Mass Participation in Evaluation and Design

Alistair Morrison, Stuart Reeves, Don McMillan, Scott Sherwood, Owain Brown, Matthew Chalmers
School of Computing Science, University of Glasgow, UK
{morrisaj, stuartr, donny, sherwood, owain, matthew}@dcs.gla.ac.uk

ABSTRACT
The emergence of 'app stores' as a means of distributing software applications on a number of platforms is a relatively recent phenomenon. Easy access to a potentially wide audience could radically alter the nature of many software trials but, as the practice is so new, the field has not yet developed guiding principles or an understanding of good practice. Here we share our experience of running trials of iPhone applications deployed in this manner, describing our findings and offering recommendations to others planning to use app store-style distribution.

INTRODUCTION
In our research into mass participation ubicomp trials, we have released a number of research applications using the app store model of distribution, experimenting with different types of application and different methods of conducting a trial. In this paper we describe some of our experiences, exploring issues such as user numbers and elicitation of user feedback.

Two applications are discussed in this paper and will be briefly introduced here. Hungry Yoshi [1], a game that uses wifi infrastructure as a game resource, was the first application we released through the app store method. It currently has around 40,000 registered users, many of whom have engaged in the trial process through completing feedback questionnaires. Our most recent application is World Cup Predictor, a game designed to run alongside the FIFA World Cup, which has £500 of prizes for top players and tries to encourage social interaction with other players through Bluetooth-based data transfers. Although only available a short time before the World Cup, it gathered 9000 users. Both apps were free to download. These systems were both instrumented with the SGLog framework [2], which regularly uploads usage logs to our servers, providing the basis for the statistics and analysis in the following sections.

WHAT IS A 'USER'?
In research papers it is common practice to state the number of users involved in a trial. Yet an interesting question emerging from our sort of trial is what exactly constitutes a user of an application. In traditional trials, the definition of a participant is usually clear, with applicants perhaps responding to a recruitment announcement, being supplied with a device to use and being paid for their time. In trials that use a repository-based distribution method, this becomes more complicated. The number of people who have launched the application does not fully explain how intensively a system is being used. How does a researcher choose a threshold for activity before the user can be said to be engaging in software use; what makes an 'active user'? As well as use of the software, which may be logged, it might also be important to consider a participant's engagement with the trial itself. Many of our applications have included feedback mechanisms through which researchers can target specific questions to users, as explained below. Trial engagement could also come in the form of telephone interviews, emailed questionnaires or communication through social networking sites.

As an example of the different ways users can be counted, some numbers are provided from our trial of Hungry Yoshi [1]. The application has had 182,714 downloads, but this figure includes software updates, so the same user might be included up to 8 times if he or she has downloaded every new version of the application. To count unique users, our SGLog database has recorded that 98,556 people have launched the application, although this number might be lower than the true value if the application was run while network connectivity was unavailable on the user's device, as SGLog would not have been able to upload to the server. On first launch, the user is presented with terms and conditions and asked to sign up for an account in order to play the game. 36,169 completed this registration process. Considering only the users who managed to score any points in the game (a non-trivial task that may involve physically walking to an area to collect 'fruit' then taking it to a Yoshi waiting at a different real-world location) the figure drops to 4,134, and looking at players who played on 5 or more different days the number is down to 3,080. It can be seen therefore that simply reporting the number of downloads an application has had is not a particularly informative statistic on its own.

ENCOURAGING USE AND GATHERING FEEDBACK
In more traditional user trials, participants are often compensated financially for their time. There could be a set number of tasks a participant is expected to complete to qualify for payment, or the pressure to use an application might be more implicit—that the user is part of the trial and they might feel that they should earn their payment. With an application downloaded from an app store, researchers are probably not going to pay for participation and a user

Digital Futures'10, October 11 – 12, 2010, Nottingham, UK.

might feel less obligated to put in 'token' hours of use or otherwise engage with the trial process. Researchers might therefore have to think of other means to motivate use. Most of the applications we have trialled have been games and, as such, it is expected that fun will be a motivating factor. It is also often possible to introduce a competitive element. Several of the games we have released have included a global scoreboard, where players can compete to rise in the rankings. Our interviews with users of Hungry Yoshi indicated that scoreboard position was an important factor for players, e.g. one user stating "*I definitely don't want to go anymore back. Actually in a week I think I will go one more up*", showing that he was defensive of his current position and was actually making reasonably long term plans about how he would rise up the table.

Although it would not be feasible to offer to pay every user who downloads and runs a trial application, researchers can still offer financial rewards as prizes for success or for fulfilling certain obligations. For example, in the Hungry Yoshi trial, participants were paid if they agreed to be interviewed over the telephone about their use of the game. A feedback question was put into the game to gather email addresses of those willing to be interviewed in this way.

With app store distribution, it cannot be guaranteed that all of a group of friends or family will download the application. Thus the situation becomes more complicated if the software under examination has a social aspect, and our applications provided social elements in different ways. In Hungry Yoshi, users were provided with an option to sign into the game using a Facebook account, in order to share progress with friends or to chat with friends or developers about the game. 16,735 of 36,169 registered users chose this option, which is 46% of the user base. Using a different approach, the World Cup Predictor game offered additional bonus points to players for playing a 'head-to-head' game with a friend where predictions were swapped locally over a Bluetooth connection. Despite the strategic advantage that could be gained from performing these head-to-heads, only 45 out of 3,847 users who played the game did so, which is only slightly more than 1%. This was in spite of the fact that there was a prize for winning the league and the fact that those players engaging in head-to-head activity were gaining an advantage—2 of the overall top 3 players were among the 45 Bluetooth users.

We explored how application usage might spread through social groups, asking players of Hungry Yoshi whether they had told their friends about the game or encouraged others to play. Roughly a third of responses indicated positively that players had spoken about the game with their friends, for example "*Ive recommend the game to a lot of my peers. They love it go yoshi!*".

These results seem to indicate that users are keen to add social aspects to their applications, but are more likely to engage in activities such as Facebook, with its flexibility to be used anywhere and at any time, than those that require users to be co-located.

We have experimented with asking more complex questions and tried to elicit more detailed responses from users. Although not answered by the same high percentages of users as more simple demographic questions, these forms of information gathering were still used by many participants and have proved very useful. In Hungry Yoshi, players were rewarded with in-game 'tokens' for answering questions, and 6,115 players did so, which is 17% of registered users and actually more than the number of players who ever scored a point in the game. In the World Cup Predictor, where a feedback section was included but no rewards offered for answering questions, 932 or 11% of registered users gave at least one response. We have gained greatly from information gained through these channels, being made aware of numerous bugs, received and subsequently implemented feature requests and setup telephone interviews with participants. Users also seemed happy that their opinion was being valued in this way, with one stating "*I find it really nice that [you are] contacting me and asking me my opinion. I guess it's a really nice thing*."

Most app stores allow users to enter reviews, to be publicly shown alongside the application in the download area. Yet despite this common mechanism for providing comment, users appear far more likely to use in-app feedback mechanisms than store-provided ones. As an example, the World Cup Predictor has had 2 reviews in the store compared to, as mentioned, comments from 932 users in the application, despite there being no reward for doing so.

CONCLUSION

Our current work explores exposing more of apps' software structure to users, so as to engage them more directly in not just evaluation but redesign, and new tools and techniques for categorising users to let us better handle large-scale trials. Although mass participation in evaluation, ideation and design continues to develop, this paper offers some preliminary points that others considering similar trials may consider. First, static or singular metrics for what is considered an 'active' user are problematic. Instead, we constantly restructure our definitions for 'user'. Second, different levels of user activity may need to be accounted for in application design. For instance, can your application design make a virtue of a large numbers of 'single shot' users? Third, integrate data collection and evaluation mechanisms into the application. The inclusion of a feedback section, where users can answer questionnaires or leave open comments is very useful. Fourth, note that the tendency for individual use rather than co-located social use of applications may be strong. In our case, social functionality in Facebook was far more widely used.

REFERENCES

1. McMillan, D., et al. Further into the Wild: Running Worldwide Trials of Mobile Systems. *Proc. Pervasive* (2010), 210-17.
2. Hall, M. et al. Adapting Ubicomp Software and its Evaluation. *Proc. ACM EICS* (2009), 143-148