COMMIT: A NEW WAY TO STUDY

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Elliot Fiske

June 2017

COMMITTEE MEMBERSHIP

TITLE:	Commit: A new way to study
AUTHOR:	Elliot Fiske
DATE SUBMITTED:	June 2017
COMMITTEE CHAIR:	Foaad Khosmood, Ph.D. Professor of Computer Science
COMMITTEE MEMBER:	Zachary Peterson, Ph.D. Professor of Computer Science
COMMITTEE MEMBER:	Hugh Smith, Ph.D. Professor of Computer Science

ABSTRACT

Commit: A new way to study

Elliot Fiske

Computer-assisted learning is older than Turing machines, and constantly evolves as technology improves. While some teachers are resistant to using technology in the classroom [22], ”e-learning” techniques are becoming more common in almost every school, from K-12 to universities [28]. As technology becomes more widespread, it becomes crucial to examine the various methodologies of computer-assisted learning and find the techniques that are most effective.

This paper explores the effectiveness of one such methodology, spaced repetition. This technique applies to homework assignments available to students online. We include an exploration of several existing apps that use this technique, and introduce our own novel app, *Polycommit. Polycommit* was developed for use with several Cal Poly classes and was deployed during the first half of the Spring 2017 quarter. With careful attention to user feedback, we created a tool that motivated students to form better study habits. While our results do not show statistically significant improvement to student grades, this tool gives insight into how modern technology and gamification can be leveraged to create an engaging app that encourages positive study habits.

ACKNOWLEDGMENTS

Thanks to:

- Andrew Guenther, for uploading this template

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Learning Technology

Learning technology is becoming nearly universally used in classrooms, to the point where some teachers refer to it as "mandatory." Yet there is still a "great variation in university lecturers' use of technology," with some teachers not even ready to replace their overhead slides with Powerpoint [22].

We will address the different methodologies teachers use to integrate technology with the classroom, paying special attention to gamification techniques and spaced repetition.

1.2 Definitions

1.2.1 Spaced Repetition

Spaced repetition is a learning technique first coined by Dr. Paul Pimsleur in 1967 [18]. Dr. Pimsleur noted the existence of a "forgetting curve." (See **Figure 1.1**) The first moment a piece of information is learned, the subject is 100% likely to be able to correctly recall it. However, as early as the first day after initially learning the information, the subject's retention rate of the information begins rapidly decaying. By reviewing the information, the retention rate is brought back to 100%, and the subsequent decay of retention is substantially *flattened*. That is, as students review information over time, it is encoded into their long-term memory, and they are less likely to forget it over a long period of time.

Dr. Pimsleur initially proposed spaced repetition as a technique solely for learning

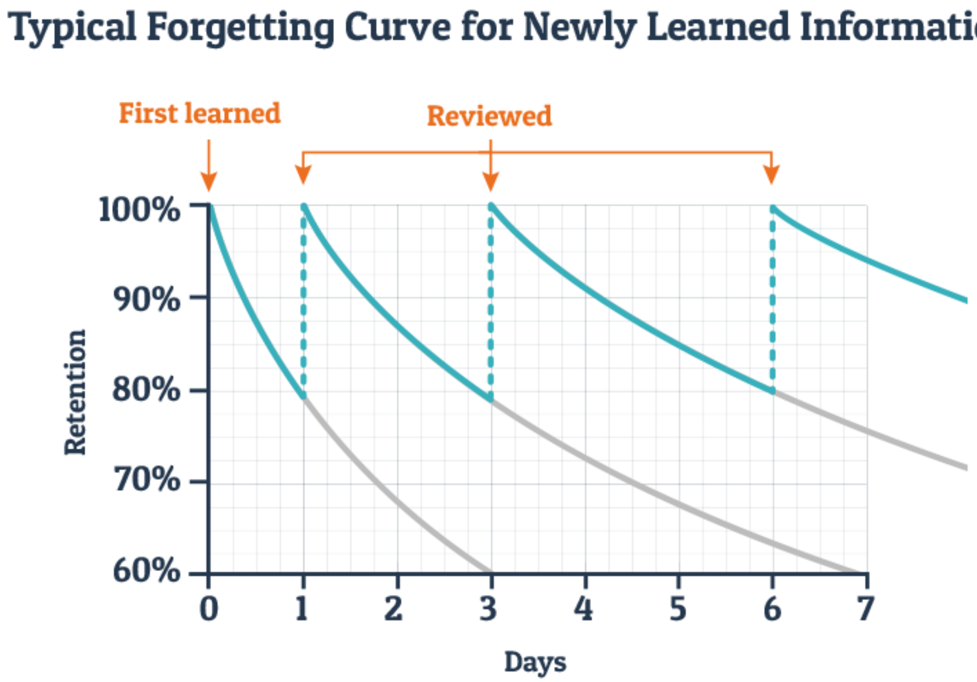


Figure 1.1: The "Forgetting Curve," as presented by Dr. Pimsleur in 1967. The y-axis represents the subject's likelihood of recalling a specific piece of information at the given time.

a new foreign language, but spaced repetition has been proven to be effective for any classroom setting. As Kang notes in [8], it is intuitive that information needs to be practiced and reviewed over time to be fully integrated into a student's understanding, but Dr. Pimsleur's techniques provide a quantifiable, scientific schedule to maximize retention.

Kang et al. [8] also attempts to find the "Optimal Spacing Lag," or the optimal time between each review session for a particular piece of information after a long period. They find that this optimal spacing lag is around 10% to 20% of the time between initial introduction of the information and the test date. That is, if the test is 10 weeks away, even the oldest information should be reviewed with at most a 1-2 week delay in between review sessions.

A simple example of spaced repetition is the Leitner system. Flashcards corre-

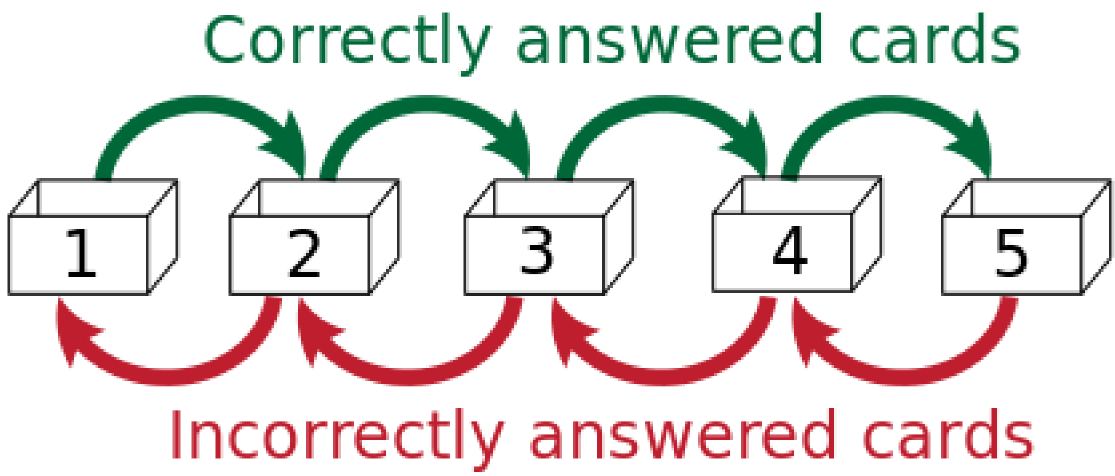


Figure 1.2: The Leitner system. If a student answers a flashcard correctly, it is moved to a higher-numbered box that is reviewed less frequently.

sponding to the desired source material are organized into numbered boxes. (See **Figure 1.2**). Each successive box is reviewed less frequently. That is, a student would review Box 1 twice per day, Box 2 once per day, Box 3 every 2 days, and so on. If a card is answered correctly, it is moved to a box that is reviewed less frequently. However, if the student answers incorrectly, the card is moved to a box that is reviewed more frequently.

Thus, tougher cards are reviewed more often, and cards the student knows will be reviewed less often. However, all cards are *eventually* reviewed, and even cards that the student always answers correctly are reviewed in the long term.

1.2.2 Gamification

The term ”gamification” first appeared in 2008 , but the concept wasn't known in the mainstream until 2011 with the advent of Foursquare [4]. Foursquare is an app that takes the mundane concept of reviewing restaurants and finding new places to visit and turns it into an engaging, popular experience. Users of the app earn points for "checking in" at locations,

Chapter 2

BACKGROUND

2.1 Early Computing

2.1.1 Pressey and Skinner

Learning applications have been used since the early days of computing. In 1958, BF Skinner [23] explained how computers could be used to facilitate learning. When Dr. Skinner wrote his paper, computers were used largely simply to test students about simple pieces of information. However, Dr. Skinner believed that computers could actually be used as teaching devices in and of themselves. Skinner argued that the main advantages of computers as teachers were that students are encouraged ”to take an active role in the educational process,” and notes that the practice of using automation dates back even to the 1920's with Sidney L. Pressey's automated testing machines. One huge advantage that Pressey noted with automated learning is that students learn at their own pace; one of the toughest challenges teachers face, according to Pressey, is that some students will fall behind while others feel bored because they already understand most of the material being taught. While Pressey's automated teaching methods eventually succumbed to the fact that technology at the time wasn't up to the task of automating education. Instead, Skinner was able to use the technology of his time to more effectively bring computers to the field of education.

2.1.2 Skinner Boxes and Operant Conditioning

Of course, BF Skinner is famous for his creation of the "Skinner Box," [24] a device that gave simple reinforcement to animals or humans in the lab in order to slowly shape their behavior, through a process known as operant conditioning. When a subject performed well according to the test, they would receive a quick reward, while if they strayed from the purpose of the test, they would not receive a reward.

2.1.3 Variable Ratio Reinforcement Scheduling

This method is combined with other interesting motivational techniques such as variable rewards scheduling [7] [9], where rewards are not actually doled out on a regular basis exactly according to the behavior of the test subject. Instead, rewards are handed out semi-randomly, varying in quantity and quality, while still mostly rewarding the desired behavior that the researcher is trying to condition. This causes the test subject to become conditioned much faster (See **Figure 2.1**).

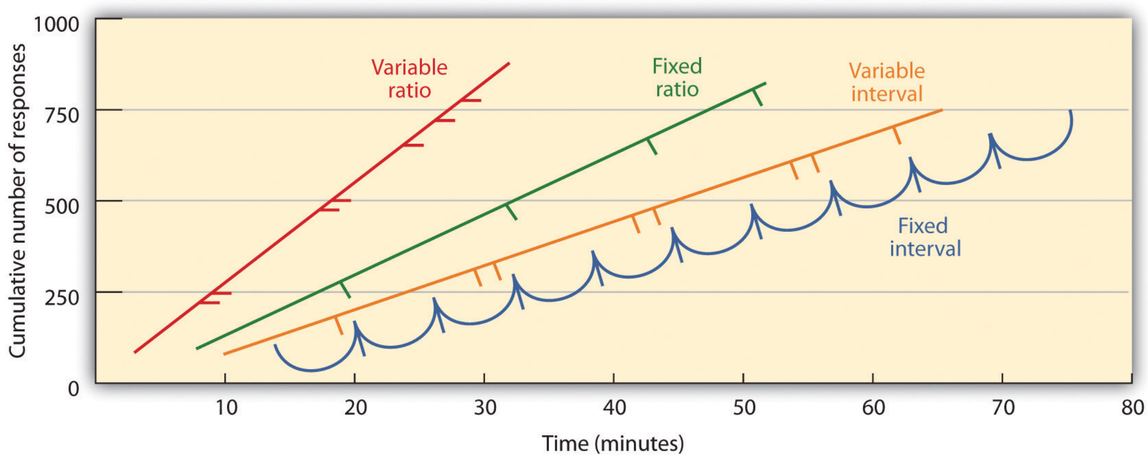


Figure 2.1: Variable ratio rewards scheduling. Note that as rewards are spaced out randomly, the desired behavior appears much more quickly.

[9]

2.2 Duolingo

As mentioned previously in the introduction, Duolingo is an app that already uses several of the processes that we are describing. Duolingo has been shown to be very effective for learning new languages, even perhaps more effective than typical classes. However, the effectiveness of Duolingo is mostly contingent on the motivation of the student. If the student is going to a foreign country soon, Duolingo is the most effective since there is a significant time pressure and pressure to learn the language in order to fit in at whatever country the student is planning to go to [25]. If, however, the student is simply learning another language for fun, there will be significantly less benefit to them. We must consider this as we develop our educational software; apparently the nature of the student's motivation and their reason for wanting to learn the subject material plays heavily into wheter or not they will be successful in using the app.

2.3 Habits

The study on Duolingo notes how difficult it is to actually get people to use the app. It is extremely hard to change people's habits, taking a lot of time and effort on the user's part to effectively change their behavior. One such strategy around this is to connect one desired "habit" behavior to an existing habit [13]. For instance, test subjects in the study by Dr. Lally determined that it was easiest to condition subjects to perform some action by instructing them to perform it right after eating breakfast in the morning. By using the powerful force behind an existing habit, it is possible to "reprogram" your behavior such that a new habit is formed with the new desired behavior.

For instance, if a subject wants to make a habit of cleaning their room, they can make a habit of picking up one piece of clothing after coming home from work for the day. In this way, the habit is driven forward by the regular schedule of coming home at a regular time each day. By associating one behavior with an existing behavior, it is possible to rewire a subject's brain to complete the second behavior far more often.

2.3.1 New Psychological Theories

It's interesting to note that the advent of "apps" and habit-forming application has created an essentially new field in psychology [20]. In a study, Dr. Renfree argues that the behavior modifications seen in apps like Lift and Memrise actually represent new advances into how habit-forming psychology operates.

Unfortunately, most of these new effects are actually negative. For instance, Dr. Renfree notes that oftentimes the new habits generated by these apps are "fragile," since they are so dependent on a tight dopamine-based reinforcement loop. As soon as the loop is broken, the brain "loses interest" and the new information is deprioritized. This will be an interesting consideration as we develop our application. We want to ensure that the habits and knowledge developed by our app is not simply impermanent, and that it won't be simply "pushed out" by new information.

Chapter 3

RELATED WORK

3.1 Spaced Repitition Apps

3.1.1 Duolingo

Several recent apps and products make use of spaced repetition to allow user to easily gain long-term recall of languages, class content, or any other information that needs to be learned. One such app is Duolingo (See **Figure 3.1**); in Duolingo, users learn a language by repeating small exercises every day.

The app encourages users to spend a small amount of time each day studying useful words and phrases, rather than cramming in a lot of knowledge at once. It encourages this behavior through the use of gamification. Each user earns "experience" in a language, eventually leveling up. Users connect their Facebook accounts and can see their friends' levels and accomplishments, adding a social element to the app.

3.1.2 Memrise

Memrise (See **Figure 3.2**) is a web app that has very similar function to *Commit.* Memrise takes a series of flashcard-based questions and answers and automatically creates a "study plan" where the application breaks up flash cards and uses spaced repitition to encode the information in the user's long-term memory. Memrise also uses gamification and social elements, as users earn points for every correct answer and can see their friends' scores.

Interesting to note is that users can input any data they choose into Memrise to receive a custom study-guide. This would allow students to easily learn flashcards if

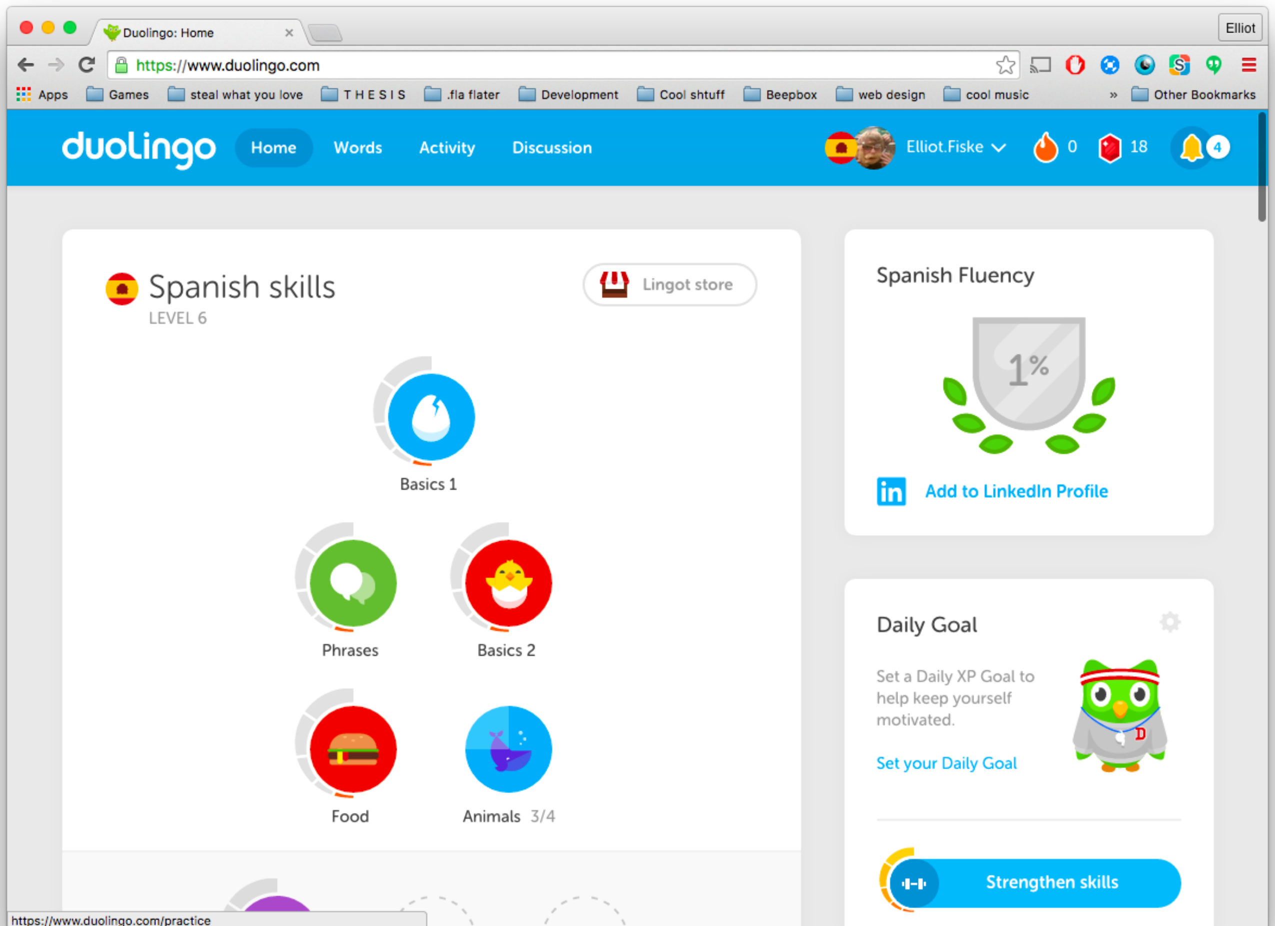


Figure 3.1: The Duolingo interface. Notice the gamification elements and the encouragement to reach a "daily goal."

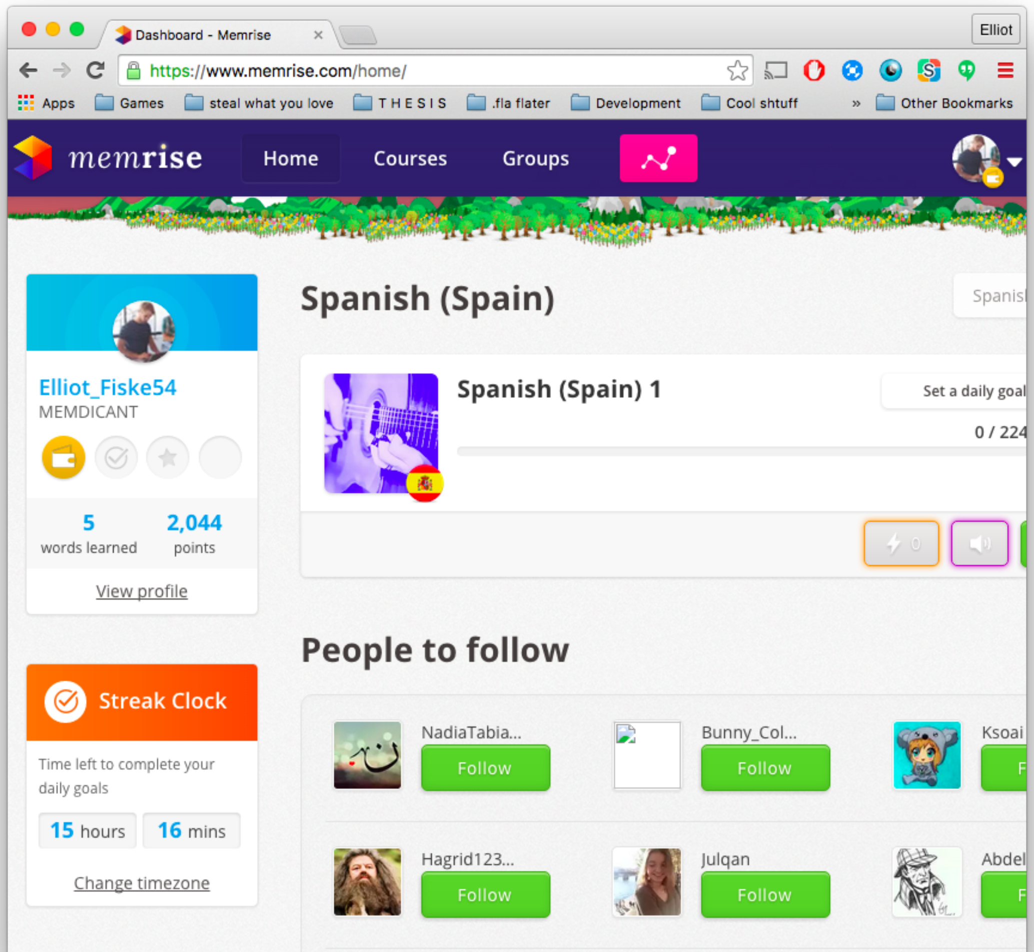


Figure 3.2: Screenshot of the Memrise interface. Note the gamification elements, the social aspect, and the "Streak" clock encouraging consistent use of the app.

they took the time to input them into the app.

Both of these applications heavily emphasize language learning, since the process of learning a language can easily be broken down into a series of small words and phrases, and re-emphasized using the process of spaced repetition. However, Commit is scoped specifically to one class, allowing students to easily learn and retain class content without the commitment of adding their own flashcards.

3.2 Habit Formation

3.2.1 Pocket Points

Pocket Points is an app that attempts to create positive habits for its users through gamification [2]. When installed, the app rewards you for keeping your phone locked while in class. It uses GPS to ensure you're on campus, then awards 1 point for every 20 minutes your phone is locked. These points can then be spent to earn discounts and coupons at local businesses.

Pocket Points has the benefit of providing a TODO: FINISH THIS SECTION ON POCKET POINTS

3.2.2 Habit Formation and User Interface

Previous work has been done to tie habit formation with online user interfaces [11]. In this book, Dr. Jeff Johnson describes how psychology ties into user interface design. He cites that familiar user interfaces lead to less mental stress, encouraging users to come back to your application repeatedly since it seems familiar and less stressful to them. We desire users to consistently use our application every day, so we must keep this in mind.

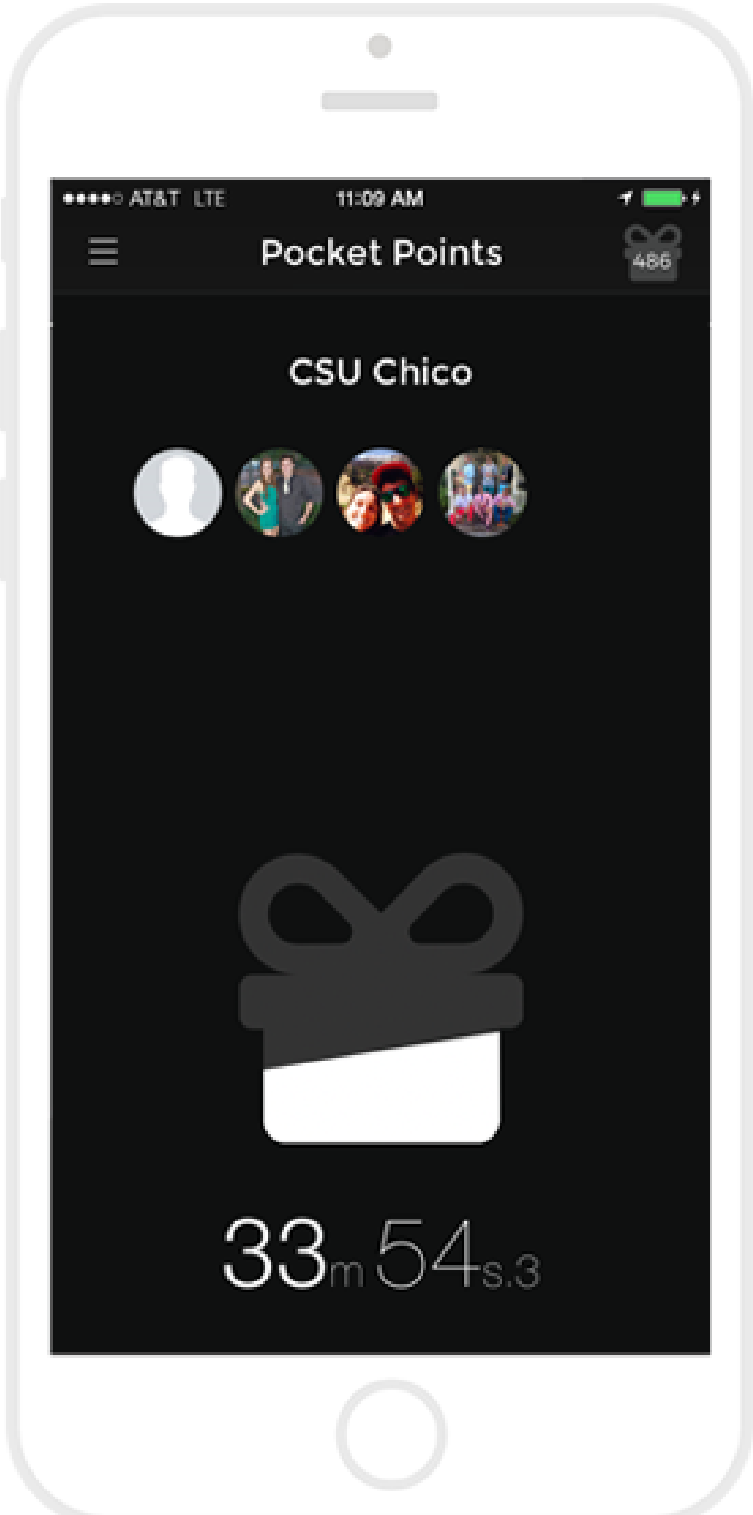


Figure 3.3: Screenshot of the Pocket Points interface.

3.2.3 Memrise

As mentioned before, Memrise is an excellent example of habit formation. In one paper, researchers gauged the ability of students to memorise simple Latin phrases in a classroom setting. They found that Memrise is an effective tool

3.3 Gamification

Gamification in the classroom has several other examples that can be compared to our application.

3.3.1 Classcraft

Classcraft is an intriguing example of gamification in education (See **Figure 3.4**). In Classcraft, students choose a "role" and go on the equivalent of a World of Warcraft raid with their fellow classmates. Classcraft is interesting in that it promotes co-operation and a variety of skills, so that students can assist each other where they might not have a certain skill.

Classcraft takes the typical elements of an RPG and converts them into an experience that supplements a typical classroom experience.

Interesting to note is the emphasis Classcraft puts on integration with existing technologies such as Google Classroom. In order for this application to be widely adopted, it is definitely necessary to have the experience of the teacher go extremely smoothly. Thus, if an application ties directly in with existing technology that the teacher is already familiar with, it will be much more readily adopted by the educational community. It's important to value the teacher's time with this application, so it is necessary to make the UI recognizable, familiar and easy to work with, as well as integrating it with existing tools and perhaps even modeling the user interface after

Figure 3.4: Classcraft user interface. Note the student's health and mana pool, as well as the list of other students in the classroom.

[9]

tools that many teachers will already be familiar with.

3.3.2 Classcraft Study

A study was done on the effectiveness of Classcraft as well as "ludicization" in the classroom in general [21]. Their study ran primarily in France, as this is the home country of the Classcraft company and application.

TODO: update classcraft stuff

3.3.3 Gamification in Other Educational Areas

So far, we have considered gamification in the classroom mostly in the context of STEM or language fields. While gamification does lend itself towards these subjects, as math problems and science problems can be easily generated by an algorithm, there have also been steps to gamify aspects of classes around liberal arts and the humanities [26].

This study by Dr. Wagner approaches a music class the same way Classcraft approaches subjects not focused on the humanity. It highlights the idea of a ”flow” where a student is fully immersed in the educational process and is fully engaged with the learning process, and notes that it is far easier to attain this ”flow” state when the student is learning in the context of a video game.

Chapter 4

EXPERIMENT

We carried out an experiment at California Polytechnic State University to test the efficacy of habit-based educational software. We created a web-based application called *Polycommit* that was connected to 4 college classes: **Introduction to Computer Networks**, **Introduction to Computer Graphics**, **Introduction to Computer Networks**, **Introduction to Operating Systems**, and **Linear Analysis I**. These classes were selected because they covered subject matter that was easy to convert to online quizzes. For instance, one staple Linear Analysis problem is to find the determinant of a matrix (often a whole number), which is easy to input into an online form. In addition, I had taken these classes in recent quarters and was familiar with the course content.

We presented *Polycommit* to each of the courses in the first 2 weeks of class. Students voluntarily signed up through a website hosted at https://polycommit.com/.

4.1 UI Overview

4.1.1 Home Scren

Upon logging in, students click "Enroll" for the classes they wish to participate in. Upon enrolling, the classes are listed under the "Enrolled Classes" section (**Figure 4.1**). This page also lists the two main "scores" that students earn by answering questions: **Commitment** and **Points**.

Commitment is a numerical value that represents how many *unique* days a student has answered a question on the website. Students could earn up to 1% extra credit on their final grade in the class by getting 15 Commitment.

Points are earned by answering questions. More points are awarded for correct answers, and bonus points are awarded based on the user's current Commitment. All participants in the experiment were placed in a raffle for $20 Amazon gift cards. Additional entries into the raffle were awarded by earning more points.

4.1.2 Course Screen

Each course page has a list of challenges (**Figure 4.2**) that are open to the student. Challenges are grouped by the week in which they are opened. If a student has completed all the challenges in a week, the week is displayed with a green check mark and does not expand. If there are open challenges in a week, the week is displayed in yellow with an "alert" icon, indicating that the student has an available challenge. This UI imparts a sense of urgency to the user, since they could potentially lose opportunities to earn Commitment by not answering a question in a 24-hour period. Each challenge also lists the date it was opened, and the number of points awarded if the challenge is already completed.

Finally, the Course page lists the student's points and Commitment for the course, along with a tooltip that explains what "Commitment" is. The Commitment and Point values are repeated from the Home page since they are central to the experience of the app, and it is inherently satisfying to watch your points and Commitment rise as you complete challenges.

4.1.3 Challenge Screen

The challenge screen is where students see the content of a challenge and input their answers. A challenge can either be Multiple Choice, Short Answer, or Numerical. An example of a simple Multiple Choice challenge is at **Figure 4.3**.

At the bottom of the Challenge screen is a link to submit feedback about a question. Through this link, students are brought to a Google form where they can provide feedback about a particular question. The feedback form is viewable at **Figure 4.4**.

Once the challenge is closed, students can see a list of their previous attempts along with the correct answer to the problem. Certain problems have their answers hidden; for instance, all Linear Analysis problems don't show their answers because many challenges on *Polycommit* are directly from the assigned homework. Hiding the correct answer prevents students from quickly inputting a wrong answer to see the homework solutions. Additionally, if a question has received feedback as being confusing, an "Explanation" field provides detailed context about the answer the problem and potential pitfalls (See **Figure 4.5**).

4.1.4 Toasts

Polycommit uses "toasts" to convey temporary, state-based information to the user. For instance, when the user answers a question, they get immediate feedback in the form of a pop-up toast. Positive information, such as a correct answer, is styled with a green background and a check mark. Negative information, such as an incorrect answer or a server error, is bright red with an alert symbol that commands the user's attention. Examples of each are at **Figure 4.6**.

4.2 Technology Used

All the code for *Polycommit* is available here: https://github.com/elliotfiske/commitment

Polycommit has its origins in a Summer 2016 section of Dynamic Web Development (CPE 437), taught by Dr. Clint Staley. The base code and overall code architecture remain, as well as some of the platforms used. The app runs on a M*EAN stack. It uses Node.js as a backend, with Express as a routing service and MySQL as a database. In addition, Sequelize is used to easily interface with the database from Javascript. Angular.js is used as a frontend framework, supplemented by Bootstrap for reactive layout.

Node.js and Javascript lends itself well to quick, iterative development. As a dynamically typed language with first-class and anonymous functions, it is easy to quickly make sweeping changes based on user feedback. However, since Javascript is an interpreted language, potential errors that a compiler could have caught may make it through to the live site. Because of this, it was extremely important to regularly run the backend code through comprehensive unit tests.

4.2.1 Test-First Methodology

All of the backend server code is thoroughly covered by various test cases. I used the service *Postman* to maintain a suite of tests that ensured all the data involved in *Polycommit* was both available and secure. *Postman* allows developers to run a series of web requests against a server, and verify that the correct response or error code is returned (See **Figure ??**). For instance, one test suite logs in as a student and attempts to complete a challenge, create a class, and query the information of another student. Only the first request should complete successfully; the student's unprivileged account should not be able to create classes or see other students' data.

Writing test cases for new server functionality *before* beginning development allowed me to see edge cases before they arose, and allowed for the satisfaction of seeing the tests pass as I completed each feature. In addition, it protected the privacy of students' answers and scores.

4.3 User Feedback

Throughout the duration of the experiment, I received a large amount of excellent feedback about the usability of *Polycommit*. At the bottom of each page is a link TODO: Finish section on how I used user feedback to improve the app. Include emails from Polycommit folder.

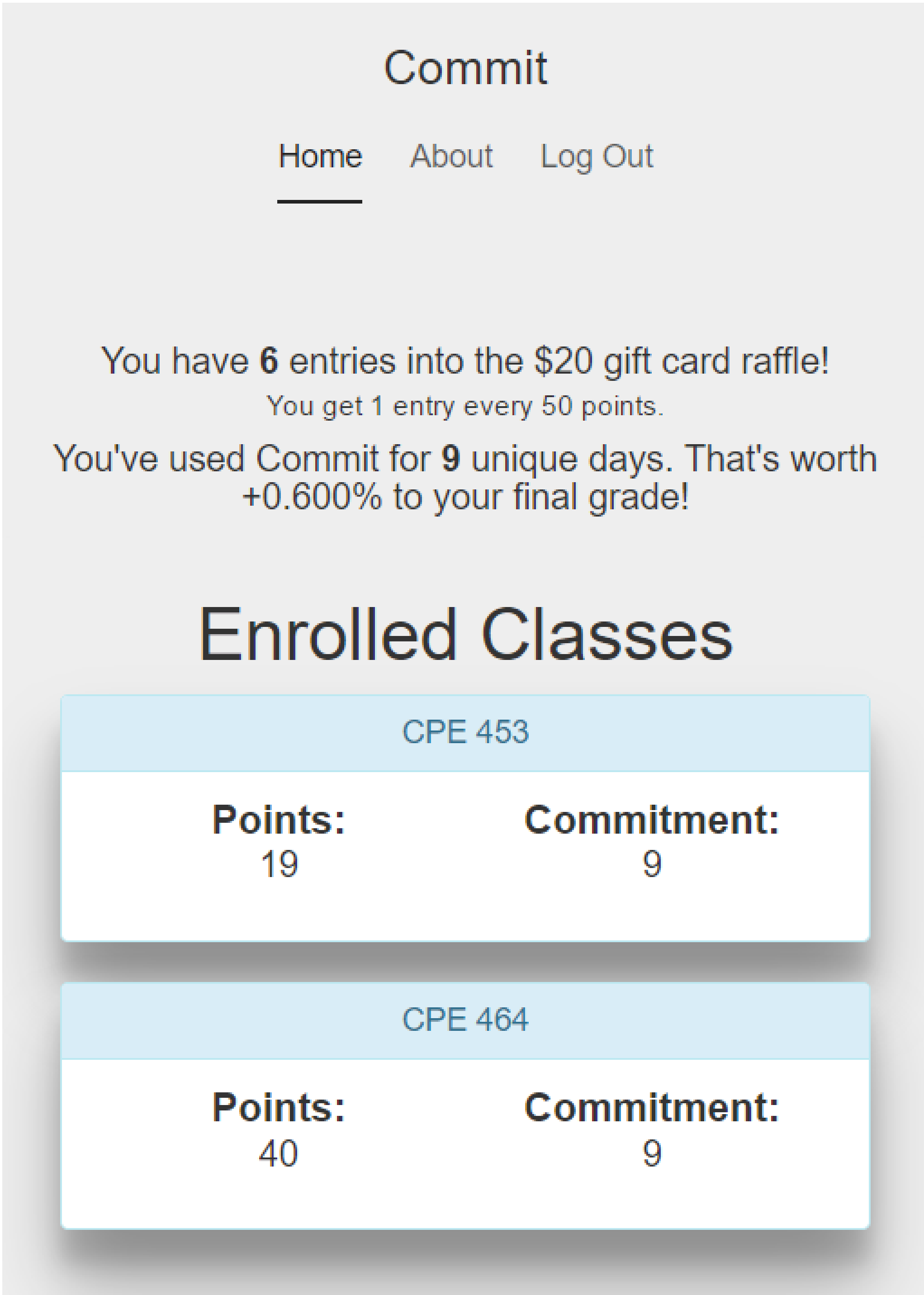


Figure 4.1: The "Home" screen for *Polycommit*. Students can see their current progress and can click on a course to answer challenges.

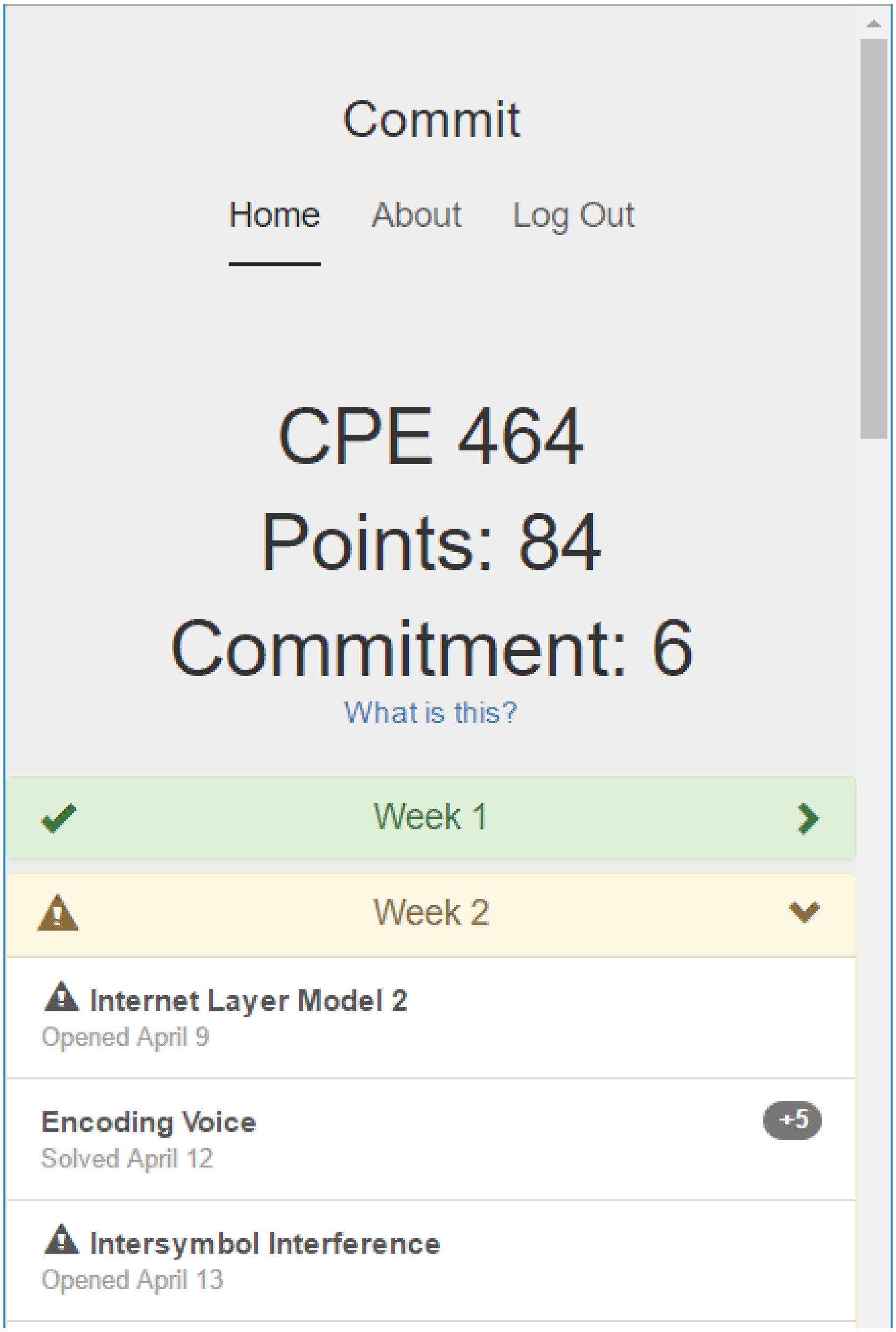


Figure 4.2: The "Course" screen for Polycommit. Students can see their current Points and Commitment and can see a list of challenges organized by date.

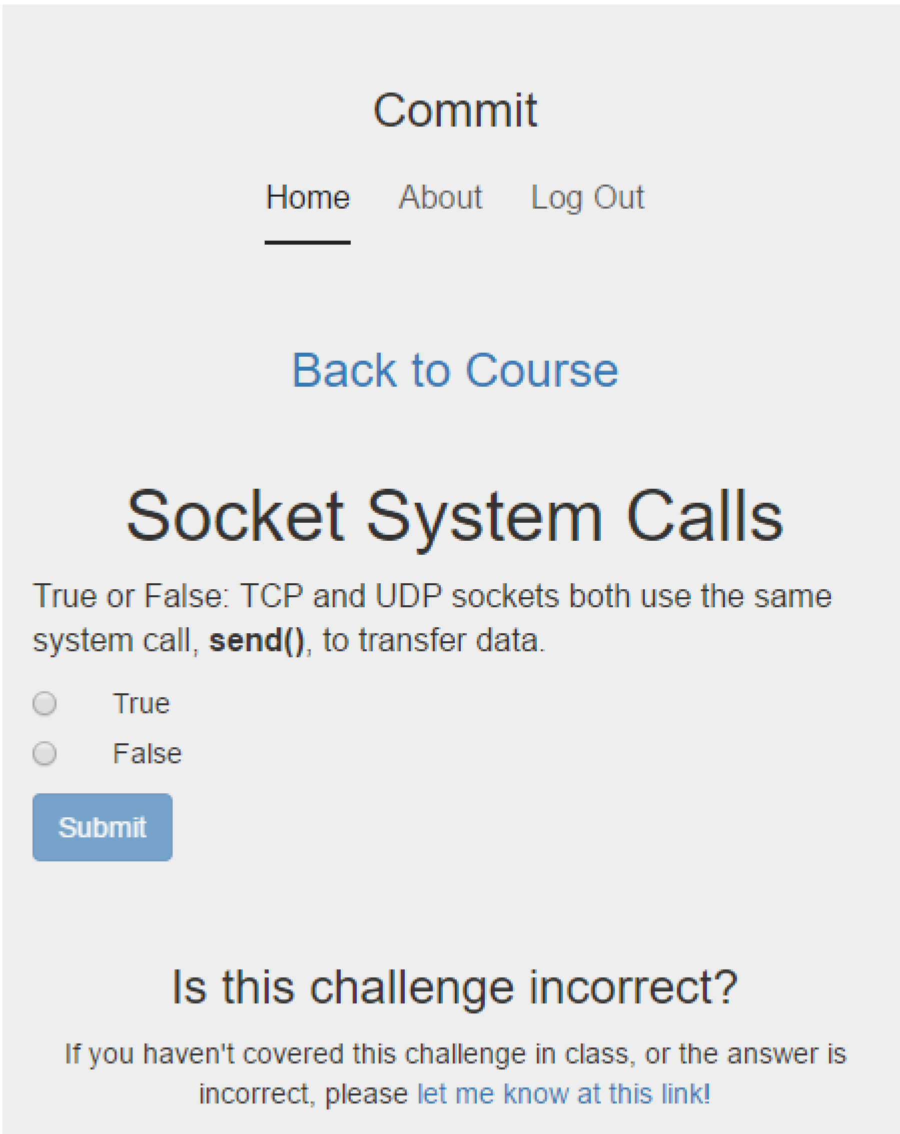


Figure 4.3: The "Challenge" screen for Polycommit. Students see the content of the challenge and can enter their answers. At the bottom is a link where students can give feedback on a challenge.

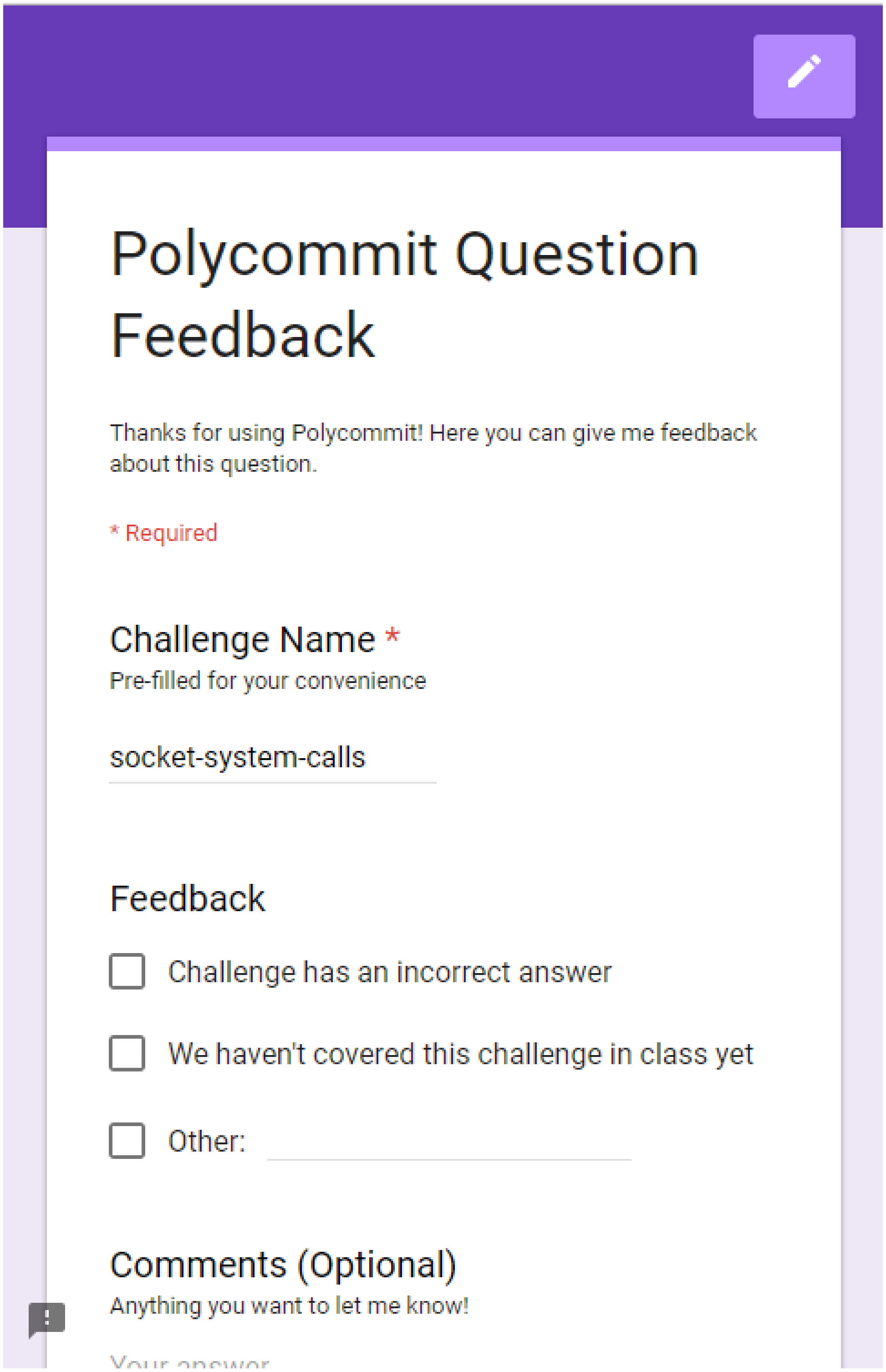


Figure 4.4: The Feedback form for questions. The id for the question is automatically filled in. The student can optionally enter their email address (off-screen) if they want to be contacted when the challenge is fixed.

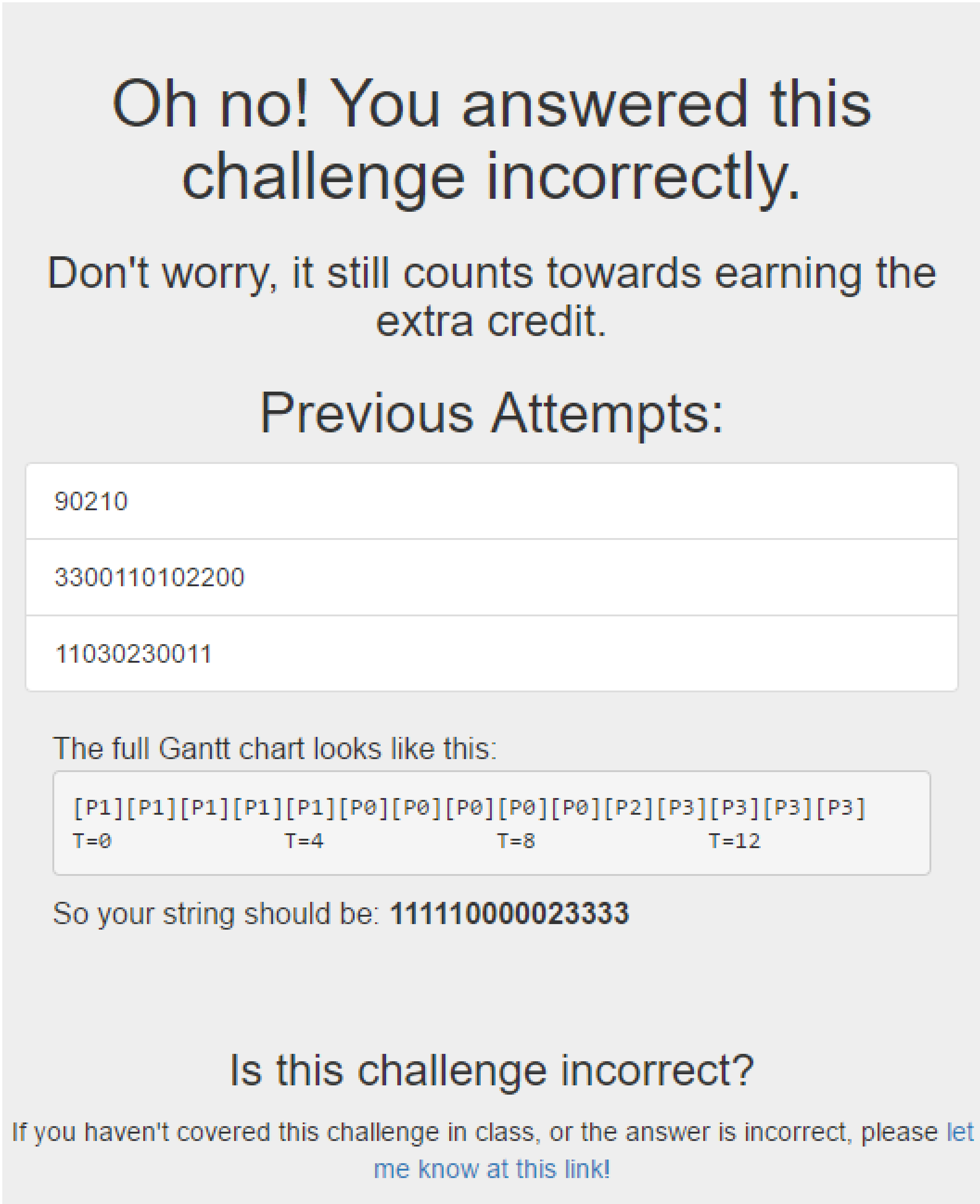


Figure 4.5: The view of a challenge once it has been answered incorrectly. The main message assures the user that it still counts for Commitment. A full explanation of the problem is under the list of the user's attempts.

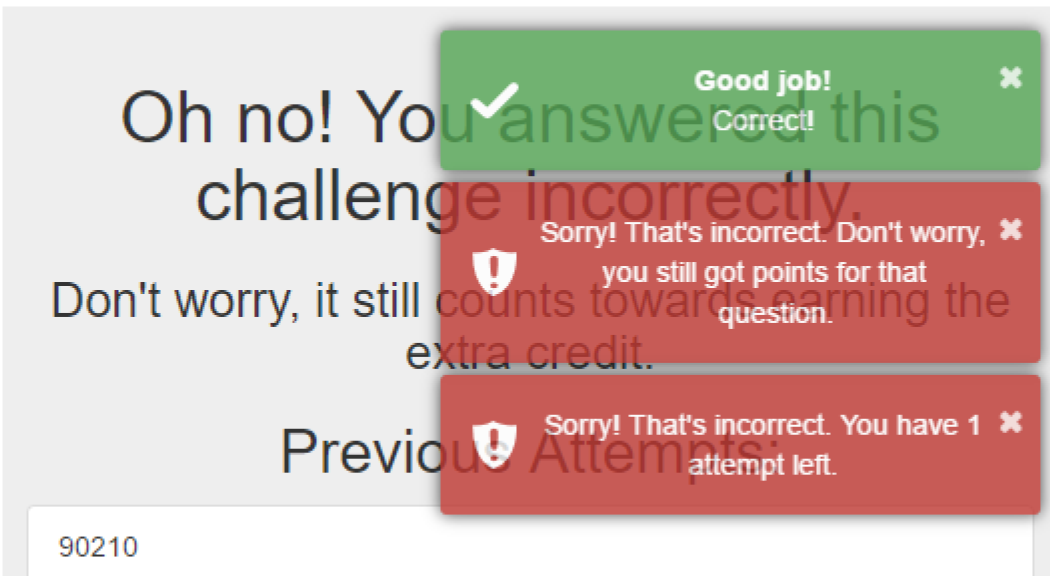


Figure 4.6: Three examples of "toast" notifications. These provide useful state-based information, such as whether a challenge was answered correctly. They also can provide feedback if an error occurred, such as a challenge not submitting due to poor network connectivity.

Chapter 5

VALIDATION

The experiment was carried out over the first half of the Spring 2017 quarter at Cal Poly. 130 students signed up across 4 classes. By the end of the experiment, just over 3200 answers had been submitted to the application.

5.1 Data Gathering

5.1.1 Midterm Scores

In order to gauge the effectiveness of *Polycommit*, we collected scores on quizzes and exams and aggregated the data to be analyzed. For **Introduction to Operating Systems**, data collection was extremely simple as all the quizzes and midterms were administered through an online exam. This allowed us to easily obtain data on how students performed on specific midterm questions that aligned with questions asked in *Polycommit*.

However, the other classes were not as simple. The other classes administered paper midterms, and we did not have access to the papers as they were being graded. As such, we administered a secondary survey to all the students in the class asking them to input their midterm scores for various questions.

5.1.2 Final Survey

After the experiment was concluded, all participants were sent a link to a survey where they could give feedback about their overall experience using *Polycommit*. As an incentive, any participant who filled out the survey received +3 Commitment and +50 points (equal to an additional entry in the gift card raffle). A total of 70 participants filled out the post-experiment survey.

5.1.3 Limitations

Certain tradeoffs were necessary in order to run this experiment.

No Control Group

Most importantly, there is no true control group for this experiment. We decided that it would be more advantageous to offer the application to all students in each section of the classes. This simplified the onboarding process, as we could present the application to all members of the class at once, and allow anyone to sign up. It also prevented potential issues where students might perceive it as unfair that certain students were give access to a study resource and others were not. In addition, this allowed more students to participate in the experiment, giving us access to more data and more user feedback.

As such, it's important to note that the students who participated in the study **self-selected** to become part of the experiment. Students who **did not** use the app can't be considered a control group, because they are not a random selection from the whole population. Allowing students to self-select may have introduced confounding variables. For example, perhaps students who are confident in their study techniques would choose to not participate, making the average midterm scores higher for non-participants.

Self-Reporting Midterm Scores

As noted above, students were asked to input their midterm scores in an online form. We have no way of determining if this data was accurately entered; students may feel tempted to enter a higher score than they actually received. Note that we *do* have access to the overall midterm scores, so the issue only arises when analyzing individual questions on the exams. Data that has been affected by this issue is clearly marked in the upcoming sections.

5.2 Overall Results

5.2.1 Cramming vs. Commitment

First, consider the relationship between a student's overall score on the midterm and their Commitment (**Figure 5.1**).

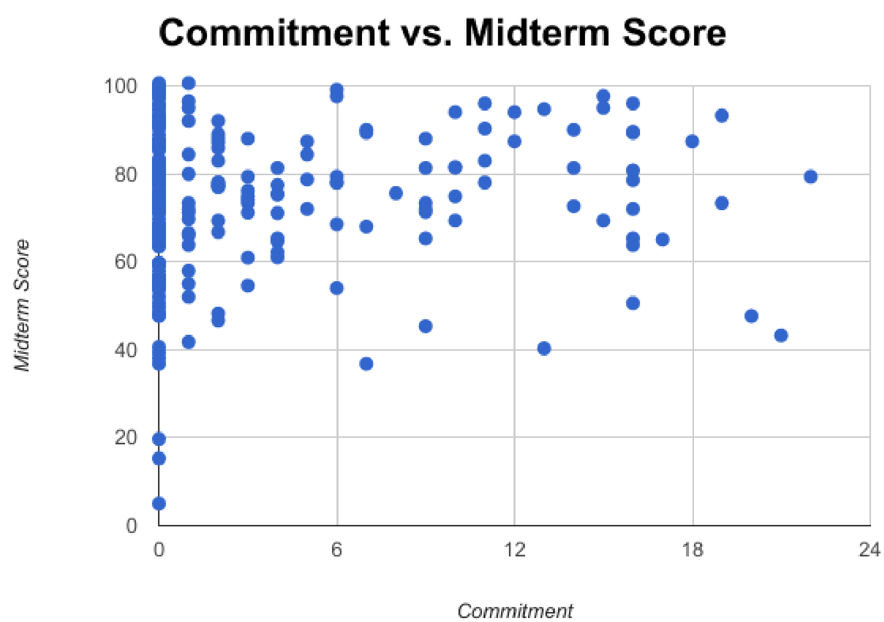


Figure 5.1: Graph of Commitment versus the student's percentage score for all classes.

Note that we are showing the midterm percentage score, since the midterms in different classes have different total score values.

There is no clear correlation between a user's Commitment value and the score they got on the midterm. Thus, this data does not support the hypothesis that using spaced repetition to study for an exam improves the user's score.

Next, we can compare the total number of questions answered by each user with their midterm score (**Figure 5.2**). If students received a benefit from the Commitment system, we might expect to see *less* of a correlation between total questions answered and midterm score. This is because if a user has a high number of questions answered but a low Commitment, this indicates they answered the questions all at once, in a "cram" session. According to our hypothesis, this would not help improve their long-term understanding of the course.

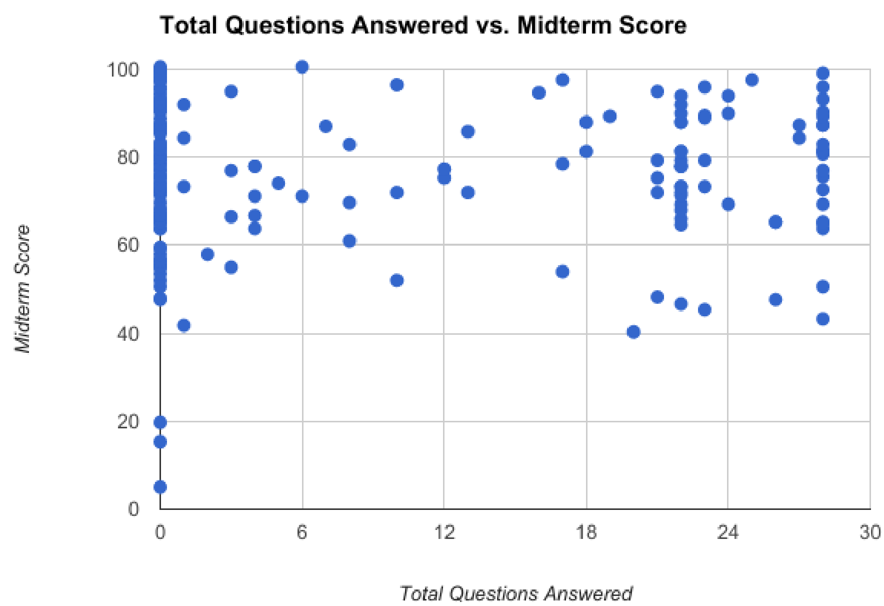


Figure 5.2: Graph of Total Questions Answered vs. the student's score on the midterm.

Next, consider the mean of all midterm scores between students that didn't answer *any* questions vs. the ones that did. For all classes except CPE 471, students that didn't answer any questions in the app scored slightly higher on average than the students that didn't. The discrepancy in CPE 471 is most likely due to the smaller

	Non-users	All Users	Active Users
All Classes			
Midterm Average	72.93	76.12	76.88
Population Size	119	103	87
CPE 453			
Midterm Average	74.38	77.56	77.40
Population Size	14	46	42
CPE 464			
Midterm Average	68.90	73.85	76.08
Population Size	47	42	34
CPE 471			
Midterm Average	87.38	82.39	83.83
Population Size	34	7	4
MATH 244			
Midterm Average	74.43	85.87	85.87
Population Size	26	11	10

Table 5.1: Average midterm scores for different samples of the population. Note that an "Active User" is a user with more than 5 Commitment.

number of students that signed up for *Polycommit*.

Note that we define "Active Users" as any user with over 5 Commitment.

5.2.2 Individual Questions

Certain questions on the midterm matched closely with the content of the questions that were repeated on *Polycommit*. We can analyze midterm results to see if the students that drilled related problems in *Polycommit* performed better on the related questions during the midterm.

Introduction to Operating Systems

One question that was shared between the midterm and *Polycommit* was a question about the 4 conditions for deadlock. The question on the midterm asked students to choose the correct 4 conditions from a series of dropdowns, while the question on *Polycommit* asked students to choose the *incorrect* condition from the list. The 4 conditions for deadlock are:

1. **Mutual exclusion**: At least one resource can only be held by one process at a time; this can result in other processes waiting for that resources
2. **Hold and wait**: A process must be holding at least one resource while waiting for other resources (held by other processes)
3. **No preemption**: Resources can only be released voluntarily by a process; Resources cannot be revoked
4. **Circular wait**: A set of n waiting process $P_0, ..., P_n$ such that P_i is waiting for resources held by $P_{(i+1)\%n}$

The content of the question in *Polycommit* and the question in the midterm are very similar. We can see if individuals who answered the Deadlock question correctly in *Polycommit* had a higher score, on average, than the individuals who did not.

The mean score for **non-users** is 3.30, and the mean score for **users** is 3.31. Thus, using *Polycommit* had a negligible effect on the average score of the participants. However, note that 23/26 **non-users** received a score greater than 3, while 34/37 **users** received a score greater than 3. This could reflect the fact that *Polycommit* only referenced 3 out of the 4 conditions for deadlock, as the 4th condition in the problem was a false plant. Although all 4 conditions are listed in the "explanation" field for the problem, by default if a student answers the question correctly they are

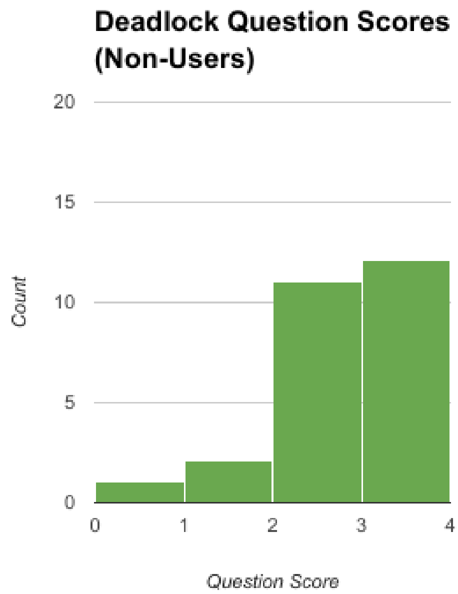


Figure 5.3: Count of scores for students that did *not* make an attempt on the Deadlock question in *Polycommit*.

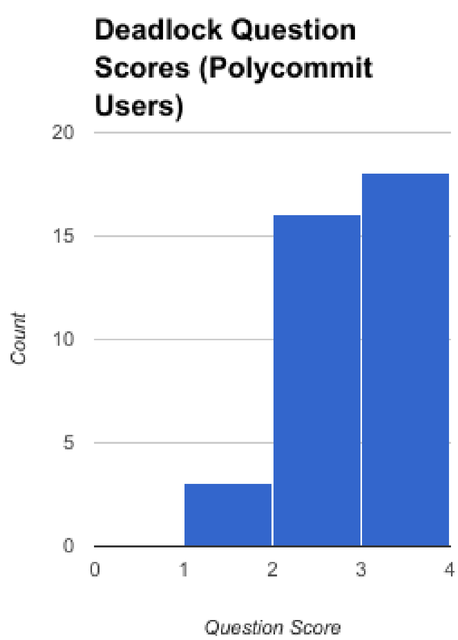


Figure 5.4: Count of scores for students that *did* make an attempt on the Deadlock question in *Polycommit*.

brought back to the course page. Thus, students who participated in *Polycommit* were only drilled on 3 out of the 4 conditions for deadlock, which would explain the overall performance on the midterm.

5.3 User Feedback

At the end of the experiment, we contacted all of the participants with a final survey that asked them questions about their usage of the app. It also included two free-response questions that asked users to offer suggestions about how the app could be improved.

5.3.1 Rating Summary

The full survey is viewable in **Figure .1**. Users responded to several statements about the app using a standard Likert scale ranging from Strongly Disagree to Strongly Agree. A summary of the user responses is available at **Figure 5.5**

5.3.2 User Self-Evaluation vs. Actual Scores

One interesting result comes by grouping midterm scores by the user's response to "Polycommit improved my scores on quizzes and exams" (See **Figure 5.6**). Students who reported that Polycommit did *not* improve their scores scored substantially higher than students who did. One explanation for this is that students who were already confident about their studying habits or had prior knowledge of Operating Systems did not feel that Polycommit had an impact on their grades, which are already high overall.

5.3.3 Free Response Summary

Before viewing the results, we predicted some categories that user responses might fall into. These categories are:

1. Problems or confusions with the user interface
2. Questions were not correct or were confusing
3. Questions did not represent test material
4. App needed more gamification, excitement, or incentive to use
5. Other

The total number of responses that fell under each category are:

Category	**# Responses**
Problems or confusions with the user interface	26
Questions were not correct or were confusing	9
Questions did not represent test material	5
App needed more gamification, excitement, or incentive to use	1
Other	29

The full list of responses is available below. Note that answers *are* shown when a question is closed, so feedback asking for answers to be displayed is filed under "UI Issues" since the problem is essentially that the app failed to present the information in a way that users could find it.

CATEGORY	RESPONSES
User Interface Problems	- There are some UI problems that could be fixed. Certain values like points and commitment weren't consistent from page to page. - Solving the can't see answers even if actually trying was kind of frustrating - I think this might be more useful if there was an app version with a notification each day. I wanted to keep up with Polycommit, but it's pretty easy to forget to check in from day to day. - Polycommit had one question a day, but the study guide had 70 questions on it. As a result, it wasn't as helpful as I expected it to be because by the time of the exam, I feel like a large portion of the class material didn't show up as a question on polycommit. - Ability to review questions would be nice - Also, explanations for answers would be helpful - Ability to review the answers to questions After missing all the allowed attempts, it would be nice to know what the answer is so we have a solution to work backwards from. Overall, however, the site was clean and easy to use. Explanation if you keep missing the answer to the question. Feedback box, more closely tie in with a professor's class fix all the bugs Give us answers! I realized that to earn unique commitment days it runs off of a 24 hour period and not actual days on a calendar. Maybe put a timer? Including the correct answer so you aren't left wondering It would be nice to be given the correct answer if I answer incorrectly too many times. Also, the worked out solution would be helpful. Make previous questions easily accessible like having an answered questions bank Mobile friendliness Provide some sort of answer system. I understand that if you provide the answers then individuals can just use them to get answers; however, for people who are genuinely using your program, it was not very helpful if you go the answer wrong.

	Providing an explanations for solutions after a student has missed it after a certain number of times could be beneficial. Quick Feedback or flagging of questions that need to be reviewed. Also the UI could be improved a bit so it has better aesthetic appeal :) Remove the floating messages after a couple seconds, tell a user how many tries they have before they guess. Reporting the correct answer was sometimes confusing See the answer if we didn't get it correct Show answers afterwards. Have a scheduled time a question will come out. Show the answer! Show the answers Showing answers after we run out of tries on a question, Having it be part of our grade (for incentive to use), having the teacher review the questions before putting them online Some of the answers were not clear and did not show you how to do it, only gave you the correct answer. Showing the steps would be much appreciated! Other than that, very useful and helpful! The software was a little buggy. The page often needed to be refreshed before displaying updated values. There were some minor issues with commitment points not being tracked on one page, but tracked on the other page.
Incorrect or Confusing Questions	Perhaps multiple choice would help a lot on questions whose answers can be more vagueAlso needs more memes Possibly offer more difficult questions as well. Maybe also add steps on how to arrive at the correct answer Some of the answer to the scheduling problems were wrong so it caused a lot of confusion. Some of the answers for scheduling problems seemed to differ from the instructor's solution after the problem was shown to him.

	some of the questions had answers that were too strict Some of the questions have wrong answers and that could be checked. Sometimes the questions could be worded better There were some issues with correctness/clarity of questions. I also thought having worked out solutions for some of the more difficult problems would have been nice. Overall, I enjoyed using Polycommit. Thank you! When the answers were incorrect or they didn't show the correct answer it really threw off my understanding of the material. I started to question a lot of what I thought I knew, and which ones of the questions were actually correct. This ended up really hurting me on the midterm, because I second guessed myself because of the PolyCommit questions that had been wrong that had not been fixed when I had last looked at them.
Questions Did Not Match Test	Feedback box, more closely tie in with a professor's class More relevant questions Questions similar to what will be asked on tests and quizzes. Some sort of feedback when we get an answer correct. Maybe not the answer itself, but a hint or a point in the right direction of where to look for right answer Some of the questions did not align correctly with the material we used in class and it would make more sense to have the weeks correspond to weeks in the quarter. I noticed that some of the questions continued to be posted in week 2 when were in week 4, etc. The app is great but the questions did not really follow the class material
Needed More Incentive	More incentive to use it every day. I frequently forgot to do it everyday.
Other	Add hints when a question is answered wrong. Adding an explanation for each solution would be very helpful. Also more questions would be good. buzz notification to try a problem

Emails or just a way to remind myself to do it every day!

Explain the answers in more depth

Helpful links to textbook pages

I know people get a lot of emails in college, but maybe a reminder system could be created.

I think it would be nice to maybe have an optional daily reminder setting or some sort of notification system.

I was never reminded to complete problems so I only did them the first week.

I would love email notifications so that I remember to use it.

Instead of an answer, maybe add in a step by step solution

It would be nice to see the solutions to answers that you got wrong, where applicable.

Maybe I missed it, but I thought maybe this was an app? I could've been imagining things, but either way, I think that would be a great extension of the site! Also, I think a good addition might be reminders. Either phone notifications or emails. For those of us who are easily forgetful :)

More clarity/ hints on difficult questions

more questions

More questions being posted

More questions per day would help. There's a lot of content to cover in CPE 464.

More questions!

More questions! I know this was just a trial period, but more of the same would be greatly appreciated

Notifications

Perhaps some sort of remind feature for the daily question

Reminder emails

Reminder emails or something like that could be really cool! There were some days when I just simply forgot to log on and do another question.

Reminders via email (optional of course). Solution explanation/breakdown.

Show the steps to the solution in order to better understand the material.

Some notification service to remind you to complete it

The favicon reminds me of the "browser tab crashed" favicon.

The timing of the submission was odd. I would do it at a certain time and my commit would not increase

Time Zone for updating score

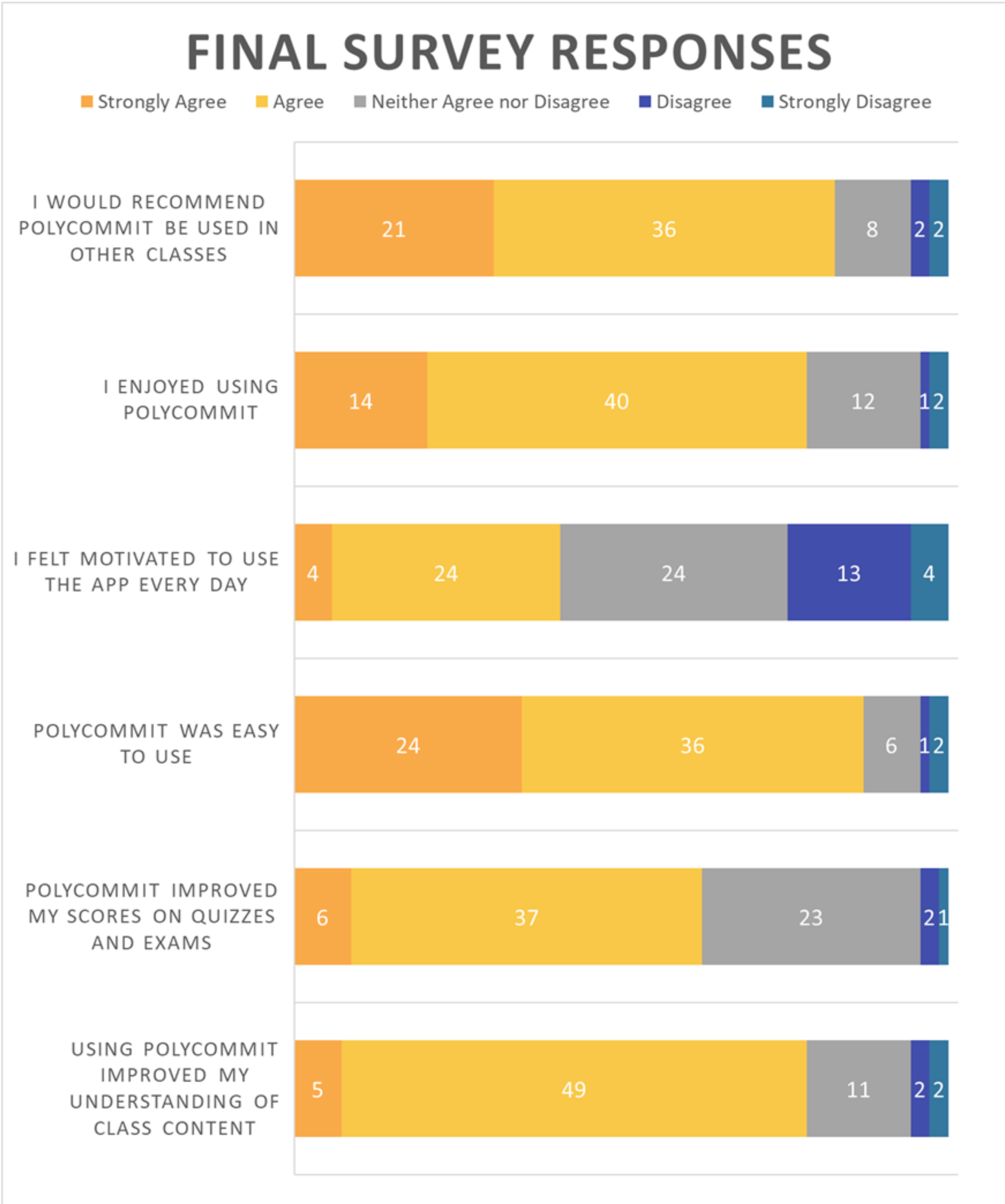


Figure 5.5: Survey responses from users of Polycommit.

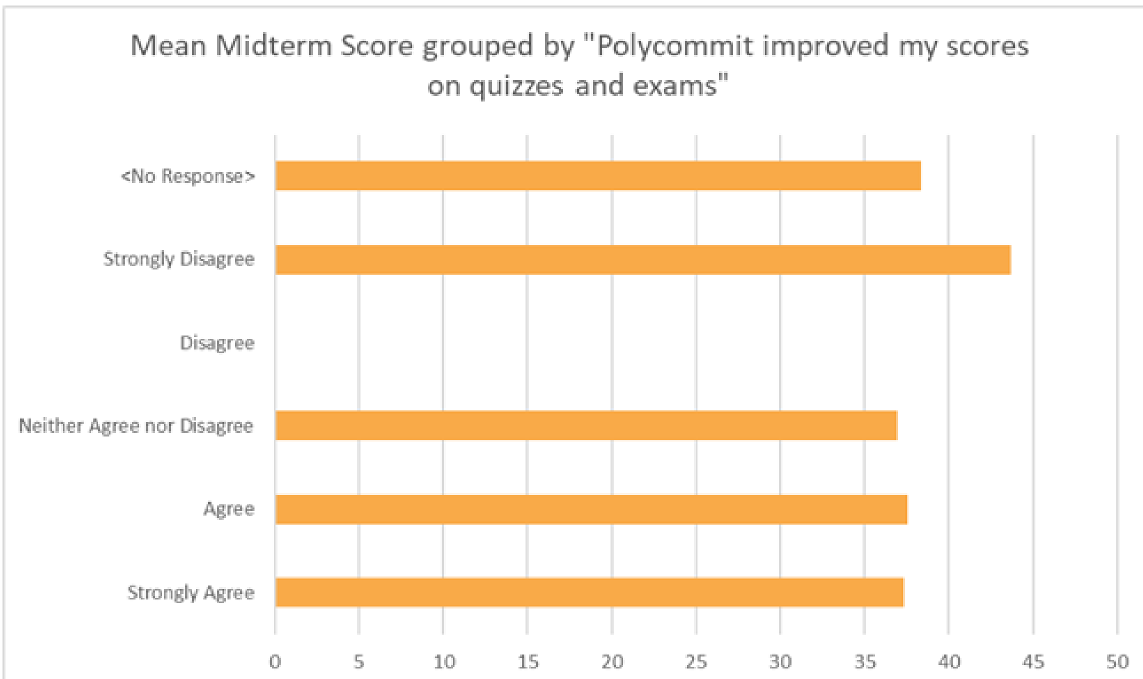


Figure 5.6: Mean midterm score grouped by response to final survey question "Polycommit improved my scores on quizzes and exams."

Chapter 6

FUTURE WORK

6.1 Potential Improvements

There are several improvements we could make to the experiment to draw a more definite conclusion about the effect of habit-forming gamification on student performance.

6.1.1 Experiment Scope

For the purposes of this thesis, we took a "breadth-first" approach to gathering data. That is, we released the app to a variety of classes (most were advanced Computer Science courses, but each covered completely separate topics) and allowed any students to participate in the experiment. This allowed us to get feedback and data from a variety of different perspectives.

However, after refining the experience based on the feedback from the "breadth-first" approach, the next step would be to take a "depth-first" approach to the experiment. That is, we would deploy the app in a more focused manner. We would choose a particular class, such as a introductory Calculus class, and deploy the app to every available section. This would give us more consistent data to work with, as we could compare the performance of different sections of the same class.

6.1.2 Control Group

A further improvement to the experiment would be implementing a control group. There are two main ways we could accomplish this.

One method would be to only offer the app to certain sections of a class, while still collecting the scores on quizzes, exams and homework from the non-using sections. This approach has the advantage of being easy to implement, but has the disadvantage of introducing confounding variables to the experiment. The population of students who choose one section over another is not randomly sampled; for instance, students with better studying habits might take a class in an earlier section, causing the average grade on quizzes and exams to be higher in certain sections.

Another method would be to only offer the app to a randomly selected sample of students in the class. This has the advantage of being more scientifically sound, as a random sample on a large population would rule out other factors that might affect student performance. However, it would complicate the onboarding process, as we would be unable to present the app to the entire class and allow anyone with a Cal Poly account to sign up. One method would be to have the app be part of a mandatory graded assignment, and send a link to each of the participating students that allows them to sign up. This would hopefully ensure that all the students that were selected to participate would actually engage with the app.

6.1.3 Experiment Duration

This experiment took place over the first 5 weeks of Spring Quarter 2017. To gain a more complete understanding of how students respond to the app and how habits are formed over time, it would be advantageous to run further experiments over longer periods of time. Data from different quarters could be compared, adjusting different factors in the experiment to determine how users best respond to gamification.

In addition, several users requested that questions continue past week 5 so they could have a study tool for the final. Future experiments could run for the duration of the whole quarter, allowing us to collect data from all the assignments throughout the quarter, including the final.

6.1.4 User Feedback

Notifications

As noted in **Table 5.2**, users requested several additional features and improvements to the UI. One of the most requested features was the ability to receive notifications each day reminding the user to answer a question and thus earn Commitment. I had considered adding this to the app, but received some feedback during the dry run in Winter 2017 that email notifications would be annoying rather than useful. Thus, I chose to focus on other features rather than adding an element that might discourage certain students from using the app. For future experiments, an opt-in email notification system with a easy-to-use "Unsubscribe" option would likely increase user engagement while not annoying users that don't wish to be notified.

Showing Answers

A number of users requested in the final survey that answers, or answer hints, be shown after completing a question. Originally, the app did not have that feature, but I added it partway into the experiment because several users contacted me about it. The number of students that requested this feature, although it already existed, indicates that the UI does not make it clear that the feature is available. Currently, if students get a question wrong and exceed the maximum number of attempts, the app stays on the page with the closed challenge, and the answer is shown at the bottom of the screen. There could be some other indication that the answer is visible, perhaps on the course page itself.

Similarly, to address some user confusion I added an *Explanation* field where I

Chapter 7

CONCLUSION

BIBLIOGRAPHY

[1] Cal Poly Github. `http://www.github.com/CalPoly`.

[2] Pocket Points. `https://pocketpoints.com/`, 2014. [Online; accessed May 31, 2017].

[3] A. Bruckman. Can educational be fun. In *Game developers conference*, volume 99, pages 75–79, 1999.

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

[5] B. Dodin and A. Elimam. Integrated project scheduling and material planning with variable activity duration and rewards. *IIE Transactions*, 33(11):1005–1018, 2001.

[6] D. Edge, S. Fitchett, M. Whitney, and J. Landay. Memreflex: adaptive flashcards for mobile microlearning. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 431–440. ACM, 2012.

[7] C. B. Ferster and B. F. Skinner. Schedules of reinforcement. 1957.

[8] S. T. Fiske and S. H. Kang. Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences*, 3(1):12–19, 2016.

[9] M. Hardy and S. Heyes. *Beginning psychology*. Oxford Univ. Press, 1999.

[10] W. H.-Y. Huang and D. Soman. Gamification of education. *Research Report Series: Behavioural Economics in Action, Rotman School of Management, University of Toronto*, 2013.

[11] J. Johnson. *Designing with the mind in mind: simple guide to understanding user interface design guidelines*. Elsevier, 2013.

[12] G. Kiryakova, N. Angelova, and L. Yordanova. Gamification in education. Proceedings of 9th International Balkan Education and Science Conference, 2014.

[13] P. Lally, C. H. Van Jaarsveld, H. W. Potts, and J. Wardle. How are habits formed: Modelling habit formation in the real world. *European journal of social psychology*, 40(6):998–1009, 2010.

[14] R. V. Lindsey, J. D. Shroyer, H. Pashler, and M. C. Mozer. Improving students long-term knowledge retention through personalized review. *Psychological Science*, 25(3):639–647, 2014. PMID: 24444515.

[15] N. W. Lingawi, A. Dezfouli, and B. W. Balleine. Physiological mechanisms of habit formation. *The Wiley Handbook on the Cognitive Neuroscience of Learning*, page 411, 2016.

[16] P. Munday. The case for using duolingo as part of the language classroom experience/duolingo como parte del curriculum de las clases de lengua extranjera. *Revista Iberoamericana de Educación a Distancia*, 19(1):83, 2016.

[17] S. Orbell and B. Verplanken. The automatic component of habit in health behavior: habit as cue-contingent automaticity. *Health psychology*, 29(4):374, 2010.

[18] P. Pimsleur. A memory schedule. *The Modern Language Journal*, 51(2):73–75, 1967.

[19] J. Rachels. The effect of gamification on elementary students spanish language achievement and academic self-efficacy. 2016.

[20] I. Renfree, D. Harrison, P. Marshall, K. Stawarz, and A. Cox. Don't kick the habit: The role of dependency in habit formation apps. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2932–2939. ACM, 2016.

[21] E. Sanchez, S. Young, and C. Jouneau-Sion. Classcraft: from gamification to ludicization of classroom management. *Education and Information Technologies*, pages 1–17, 2016.

[22] C. Shelton. virtually mandatory: A survey of how discipline and institutional commitment shape university lecturers perceptions of technology. *British Journal of Educational Technology*, 45(4):748–759, 2014.

[23] B. F. Skinner. Teaching machines. *Science*, 1958.

[24] B. F. Skinner. Operant behavior. *American Psychologist*, 18(8):503, 1963.

[25] R. Vesselinov and J. Grego. Duolingo effectiveness study. *City University of New York, USA*, 2012.

[26] C. Wagner. Digital gamification in private music education. *Antistasis*, 7(1), 2017.

[27] L. Walker. The impact of using memrise on student perceptions of learning latin vocabulary and on long-term memory of words. *Journal of Classics Teaching*, 16(32):14–20, 2016.

[28] P. Wastiau, R. Blamire, C. Kearney, V. Quittre, E. Van de Gaer, and C. Monseur. The use of ict in education: a survey of schools in europe. *European Journal of Education*, 48(1):11–27, 2013.

[29] R. M. Yerkes and J. D. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, 18(5):459–482, 1908.

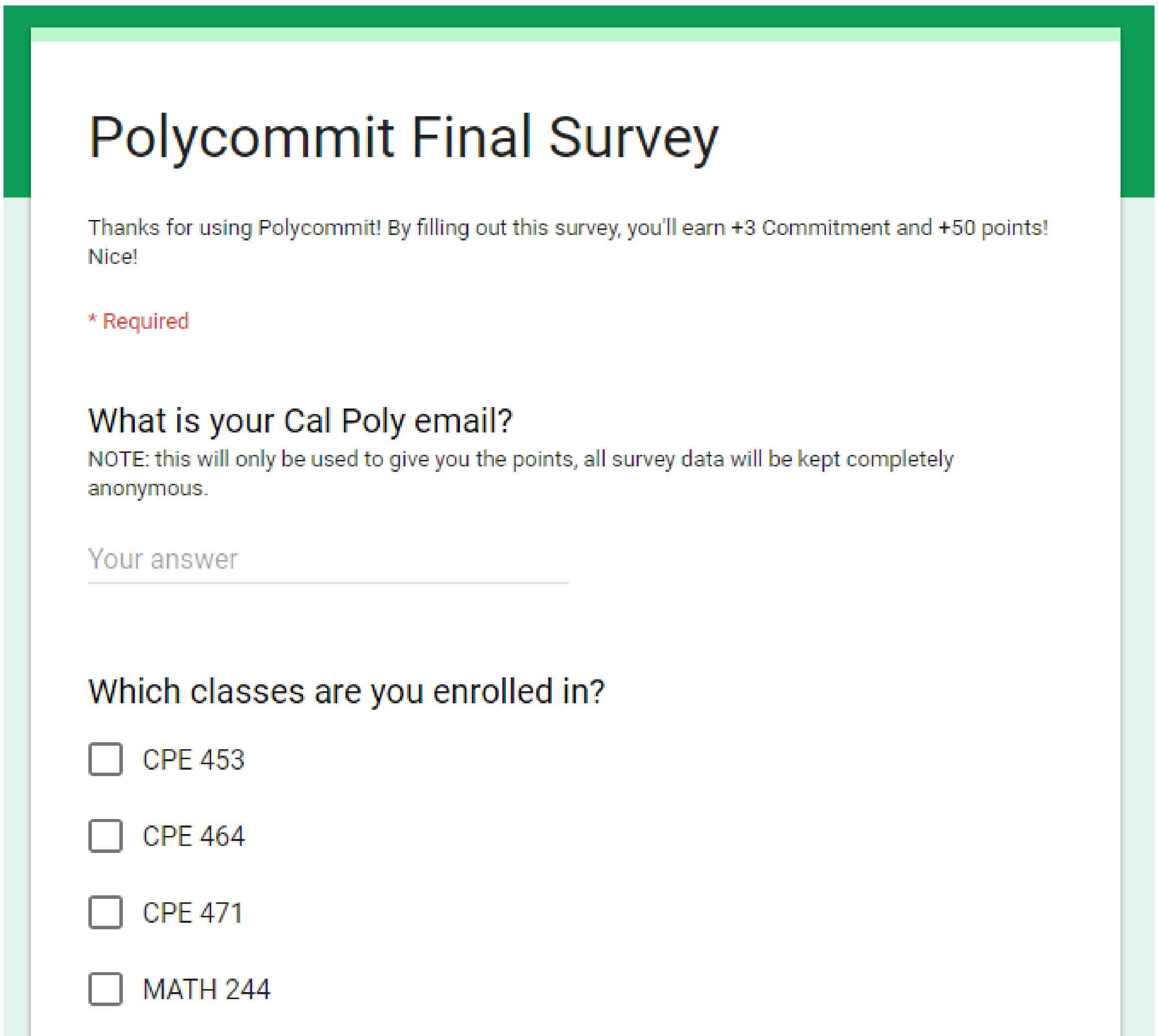
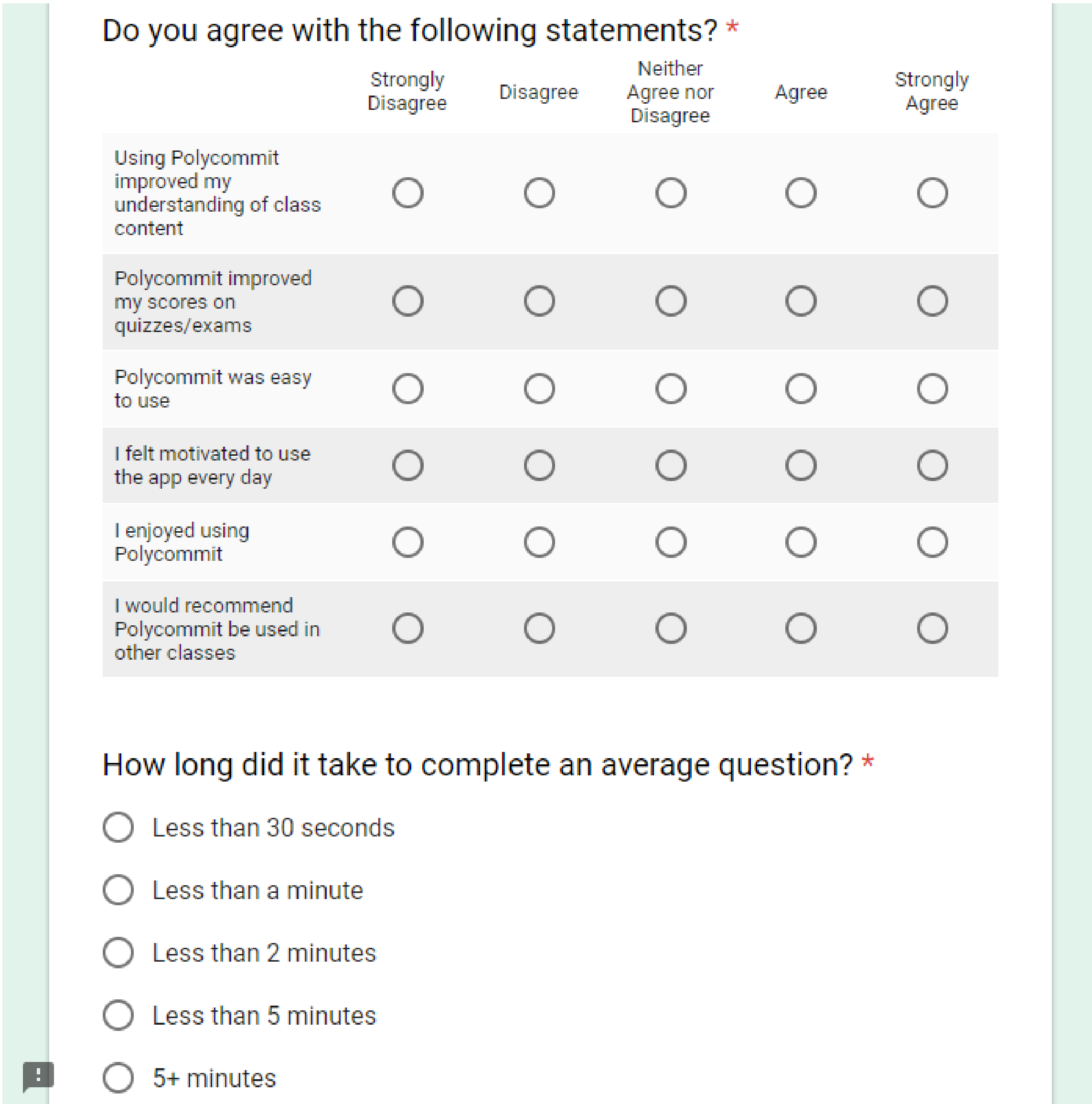


Figure .1:



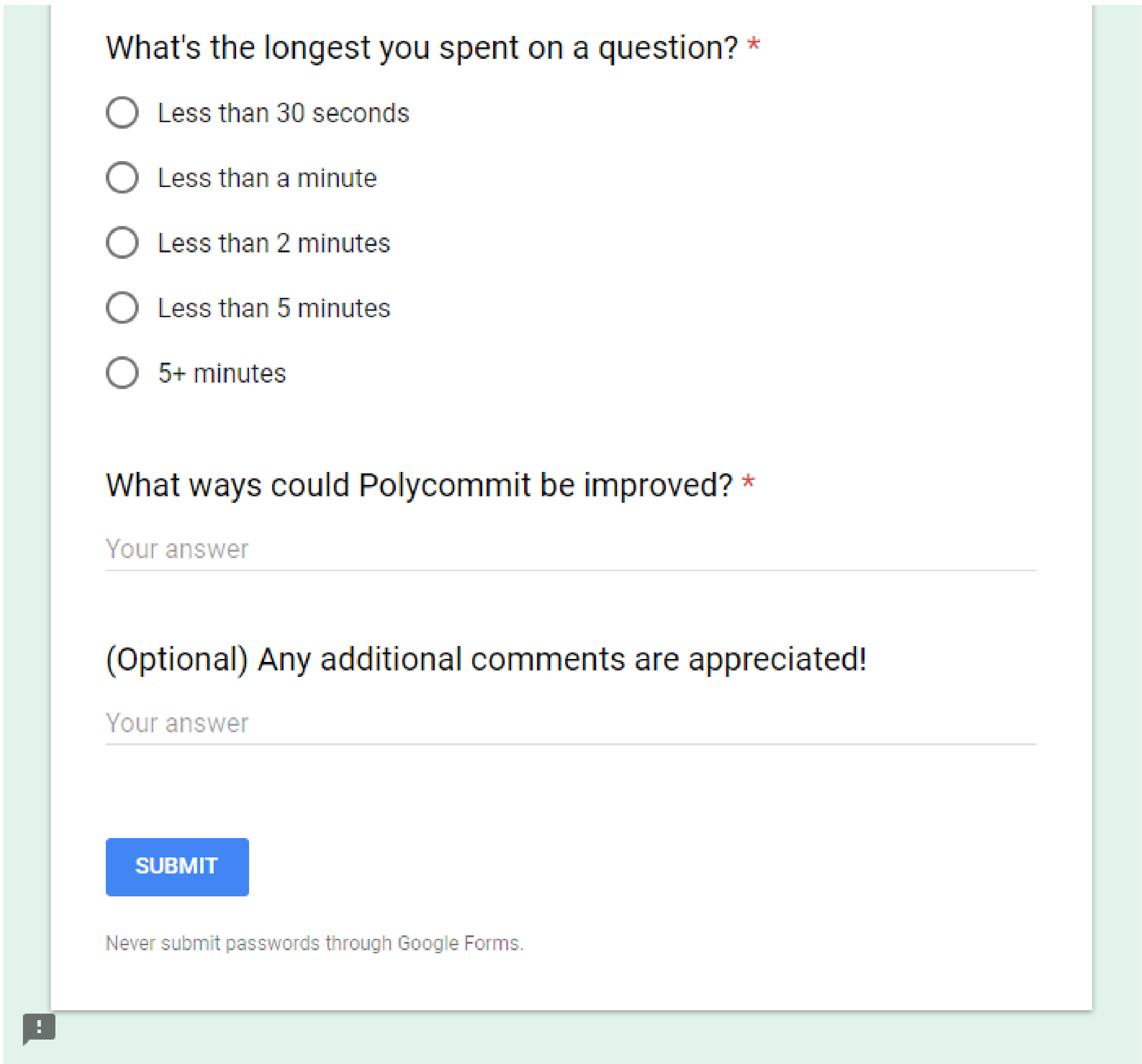
Do you agree with the following statements? *

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
Using Polycommit improved my understanding of class content	○	○	○	○	○
Polycommit improved my scores on quizzes/exams	○	○	○	○	○
Polycommit was easy to use	○	○	○	○	○
I felt motivated to use the app every day	○	○	○	○	○
I enjoyed using Polycommit	○	○	○	○	○
I would recommend Polycommit be used in other classes	○	○	○	○	○

How long did it take to complete an average question? *

- ○ Less than 30 seconds
- ○ Less than a minute
- ○ Less than 2 minutes
- ○ Less than 5 minutes
- ○ 5+ minutes

Figure .2:


What's the longest you spent on a question? *

- Less than 30 seconds
- Less than a minute
- Less than 2 minutes
- Less than 5 minutes
- 5+ minutes

What ways could Polycommit be improved? *

Your answer

(Optional) Any additional comments are appreciated!

Your answer

SUBMIT

Never submit passwords through Google Forms.

Figure .3: