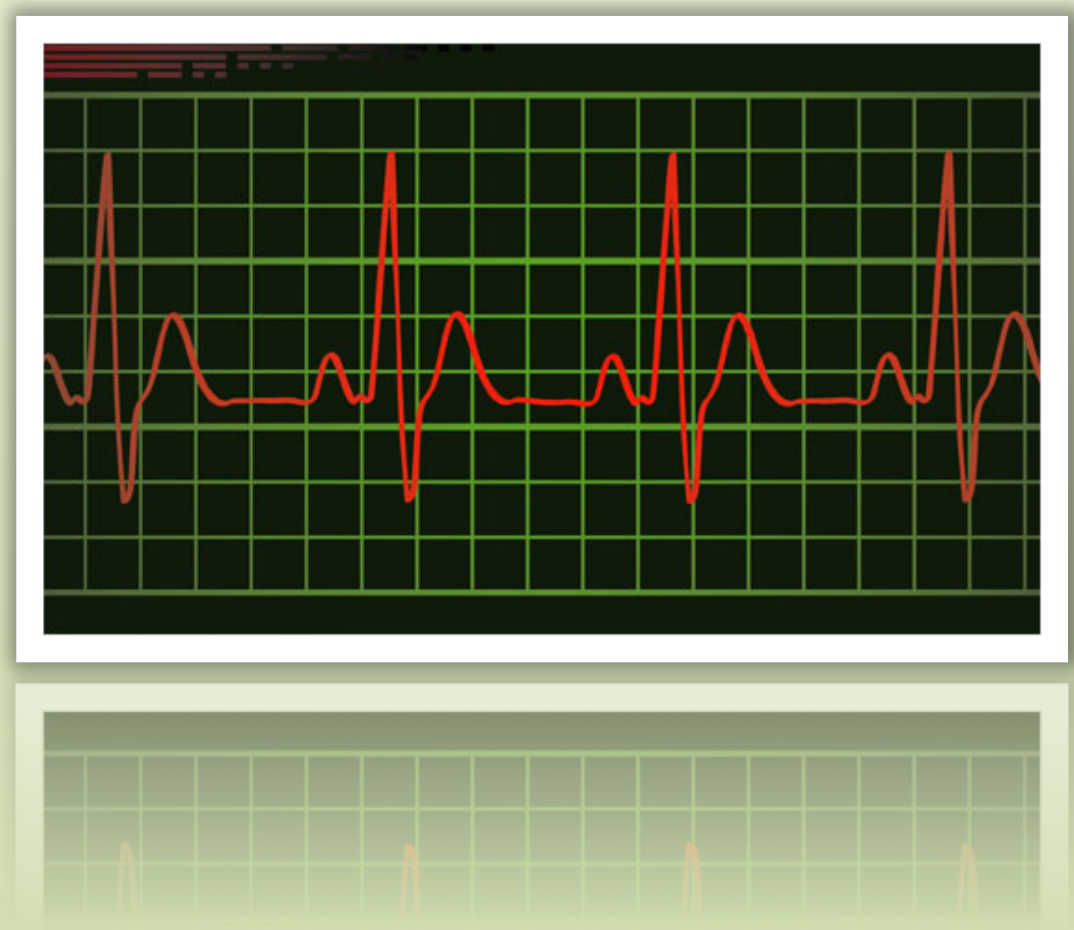
Final Submission (INFO 213)

Team Pulse

Alex Chung, Marco Cozzi, Galen Panger, Karen Rustad

Submitted: May 10th, 2011



Abstract

Pulse is a simple mobile-based application for large lecture courses that allows students to submit their current emotion and monitor the sentiments of other students around them during class in real time. Pulse emerged as a potential solution to the chronic problems of large lectures, including students feeling afraid of asking questions when they're confused, not knowing how the students around them are feeling (whether confused, frustrated or curious), and the typical lack of quantitative course evaluations that allow for mid-course corrections. Through rapid prototyping and iterative development techniques coupled with contextual and "think aloud" interviews, heuristic evaluations and a formative evaluation, we produced a final pre-field testing prototype that was then evaluated experimentally. Results from the experiment, in which Facebook served as a "control" for our alternative design, suggest not only that Pulse is an effective tool for harnessing and sharing class sentiment, but that it's an appreciated tool that could potentially enhance feelings of unity and student solidarity within the classroom. Our next steps would include further enhancements to Pulse's functionality to address weaknesses highlighted in the experiment, followed by field deployment and testing in a live lecture setting.

Introduction and problem statement

Large lecture courses at universities have endemic problems that interfere with students' learning and satisfaction with the class. With hundreds of students in the room, it is difficult for any one student to have their questions answered and needs met. This is compounded by the mere presence of so many other students; many, perhaps most students are hesitant to speak up in class when they do not fully understand something because they fear looking stupid in front of the professor and the rest of the class.

Students also have few viable options for giving feedback to the professor. Telling a professor about an annoying speaking tic, or a topic students didn't find adequately covered, to the professor's face during office hours is highly intimidating. Students are also hesitant to tell TAs about problems because they don't want to be seen as whining and fear repercussions in their grades. Yet by the time students get to the anonymous, required end-of-semester feedback surveys, students have little incentive to provide detailed feedback since they will no longer benefit from it. They also may not remember lectures that happened earlier in the semester and thus fail to include them in their assessment. For similar reasons many professors also fail to see value in the end-of-semester surveys; often they do not even read their feedback. The end result is bored, confused, and frustrated students and professors who, despite meaning well, have no idea how to improve, semester after semester.

Our project, Pulse, aimed to ameliorate these issues by enabling continuous feedback as the semester proceeds and promoting student engagement and solidarity in the classroom. Initially, we targeted professors with our app and imagined what sorts of detailed classroom sentiment feedback would be useful for them to hone their lectures and address student concerns, perhaps even in real-time, during lecture. However, our contextual inquiry and interviews found

that most professors felt little need or desire for such feedback. They were confident that they were in touch with their class's feelings, did not want unnecessary distractions and interruptions during lecture, and were skeptical of additional classroom IT. While each professor we talked to said that they valued teaching, all of them were sure to point out that "many" professors didn't care about teaching and that they had essentially no institutional incentive to teach well. Even if a professor were interested in teaching more effectively and thought our application would help them do so, the incentives simply were not in place to make them likely to put forth the effort of learning and using the application. Thus, instead we turned our focus to students' needs.

We wanted to give students a low-distraction, course-focused backchannel to express their feelings, both for their own catharsis and to let them see if other students felt the same way. We hypothesized—and had confirmed by our user interviews—that students might be more willing to speak up about confusion or class problems if they saw that they were not the only one who felt confused or frustrated. The data would serve not only as information for them, but also as a form of proof and therefore leverage for asserting that they're not alone in being confused. Thus, instead of creating a whole new channel for feedback for professors (though nothing would preclude them or a TA from monitoring Pulse), our application would enhance and increase usage of existing feedback channels, with hopefully the end result of better lectures and learning.

Design process

The first step in our design process was engaging in contextual inquiry by watching students' and professors' current behaviors and workflows in live lectures. In one class, laptops were not allowed so students took notes on paper. The fold-out desks in the room were tiny, with little surface space for more than just a notebook. The professor presented the material using PowerPoint and a tablet PC which enabled her to draw graphs and the like directly on the slides. Given the size of the class (roughly 400 students), the professor's main source of feedback was pop-quiz questions between slides, which students answered via iClicker. In the other class we watched, there were fewer than 40 students, enabling students to ask questions directly via raised hands. Most students had laptops out, usually with multiple tabs and applications open, many of which were not relevant to the class at hand and which appeared to be distracting. The professor also used a PC with PowerPoint to accompany his lecture, but he wandered far from the computer in the course of his speaking.

After viewing each class, we interviewed each professor and a couple students to better understand their feelings about the current classroom feedback channels and see what their needs and concerns were. We learned that professors believe themselves to be highly-attuned to student sentiment and understanding and that they often want students to struggle in class as part of the learning process. Some professors disliked the idea of our application and believed it would be distracting, while a few thought it could improve their performance. Nearly all preferred getting feedback post-lecture to during lecture in real time; they felt the latter would be both distracting and difficult to usefully act upon. On the student side, most students could point to a bad experience with a poor lecturer that they wished they could have done more about. They

were aware of existing feedback channels (professor office hours, questions in-class, TAs, and course evaluations) but found them difficult to use due to a varying combination of student embarrassment, laziness, and intimidation.

Both professors and students liked the iClickers and thought that they increased classroom engagement. Both were also sensitive to egos and hurt feelings; professors were wary of overly negative or unhelpfully harsh feedback, while students wanted both anonymity and incentives to discourage flaming. Finally, both sides saw end-of-semester course evaluations as a joke; students rarely put much effort into writing them, and professors often didn't read them.

Through our interviews, we built several generations of lo-fi prototypes. Our first paper sketch prototypes explored several different concepts—some interfaces targeted primarily at professors, others mainly student-facing. We determined that we wanted to track multiple student emotions, though we were not sure if we wanted those emotions to be pre-set or user-generated. We also sought to display an intuitive graph of those emotions over time. We also drew up plans for a question and/or comment handling facility (for both students and professors) and a course search facility.

Based on our interview data, we eventually chose one of the student-facing concepts to expand into a semi-interactive paper prototype with multiple "screens," with which we did our initial think-aloud. We decided to make the set of possible emotions pre-determined, to limit unnecessary distraction of students. We also drew up a more concrete plan for the question handling system, adding the ability to rate questions up or down.

After that, we built an interactive digital lo-fi prototype in Balsamiq which enabled testers to actually click on buttons to move between views. Based on the feedback from our user-based think-aloud testing of the previous prototype, we moved our sentiment visualization from a series of line graphs over time to a simpler bar chart, with a bar for each emotion. While the bars would move up and down in real-time, you could no longer view what emotions the class had had over time directly in the dashboard. The question handler now became a comment handler (with questions merely being another type of comment). We also ironed out some basic application flow problems, like missing back buttons or unclear close buttons. With the Balsamiq prototype we conducted two heuristic evaluations—one by us, the other by another group of user interface designers—the results of which informed the design of our final working prototype.

Final prototype implementation

Process: Incorporating Balsamiq heuristic evaluations

The heuristic evaluation conducted by Team "User May I" for our Balsamiq prototype informed us that people were confused with the dashboard layout, especially the disconnect between the bar chart and the emotion (choice) buttons. Additionally, we got feedback that the some of the user-navigable screen interfaces were crowded with too many "clickable" options and led to accidental button presses and frustration. Lastly, our commenting feature did not conform

to how the users expected it to function. Based on this feedback, we incorporated many UI revisions and information layout changes into our Android-based final prototype.

Implementation with Eclipse and Android

Given that we had chosen the students to be our primary focus, the functional prototype focused on enabling the following three tasks: (1) creating a list of classes in which a student can participate; (2) entering a student's emotion and viewing the class's overall sentiment; and (3) posting a comment or replying to a specific question. We also decided to develop on the Android phone OS using Eclipse IDE. Although our Balsamiq prototype was based on Apple iOS, developing on the Android allowed us to leverage existing development skill and to avoid the iOS developer license fee of $100 per license.

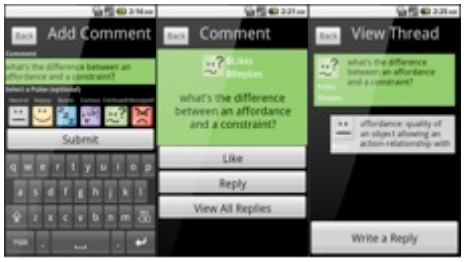
In mobile app development, where the screen resolution averages around 480 x 320 pixels, space is the greatest UI constraint. Thus we used the Simplicity Design Theory (Approachable, Recognizable, Immediate, and Usable) as our guide to pare down interface actions to just the essentials. For example, we consolidated the emotion buttons with the graph legend by lining up the buttons with their respective bars on the bar graph. This change satisfied the labeling issue, spatial relation, and physical constraint. Yet we still had to fit all five emotion buttons within the 320-pixel width because we felt all five choices were necessary to cover the minimal range of emotion. To resolve this issue, we had our graphic designer Karen design simple, colored "emoticons" to represent each of the emotions. We then used the unique color of each emoticon to highlight the comments in order to provide additional contextual information about the mental state of each commenter.

The Android OS's SDK came with many core applications built-in, making the task of the application developer relatively easy. It also came with SQLite as a database system to handle data storage. The SDK does not have the refined look of Apple iOS but its open source nature with the basis of the Java language make the transition from desktop programming less daunting. In terms of using the Eclipse IDE to develop user interfaces, sadly we found that its best feature is merely its price: free. The graphical layout of Android is XML-based and the Eclipse IDE came with a GUI allowing developers to drag and drop components such as buttons, tables, frames, etc. However, the GUI display never scaled correctly and its layout display did not accurately represent how the application looked on the phone. After a couple of tries, our developer gave up on Eclipse's GUI and coded the XML for the app's layout by hand.

Final set of features

In Figure A, the task is creating a list of new classes. The available classes have been pre-populated into the SQLite database affording the auto-complete function after users enter the first two characters (right image). There are three basic views: "My Class List," "Dashboard" and "Comments." When the users first open the app, they will see an empty "My Class List" and be encouraged to add a class (left image). Tapping the "Add Class" button takes the user to a search box (center image). In Figure B, when the users tap on a course and the course is in session, they are brought to the course "Dashboard." Users are allowed to enter any combination of their emotions by clicking on the emoticons. Each entry represents a single vote;

however, users are only allowed to enter one vote within a given period of time. Thus a second click of the same emoticon will subtract the vote that has been submitted. The design is meant to prevent “trolling” where one person can skew the results by repeatedly entering the same emotion. Then there is the top comment placed below the class sentiment graph based on its number of “Like” votes. Finally, tapping on “Comments” brings the users to the “Comments” view (Figure C). In Figure C, Tapping “New Comment” allows the users to enter a new comment thread. Moreover, users can tag the comment with an emotion by choosing an emoticon. Replies to a single thread are logged and can be viewed when choosing the “View All Replies.”

<table>
<tr>
<td> 
</td>
<td> 
</td>
<td> 
</td>
</tr>
<tr>
<td>Figure A</td>
<td>Figure B</td>
<td>Figure C</td>
</tr>
</table>

Future features

In the future, we would like to implement the following features: (1) implement a back-end web engine so that multiple students can simultaneously use the app and see each other’s input reflected in the dashboard; (2) improve the commenting interface to match the commenting system of Facebook because our usability testing suggests that they are most comfortable with that model; (3) Add a retrospective view to allow users to go back and review their previous emotion entries and those of the class as a whole; and (4) add a horizontal layout for when users rotate the orientation of their phones.

Why Pulse is unique

Pulse is unique because it addresses the broken system of class evaluation with its real-time “sentiment analysis.” Both students as well as faculty and staff can sign on to view the sentiment of student body at that specific time. Instead of a “pull” model where students’ feedback is expressed when prompted, the initiative is passed to the student body to “push” or assert their sentiments. Students benefit from having a channel to express themselves, and faculty and staff benefit from better analytics and more data to adjust their teaching styles and materials. Either way, we think Pulse can improve the quality of teaching and the satisfaction level of both parties.

Evaluation method

We chose a “task-oriented” evaluation as opposed to evaluating design variations because of resource constraints (it was difficult for us to complete the one design variation we did on time). We chose Facebook instead as our “control” condition because it’s what students are used to, because all of these tasks are possible on Facebook, because it’s generally regarded as well-designed and easy to use, and because as the leader in social networking, it also provides a kind of “market test” for Pulse, in terms of whether Pulse might actually succeed in the market or not. It also helps answer the question, “Why not do this on an existing platform?” if indeed the Pulse model outperforms Facebook. On Facebook, we created a group for the student’s class, and students worked within the group and their News Feed to accomplish each task.

After using each tool, we asked our students (10 total) to take a short survey to assess the usefulness, intuitiveness, satisfaction, solidarity and overall distraction of the platform tested.

We asked our users to provide a Likert scale rating (1 - Strongly disagree to 5 - Strongly agree) to the following questions:

1. *I felt that Pulse was easy and intuitive to use.*
2. *Using Pulse would be useful to me in class.*
3. *Using Pulse would make me feel more satisfied with my classroom experience.*
4. *Using Pulse would distract me in class.*
5. *If other students used it, too, using Pulse would give me a greater sense of solidarity with them.*

For each of the 3 tasks we recorded the time to completion, as well as the number of superfluous or unnecessary steps taken by the user. We then plotted the means for all variables and compared the results to get a sense of which tasks were easier to perform on Pulse vs. Facebook.

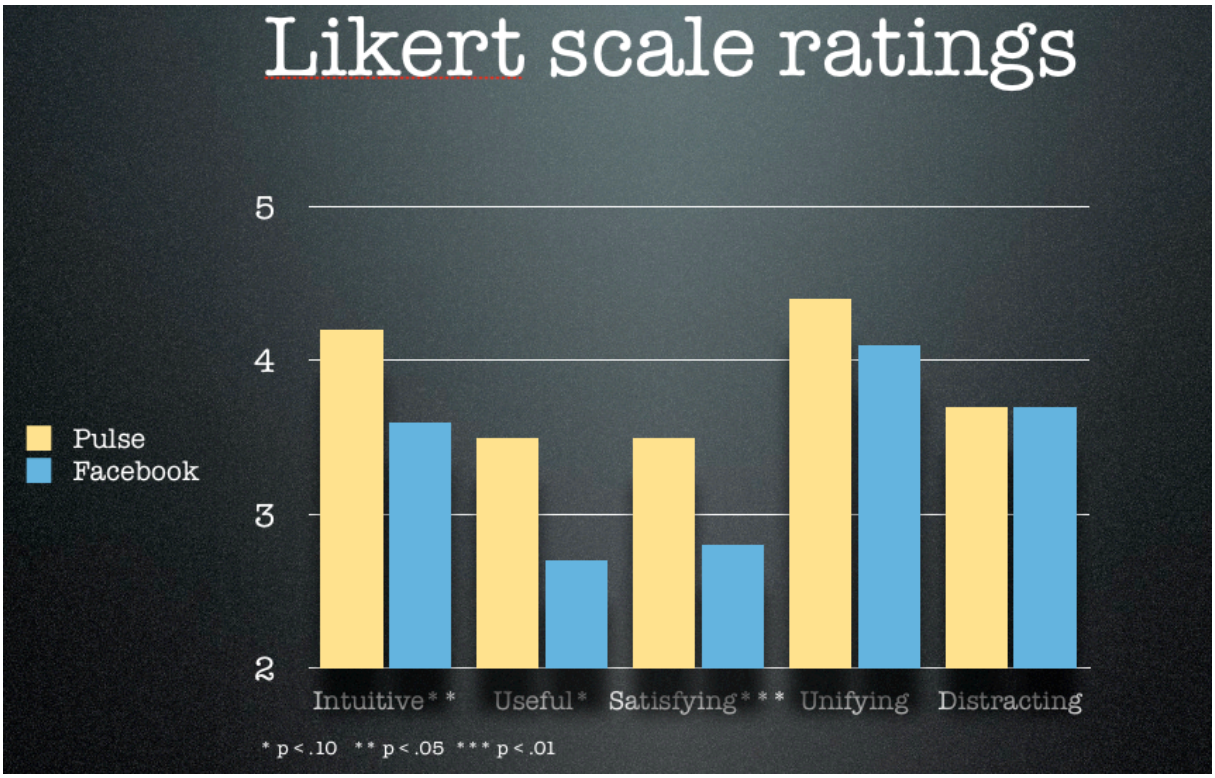
Because the same users tested both Pulse and Facebook (5 using Pulse first and 5 using Facebook first), we conducted paired-sample t-tests using one-tailed hypotheses to assess the significance of our results. The hypotheses generally took the form of, “Users took less time/made fewer errors on Pulse than Facebook for task X” and “Users gave Pulse higher [intuitiveness/satisfaction/etc.] ratings than Facebook,” We also did single-sample means tests for Pulse to assess whether it exceeded neutral ratings on all positive dimensions (intuitiveness/ satisfaction/etc.), and was below neutral on the negative dimension (distraction).

Lastly, we also used qualitative measures to evaluate the Pulse prototype. In the survey, we gave users the opportunity to respond to the following open-ended questions:

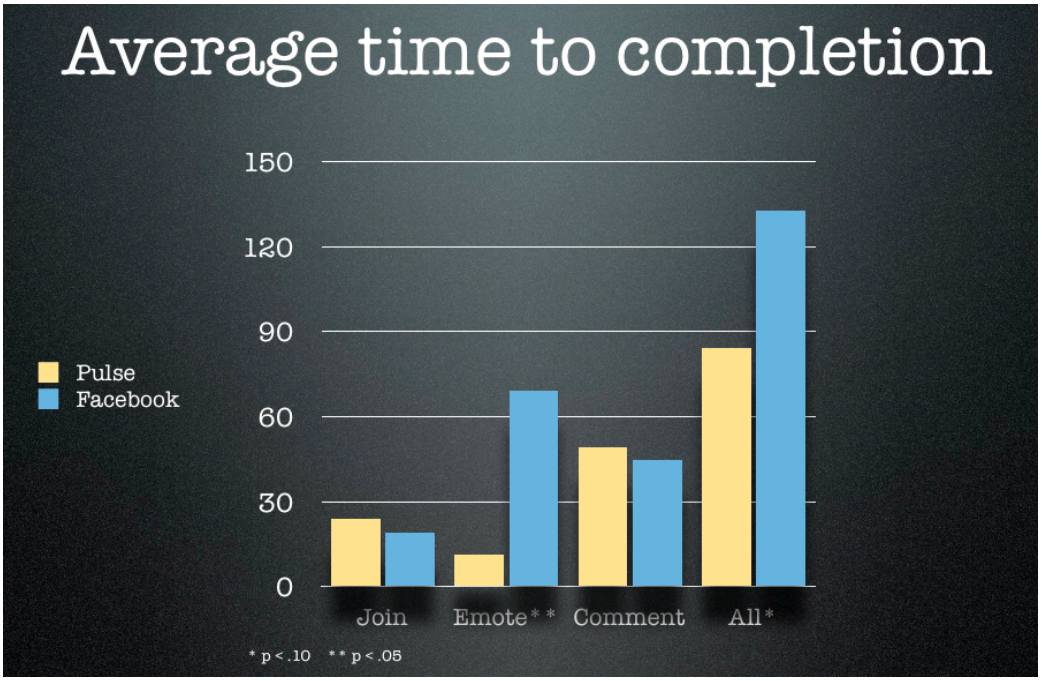
1. If there were aspects of Pulse that were difficult to use, please describe them and the issues you encountered.
2. Describe how using Pulse would make you feel in class.
3. Do you have any other thoughts or suggestions about Pulse to share?

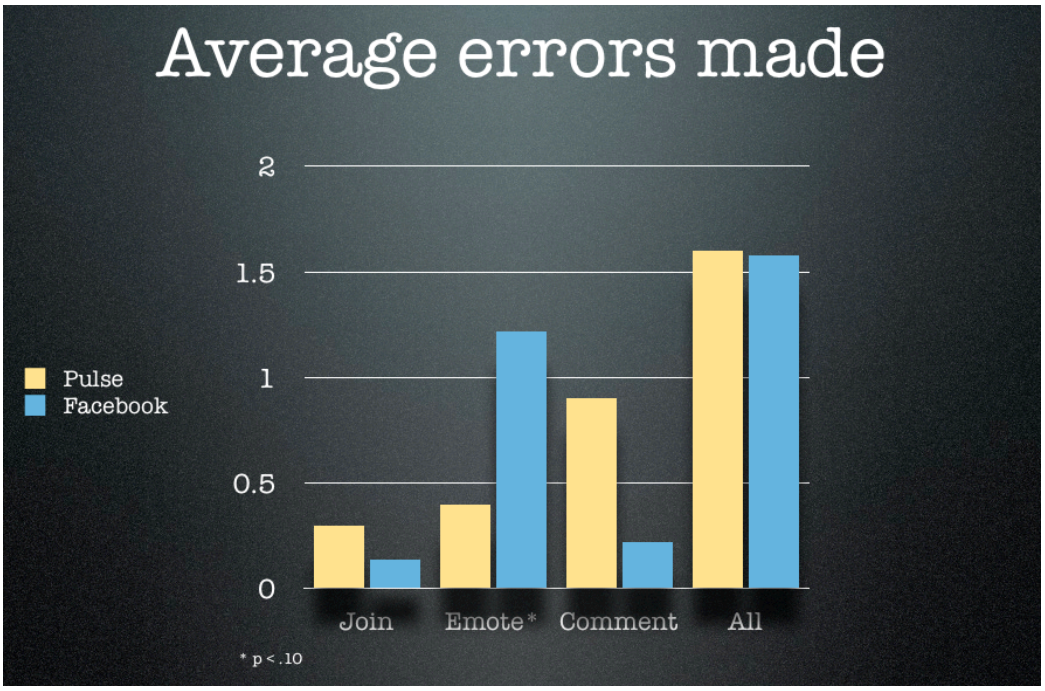
Evaluation results

Focusing on the student ratings of the different platforms, we found that students rated Pulse more intuitive, useful, satisfying, and unifying (solidarity-producing) compared to Facebook. With the exception of solidarity/unity, these results were statistically significant. We also found that Pulse was no more distracting than Facebook.



To evaluate the efficiency and ‘ease of use’ of Pulse, we infer the level of difficulty of a task is related to the time to completion and number of errors made. Specifically, an easy or more efficient task should have a shorter completion time and fewer number of errors made, while a more difficult task would take longer and be prone to more errors.





The recorded quantitative results aligned well with what users reported were the easiest and hardest tasks on Pulse and Facebook (this was a specific question asked on the survey). Submitting an emotion was the easiest task on Pulse, and the hardest on Facebook. Submitting a comment was the hardest task on Pulse, while users said (a) commenting and (b) searching for a course were the easiest on Facebook (they were tied).

There were order effects that you might expect, too: people took less time and made fewer

errors with the second tool they used (probably due to their familiarity with the tasks), but reported the second tool they used to be more distracting. This difference in distraction from first to second tool was statistically significant at the 5% level (though admittedly the hypothesis was generated after the difference was observed).

Finally, comparing the Likert scale ratings for Pulse to the neutral baseline (a 3 rating), we found that Pulse was on average more intuitive, useful, satisfying, unifying and distracting than a neutral rating, which is positive except with regard to distraction. The differences from neutral on all counts were tested for significance (except for distracting, because its raw mean already violates our hypothesis), and all but satisfaction were significant at (at least) the 10% level. Satisfaction was just barely over 10% significance ($p = .106$). Please refer to the References section for the Stata output from all t-tests.

Qualitative feedback

From the survey comments, a number of issues were highlighted. Many students recognized the potential usefulness of the tool, with one person saying specifically that they felt it would give students a “greater voice... we would be more inclined to speak up.” Students expressed concern for the professor’s feelings if their emotion ratings were negative. Distraction was still a concern, though in the qualitative feedback most preferred Pulse over Facebook on this point because options for distraction were more limited in Pulse. Lastly, several students were not “completely sure what the bar graphs for different emotions represented.” We acknowledge that there is future work to be done to explore and test the effectiveness of alternative visualizations. However, our experiment lacked any real-time dynamic, so we believe it was not readily apparent that the graph represented the emotions of the entire class.

Discussion and conclusion

Overall, we were pleased with the results of the experiment, both in terms of how Pulse performed against Facebook and in terms of the ratings Pulse received relative to neutral on the Likert scale. We were also pleased that many of the results we received were statistically significant (or almost significant), especially in light of the typical variance in user expectations and our small sample sizes. Generally, the results conformed with our expectations, though we were a bit surprised that Pulse was slower and more error-prone on two of the three tasks. However, because not all of the Facebook features we needed for our tasks are available on the Facebook mobile applications, we ended up conducting our tasks for Facebook on a laptop, meaning that the longer time to complete a comment on Pulse was in part due to the general difficulty of typing on a virtual keyboard. On average, users made about one error using Pulse’s comment system, meaning confusion was also a part of the slowdown.

Overall then, our Pulse prototype was as or more effective for the given tasks and measured outcomes than Facebook, which is an important proof point not just against a market leader, but to also to answer those pesky questions like, “Why not just use Facebook Groups to do all of this?” Clearly there is an opportunity to do better, and these results suggest that we have.

One result that wasn't statistically significant—that users felt higher levels of solidarity using Pulse than Facebook—in fact is still a very positive result for us. Solidarity ratings were higher than any of the other Likert scale ratings—an average of 4.4 out of 5 for Pulse. That Facebook was right behind at 4.1 out of 5 doesn't diminish what we're attempting to do here; rather, it emphasizes an important benefit that this student-to-student communication during class might bring—helping students feel more tightly bonded as a group overall. This in turn might diminish the sense of isolation that students feel during lectures, where they aren't sure if other students feel as confused as they do, and thus don't feel comfortable (or justified) in speaking up on behalf of the class.

Finally, the experimental results also leave clear room for improvement, and highlight some remaining doubts about the application. The most obvious risk for Pulse is that it will prove to be too distracting in class. But less obvious, and perhaps more worrisome, is the tepid approval students gave Pulse on the dimensions of usefulness and satisfaction in class. While the results are better than "neutral," they are not enthusiastically better, and enthusiasm will be needed if the tool is to succeed in the field.

That brings us to the most important next step for the application: field testing. Were the semester longer, we would have a chance to give the app the real test it needs, which is in an actual, live classroom setting. There, with live data coming in from other students, we'd have a real chance to see what kinds of feelings students have on the dimensions of usefulness, satisfaction, ease of use, solidarity and distraction. In our experimental evaluation, students *projected* how they'd feel in the classroom with Pulse. Most important on the key dimensions of solidarity and distraction is how students *actually* feel in the classroom with Pulse.

References

- Survey/task walkthrough (No sign-in required. No data is submitted until you hit the final "Submit" button, so feel free to explore.)
- Survey results and summary statistics (You should be able to access this with your Gmail credentials; if not please let us know.)
- Stata output for t-tests (No sign-in required.)
- Informed consent and records release forms (TXT)

Appendix

Figure 1: Our group presents Pulse at the May 6 Design Tradeshow

On May 6, 2011, Team Pulse presented their poster at the UI Design Tradeshow in Wozniak Lounge of Soda Hall, Berkeley. There were a total of 21 posters with our team as the only entry from I School. The host was Bjorn Hartmann and the participants were his students from CS160 - UI design. Our team fit in perfectly as the rest of other groups were also building functional prototypes on the Android OS. However, Team Pulse were the only one who spent time conducting field research and usability testing. The poster session was two hours long and we received many thoughtful feedback and compliments. In general, the students reacted very positively to the project's idea on the "push" commenting model while non-students thought the simplicity of our design was intuitive and not distracting. All in all, the team thought the experience was satisfying, as we had a chance to explain our project to a large audience. Also, we thought the collected data from our usability testing and contextual inquiry made the project special because we were able to point to those tests and interviews when someone had questions about the applicability of Pulse in the classroom.

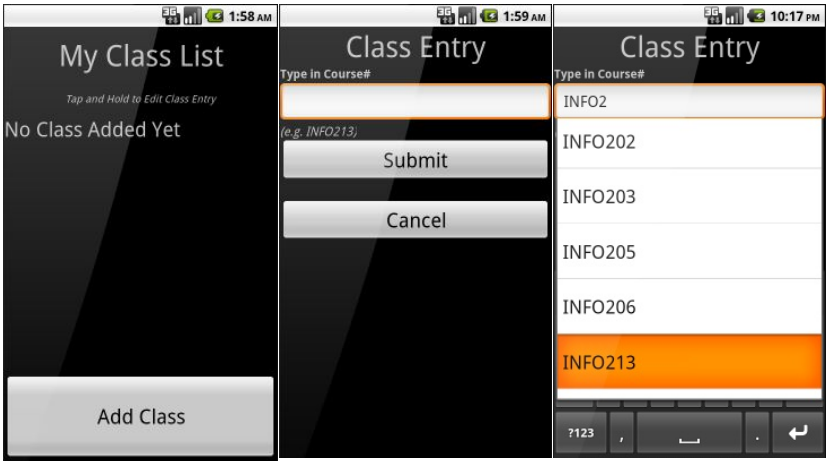


Figure A.

When users first open the app, they will see an empty “My Class List” and be encouraged to add a class (left image). Tapping the “Add Class” button takes the user to a search box (center image). The available classes have been pre-populated into the SQLite database affording the auto-complete function after users enter the first two characters (right image). Tap on a course name after adding it to enter the course “Dashboard.”

Figure B.

Users are allowed to enter emotions by tapping on the emoticons. Each entry represents a single vote; users are only allowed to enter one vote within a given period of time. Thus a second tap of the same emoticon will subtract the vote that has been submitted. Then there is the top comment placed below the class sentiment graph based on its number of “Like” votes.

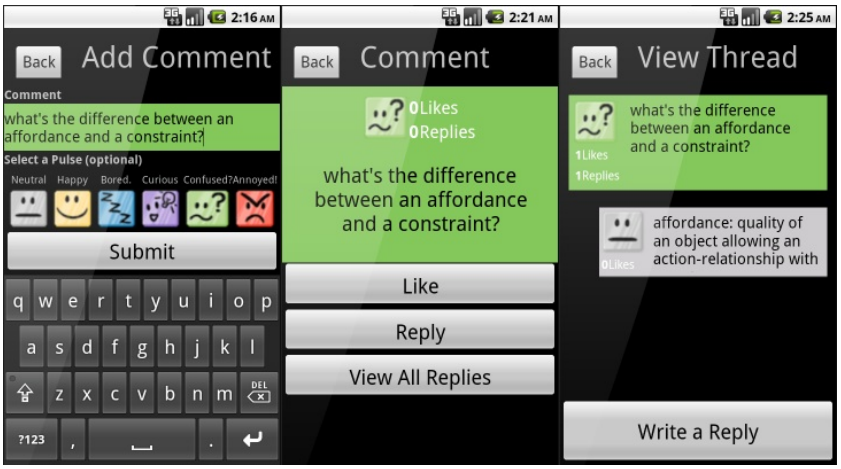


Figure C.

Tapping “New Comment” allows users to enter a new comment thread. Moreover, users can tag the comment with their emotion by choosing one of the emoticons (left image). Replies to a single thread are logged and can be viewed when choosing “View All Replies.”