HealthyTogether: Exploring Social Incentives for Mobile Fitness Applications

Yu Chen
Human Computer Interaction Group
Swiss Federal Institute of Technology (EPFL)
CH-1015, Lausanne, Switzerland
yu.chen@epfl.ch

Pearl Pu
Human Computer Interaction Group
Swiss Federal Institute of Technology (EPFL)
CH-1015, Lausanne, Switzerland
pearl.pu@epfl.ch

ABSTRACT

A crucial element in many mobile fitness applications is *gamification* that makes physical activities fun. While many methods focus on competition and individual users' interaction with the game, the aspect of social interaction and how users play games together in a group remains an open subject. To investigate these issues, we developed a mobile game, HealthyTogether, to understand how users interact in different group gamification settings: competition, cooperation, or hybrid. We describe the results of a user study involving 18 dyads (*N*=36) over a period of two weeks. Results show that users significantly enhanced physical activities using HealthyTogether compared with when they exercised alone by up to 15%. Among the group settings, cooperation (21% increase) and hybrid (18% increase) outperformed competition (8% increase). Additionally, users sent significantly more messages in cooperation setting than hybrid and competition. Furthermore, physical activities are positively correlated with the number of messages they exchanged. Based on the findings, we derive design implications for practitioners.

Author Keywords

Health; mobile fitness applications; group interaction; competition; cooperation; gamification.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Citizens of the industrialized nations are susceptible to inactivity and stress due to sedentary and highly competitive lifestyles, which can lead to serious health problems such as cardiovascular diseases, diabetic type II, depression, and cancer. The World Health Organization defines health as a "state of complex physical, mental and social well-being and not merely the absence of disease or infirmity". Both research communities and commercial sectors are putting an 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.

Wellness and lifestyle management systems using wearable sensors and mobile apps are on the rise [22]. Many of these mobile fitness applications use gamification to motivate users to exercise more. *Gamification* is the use of game elements in non-game context in order to engage users [7]. Traditional methods have focused on competition such as leaderboard for community users [2], self-reflection such as metaphorical visualization of garden flowers for individual users [8, 13, 14], or a combination of the two [15]. Recent work has identified crucial evidence supporting social interaction as a key element to motivate users to perform physical activities in designing pervasive fitness apps [4, 5]. Such social interaction schemes are called **social incentives**. In particular, such social interaction among group members includes sharing physical activities, cooperating and competing. However, a detailed examination of users' behavior in a group environment remains an open subject of study.

In our work, we are interested in how users in a group environment behave in various settings of gamification methods: competition, cooperation and a hybrid model of competition and cooperation. 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-situ user study (*N* = 36) to evaluate how the various rewarding schemes influence users' exercises and social interactions.

Our analysis shows that users significantly enhanced physical activities using HealthyTogether compared with when they exercised alone by up to 15.1%. Among the social incentives, cooperation and hybrid of cooperation and competition by far outperformed competition. Additionally, messages exchanged between participants in cooperation setting significantly outnumbered that in

Chinese CHI '14, April 26 - 27 2014, Toronto, ON, Canada

ACM 978-1-4503-2876-0/14/04…$15.00.
http://dx.doi.org/10.1145/2592235.2592240

competition setting and in hybrid setting. Furthermore, users' improvement in physical activities is found to be strongly associated with the number of messages users exchanged.

This work offers multiple contributions in the field of pervasive health. First, it provides an environmental platform to compare social incentives as gamification strategies. Second, it extends research findings related to social incentives in mobile fitness applications using controlled studies. Third, the design implications derived from the findings can be directly applied in mobile fitness apps to enhance user social experience.

RELATED WORK

Gamification component is frequently used in designing mobile fitness applications to make sports activities *fun*. Methods such as leaderboards and metaphorical visualization have been used to encourage individuals to work out more. Most methods relied on either community- or self-perception for the implementation of the games. Many commercial fitness products, such as Nike+[1], Fitbit[2], miCoach[3], adopt community-based competition. A number of systems also use metaphors to present physical activities. UbiFit Garden [6] is a mobile application that visualizes users' daily steps by the growing status of plants. The more activities a user takes, the healthier his/her plant looks. UbiFit Garden also encourages users to take various types of activities by displaying butterflies. Fish'n'Steps [15] uses the metaphor of fish tank to visualize the step count of users. Using metaphor connects a user's physical activities with a living creature. The empathy to take care of a plant or an animal is shown to motivate a user to take exercise. Fitster [2] accumulates users' exercise distances and translates them to the routes on geographical maps, so that users can treat their exercises as traveling around cities. IFitQuest [16] uses several metaphors, such as escaping the wolf and collecting the coins, to enhance different aspects of physical performance: speed, endurance and etc. The above work mainly motivates users in an individual setting.

Researchers have identified the leverage of social interaction as an essential element in designing pervasive fitness applications [4, 5, 21]. Social interaction, such as peer-support, competition, cooperation, has been a clear motivator for wellness activities [1, 15]. Sharing physical activity data with buddies who exercise together have been reported to be effective in motivating users. This is reflected in users' qualitative feedback from the study of both Consolvo et al. [5] and Toscos et al. [21]. Even sharing exercise information with other people who are *not* in a fitness plan can help promote fitness activities. Ahtinen et al. [1] have found out that connecting with family members and loved ones can help motivate users. Stickk.com[4] is such a website that 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 or cheer them by putting comments on their progress journal. Currently, many mobile apps allow users to share their physical activities using social media. Typical examples include Nike+, RunKeeper[5], Runtastic[6], miCoach and etc. However, Munson and Consolvo [18] showed users' hesitation in sharing in social networks due to the concerns about whom to share and over-sharing.

Many applications allow users to form a team and compete with each other. Fitster [2] visualizes users' steps in a social group and motivates users through a virtual competition environment. Kukini [4], Fish'n'Steps [14] and Life Coaching Application [9] support competition by helping users to form a team and explicitly introducing social interaction and social pressure. Cooperation, which binds users' performance with that of their team members, not only promotes users' physical activities, but also brings social benefits such as group enjoyment and socialization with friends [17]. In Fish'n'Steps, any insufficient performance of group members deteriorates the environment of the fish tank, such as darker water, removal of tank decoration and etc. Users are thus motivated to enhance their performance for the sake of group responsibility. Chick Clique [20] calculates and displays the average steps of a team. This makes users aware of which members are below average and stimulates them to encourage teammates who are left behind.

Several studies have documented users' qualitative feedback about competition and cooperation in mobile fitness applications. For example, Lin et al. [15] found that some users considered competition engaging while others believed it was unnecessary and incompatible with the theme of Fish'n'Steps. They also discovered that cooperation did not produce any significant improvements if team members were anonymous. Qualitative results of Ahtinen et al. [1] indicated that users were motivated by workout as a group, which they referred to as cooperation. As for competition, users would rather compare with themselves than competing with others. Macvean [16] demonstrated that users' background should be considered carefully in competition setting. Findings of the above studies were based on evidence collected by users' qualitative feedback about social incentives. A controlled study comparing competition and cooperation is lacking. Halko and Kientz [11] found, through large-scale online

[1] http://nikeplus.nike.com/plus/

[2] http://www.fitbit.com/home

[3] http://micoach.adidas.com/

[4] http://www.stickk.com

[5] http://runkeeper.com/

[6] http://www.runtastic.com/

survey, that users of conscientious personality traits are more likely to use both competition and cooperation strategies and that people of agreeableness and openness traits tend to favor competition. However, their study focuses on users' proposition to social incentives and the results are not validated empirically. The closest one to our study is conducted by Peng and Hsieh [19], who found that cooperative goal structure lead to greater effort put into motor game than the competitive goal structure through a controlled user study. However, they did not generalize their findings in mobile fitness apps.

Besides the above social interaction schemes, researchers also uncovered the motivating effect of social communication. Such benefits include 1) providing a channel to give support to group members [5], 2) raising the awareness of each other's effort [17], 3) increasing a user's responsibility to exercise [3] and 4) enhancing players' intimacy, which builds a solid basis for long-term play experience [4]. However, the above findings were mainly collected from qualitative user feedback.

In order to investigate and compare different social interaction schemes ***in a controlled setting***, we designed a prototype as an experimental platform. The prototype design followed the requirements for mobile fitness applications derived by Consolvo et al. [5] and Toscos et al. [21]: 1) give user proper credit for activities, 2) provide personal awareness of activity levels, 3) support social influence, 4) consider the practical constraints of users' lifestyles and 5) encourage text message exchange. The next section presents the design of our prototype.

HEALTHYTOGETHER

We started investigating various group gamification incentives by examining their effects on two-person groups, namely *dyads*. To this end, we developed a mobile application on Android platform called HealthyTogether that involves *a pair of users* to exercise together and earn badges based on their walking and climbing performance.

To measure users' activities, we chose Fitbit among a number of off-the-shelf activity trackers. First, it is unobtrusive and convenient to wear. Second, compared to other competitor products, Fitbit is a comprehensive tracker that records not only steps (Figure 1a)) and distance, but also floors (Figure 1d)) and sleep. Last but not least, its API allows us to interact with Fitbit data in HealthyTogether. HealthyTogether automatically retrieves users' step and floor information from Fitbit server every two minutes. It also supports manual synchronization.

The main interface of HealthyTogether consists of two parts: game and log. The game area contains a 'Steps' tab (Figure 1b)) and a 'Floors' tab (Figure 1c)), which display the dyad's performance in taking steps and climbing the

Figure 1. HealthyTogether main interfaces and Fitbit tracker interfaces: a) Fitbit displaying step count, b) HealthyTogether step interface, c) HealthyTogether floor interface, d) Fitbit displaying floor count.

stairs, respectively. Users' effort in taking stairs is counted as the number of floors. In each tab, a progress bar visualizes a user's current number of steps or floors in orange color, as opposed to that of his/her buddy in blue – the opponent color of orange. This is followed by another progress bar which presents how the dyad's step or floor count contributes to the badges and how many steps or floors the user needs to gain the next badge. The area between the two progress bars illustrates the game rule, which is explained later. The dyad can send messages to each other by pressing either of the two blue icons: cheering face and taunting face. The subsequent badge area shows badges the user has earned during the day. Moreover, HealthyTogether allows users to view their performance history of the recent seven days. In the log area, users can log their mood, social interactions, food intake and activities as well as view log history. We do not present this area in detail, since it is out of the scope of this paper.

Game Rules

We designed a series of rewarding mechanisms for HealthyTogether in order to investigate the impacts of different social settings in mobile fitness application. A user can win step and floor badges based on his/her *Karma points*, which are calculated as following.

$$kp_{step}(u) = a \cdot step(u) + b \cdot step(u')$$

$$kp_{floor}(u) = a \cdot floor(u) + b \cdot floor(u'),$$

where u stands for a user and u' stands for his/her buddy.

Based on different values of a and b, HealthyTogether provides the following three social settings.

- **Competition setting**, where a = 100%, b = 0;
- **Hybrid setting**, where a = 80%, b = 20%;
- **Cooperation setting**, where a = 50%, b = 50%.

In **competition** setting, a dyad competes to gain more points and badges. Thus, a user's Karma points are calculated purely by his/her steps or floors. To gain more badges, both users only needs to improve their own activities compared with each other. In **cooperation** setting, the dyad contributes equally to win points and badges. Thus, we set a = 50% and b =50% so that a user's own performance and that of his/her buddy are equally important for winning badges. In **hybrid** setting, users both cooperate and compete. We set a = 80% and b = 20%, and thus a user's own performance weights more than that of his/her buddy. We designed a hybrid setting in order to understand the effects of different proportions of competition and cooperation on motivating users.

Badges

HealthyTogether provides incentives by badges. We considered *badges* to be more suitable for two-membered groups than *leaderboard*, which is usually applied in larger groups. The visual design of HealthyTogether badges follows the style of Fitbit badges. These badges are issued based on $kp(u)$. We set a low threshold for the first badge to help users get started easily. For steps, the first badge is issued when $kp_{step}(u)$> 500. This number is followed by 1000 and 2000 and then increases by every 2000 points. For floors, the first badge is issued when $kp_{floor}(u)$ > 2, followed

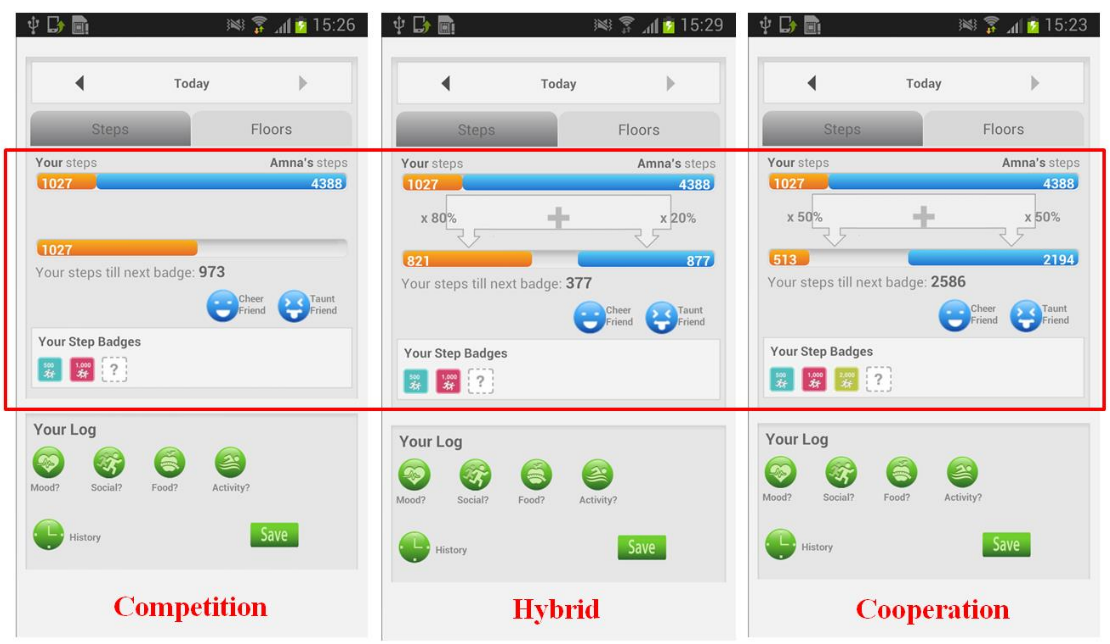


Figure 2. Interfaces of the three social settings: left: competition setting; middle: hybrid setting; right: cooperation setting.

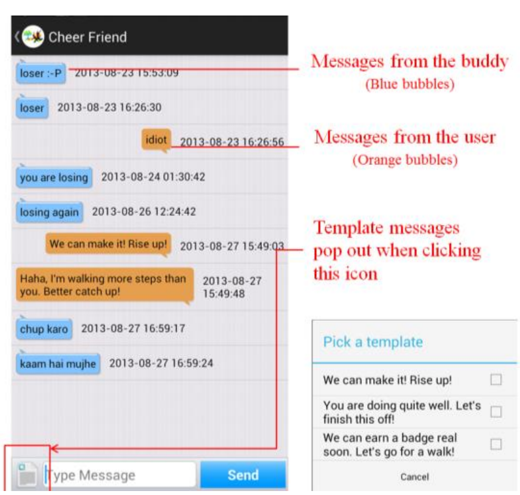
by 5 and then increases by every 5 points.

Social Interfaces

The social interfaces of HealthyTogether in three settings are the same except the explanations on how to earn the next badge. Figure 2 illustrates an example of *step* interface. In all three versions, the dyad's current steps (1027 vs. 4388) are visualized as opponents in the first progress bar. However, the second progress bar shows how the dyad's performance contributes to the next badge in different ways. In competition setting, the second progress bar only shows current steps (1027) relative to points for the next badge since the first progress bar already contains necessary information for competition within the dyad; in hybrid badge, the progress bar shows 80% of the user's steps and 20% of the buddy's steps relative to points for the next badge (821 vs. 877); in cooperation setting, the progress bar shows 50% of both the user's and the buddy's steps relative to points for next badge (513 vs. 2194). The area in between of the two progress bars explains how to earn badges. In competition mode, the rule is most straight-forward and thus no additional information is displayed; in hybrid and cooperation settings, this area illustrates the weights of each member's steps, i.e., 80% vs. 20% or 50% vs. 50%. In the example of Figure 2, the user needs 973 steps and 377 steps for the third badge in competition and hybrid settings respectively. By contrast, the user has already earned the third badge in cooperation setting.

Messaging Function

Users can send messages to each other using HealthyTogether (Figure 3a)). They can also choose a template message when they click the template icon located left to the text field. These templates include two themes: cheering and taunting buddies. Figure 3b) is an example of template messages for cheering buddies. Users can access



a) b)

Figure 3. Messaging functions in HealthyTogether: a) messaging interface; b) templates for cheering buddies.

messaging interface (Figure 3a)) by clicking either cheering or taunting icons, as shown in Figure 1. The results of clicking the two icons are the same, but the provided templates are different: either cheering or taunting themes.

USER STUDY

Participants

We recruited participants by posting advertisement on campus (in Switzerland) seeking users who were interested in using Fitbit. After one person signed up, we asked him/her to invite a buddy of his/her choice to join. In this way, we recruited 36 participants, consisting of 18 *dyads* (two individuals). We required that each dyad should not work in the same office or too close to each other, so that they are less likely to discuss about their experiences face-to-face. We randomly assigned the 18 dyads to one of the three conditions, which led to six in competition setting, six in hybrid setting and six in cooperation setting. Hereafter, we refer to participants in the different settings as competition group, hybrid group and cooperation group respectively.

The participants (15 males and 21 females) are from 17 different countries (Switzerland, France, Iran, China, US, and etc.). Most of them are students pursuing different levels of educational degrees (bachelor, master or Ph.D.), except four lab technicians, one researcher and one engineer. All but three age between 20 and 30: one below 20 and two above 30. All of them hold normal body mass indexes (BMI) (18.5-25), except one below normal (15-18.5) and another obese (30-40). They are all currently using a smartphone. No one has used Fitbit trackers before, but six of them have used various mobile fitness apps, such as RunKeeper and Runtastic. Each dyad was made up of close friends. The distribution of gender composition (female-female, female-male, and male-male) was equivalent across dyads in the three social settings (see Table 1).

Materials

We provided users with an Android phone with 3G SIM card and a Fitbit tracker, either Fitbit Ultra or its predecessor Fitbit One, but the differences of the two models are irrelevant with our study. Ideally, users could use HealthyTogether on their personal phones. However, as some participants complained either themselves or their buddies did not have an Android phone, we decided to provide Android phones to all of them.

Procedure

The study consists of a two-day warm-up session, a one-week control session and a following one-week experimental session. Both control and experiment sessions contain five continuous working days of a week. We refer to the control session as *Phase I* and experiment session as *Phase II*. We chose a period of two weeks so that the control and the experimental sessions span over identical

Groups	F-F	F-M	M-M	Total
Competition	3	1	2	6
Hybrid	3	2	1	6
Cooperation	2	3	1	6

Table 1. The distribution of *dyads* with different gender combinations (F-F: female-female, F-M: female-male and M-M: male-male) in three social settings.

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. During the warm-up session, participants filled in a demographic questionnaire, signed an informed consent, set up Fitbit and familiarized themselves with the device. In Phase I, all participants used Fitbit alone without connecting with buddies. Phase I allows us to collect users' baseline activity data. Participants were not informed of HealthyTogether until the end of Phase I, when we visited them individually and distributed the phones installed with HealthyTogether. In Phase II, participants started using HealthyTogether with their buddies in the respective social rules: competition, hybrid *or* cooperation.

We requested participants to fill in a daily diary in both phases. At the end of each day, we sent a reminder email with the diary link to participants asking them whether they had anything to share with us about their experience using Fitbit or HealthyTogether. The daily diary not only helps us to gain an in-depth understanding of users' experience, but also provides contextual information to participants' activity data during the corresponding day. At the end of the study, participants returned to our laboratory. After collecting back devices, we conducted and audio-recorded an unstructured interview about their overall experiences. We then offered each participant a 40CHF gift card as compensation for their time.

Results

In this section, we report results in four aspects: 1) a comparison between physical activities in control session (Phase I) and experiment session (Phase II); 2) a comparison of exercise performances among the three social settings; 3) a comparison of exchanged messages among the three social settings and 4) the relationship between exercise performance and interactive messages. The results were analyzed based on four sources of data: 1) Fitbit activity data, 2) messages exchanged between exercising buddies, 3) daily diary entries and 4) post-study interview recordings.

We prepared a PHP script to Fitbit data of the 36 participants within the two weeks of study. Participants reported two accidents when they forgot to wear Fitbit on a certain day during Phase I, because they changed their trousers. We discarded their step and floor data of the two days from our analysis. The mean daily step count is 9124.3 (*Max* = 28816, *Min* = 614) and mean floor count is 20.5 (*Max* = 85, *Min* = 0). We coded each entry with the phases, i.e., Phase I or Phase II. We also coded them with users' social setting, i.e., competition, hybrid or cooperation.

Additionally, we gathered a total of 370 messages that participants sent when using HealthyTogether in the five working days of Phase II. Most of the users sent messages in English except eight in French, two in Urdu and two in Persian. We recruited volunteers with the above as native language to translate the messages into English. Afterwards, two researchers inductively coded the messages.

Finally, we deductively analyzed daily diaries and post-study interviews, based on findings from quantitative analysis. The length of diary entries ranges from a short sentence (e.g., 2 words) to an essay (e.g., 1272 words). The length of the post-study interview lasted up to 20 minutes. The audio-recording was transcribed by the researchers. Qualitative data, such as the diary entries and post-study interview, not only provide complementary and contextual information to quantitative data, but also help us understand *why* and *how* our system and social settings work [12].

Individual versus Team Exercise

We first compared participants' exercise in control session and experiment session. Results of a paired-samples t-test showed that users' steps have significantly increased by 11.8% from Phase I (M = 8828) to Phase II (M = 10340), $t(36) = 5.12$, $p < .001$. Similarly, the average floor count has significantly risen by 15.1% from Phase I (M = 20.00) to Phase II (M = 23.03), $t(35) = -3.22$, $p = .003$. If we assume that Phase I is a baseline for participants, then the above results indicate that **HealthyTogether helps users to increase their activity levels** in terms of steps and floors.

Daily diary also reflects the above findings. When users started to use Fitbit, they were highly motivated, because Fitbit helped them to quantify themselves and increase their awareness of physical exercises. "*The device is cool, and it did increase my incentive to walk around. I didn't know that I can make so many steps in a single day!"* -- P12. However, some issues arose when they used Fitbit individually. Some reported that they forgot paying attention to Fitbit when they were busy, since Fitbit was unobtrusive to wear. "*Unfortunately today was such a busy day that I didn't even have time to think about Fitbit.*" -- P25. Several participants also mentioned that the novelty effect of Fitbit wore off after a few days. "*Not as great as the first two days."* -- P9.

Meanwhile, a few advantages emerged when they were using HealthyTogether. First, HealthyTogether has helped them to compare with each other. Participants widely appreciated that they were motivated to pay attention to their physical activities because someone else was present. "*From today on, my buddy is watching me. I was motivated* *to spend the day with more vigor and less laziness*." -- P34. Additionally, participants were triggered to compare with their buddies. "*I have been frequently checking and comparing with how she was doing*." – P7. Furthermore, they could send messages to each other. "*She made me aware that it was coffee time and I should take a break, go upstairs and take a coffee*" -- P33.

The above findings confirm the qualitative results from Consolvo et al. [5] and Toscos et al. [21] that applications that integrate social influence and message exchange can motivate users compared when they exercise alone.

Competition versus Cooperation

We then examined the impact of the three social settings in HealthyTogether. A paired-samples t-test was conducted to compare mean step and floor count in Phase I and Phase II in three groups respectively (see Table 2). There was a significant step increase in both hybrid groups (*diff* = 1084, $t(11) = -2.09$, $p = .061$) and cooperation groups (*diff* = 1728, $t(9) = -4.29$, $p = .001$). The floor increase also tends to be significant in hybrid groups (*diff* = 3.5, $t(11) = -2.04$, $p = .066$) and cooperation groups (*diff* = 3.9, $t(9) = -2.34$, $p = .039$). However, no significance was found when comparing the two phases in competition groups. This suggests that **both cooperation and hybrid setting could effectively motivate users to do more physical activities**, i. e., taking more steps and climbing more stairs.

We further compared whether the effects of cooperation and hybrid settings were significantly different. We were aware that participants had diverse lifestyles which could lead to different levels of steps and floors between individuals. Thus, we used **percentage increase** from Phase I to Phase II as the metric for comparative analysis. Results of an independent-samples t-test showed no significant difference in percentage increase between cooperation and hybrid groups (21.1% vs. 10.8%, $t(22) = -1.46$, $p = .159$). Nor did we find significance for floor percentage increase between the two settings, (18.3% vs. 18.2%, $t(22) = .095$, $p = .925$).

Groups	Data Type	Phase I (Individual)	Phase II (Team)	% Inc.	p-value
Overall	Step	9332	10433	**11.8**	***.002***
	Floor	20.00	23.03	**15.1**	***.003***
Competition	Step	9687	10181	5.1	.507
	Floor	19.74	21.48	8.8	.300
Hybrid	Step	10064	11148	**10.8**	***.061***
	Floor	19.05	22.51	**18.2**	***.066***
Cooperation	Step	8243	9971	**21.1**	***.001***
	Floor	21.22	25.10	**18.3**	***.039***

Table 2. A comparison of different groups: average steps and floors in Phase I and II, percentage increase from Phase I to II, and p-value of 2-tailed paired-samples t-test between Phase I and II.

Additionally, we analyzed the performance between two participants in a dyad. The percentage increase of users' steps did not correlate significantly with that of their buddies, and nor did floors. Interestingly, in competition groups, percentage increase of steps between teammates tended to be negatively correlated, $r = -.908$, $p = .012$, and so did floor increase, $r = -.57$, $p = .085$. This indicates **a reverse tendency of step enhancement between teammates in competition setting**: the more one increase, the less likely his/her buddy would do.

Analysis from users' diaries and post-study interviews showed evidence of the above findings. Generally, participants in **cooperation group** reported positively about the experience of exercising with their buddies. Firstly, they not only focused on their own performance, but also attempted to encourage each other. For example, P34 reported, *"Usually I prefer to stay at my office to work. But because I knew I'm playing with someone, I went out with my papers going to the farthest seats in the building and studied there. Moving around a little bit refreshed my mind. So I shared my experience with my buddy hoping she could walk more."* Additionally, participants felt obliged to increase their activity levels when their performance was connected with their buddies. As is written by P33, *"I realize that i am always winning the same badge as my buddy! We live almost same distances from the university. Seeing her always having more steps makes me a bit guilty. Maybe i should take more breaks and move around."* In addition to helping each other, participants also tended to be comparable with their buddies in **hybrid group**. As is reported in P24's diary, "*i ran after work because during the day i had course from morning till afternoon so that i was not really moving. When my partner saw that i achieved 19000 steps he also went out to run...he was jealous*."

Several participants in **competition group** also enjoyed the experience with their buddies. For example, P9 lived in the same community as his buddy (P10), but always lost the competition. So he had analyzed the causes: *"I think I'll lose the competition against my buddy for some ergonomic and physical reasons: she is shorter than me; hence, with smaller legs. Of course, she needs more steps to get the same places as me. Therewith, by considering a similar daily routine of "in-out", I'm very prone to lose."* Despite of this, he had been industriously boosting his climbing performance. *"I was really motivated to increase my stairs and really committed to strive some success, and I made it"*. His buddy also reported enjoying winning the competition and exchanging messages with each other.

Design implication 1: Consider designing cooperation in gamified exercise applications.

As is consistent with quantitative results, some participants reported some demotivating factors of **competition**. For example, P2, an active person who vowed to win the competition at the beginning, stated in post-study interview, *"I was very enthusiastic to compete with Ken. On the morning of the first day, I found his step more than mine. No way! I will beat him! But later it was not challenging at all. He was always sitting in front of his computer. I want to compete with some students who are active in the sport center and have regular exercises."* On the other hand, his buddy (P1) described: *"He was so competitive and energetic. I don't think I can ever compete with him."* Similarly, P11 wrote in her diary: *"I expected having a buddy to encourage me to be more active. It didn't happen so far :("* Her teammate P12 reflected in her post-study interview, *"Competition is not helpful if one day your buddy has no time for exercise. Competition is good only if you have competitors who are more or less in the same situation as you. If I have a friend who runs 20 Km per day, I'm not interested in competing with him/her."* Similar stories were reported by seven participants among the twelve in competition setting. Competition groups' experience shows that **competition motivates dyads if they have equivalent performances and availabilities, but is likely to demotivate them if otherwise**.

Design implication 2: When designing playful competition, consider paring partners with equivalent abilities.

Messages

Participants sent a total of 370 messages when using HealthyTogether in five working days, shared by competition groups (*n* = 58), hybrid groups (*n* = 114) and cooperation groups (*n* = 198). Their average number of messages in different social settings is listed in Table 3. A one-way ANOVA test showed significant difference among the three settings, $F(2,33) = 7.17$, $p = .03$. Post hoc (LSD) test revealed that cooperation groups sent significantly more messages than competition groups ($p = .001$) and hybrid groups ($p = .030$). More concretely, the number of messages in cooperation groups is 73.6% more than hybrid groups and twice more than competition groups. However, no significance was found between hybrid and competition groups. This indicates that **cooperation setting is more likely to stimulate users to interact with each other via messages than competition and hybrid settings**.

We then examined the content of the messages that users have exchanged. Table 4 shows six themes that emerged. Cheering buddies had the largest share (*n* = 126, *pct* = 34.1%), which was three times the number of taunting messages (*n* = 42, *pct* = 11.4%). Noticeably, users also shared their experience in physical exercises (*n* = 75, *pct* = 20.3). This included how they have successfully gained more steps, what caused their step count to drop and how they should improve in the next day.

Groups	Mean	SD	Max	Min
Overall	10.50	9.22	36	0
Competition	4.83	5.20	19	0
Hybrid	9.67	7.81	26	2
Cooperation	17.00	10.02	36	3

Table 3. Descriptive statistics of messages.

Themes	Examples	%
Cheer	*"u r doing great!"*	34.1
Share experience	*"2k steps more by walking home instead of bus."*	20.3
Chat	*"Good morning!"*	18.6
Compare	*"You are still better than me :("*	11.6
Taunt	*"Loser!"*	11.4
Workout together	*"Coffee together?" together*	4.1

Table 4. Message themes sorted by proportion.

We also compared the distribution of messages within the six message themes by the three social groups. The distribution, as shown in Figure 4, reveals the following phenomenon: 1) for the theme of cheering buddies and sharing experience, cooperation groups have obvious larger proportions, followed by hybrid groups, while competition groups hardly shared experiences with their buddies; 2) users in all the groups chatted and compared with their buddies; 3) competition groups sent more taunting messages than cooperation and hybrid groups; and 4) competition groups hardly organized activities together via messages. The above analysis showed that the three social incentives were likely to drive users to interact with their buddies in different ways.

We further examined the messages about steps and floors among all messages except the theme about chatting. Messages about general activities, steps and floors took shares of 42%, 48% and 10% respectively, indicating that users sent more messages about steps than floors. This might be due to the fact it is more convenient to encourage users to take a few steps than climbing the stairs. For example, one can easily stand up from his/her chair and walk a few steps around the corridor. By contrast, s/he might not have the opportunity to immediately find stairs nearby. However, this surmise needs further validation.

Design implication 3: Consider providing templates for messaging and tailoring the content for the respective goals of the strategies.

In addition to themes of messages, we also observed several interesting patterns. Firstly, users frequently used **emoticons**, such as ':o)', ':P', 'T_T'. Messages with emoticons constitute 24.3% (*n* = *90*) of all messages. Secondly, template messages were mainly used to help users get started with communication. Template messages, which constituted 7.6% (*n* = 28) of the total number, were primarily sent out during the early rounds of message

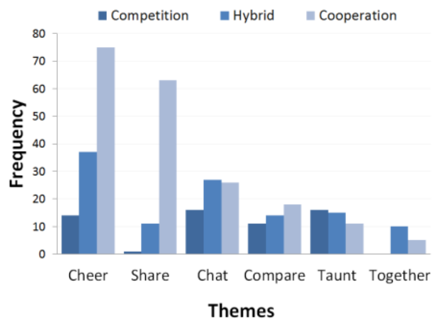


Figure 4. Message themes and distributions by group.

exchange. This reflects several suggestions that participants proposed in order to facilitate them in sending messages. For example, P9 suggested including emoticons in sending messages. "*I like the taunting face (icon). If would be nice to use it in the messages. Or perhaps sending a taunting face is convenient.*" For another example, P23 favored templates especially when she did not know how to encourage her buddy. "*I don't know what to say, but the template messages helped me to get started*." This indicates that **designing user-friendly messaging interface could help users to communicate more actively with their buddies**.

Design implication 4: Consider providing emoticons in messaging components to encourage users to express their feelings while playing exercise games with their buddies.

Physical Activities and Communication

Communication between teammates has been identified as beneficial for enhancing activities by users' *qualitative* feedback [4, 5, 18]. This subsection presents *statistical* analysis between users' activities and communication. We used **percentage increase of activities** from Phase I to Phase II to measure a user's improvement in physical activities and **message count** to measure his/her communication volume. More concretely, percentage increase includes two facets: steps and floors. For message count, we refer to the messages a user sent as sent messages and messages that the buddy sent as received messages. We conducted a bivariate analysis to examine the association between percentage increase and message count. Table 5 shows the Pearson correlation matrix of the variables.

We first examined the sub-metrics that describe percentage increase and message count. Step percentage increase is significantly correlated with that of floors (r = .405, n = 36, p = .032), indicating that users' improvements in steps and floors associated with each other. Users' number of sent messages and that of received messages were also significantly related, r = .877, n = 36, p < .001. This hints that the level of a user's social interaction (in terms of communication) is associated with that of his/her buddy.

We then investigated the relationship between users' number of messages and improvement in physical activities. Percentage increase of steps was positively correlated with number of sent messages (r = .398, n = 36, p = .036, two-tailed) and number of received messages (r = .394, p = .038, two-tailed). Thus, the more a user received and sent messages to his/her exercise buddy, the more likely he was motivated to take more steps compared to his baseline performance; similarly, the more progress s/he made, the more likely s/he was motivated to send messages to his/her buddy. However, we did not find strong correlation between percentage increase in floors and the number of message. This could be due to the fact that messages about steps outnumbered those about floors, as presented in previous subsection. Therefore, **exchanging messages help users to increase their steps.**

	1	**2**	**3**	**4**
1. Step % increase	1			
2. Floor % increase	.566**	1		
3. Message sent	.339*	.084	1	
4. Message received	.335*	.080	.876**	1

Table 5. Correlation Matrix among percentage increase in physical activities and message count (* p < .05, ** p < .01)

As is consistent with quantitative results, users' feedback confirmed that sending messages helped users to become more active in all three social settings. For example, P4 in competition group wrote, "*Every day she won, she would send me one message: loser. Finally, I beat her during the weekend because of my hiking trip*". P20 in hybrid group also logged, "*I told my buddy that I ran along the lake today. It motivated him to go to the gym.*" *"Today, I was more eager to take more steps and don't feel guilty anymore! A few messages were enough to persuade me!" said* P34 in cooperation group. Thus, it is essential to stimulate users to communicate with their buddies.

CONCLUSION AND FUTURE WORK

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-situ user study with 18 dyads, over a period of two weeks and compared participants using HealthyTogether in three social settings. The two-week study was divided into a one-week control session and a one-week experiment session using HealthyTogether. Results show that users' activities in experiment session have increased *significantly* compared to control session. Among the three social settings, cooperation and hybrid outperformed competition in motivating users to enhance their physical activities. More specifically, users have *significant* step and floor increase in both cooperation (by up to 21.1%) and hybrid setting (by up to 18.2%), but not in competition setting (by up to 8.8%). In addition, increase in physical activities significantly associates with the number of exchanged messages, and users in cooperation groups sent nearly twice the number of those exchanged in hybrid groups and three times more in competition groups. Based on the findings, we summarize four design implications from our studies.

This work can provide researchers with an environmental platform for testing social schemes and serve as a starting point to investigate social incentives. It can also help practitioners to design mobile fitness apps that enhance social user experience. In the future, we will examine more diverse social incentives and investigate group dynamics in larger groups.

ACKNOWLEDGMENTS
This research was supported by Swiss National Science Foundation. We thank Onur Yuruten and Jiyong Zhang for their support and feedback. We are also grateful for anonymous reviews for their insightful comments.

REFERENCES

1. Ahtinen, A., Isomursu, M., Mukhtar, M., Mäntyjärvi, J., Häkkilä, J., and Blom, J. Designing social features for mobile and ubiquitous wellness applications. In *Proc. MUM 2009*, ACM Press (2009).
2. Ali-Hasan, N., Gavales, D., Peterson, A., and Raw, M. Fitster: social fitness information visualizer. In *Ext. Abstracts CHI 2006*, ACM Press (2006), 1795-1800.
3. Anderson, I., Maitland, J., Sherwood, S., Barkhuus, L., Chalmers, M., Hall, M., and Muller, H. Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. In *Mobile Netw Appl* (2007), 185-199.
4. Campbell, T., Ngo, B., and Fogarty, J. Game design principles in everyday fitness applications. In *Proc. CSCW 2008*, ACM Press (2008), 249-252.
5. Consolvo, S., Everitt, K., Smith, I., and Landay, J. A. Design requirements for technologies that encourage physical activity. In *Proc. CHI 2006*, ACM Press (2006), 457-466.
6. Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., and Landay, J. A. Activity sensing in the wild: a field trial of UbiFit garden. In *Proc. CHI 2008*, ACM Press (2008), 1797-1806.
7. Deterding, S., Sicart, M., Nacke, L., O'Hara, K., and Dixon, D. Gamification: using game-design elements in non-gaming contexts. In *Ext. Abstracts CHI 2011*, ACM Press (2011), 2425-2428.
8. Fan, C., Forlizzi, J. and Dey, A. A Spark Of Activity: Exploring Informative Art As Visualization For Physical Activity. In *Proc. Ubicomp 2012*, ACM Press (2012), 2–5.
9. Gasser, R., Brodbeck, D., Degen, M., Luthiger, J., Wyss, R., and Reichlin, S. Persuasiveness of a mobile lifestyle coaching application using social facilitation. In *Prof. of Persuasive 2006*, Springer Press (2006), 27-38.
10. Gockley, R., Marotta, M., Rogoff, C., and Tang, A. AVIVA: a health and fitness monitor for young women. In *Ext. Abstracts CHI 2006*, ACM Press (2006), 1819-1824.
11. Halko, S. and Kientz, J. A. Personality and persuasive technology: An exploratory study on health-promoting mobile applications. In *Persuasive* 2010. Springer Press (2010), 150-161.
12. Klasnja, P., Consolvo, S., and Pratt, W. How to evaluate technologies for health behavior change in HCI research. In *Proc. CHI 2011*, ACM Press (2011), 3063-3072.
13. Li, I., Dey, A. K., and Forlizzi, J. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proc. UbiComp 2011*, ACM Press (2011), 405-414.
14. Li, I., Dey, A., and Forlizzi, J. A stage-based model of personal informatics systems. In *Proc. CHI 2010*, ACM Press (2010), 557-566.
15. Lin, J., Mamykina, L., Lindtner, S., Delajoux, G., and Strub, H. Fish'n'Steps: Encouraging physical activity with an interactive computer game. In Proc. *UbiComp 2006*, Springer Press (2006), 261-278.
16. Macvean, A., and Robertson, J. Understanding exergame users' physical activity, motivation and behavior over time. In *Proc. CHI 2013*, ACM Press (2013), 1251-1260.
17. Mueller, F., Vetere, F., Gibbs, M., Edge, D., Agamanolis, S., Sheridan, J., and Heer, J. Balancing exertion experiences. In *Proc. CHI 2012*, ACM Press (2012), 1853-1862.
18. Munson, S. and Consolvo, S. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *Prof. PervasiveHealth 2012*, IEEE Press (2012), 25–32.
19. Peng, W. and Hsieh, G. The influence of competition, cooperation, and player relationship in a motor performance centered computer game. *Computers in Human Behavior* (2012), 28(6), 2100-2106.
20. Toscos, T., Faber, A., An, S., and Gandhi, M. P. Chick clique: persuasive technology to motivate teenage girls to exercise. In *Ext. Abstracts CHI 2006*, ACM Press (2006), 1873-1878.
21. Toscos, T., Faber, A., Connelly, K., and Upoma, A. M. Encouraging physical activity in teens Can technology help reduce barriers to physical activity in adolescent girls? In *Prof. PervasiveHealth 2008*, IEEE Press (2008), 218-221.
22. Yumak, Z. and Pu, P. Survey of Sensor-Based Personal Wellness Management systems. In *BioNanoSci.* (2013) 3:254-269.