Playing while Testing: How to Gamify a User Field Evaluation

Amon Rapp

University of Torino

C.so Svizzera, 185

Torino, 10149, Italy

amon.rapp@gmail.com

Alessandro Marcengo

Telecom Italia

Via Reiss Romoli, 274

Torino, 10148, Italy

alessandro.marcengo@telecomitali
a.it

Rossana Simeoni

Telecom Italia

Via Reiss Romoli, 274

Torino, 10148, Italy

rossana.simeoni@telecomitalia.it

Luca Console

University of Torino

C.so Svizzera, 185

Torino, 10149, Italy

lconsole@di.unito.it

Abstract

In this paper we claim that gamification can lead to interesting results in the evaluations of interactive systems. A case study will be illustrated.

Author Keywords

Gamification, Engagement, Evaluation, User Test.

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – evaluation/methodology.

General Terms

Human Factors, Experimentation.

Introduction

Since now, HCI community has paid attention at the gamification practices mainly as tools to enhance user engagement in the usage of a given application: "the use of game design elements in non-game contexts" [2] is intended as a set of design techniques that can improve the ability of a service to generate a long lasting involvement in its users. However, it has not yet given sufficient attention to the possibility of using gamification within the design process of digital artifacts: while some methods such as role play gaming and make-believe are tools available to the designers since long time, it does not seem a common practice to

use game mechanics during the evaluation stages of an interactive system. Laboratory tests allow researchers, through the experimental protocol, to guide users in the evaluation of specific and critical functionalities of a system: tasks motivate participants to use certain features rather than others, that perhaps would not be used in a less controlled situation. However, they create artificial contexts that could twist the results when we test applications with "social features" (e.g. comment, post, share, etc.). In fact, their usage requires intrinsic motivations that can be generated only in connection with social relationships and communication habits that are not easily replicable within a laboratory setting: they need a pre-existing community to run at their best and continuous social feedback to be meaningful for their users (e.g. the use of a comment feature has a sense only if the post has the possibility to receive a reply in a short time from another user). Therefore, the laboratory usability tests, as Greenberg and Buxton [3] had already pointed out, do not seem suitable for all contexts. Field tests, otherwise, allow users to try a system without constraints, in a context close to that of the "everyday life". The lack of punctual tasks, however, subtracts control from the researchers' hands, not allowing them to obtain reliable data on all the critical points of the system under examination. From these premises, gamifying a field evaluation session could be an optimal solution to balance the need to test an application "in the wild" [5] and, simultaneously, to motivate testers to use every features of a system, even in the absence of an experimental protocol: game missions can substitute the laboratory tasks, with the advantage that they can be internalized more easily and perceived in a less abstract way by the players, since they could be directed to the goal of winning the game or obtaining a

certain reward. Within the WantEat project [1], a suite of applications that aims to increase communication between objects and people in the food and wine domain[1], we tried without success, using laboratory and field tests, to gather significant insights on the social features of WantEat Mobile App[4]. Hence, we set a gamified field test, inserting game mechanics in traditional evaluation methodologies, in order to create a large-scale engaging experience, in which multiple users, at the same time, were able to stress the system while enjoying themselves.

Gamifying a Field Test
The gamified evaluation of WantEat Mobile App took place during the international food exhibition "Cheese 2011" (a huge event that appealed about 300,000 visitors), held in the town of Bra (Italy) between 16 and 19 September 2011. During the four days of the fair, the application has been installed on the users' iPhones. Scattered through the fair were available 10 cheeses, recognizable by WantEat: main purpose of the evaluation game was to recognize at least five of these cheeses with the mobile camera, taste them and use the social features of the app to add information to these products. Every action performed (e.g. writing a review, applying a tag, etc.) allowed the user to earn 500 points: at 6000 points she was awarded with a T-shirt with the Application logo. Each participant received the game instructions and a map of the fair that highlighted the areas in which the application was

---

[1] WantEat is an interactive cross-media system, consisting of smartphone, tablet and web applications. With WantEat mobile app users can a) frame with the mobile camera a product label, b) obtain information about the product and how this product is in relationship with other objects and people, c) add information to the product (e.g. with tags, comments, votes)

fully working. Contextually at the claim of the prize, users were asked to answer a questionnaire, through which, using 4 Likert scales accorded to four different dimensions (ease of usage, efficiency, engagement and usefulness), we tried to gather feedback about the experience with the application. In addition to the basic actions of 500 points, users could accomplish special missions that required a lot of energy and time, but allowed them to earn from 10.000 to 20.000 points: these objectives stimulated social cooperation between users, raising the level of challenge by promoting the exploration of the entire fair area. For example, users were suggested to exchange a special identification coin (provided with game instructions) using the application communication features, or to discover hidden objects and secret places that could have been recognized by the application. In this way, we encouraged everyone to play, offering a relatively easy goal to reach (6000 points) and, at the same time, optional objectives and incremental problems, that could motivate harder players to satisfy their willingness to play. A live leaderboard at the installation base maintained all participants informed of their current score. Furthermore, a web application was deployed to support the whole game: accessing with their accounts, users was able to retrace their application usage experience within the fair, seeing the earned points, the actions taken, the products tasted and the people met. The evaluation led to interesting results in terms of user engagement and participation: 157 people attended the field test. Users performed a total of 2134 actions. Analyzing the nature of the actions it was possible to find out that game mechanics have led the testers to deeply use the application: users commented, voted, tagged the products recognizable by the system, generating a social network from

scratch, in a limited amount of time (4 days). Participants also explored and used features that were not specifically required in the game instructions: this shows an interest for the service as a whole. More than half of the sample did not stop at 6000 points (the minimum amount to obtain the prize), continuing to play even if no reward was expected for the winner of the competition. The leaderboard and the structure of the available missions were sufficient to create a challenge able to motivate the users in the game. Moreover the comments posted were always relevant to the type of the product on which they were made: although it was possible to game the system, through the posting of empty comments or the mechanical repetition of the same action, users performed actions coherent to the product on which they were applied. These aspects suggest that the format of the game was internalized by the participants and that the challenge was welcomed in a profound way. The questionnaires gathered at the end of the test made possible to collect useful data and requirements for the further improvement of the application design. For example users rated positively the intuitiveness of the opinion expression mechanism about the products recognizable by the system (3.39 out of 4 SD 0.62) and the quickness of the tag and vote features (3.64 out of 4 SD 0.56), while expressed a low interest for capabilities facilitating new people encounter through products (2.43 out of 4 SD 0.83). It also emerged how users prefer to provide opinions and comments about the products they tasted rather than consider other opinions when they have to make a choice: it seems to be privileged the expression of personal identity through the posting action, rather than the use of information provided by others.

Suggestions for the Gamification of a user field test

WantEat at Cheese 2011 highlighted how the gamification of a field test can yield to excellent results. In particular, it allows to overcome the cold start effect, a social network typical problem, which can prevent the participation of users, and the collection of useful and valid requirements during an evaluation session. It is possible to think that the gamification of an user test may also be a good remedy about the artificiality of the laboratory context: although the game situation does not re-create the daily experience of use with a non recreational service, which commonly takes place in not-playful contexts, it manages to generate genuine motivations that could be similar to those experienced by people during their everyday life, since they are moved by practical objectives and concrete aims (e.g. the desire to excel in a competition). In conclusion, from our fieldwork we can provide a first set of guidelines in order to gamify the evaluation session of an interactive system. First, the playful part has to be perfectly integrated with the non-recreational part of the evaluation experience: rewards, objectives and game mechanics have to be melted with the purposes of the testing plan, without appearing as an external layer added to a situation already self-contained. Then, the design of missions, constraints and opportunities should be carried out with a deep understanding of the context in which the experience will take place: understanding always precedes the design of the experience itself. Secondly, the gameplay should be well balanced with regard to the difficulty of achievable goals: an incrementally complexity of the game objectives, structured in easier mandatory tasks and in optional more challenging missions, provides the necessary motivation to all users to reach a minimum

level of participation and to continue the game experience if they like it. In addition, the definition of the missions of the game should promote cooperation and sharing among the participants. Creating differentiated targets that leverage on the competition, but need for their fulfillment of a social co-operation between users, is the right way to get active participation in the social network of the application under test. Finally, the importance of the gaming stage in the design of truly immersive gaming experience should not be underestimated: the careful articulation of sub-areas, in which sub-objectives can be achieved, and the balance between exploration and control on the surrounding spaces must never be lacking. From these first guidelines it could be possible to think a new method for evaluating interactive systems in the wild that gives a central role to game mechanics.

References

[1]   Console L. et al. Interacting with Social networks of intelligent things and people in the work of gastronomy, to appear ACM Trans on TIIS, 2013, Short version in Proc. of ENTER2012, eTourism Present and Future Services and Applications 2012, Springer Wien.

[2]   Deterding, S., Dixon, D. Khaled, R. and Nacke, L. 2011. From Game Design Elements to Gamefulness: defining "gamification". In Proc. of Mindtrek *2011*.

[3]   Greenberg, S. and Buxton, B. Usability evaluation considered harmful (some of the time). In Proc. CHI 2008, ACM Press (2008), 111-120.

[4]   Marcengo, A. Rapp, A., Console, L., Simeoni, R. Evaluating WantEat: A social network of people and objects. Book Chapter in Rebelo, F., Soares, M. Advances in Usability Evaluations. Part 2. CRC Press, Taylor & Francis, 2012.

[5]   Rogers, Y. Interaction Design Gone Wild: Striving for Wild Theory. Interactions 18, 4, 58-62, 2011.