Exploring Social Accountability for Pervasive Fitness Apps

Yu Chen
Human Computer Interaction Group
Swiss Federal Institute of Technology,
Lausanne, Switzerland
yu.chen@epfl.ch

Jiyong Zhang
Artificial Intelligence Laboratory
Swiss Federal Institute of Technology,
Lausanne, Switzerland
jiyong.zhang@epfl.ch

Pearl Pu
Human Computer Interaction Group
Swiss Federal Institute of Technology,
Lausanne, Switzerland
pearl.pu@epfl.ch

***Abstract*—Mobile fitness applications have gained increasing popularity to help users walk and exercise more. A key component in such apps is its ability to motivate users. Traditional gamification methods have focused on competition such as leaderboard for community users, self-reflection for individual users, or a combination of the two. Motivated by recent work showing a promising effect of social capital, we have designed and developed a mobile game, HealthyTogether, based on such ideas. We are further interested in how users behave in different settings of gamification methods compared to a baseline. To this end, we have designed and conducted an in-depth user study (N=24) involving 12 dyads playing these games in 4 conditions over a period of two weeks. We report here the design of the application as well as the user study. Among the various rewarding schemes, one that uses a hybrid concept of competition and social accountability gives the most desirable outcome.**

Keywords-health; pervasive fitness applications; gamification; competition; social accountability.

I. INTRODUCTION

Wellness and lifestyle change have gained significant attention in recent years. Both research communities and commercial sectors are putting increasing effort to develop wearable sensors and mobile applications that help and “nudge” individuals to increase their physical activities, eat healthier diet, better manage their sleep and stress, and engage in social lives with family and friends.

Many of the applications use gamification -- the use of game elements in non-game context [14] -- to motivate users to exercise more. Concrete methods include competition such as leaderboard for community users [9], self-reflection such as visualization for individual users [2][5][6], or a combination of the two [7][12]. Recent work shows a promising effect of social responsibility for the sake of helping each other especially among family members, friends, and people who share same interests and goals [1]. We define this concept as social accountability, which refers to a person’s awareness of another person’s goal and rendering himself/herself responsible to the goal’s successful fulfillment.

In this work, we are interested in how users behave in various settings of gamification methods: competition, social accountability, a hybrid model of a mixture of competition and social accountability, and a baseline non-social setting. To this end, we have developed a mobile application, HealthyTogether, which enables dyads to participate in physical activities together, send each other messages, and earn badges. We use this application as an experimental platform for an in-depth user study (N=24) to evaluate how the various reward schemes influence users’ exercises and social interactions, both quantitatively and qualitatively.

The rest of the paper is organized as follows. After covering related work in Section II, we present HealthyTogether in Section III, user study design in Section IV and results in Section V. We conclude this paper in Section VI.

II. RELATED WORK

Self-reflection is considered as a self-motivation and successful strategy for pervasive health applications. Research prototypes such as Shakra [4] and Houston [13] and commercial products such as Fitbit and Nike+ all visualize users’ daily activities to achieve self-reflection. A number of systems also present physical activities using metaphors. UbiFit Garden [12] visualizes users’ daily steps by the growing status of plants. The more activities a user takes, the healthier his plant looks. Fish’n’Steps [7] uses the metaphor of fish tank to visualize users’ step count. Recent work has employed informative art as a visualization tool, such as research prototype Spark. The above work mainly motivates users in an individual setting.

Social interaction, including peer-support, cooperation, competition and belonging to a group has been a clear motivator for wellness activities [1][8]. Commercial products have widely adopted competition to motivate user, such as Nike+ and Fitbit. Fitster [9] is a research prototype that visualizes users’ steps in a social network and places users in a virtual competition environment. Kukini [16], Fish’n’Steps [7] and Life Coaching Application [11] support competition by helping users to form a team and explicitly introducing social interaction and social pressure.

Research also shows that social communication can motivate users to exercise. Consolvo et al. [13] show that message exchange can help users to increase the responsibility and give support to group members [9]. Champbell et al. [16] suggest that communication using everyday fitness games can help enhance players' social relationship and sustainability in everyday fitness.

Social accountability has been shown to be effective in helping users to achieve goals. Ahtinen et al. [1] have found out that connecting with family members and loved ones can

Figure 1. a) The FitBit tracker, b) FitBit in use, and c) the Samsung Galaxy.

help motivate users; connecting with people with similar wellness targets from communities within short distances can also increase motivation towards wellness activities. Stickk [15] helps users to achieve their goals by allowing them to appoint another person to monitor the progress and verify the accuracy of progress report. They can add supporters who can encourage them by commenting on their progress. Users can also put stake on the goal and specify where the stake would go if they fail in the goal. GoalSponsor [3] allows users to set up goals and sponsors whom they should be accounted for. A sponsor can be a friend, a professional in healthcare, or someone who has accomplished the goal successfully. Users are more committed in fulfilling the goals either because they do not want to let others down or because they do not want to lose reputation in front of others [3]. In the above work, the structure includes one person who has a goal to fulfill and another person who monitors the progress.

To the best of our knowledge, current fitness applications have not well studied interaction schemes in which users mutually account for each other's progress. We are motivated to investigate social accountability factor in pervasive fitness applications using an experimental platform called HealthyTogether.

III. HEALTHYTOGETHER

HealthyTogether is a mobile application that involves a pair of users to exercise together, and it is implemented on the Android platform. To measure users' physical activities, we choose the Fitbit sensor (as shown in Figure 1 a) and b)) among many off-the-shelf sensors as the activity tracker for our HealthyTogether system. Here, we describe the user interface design and the underlying rewarding mechanisms.

A. Game Rules

We designed a series of rewarding mechanisms for HealthyTogether in order to investigate the impacts of different social settings in pervasive fitness application. A user can win badges based on *Karma Points*, which are calculated as below.

$$kp(u) = \alpha \cdot steps(u) + \beta \cdot steps(u')$$

Based on different α and β values, HealthyTogether provides the following three reward settings:

- **Competition setting**, where $\alpha = 100\%$, $\beta = 0$;
- **Accountability setting**, where $\alpha = 0$, $\beta = 100\%$;
- **Hybrid setting**, where $\alpha = 80\%$, $\beta = 20\%$.

In **competition** setting, a user's Karma points are calculated purely by his or her steps. To gain more badges, a user only needs to focus on his or her own activities even if he is exercising with a buddy. Thus, we name this rule competition setting. In **accountability** setting, a user's Karma points are calculated by the steps of the buddy. Therefore, the more he encourages his buddy to exercise, the more points he earns. Thus, we name it accountability setting. On the other hand, even if a user does not move at all, he can still gain badges from the buddy's activities. In the **hybrid** setting, a user's Karma points are calculated based on both his (her) own and that of the buddy, proportionally. The idea behind this reward scheme is to encourage competition while also motivating users to cheer each other. Initially, we set $\alpha = 80\%$ and $\beta = 20\%$ based on the well-known Pareto Principle. In the future, we will also experiment different ratios of competition and social accountability, such as 50%-50% and 20%-80%.

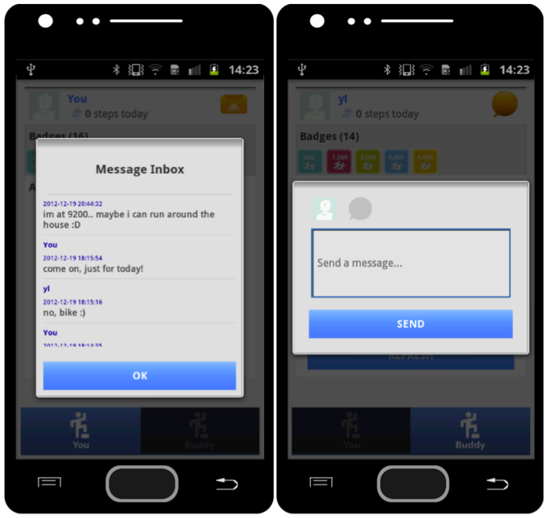
B. Badges

HealthyTogether issues badges based on *kp(u)*. The first badge is issued if *kp(u)>500*, to help users get started in a short time. This number is followed by 1,000 and 2,000 and then increases by every 2,000 points. HealthyTogether calculates Karma points in a daily basis but accumulates badges over time. For example, if a user earns 5,353 Karma points in a day, he can gain 4 badges, i.e., 500, 1000, 2000 and 4000. If a user earns 5,353 and 6,086 points in the first two days, he can gain 4 and 5 badges respectively and a total of 9 badges.

C. Interaction Design

The main interface of the HealthyTogether system is shown in Figure 1 c). It contains a 'self' tab and a 'buddy' tab. Each tab displays information about step count, active time and badges of the current day. We use a pie chart to visualize the proportion of time that a user is in various

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-353-7



a) b)

Figure 2. Screenshots of messaging components.

activity modes, i.e., sitting, lightly active, fairly active and very active.

The badge area displays the total number and the badges that a user has earned. The badges are accumulated over time. In Figure 1 c), the user has earned 6 types of badges with a total number of 16. When he/she clicks on a badge icon, a dialog box pops out explaining the details of this badge type, including how many badges the user has earned and how he/she earned the badges.

There is a messaging button on the top-right corner of each page. When it is clicked, users can either view message history (Figure 2 a)) or send messages to their buddies (Figure 2 b)). Users will receive a vibrated notification when buddies send them new messages.

IV. USER STUDY

To study different game settings in real situations, we designed an exploratory deployment study. We first conducted a user study (Study 1) that spans for six continuous working days, which was divided into a three-day control session and a three-day experimental session. After conducting the study, we were able to discover some interesting results. For example, participants suggested that we extend the study to two weeks, excluding the weekends, so that the control and the experimental sessions span over identical days of the week, thus minimizing the influence of a given day's schedule to the physical activities being monitored. For example, a user may work in the office on Mondays but conduct experiments in the laboratory on Wednesdays. We therefore conducted the second study (we name it Study 2) with duration of two weeks. We refer to the control session as Phase I and experiment session as Phase II in both studies.

A. Participants

We recruited the participants on campus via word-of-mouth. After one person signs up, we asked her to invite a buddy of her choice to join. Their ages range between 22 and 33 and they never used Fitbit before. We required that each dyad should not work in the same office or too close to each other. We offered all participants a 50CHF gift card as compensation for their time.

B. Materials

We provided users with an Android phone with 3G SIM card and a Fitbit. Three users requested to use their own Android phones because it would be more convenient for them. We checked that their phones are compatible for installing HealthyTogether.

C. Procedure

Both Study 1 and Study 2 were structured as a two-phase, within-subjects design. Phase I allowed participants to become accustomed to using Fitbit and allowed us to collect baseline fitness data. In this phase, all participants use Fitbit alone without connecting with buddies. In Phase II, participants in baseline groups (Group A1- A3) continued to use only the Fitbit while groups in social settings (Group B1-B3 in competition setting, C1- C3 in accountability setting, D1- D3 in hybrid setting) started to use Fitbit and HealthyTogether with buddies. The structure of Study 1 was the same as the Study 2, except the duration was extended to two weeks, with both phases extended from three days to five days.

At the beginning of the study, we invited each pair of participants to our laboratory and helped them to set up their Fitbit accounts. We also had a short interview with them on their experience in using fitness sensors. At the end of Phase I, we invited participants in social settings to our laboratory again to install HealthyTogether with different game rules.

Since our user study lasts for up to two weeks, we requested participants to fill in a daily experience survey related the study. At the end of each day, we sent a reminder email with the survey link to participants asking them whether they have anything to share with us about their experience using Fitbit or HealthyTogether. The survey only contains one question: "Do you have anything to share with us on your experience using Fitbit/HealthyTogether today?" Daily survey not only helps us to gain an in-depth understanding of users' experience, but also facilitates us to explain their step data with activities during that day.

At the end of the study, we organized a semi-structured interview. We invited two participants in each group to attend the session together, so that they could share their stories. We did not ask a fixed set of questions, but mediated the session with the following aspects: overall impression, experience, attitudes and aptitudes, motivation of usage, social relationship.

V. RESULTS

In this section, we report both quantitative and qualitative results collected in Study 1 and Study 2. To facilitate describing results, we encoded the two participants in each

 ISBN: 978-1-61208-353-7

dyad with 'a' and 'b' together with their group ID. For example, we encode the two participants in Group C1 as 'C1a' and 'C1b' respectively.

A. *Quantitative results*

Study I

We first investigate users' step count across the 6 days. The overall average daily step count is 7,439 (min=3,185, max = 11,490). We then compare the average daily step count between baseline group (A1) and groups who used HealthyTogether (B1—D1) to evaluate the effectiveness of social interaction incentives (see Figure 3). Results show a slight decrease of steps from Phase I to Phase II across all groups. One explanation is the novelty effect of using Fitbit in the first 1—2 days, as reflected in daily survey. One interesting finding is that in Group A1 the average step count decreased by 20.4% but in Group B1—D1 it decreased by only 10.6%. This implies that HealthyTogether with social settings could help users to persist in physical activities.

We further compare Group B1—D1 with different rules of calculating the accumulative Karma points of the three days. Results show that users in Group B1 (competition setting) have largest difference of Karma points (Δkp = 14,556) compared with Group C1 (Δkp = 839) and Group D1 (Δkp = 2,982). One explanation is that participants in Group B1 focus more on their own performance compared with other groups. In other words, it implies that the social accountability factor, applied in Group C1 and D1, could lead to more balanced performance of physical activities between buddies.

Meanwhile, users exchanged 72 messages, shared by Group B1 (N=43), Group C1 (N=27) and Group D1 (N=2). The distribution shows that Group C1 (accountability setting) and D1 (hybrid setting) interact more compared with Group B1 (competition setting). Particularly, participants in Group D1 exchanged 59.2% more messages than Group C1. This implies that hybrid setting is most useful to encourage participants to interact with each other.

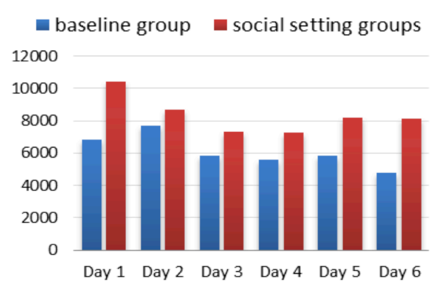


Figure. 3. Distribution of average daily steps for groups with non-social setting vs. social setting in Study 1.

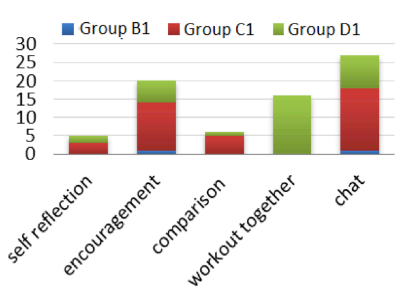


Figure 4. Topic distribution of messages exchanged using HealthyTogether in Study 1.

TABLE I. MESSAGE TOPICS AND EXAMPLES

Topics	Examples
Self-reflection	*"im at 9200.. maybe i can run more"*
Cheering	*"you should make it 8k for a new badge"*
Comparison	*"the first time i am higher than you!!!"*
Workout together	*"we should walk around the floor together to take a break:)"*
Chat	*"feeling so tired now, go to bed soon"*

Several topics emerged from the analysis of message content, and we present them with sample text in Table I. The distribution of each topic shared in Group B1—D1 is shown in Figure 4. It reveals the following phenomenon: 1) in total, there are 27 chat messages, which have the largest share; 2) encouragement is the main topic that is relevant with physical activity (20 messages); 3) Group C1 has the largest share (13 messages) in encouraging messages; 4) the major topic of Group D1 is workout together (16 messages), and it only appears in Group D1. The results imply that hybrid setting introduces most conversation in the topic of workout together.

Study 2

The deployment of Study 2 is the same as Study 1 except for the duration. We first verify whether discoveries in Study 1 still exist in Study 2. The average daily step count is 9,501 (min=3,200, max = 24,334). Figure 5 is a distribution of average daily steps between baseline groups (Group A2, A3) and social setting groups (Group B2, B3, C2, C3, D2, D3) in two weeks. The distribution shows a steady increase of average daily steps in social groups from Phase I to Phase II. Comparing social setting groups and baseline groups, we found average steps increased by 9.8% from Phase I to Phase II in social setting groups but decreased by 10.1% in baseline groups. This is consistent with implication in Study 1 that social settings could motivate users to exercise compared to when they walk alone.

We then compare groups using HealthyTogether in different social settings. Figure 6 shows each participant's average daily steps in Phase I vs. Phase II. The average daily steps in competition groups (B2 and B3) have increased from 9,747 to 10,128 (Δ=381). In accountability groups (C2 and C3), this number increased from 8,888 to 9,717 (Δ=829), and in hybrid setting (D2 and D3) from 10,762 to 12,437 (Δ=1,675). The average daily step increase of hybrid group is 51% more than that of accountability group and three times more than that of competition group. If we assume that participants have the same schedule of the same workday in different weeks, and that Phase I is a baseline for participants, then the above results suggest that hybrid setting encourage users to walk more.

 ISBN: 978-1-61208-353-7

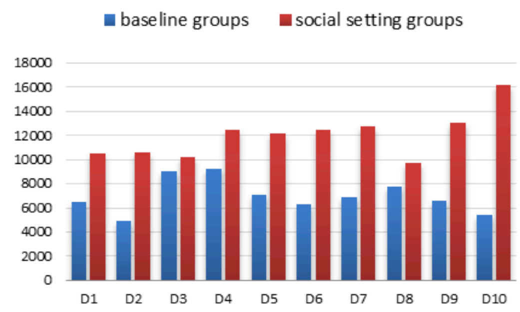


Figure 5. Distribution of average steps non-social setting vs. social setting in Study 2.

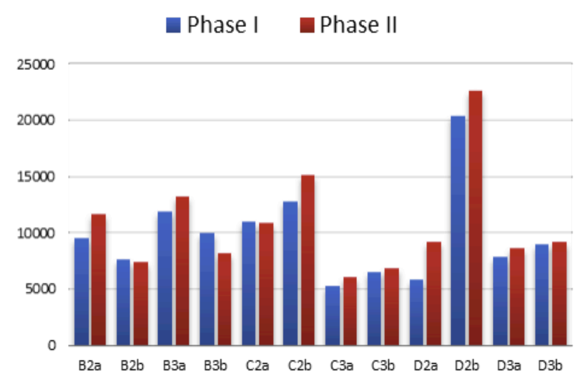


Figure 6. Average step count Phase I vs. Phase II in Study 2.

We further compare the steps between participants in a dyad. As shown in Figure 6, dyads from both accountability groups (C2 and C3) and hybrid groups (D2 and D3) have increased steps together from Phase II to Phase I. On the other hand, in competition groups, average step count of B2a and B3a have increased 22.5% and 11.0% respectively from Phase I to Phase II while this number decreased for their buddies (2.1% for B2b and 17.8% for B3b). This implies that accountability factor helps dyads to walk more steps together. In Study 1, we found participants in accountability group and hybrid group are trying to achieve a balanced number of badges. Even though we do not have the same finding in Study 2, the results concur with the implication in Study 1 that users have more balanced working performance that both users in a dyad improve together.

We then analyze the 86 messages sent between the dyads in Study 2. Participants in hybrid groups sent 58 messages, which is more than twice the number of accountability group (N=21) and seven times more than that of competition group (N=7). Figure 7 shows the distribution of message topics within groups of the three social settings. Different from Study 1, messages with topics about self-reflection and encouragement have the largest share in the total of messages (27.8% for both topics). We also discover that hybrid groups have the largest share of messages (81%) in the topic of self-reflection. The distribution accords with what we have found from Study 1 that 1) hybrid groups have most share of messages (75%) in the topic of workout together, and 2) encouragement is the major topic (54%) in accountability groups. If we consider the number of messages as one metric to evaluate social interaction, results in Study 2 further provide evidence that hybrid setting is more likely to stimulate social interactions.

B. Qualitative analysis

In this section, we report the results we found through the logged daily survey and post-study interviews from both Study 1 and Study 2.

During the user interview, we found some qualitative results that are related to our findings from quantitative analysis. Overall, the feedback about HealthyTogether was very positive. First, HealthyTogether has helped them to compare with each other. As B1b said in the interview, "to check her (buddy's) steps and compare with mine is most important for me". Second, they could interact with each other via HealthyTogether. B1a reported in survey: "I received message on the first day, so the next day I intentionally walked more between the buildings."

We also found some evidence that the accountability factor (applied in accountability and hybrid setting) could help users to care about each other. As C1b reported in her daily survey, "I discovered his step is twice more than mine. As his badges depend on my steps, I feel I should walk more in order not to discourage him." This supported what we found in Study 1 that the participants in C1 have more balanced performance when using HealthyTogether. Additionally, when we asked whether users about their social relationship before and after using HealthyTogether, participants in D3 revealed that they were already very close friends. Participants in C1, C2, and D1, and D2 reported that they had developed further relationship with buddies after using HealthyTogether. For example, D1b said: "Even though we are colleagues, we did not talk much. Finding something to do together rapidly brought us closer." For another example, C2a reported that she knows more about her buddy: "When I woke up, my buddy already had 3,000+ steps... She is already on campus..." However, we did not find the same report from competition groups. This suggests that social accountability could be helpful for a user to enhance social relationship with the buddy.

Participants have reported their concerns regarding competition setting. Both B1a and B1b reported competing with each other cause demotivation. "I knew I would never beat him because he needs to walk a lot from home to

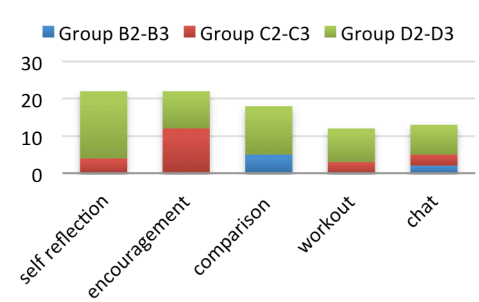


Figure 7. Topic distribution of messages in Study 2

work…" –B1b. "I clearly had advantage in winning the game and sometimes I'm afraid walking too much makes her pressure…" –B1a. B3a also reported that when he noticed his buddy walking less, he also became lazier. Admittedly, the demotivation effect could be caused by the unbalanced abilities in physical exercise. This suggests that choosing a suitable exercise buddy can be important in order to maximize the effect of competition.

Evidence shows that the hybrid setting is more preferred than the accountability setting. "I feel a little bit wired if my badges only depend on my buddy's steps. Is it a little bit de-motivating me?" mentioned by C3a. By comparison, instead of 'depending on' others, both D1 and D3 have reported they arranged activities together, such as walking to a farther away cafeteria to have lunch (D1), and going to Zumba course together (D3). D2a reported in his survey that his buddy gave him suggestions to increase the steps: "Without trying too hard, I almost reached 10k steps. Seeing his progress during the day motivated me to move more. It was also useful to talk to him (via message) - he gets a high step count by walking around while reading articles. I also did this walking in circles thing for about 1h, which helped my brainstorming." Even though D2b (avg=20,372) have clear advantages over D2a (avg=5,852) in Phase I, this discrepancy did not demotivate either of them. Instead, D2a's average daily step has increased by 57.3% (avg=9,207) from Phase I to Phase II and D2b has increased by 11.2% (avg=22,661). As D1a said: "Helping others to become better is a `plus' rather than `minus' to your life."

VI. CONCLUSIONS

In this work, we have developed a mobile application called HealthyTogether that allows dyads to participate in daily physical exercises as a game. We conducted an in-depth user study with 12 dyads with a period of up to 2 weeks and compared participants using HealthyTogether in 3 social settings and a baseline non-social setting. Results show that social settings, even in the competition mode, can help users to persist more in physical activities compared with baseline group. Additionally, the hybrid setting is more likely to motivate users to walk more and more actively help others. Furthermore, the number of messages sent between participants in the hybrid settings is 8 times more than those in the competition setting and twice of those in the accountability setting. Integrating social accountability factor is also promising to enhance social relationship between buddies.

In the future, we plan to conduct longitudinal studies with more users in various conditions to validate our findings with statistical analysis.

ACKNOWLEDGEMENT

We thank Swiss National Science Foundation for sponsoring this research work. We are also grateful for all anonymous reviewers for their valuable feedback.

REFERENCES

[1] A. Ahtinen, M. Isomursu, M. Mukhtar, J. Mäntyjärvi, J. Häkkilä, and J. Blom. Designing social features for mobile and ubiquitous wellness applications. The 8th international Conference on Mobile and Ubiquitous Multimedia (MUM'09), ACM. Nov. 2009.

[2] C. Fan, J. Forlizzi, and A. K. Dey. Spark Of Activity: Exploring Informative Art As Visualization For Physical Activity. Proc. of the ACM Conference on Ubiquitous Computing (Ubicomp'12), ACM. Sept. 2012, pp. 2–5.

[3] GoalSponsor. https://www.goalsponsors.com. Accessed: August 6, 2014.

[4] I. Anderson, J. Maitland, S. Sherwood, L. Barkhuus, M. Chalmers, M. Hall, and H. Muller. Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. Mobile Networks and Applications, 12(2-3), 2007, pp. 185-199.

[5] I. Li, A. K. Dey, and J. Forlizzi. A stage-based model of personal informatics systems. Proc. of the SIGCHI conference on Human Factors in computing systems (CHI'10), ACM. Apr.2010, pp. 557-566.

[6] I. Li, A. K. Dey, and J. Forlizzi. Understanding my data, myself: supporting self-reflection with ubicomp technologies. Proc. of the ACM Conference on Ubiquitous Computing (Ubicomp'11), ACM. Sept. 2011, pp. 405-414.

[7] J. Lin, L. Mamykina, S. Lindtner, G. Delajoux, and H. Strub. Fish'n'Steps: Encouraging physical activity with an interactive computer game. UbiComp 2006, pp. 261-278.

[8] J. Maitland, and M. Chalmers. Designing for peer involvement in weight management. Proc. of the SIGCHI conference on Human Factors in computing systems (CHI'09), ACM Press (2011), pp. 315-324.

[9] N. Ali-Hasan, D. Gavales, A. Peterson, and M. Raw. Fitster: social fitness information visualizer. In CHI'06 Extended Abstracts on Human Factors in Computing Systems. Apr. 2006, pp. 1795-1800.

[10] P. Klasnja, S. Consolvo, and W. Pratt. How to evaluate technologies for health behavior change in HCI research. Proc. of the SIGCHI conference on Human Factors in computing systems (CHI'11), ACM. May 2011, pp. 3063-3072.

[11] R. Gasser, D. Brodbeck, M. Degen, J. Luthiger, R. Wyss, and S. Reichlin. Persuasiveness of a mobile lifestyle coaching application using social facilitation. Persuasive Technology 2006, pp. 27-38.

[12] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, and J. A. Landay. Activity sensing in the wild: a field trial of UbiFit garden. Proc. of the SIGCHI conference on the SIGCHI conference on Human Factors in computing systems (CHI'08), ACM. Apr. 2008, pp. 1797-1806.

[13] S. Consolvo, K. Everitt, I. Smith, and J. A. Landay. Design requirements for technologies that encourage physical activity. Proc. of the SIGCHI conference on Human Factors in computing systems (CHI'06), ACM. Apr. 2006, pp. 457-466.

[14] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, and D. Dixon. Gamification: using game-design elements in non-gaming contexts. In CHI'11 Extended Abstracts on Human Factors in Computing Systems, ACM. May 2011, pp. 2425-2428.

[15] Stickk. http://www.stickk.com. Accessed: August 6, 2014. 2

[16] T. Campbell, B. Ngo, and J. Fogarty. Game design principles in everyday fitness applications. Proc. of the 2008 ACM conference on Computer supported cooperative work (CSCW'08), ACM. Nov. 2008, pp. 249-252.