skills. In this paper, we introduce a novel smiley-based peer assessment tool, Daily Smirk, which strives to tackle some limitations of existing peer assessment approaches and platforms, including the heavyweight workload to complete assessing peers, lack of experience in providing helpful feedback, and inflation of assessment results. The Daily Smirk is designed for conducting high-frequency, lightweight formative peer assessment that enables students to provide and receive continuous and timely feedback, as well as instructors to inspect assessment results and monitor ongoing teamwork effectively.

The preliminary evaluation of the Daily Smirk was conducted in a software design course with team-based projects. The results collected from the survey are encouraging in terms of students' experience with using the Daily Smirk to assess their peers including the involvement, workload, learning experience, usability, and overall experience. The primary limitation of this study is the fact that the Daily Smirk was evaluated in one course, which is a relatively insufficient sample size for statistical measurements. In the future, we plan to evaluate the impact of the Daily Smirk in other software engineering and computing courses at our university and other universities, in order to obtain a more comprehensive study of peer assessment in team-based projects.

Second, we recognize that the instructor dashboard is limited in providing merely on overview of rating results collected from the student side of the Daily Smirk. In future work, we aim to develop a more analytical dashboard to provide instructors with the summary of ratings from both a team and

an individual perspective. We also aim to evaluate the Daily Smirk from the instructors' points of view, such as how the Daily Smirk assists them in identifying dysfunctional teams or whether it allows them to intervene early and effectively when issues arise within teams.

Finally, our results also show the need to make small improvements to the tool, including adding or editing previous ratings with a reasonable explanation, optional comments on the other three faces (light green, yellow, and light red), and providing an improved calendar visualization of how users rate their peers and how they get reviewed.

REFERENCES

[1] J. Hayes, T. Lethbridge, and D. Port, "Evaluating individual contribution toward group software engineering projects," in *25th International Conference on Software Engineering, 2003. Proceedings.*, pp. 622–627, May 2003. ISSN: 0270-5257.

[2] E. Hammar Chiriac, "Group work as an incentive for learning – students' experiences of group work," *Frontiers in Psychology*, vol. 5, p. 558, June 2014.

[3] S. Isaac and R. Tormey, "Undergraduate group projects: Challenges and learning experiences," *QScience Proceedings*, vol. 2015, p. 19, June 2015.

[4] K. S. Double, J. A. McGrane, and T. N. Hopfenbeck, "The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies," *Educational Psychology Review*, vol. 32, pp. 481–509, June 2020.

[5] T. Wanner and E. Palmer, "Formative self-and peer assessment for improved student learning: the crucial factors of design, teacher participation and feedback," *Assessment & Evaluation in Higher Education*, vol. 43, pp. 1032–1047, Oct. 2018. Publisher: Routledge _eprint: https://doi.org/10.1080/02602938.2018.1427698.

[6] T. Tenório, I. Bittencourt, S. Isotani, and A. Silva, "Does peer assessment in on-line learning environments work? A systematic review of the literature," *Computers in Human Behavior*, vol. 64, pp. 94–107, July 2016.

[7] 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*, vol. 21, p. 426, Aug. 2021.

[8] H. Li, Y. Xiong, C. V. Hunter, X. Guo, and R. Tywoniw, "Does peer assessment promote student learning? A meta-analysis," *Assessment & Evaluation in Higher Education*, vol. 45, pp. 193–211, Feb. 2020. Publisher: Routledge _eprint: https://doi.org/10.1080/02602938.2019.1620679.

[9] C. M. Brooks and J. L. Ammons, "Free Riding in Group Projects and the Effects of Timing, Frequency, and Specificity of Criteria in Peer Assessments," *Journal of Education for Business*, vol. 78, pp. 268–272, May 2003. Publisher: Routledge _eprint: https://doi.org/10.1080/08832320309598613.

[10] A. Planas, L. Soley, R. M. Fraguell Sansbelló, G. Arbat, J. Pujol, N. Roura, J. Suñol, and L. Montoro, "Student perceptions of peer assessment: An interdisciplinary study," *Assessment & Evaluation in Higher Education*, pp. 1–19, Oct. 2013.

[11] 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*, vol. 5, pp. 432:1–432:20, Oct. 2021.

[12] R. S. Hansen, "Benefits and Problems With Student Teams: Suggestions for Improving Team Projects," *Journal of Education for Business*, vol. 82, pp. 11–19, Sept. 2006. Publisher: Routledge _eprint: https://doi.org/10.3200/JOEB.82.1.11-19.

[13] I. Čavrak, M. Orlić, and I. Crnković, "Collaboration patterns in distributed software development projects," in *2012 34th International Conference on Software Engineering (ICSE)*, pp. 1235–1244, June 2012. ISSN: 1558-1225.

[14] P. Aggarwal and C. L. O'Brien, "Social Loafing on Group Projects: Structural Antecedents and Effect on Student Satisfaction," *Journal of Marketing Education*, vol. 30, pp. 255–264, Dec. 2008. Publisher: SAGE Publications Inc.

[15] C. L. Colbeck, S. E. Campbell, and S. A. Bjorklund, "Grouping in the Dark: What College Students Learn from Group Projects," *The Journal of Higher Education*, vol. 71, no. 1, pp. 60–83, 2000. Publisher: Ohio State University Press.

[16] H. Shema, "The Birth of Modern Peer Review." https://blogs.scientificamerican.com/information-culture/the-birth-of-modern-peer-review/, Apr. 2014.

[17] K. J. Topping, "Peer Assessment," *Theory Into Practice*, vol. 48, pp. 20–27, Jan. 2009. Publisher: Routledge _eprint: https://doi.org/10.1080/00405840802577569.

[18] J.-W. Strijbos, T. A. Ochoa, D. M. A. Sluijsmans, M. S. R. Segers, and H. H. Tillema, "Fostering Interactivity through Formative Peer Assessment in (Web-Based) Collaborative Learning Environments," 2009. ISBN: 9781605663920 Pages: 375-395 Publisher: IGI Global.

[19] 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*, vol. 4, Oct. 2005. Number: 1.

[20] S. Lindblom-ylänne, H. Pihlajamäki, and T. Kotkas, "Self-, peer- and teacher-assessment of student essays," *Active Learning in Higher Education*, vol. 7, pp. 51–62, Mar. 2006. Publisher: SAGE Publications.

[21] B. Friedman, P. Cox, and L. Maher, "An Expectancy Theory Motivation Approach to Peer Assessment," *Journal of Management Education*, vol. 32, pp. 580–612, Oct. 2008.

[22] L. Li, X. Liu, and A. L. Steckelberg, "Assessor or assessee: How student learning improves by giving and receiving peer feedback," *British Journal of Educational Technology*, vol. 41, no. 3, pp. 525–536, 2010. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8535.2009.00968.x.

[23] M. Freeman and J. Mckenzie, "SPARK, a confidential web-based template for self and peer assessment of student teamwork: Benefits of evaluating across different subjects," *British Journal of Educational Technology*, vol. 33, pp. 551–569, Dec. 2002.

[24] 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*, vol. 36, pp. 5–19, Apr. 2014. Publisher: SAGE Publications Inc.

[25] D. Bacon, K. Stewart, and W. Silver, "Lessons from the Best and Worst Student Team Experiences: How a Teacher can make the Difference," *Journal of Management Education*, vol. 23, Oct. 1999.

[26] E. B. Nuhfer, "The Place of Formative Evaluations in Assessment and Ways to Reap Their Benefits," *Journal of Geoscience Education*, vol. 44, pp. 385–394, Sept. 1996. Publisher: Routledge _eprint: https://doi.org/10.5408/1089-9995-44.4.385.

[27] C.-h. Chen, "The implementation and evaluation of a mobile self- and peer-assessment system," *Computers & Education*, vol. 55, pp. 229–236, Aug. 2010.

[28] J. R. Beatty, R. W. Haas, and D. Sciglimpaglia, "Using Peer Evaluations to Assess Individual Performances in Group Class Projects," *Journal of Marketing Education*, vol. 18, pp. 17–27, Aug. 1996. Publisher: SAGE Publications Inc.

[29] V. Farrell, G. Ravalli, G. Farrell, P. Kindler, and D. Hall, "Capstone project: fair, just and accountable assessment," in *Proceedings of the 17th ACM annual conference on Innovation and technology in computer science education*, ITiCSE '12, (New York, NY, USA), pp. 168–173, Association for Computing Machinery, July 2012.

[30] 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*, vol. 9, pp. 652–662, Dec. 2010. Publisher: Academy of Management.

[31] T. Rotsaert, E. Panadero, and T. Schellens, "Anonymity as an instructional scaffold in peer assessment: its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills," *European Journal of Psychology of Education*, vol. 33, pp. 75–99, Jan. 2018.