# Understanding the Effect of Existing Positive Relationships on a Social Motion-based Game for Health

**Karina Caro**
Drexel University
karinacaro@drexel.edu

**Yuanyuan Feng**
Drexel University
yf93@drexel.edu

**Timothy Day**
Drexel University
tjd62@drexel.edu

**Evan Freed**
Drexel University
efreed52@yahoo.com

**Boyd Fox**
Drexel University
seafordtbf@gmail.com

**Jichen Zhu**
Drexel University
jichen.zhu@gmail.com

## ABSTRACT

In this paper, we present the iterative design of *StepQuest*, a Fitbit-based social motion-based game for health (MGH) to sustain physical activity (PA) and support extended play. We conducted two 6-week user studies ($n$=24) to evaluate the effectiveness of the game to promote PA for an extended period of time as well as the role of existing social relationship. Our findings indicate that a pre-existing positive relationship (e.g., friendship) has a positive impact on players' PA levels when they play a social MGH, compared to strangers, and that this effect was amplified when more gameplay actions were available. However, our results also show that overall PA levels declined for both groups in week 4, and that pre-existing social relationship and a variety of gameplay actions are not enough to sustain long-term motivation for PA. Based on these results and drawing from game design literature, we present a list of design implications including less-discussed key topics such as game balancing.

## CCS CONCEPTS

• **Human-centered computing** → **User studies**; • **Applied computing** → **Computer games**;

## KEYWORDS

Social game; Motion-based game for health; Social relationships.


*PervasiveHealth '18, May 21–24, 2018, New York, NY, USA*

ACM ISBN 978-1-4503-6450-8/18/05…$15.00
https://doi.org/10.1145/3240925.3240942

**ACM Reference Format:**
Karina Caro, Yuanyuan Feng, Timothy Day, Evan Freed, Boyd Fox, and Jichen Zhu. 2018. Understanding the Effect of Existing Positive Relationships on a Social Motion-based Game for Health. In *PervasiveHealth '18: 12th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '18), May 21–24, 2018, New York, NY, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3240925.3240942

## 1 INTRODUCTION

Approximately two thirds of the U.S. adult population is overweight or obese [25]. More alarmingly, the rate of child and adolescent obesity is rapidly growing, which is partially caused by sedentary behavior. According to a recent study, only 8% of adolescents and young adults get the recommended 60 minutes of daily physical activity (PA) [32].

Among existing approaches to promoting PA, a promising one is motion-based games for health (MGHs) where physical exertion is integrated with gameplay. Research has shown that the energy expenditure of existing MGHs is a sufficient form of exercise that may positively influence the players' health [16]. As a result, MGHs have been studied as part of the treatments for patients [18] and for increasing the well-being of the non-patient population [27]. They have been designed for children [6], and for senior citizens [14]. Our research focuses on social MGHs. Existing work has shown the positive impact of incorporating social elements into MGHs, including increased PA [4], higher game engagement [24], and both physical and mental benefits for specific populations [6, 14]. However, current research is limited to social MGHs with relatively simple social interaction and game mechanics, and evaluations that are relatively short (2-3 weeks). We believe that advancing the knowledge of designing richer social interactions and more engaging game mechanics can lead to longer engagement and potentially greater health benefits.

We designed *StepQuest*, a six-player social MGH that integrates game design and the social interaction mechanisms of competition and collaboration to promote PA for an extended

period of time. It utilizes the *Fitbit* activity tracker to monitor players' PA and use the *Fitbit* as the in-game currency. In this paper, we present results from two 6-week studies, focusing on the effect of existing positive relationships (i.e., friendship) on participants' ($n$=24) gameplay experience and their PA levels. Both studies showed the friends group had significant higher PA levels than the strangers group for the first three to five weeks. Furthermore, based on the results from the first user study, we increased gameplay actions by adding a new power-up game mechanism in the second study. Our results indicate that the additional gameplay complexity increases the significant difference between the friends group and strangers group by additional two weeks. Although our findings suggest that a pre-existing social relationship and additions of in-game actions to a MGH do not ensure player PA motivation beyond four weeks, our results yield design implications and future research directions for designing more sustainable social MGHs.

This paper provides several contributions. First, we provide quantitative and qualitative evidence showing that a pre-existing positive relationship (e.g., friendship) has a positive impact on players' PA levels when they play a social MGH, compared to strangers. Second, we provide empirical evidence that the difference between strangers and friends widened when more gameplay actions were available. Third, we present design implications for social MGHs aimed at sustaining motivation for PA over time. Finally, our study is among the first to provide empirical data of the effectiveness of a social MGH among non-patient population over an extended period of time (6 weeks).

## 2 RELATED WORK

### Social MGHs and Extended Play

Due to the increasingly various means to combine games and PA, different terminologies have been proposed (e.g., exergames, motion-based video games [14], exercise gamification [17]) to highlight certain differences. In exergames, for example, PA and gameplay typically occur at the same time. Built on existing work [23], we use the term motion-based game for health (MGH) in this paper to include a broad range of approaches where physical motions and gameplay are combined to improve health and well being.

Recent research has shown that incorporating social interactions into MGHs may foster player engagement and increase motivation to exercise [24]. Despite the potential benefits of social MGHs, sustaining PA through extended gameplay remains an open problem [10, 29]. So far, most of these social MGHs for non-patient population were evaluated in a relatively short period of time (typically 2 to 3 weeks) [1, 8], which could be not enough for positive behavioral change. Among the few exceptions, studies of the popular game *Pokemon GO* show that its exercise promotion benefit wore off after about a month [2, 19]. Therefore, our research aims to contribute new data points on social MGHs by conducting two controlled user studies of *StepQuest* over 6 weeks.

### Gameplay Complexity

To support extended use, social MGHs need to provide a sustained, engaging gameplay experience for players. Currently, most social MGHs incorporated relatively simple social designs, often including competition among players (e.g., *Kukini* [5], *Fish"n"Steps* [21]) and collaboration with other players to achieve a common goal (e.g., *Liberi* [18]). However, very few MGHs incorporated complex social interactions such as the ability to support team strategies. In addition, most evaluation studies tested their MGHs with a small number players (often two players) per game [8, 14, 15]. Our research on *StepQuest* hence explores social interactions and gameplay afforded by a larger number of players (i.e. six). Moreover, most available social MGHs in literature involve one similar dynamic — whoever exercises more wins [1, 3, 21]. Since it is common that players have different levels of precedent PA, this dynamic leads to predictable game results. Game design literature has shown that expected outcome lead to boredom and decreased enjoyment [9, 13]. We hence explore new game design that will break this predictable pattern.

### Existing Social Relationship: Friends vs Strangers

The effect of pre-existing positive relationships, such as friends versus strangers, has been previously explored in digital games. Some studies suggest that playing with friends results in increased motivation and enjoyment to play [7, 11, 28]. For example, the work of [7] examines the effect of pairing people with similar attitudes within the context of videogames. This literature suggests that there is a positive impact on motivation and enjoyment when participants with a relationship of friendship play digital games, compared to strangers. However, more research is needed to investigate if these results extend to MGHs, as well as if the difference persists in more than one game session (over an extended period of time).

Initial research shows that social dynamics among players could affect players' engagement with MGHs, as a few MGHs tested on groups of friends or co-workers led to positive game experience as participants reported excitement to play with people they knew [3]. Later research has compared single-play versus multi-play (i.e., pairs of friends) [12, 28]. Their results suggest that there is a difference in PA levels in MGHs when participants play in pairs, compared to single mode. However, to the best of our knowledge, research investigating the effects of social relationships such as friends and strangers of more than two players in the context of MGHs is scarce. The current literature address this difference only

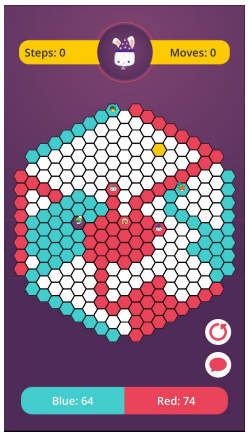



**Figure 1: Two teams of 3 players compete to capture spaces in *StepQuest***

in laboratory conditions, with at the most one game session. Our work investigates this effect in the context of MGH for a longer period (6 weeks).

## 3 STEPQUEST

*StepQuest* is a mobile social MGH designed to increase and sustain players' PA. Our gameplay mechanics allow for extended play by encouraging social interactions between players through competition and collaboration. In *StepQuest*, two teams of three players compete to capture spaces on a board using their real-world steps, recorded by *Fitbit* activity trackers. At the end of a predetermined period of time, the team that controls the most spaces wins. Spaces on the hexagon-shaped board (Figure 1) can be captured in two ways. When a player moves onto a space, she captures that space, regardless of whether the space was previously occupied. Additionally, when a team (red vs. blue) captures contiguous spaces that *encircle* an area, the team automatically captures every single space inside that area.

### Social Interaction and Play

The core mechanics of *StepQuest* are designed to support both collaboration and competition. We followed the guiding principles in literature for incorporating these two forms of social interaction in our game [23, 24, 31]. Specifically, opposing teams compete to capture more spaces. Within each team, collaboration is reinforced by a team-based scoring system and facilitated by a team chat system.

Different from many similar MGHs, *StepQuest* is designed to balance PA and gameplay engagement through the above mentioned encircling mechanic. In a typical multiplayer MGH, the player who has more PA wins. Although this design incentivizes PA, it makes the game outcome predictable and hence make the game less engaging [13]. The encircling mechanic in *StepQuest* breaks the *"whoever walks more wins"* pattern by incentivizing collaboration. Because larger areas are more easily captured when the team works together, a team with less PA but better team collaboration can still win.

### Designing for Extended Play

Designing a social MGH for extended play for groups of six players is not well understood in the literature. Therefore, we extend the discussion of this part of our design rationale. The following design decisions are results from three iterations of design and evaluation.

*Flexibility and Asynchronous Design.* To make extended play more sustainable, it is important *StepQuest* can be seamlessly integrated into players' daily life. We therefore designed *StepQuest* to track player's on-going PA throughout the day, using *Fitbit* activity trackers, while only requiring minimal time investment each day (about 3-5 minutes). If a player wants to play more, there are in-game activities she can engage with. An additional challenge is to provide meaningful social gameplay to six players with different schedules. Different from many existing social MGHs, *StepQuest* is asynchronous — players do not have to play at the same time. We designed our game in such a way that when opening the game at any time, a player is always able to execute their moves regardless the states of other players.

*Variety and Complexity.* Each game in *StepQuest* lasts between 2 and 3 weeks. Similar to classic board games such as *Reversi* and *Checker*, the board space in *StepQuest* supports high variability and thus each game could be vastly different from another. In addition, we added a randomly-placed bonus space on the board (see the gold spot in Figure 1) in order to add further variability. When captured, the bonus space gives the player extra amount of moves.

*The Temporal Structure.* For the gameplay experience to be less repetitive, we designed for a loose progression of three phases in each game, each with different dynamics and strategies. The focus of the first phase, roughly the first few days, is the initial capturing of empty spaces. When the game starts, each player is placed on an edge of the board, alternated by team. Thus, each player starts in between two players from the opposing team, in order to introduce conflict and competition early on. In the second phase, players begin colliding with one another more frequently. As players move toward the center of the board, the proximity between team members allows for team-based collaborative actions (e.g., enclosing a large space). Similarly, being close to the opposing team increases the chance for competitive actions, such as capturing occupied spaces. The focus of the third phase, in the last few days of the game, is a race to keep as many spaces for the team as

possible. The gameplay shifts from capturing empty spaces to recapturing spaces occupied by the opposing team. As in previous phases, teams communicate to balance actions with short-term and long-term benefits. Overall, *StepQuest* extends existing literature by exploring how to use more complex game mechanics to sustain longer gameplay.

## 4 INITIAL EVALUATION

To evaluate the design of *StepQuest* and the role of existing relationship on how players interact with our game, we conducted a 6-week user study with a group of friends and a group of strangers. We aim to answer two main research questions.

- *RQ1: Physical activity*. How a pre-existing social relationships affect StepQuest players' steps across six weeks by comparing a group of friends and a group of strangers?
  - A. Which group performs a higher number of steps across the six weeks?
  - B. Which group perceives a higher level of motivation for performing PA using *StepQuest* across the six weeks?
- *RQ2: Gameplay*. How a pre-existing social relationships affect StepQuest players' captured spaces across six weeks by comparing a group of friends and a group of strangers?
  - A. Which group is more motivated to capture spaces on the board of *StepQuest* across the six weeks?
  - B. How does each group perceive the general gameplay in *StepQuest*?

### Study Design

*Participants.* We recruited 12 college students between the ages of 18 to 22 from undergraduate classes in a university located in a large U.S. city. The evaluation followed a between subject design. The strangers group (i.e., control group) consists of 6 randomly selected participants (4 males and 2 females, average age = 19.2). The participants did not know each other before the study. The friends group (i.e., experimental group) consists of 6 participants (5 males and 1 female, average age = 21). The inclusion criteria for the friends group included that they should know each other for at least 2 years (average = 3.5 years) and have a social relationship both at and outside of the school (e.g., attend social events together).

*Procedure.* All participants were asked to wear a *Fitbit Charge HR* every day, provided by the research team, and play *StepQuest* on their mobile phones. Participants in both groups played two games of *StepQuest* that last around 3 weeks to reach a total study length of 6 weeks. In each game, a participant was randomly assigned to one of the two teams (blue or red) to play 3-versus-3 in *StepQuest*.

*Data Collection and Analysis.* We collected a range of quantitative and qualitative data from participants in both groups. From the *Fitbit* device, we collected participants' daily step count. *StepQuest* records all gameplay actions and metrics including the spaces[1] captured by individual players for each team. We refer to this data as *in-game actions.* We also distributed a web-based weekly questionnaire using a 5-point Likert scale to gather data on participants' experience with *StepQuest* over 6 weeks, inquiring about their motivation for performing PA and playing the game. Finally, at the end of the evaluation, we conducted a focus group with each group to gather qualitative data about their overall gameplay experience and feedback.

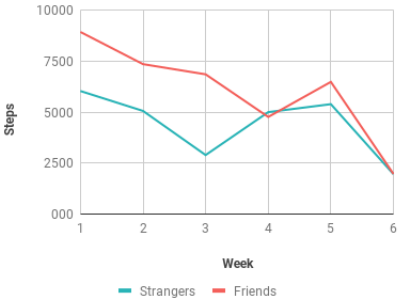
We used a mixed-method approach to analyze the collected data. For *Fitbit* step, in-game actions, and questionnaire data, we used descriptive and inferential statistics to compare the differences by week between and within groups. To investigate if there is a significant difference between steps or in-game actions by week, we performed a independent t-test as the data followed a normal distribution. Also, we conducted an intra-group analysis to investigate how each group changed over time, i.e., differences in the steps or in the in-game actions across the six weeks. An ANOVA of repeated measures following for post hoc tests was performed. Finally, the qualitative data obtained from the focus groups was analyzed using qualitative techniques such as open and axial coding [30].

### Results

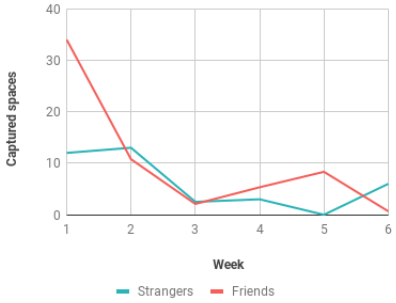
*RQ1: Physical Activity. Steps.* In general, we found a decrease in the number of steps of each group across the six weeks. However, the friends group started with a significantly higher number of steps the first three weeks compared to the strangers group. The average number of steps of week 1 was significantly higher for the friends group (mean = 8,947) than for strangers group (mean = 6,043)(t = -2.6521, p-value = 0.00481, right tailed). The same trend was observed for week 2 and 3 ($p$ <0.05). However, in weeks 4 to 6, a significant difference was not observed (Figure 2–(a)). The intra-group analysis showed that the friends group decreased the number of steps in week 4, while the strangers group decreased significantly in week 3 (Figure 2–(a)).

*PA Motivation.* There was no significant difference in the average score of the perceived degree of motivation for performing PA between both groups ($p$ >0.05). For the friends group, the average score was 2 (Slightly Motivated) during all the six weeks. For the strangers group, the average score

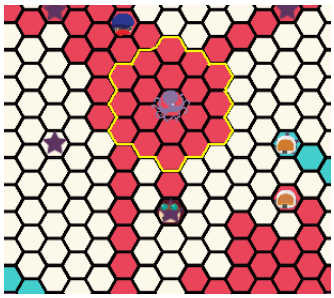
[1] A captured space is worth 1 point. 1000 steps to capture a friendly space, 2000 a neutral, 3000 on an enemy space. The team with the most points wins the game.




(a) Steps.




(b) Captured spaces.



**(c) A zoomed-in view of a power-up that captures a ring of spaces (highlighted yellow for clarity).**

**Figure 2: Average number of steps and captured spaces by week by group during the initial evaluation. A screenshot of the "octopus" power-ups.**

of the perceived motivation was 2 (Slightly Motivated) for weeks 1, 3 and 6. For weeks 2, 4 and 5, the score was 1 (no motivation). Finally, the overall score for both groups was 2 (Slightly Motivated).

*RQ2: Gameplay. Captured Spaces.* In line with the number of steps, we observed that the number of captured spaces during week 1 was significantly higher in the friends group (mean = 34) than the stranger group (mean = 12) (t = 4.2019, p-value = 0.00841, left tailed). There was no significant difference between the captured spaces for the weeks 2-6 (Figure 2–(b)). The intra-group analysis showed that the friends group decreased significantly the number of captured spaces from week 1 to week 3 (p-value = 0.0069). A significant difference was not observed across the six weeks in the strangers group (Figure 2 – (b)).

*Gameplay Experience.* The analysis of the weekly questionnaires showed no significant difference in the average score of the perceived degree of gameplay experience with *StepQuest* between both groups ($p$ >0.05). For the friends group, the average score was 3 (neutral) for all the six weeks. For the strangers group, the average score was 3 (neutral) in most cases (weeks 1, 2, 3 and 5) and 2 (negative) for weeks 4 and 6.

## Discussion

Our quantitative results show the change of steps and in-game actions over time. Regarding steps, we found that friends group decreased significantly in week 4, while strangers did in week 3. From there, both groups maintained the number of steps until week 5, but decreased again in week 6 (2–(a)). These results indicate that *StepQuest* can sustain motivation for performing PA (e.g., walking) for at least four weeks in a group of friends, and at least three weeks in a group of strangers. Moreover, a significant difference in the number of steps between friends and strangers in the first three weeks of the study was also found, indicating that a positive social relationship such as friendship affects the PA levels performed with a MGH at least during the first three weeks.

On the other hand, in-game actions decreased significantly from week 1 to week 2 for the case of friends group, and no difference was observed from there over the six weeks. In the case of the strangers group, it was not difference over in the number of in-game actions (Figure 2– (b)). We attribute this result to the game complexity of *StepQuest*: as Fitbit steps do not translate directly into in-game actions, we did not find a correlation between steps and in-game actions. However, the difference between both groups was only observed in week 1, indicating that the effect of a positive social relationship impacts the in-game actions that participants take into a MGH such as *StepQuest* at least in the first week.

The difference found between friends and strangers regarding number of steps and in-game actions was supported by our qualitative analysis. During the exit focus groups, the friends group expressed that they would be uninterested to play *StepQuest* with strangers. They enjoyed to play with people that they already know:

> *"I definitely wouldn't try to... it would be even harder for me to try to interact with teammates if I didn't already know them."* P2, friends group.

Similarly, the strangers group indicated that they would prefer the gameplay available in *StepQuest* of compete and collaborate with others, if they were playing with people they knew:

> *"If the people already came in and knew each other, [and] they get a group of friends that want to compete against other group of friends, the*

> *competitiveness would probably be better."* P3, strangers group.

Furthermore, both groups expressed that the core gameplay of *StepQuest* was too simple for the duration of the study and expressed the wish for more variety in game actions. Both groups expressed that their motivation for playing the game was higher at the beginning than the last week.

> *"During the first weeks [the gameplay] worked out pretty well, then I kind of lost motivation after that."* P2, strangers group.

Both groups felt that more game mechanics could help balance the game rules with the PA requirement to play the game. For example, having more diverse items or power-ups, the ability to donate steps to their teammates, as well as changes in maps and move costs to sustain replayability.

> *"After a certain point the core gameplay of just moving your character around the board got very old... so maybe having some different pick ups that do different things besides just picking more steps to do more of the same thing."* P5, friends group.

In summary, we found a difference between friends and strangers in PA levels and in-game actions when playing *StepQuest*. The gameplay of *StepQuest* was enough to sustain PA levels for 3 weeks in the case of strangers, and 4 weeks in the case of friends. However, the lack of variety of in-game actions was reflected in the captured spaces in both groups, as well as, in our qualitative analysis.

### Game Improvements

Reflecting on our results and to increase the variety of in-game actions and social interactions, we added six types power-ups to *StepQuest*. For example, the *"octopus power-up"* (Figure 2–(c)) can capture a ring of spaces around it over time, unless the opposing team stops it. Another power-up lets players donate their steps to their teammates. These power-ups are designed to create more social interactions among players. They also encourage players to think more strategically about how to use their *Fitbit* steps, whether on moving on the board or on purchasing power-ups, creating more gameplay options.

To address the feedback that the games became stagnate, we reduced each game's length from three to two weeks.

## 5 SECOND EVALUATION

We conduct a second evaluation study with the new version of *StepQuest* to investigate players' PA levels and whether the significant difference between the friends group and strangers group can still be observed.

### Study design

*Participants.* We followed the same recruitment process of the initial evaluation (details above). In this study, the strangers group (i.e., control group) consists of 3 males and 3 females, average age = 19.7 years. The friends group (i.e., experimental group) consists of 4 males and 2 females, average age = 21.6 years. None of the participants were part of the initial evaluation.

*Procedure.* The procedure was the same as the initial evaluation with the following extensions:

- Participants of both groups played three games (instead of two) that lasted 2 weeks each. Thus, the study length was also 6 weeks.
- Participants of both groups filled a questionnaire about their game experience at the end of the evaluation.

*Data collection and Analysis.* Our methods for data collection and data analysis were the same as the initial evaluation with the following additions.

In data collection:

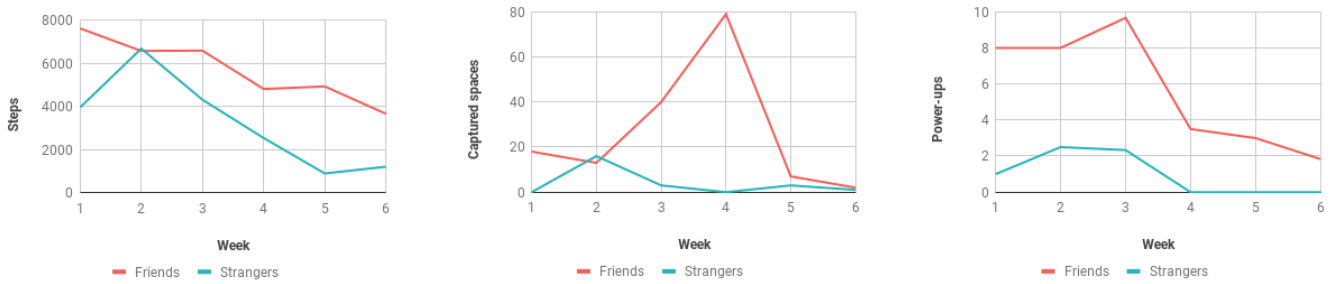
- We used the in-game module of the Game Experience Questionnaire (GEQ) [20] to investigate the perceived game experience of our MGH between both groups. The GEQ measures game experience in seven dimensions using a 5-point Likert scale: competence, immersion, flow, tension, challenge, positive and negative affect.
- New in-game actions (power ups) were collected.

In data analysis:

- We performed an independent t-test to investigate differences in the dimensions of the game experience of GEQ between both groups.
- Power-ups were analyzed using independent t-tests to investigate differences in both groups. Also, An ANOVA for repeated measures with possible post hoc tests was performed to investigate differences within each group across the six weeks.

### Results

*RQ1: Physical Activity. Steps.* Similar to our initial evaluation, the decreasing trend in the number of steps of each group across the six weeks was also observed. However, contrary to the initial evaluation, the friends group performed a significantly higher number of steps during all weeks ($p$ <0.05), except week 2, comparing to the strangers group (Figure 3–(a)). The intra-group analysis showed that the number of steps in both groups decreased significantly in week 4 ($p$ <0.05). From there, the friends group maintained the number of steps across week 6, and strangers group decreased again in week 5 ($p$ <0.05), but maintained from week 5 to week 6 (Figure 3–(a)).




**(a) Steps** **(b) Captured spaces** **(c) Power-ups**

**Figure 3: Average number of steps, captured spaces and power-ups by week by group, during the second evaluation study.**

*PA Motivation.* There was no significant difference in the average score of the perceived degree of motivation for performing PA between both groups ($p$ >0.05). For the friends group, the average score was 3 (moderate motivated) for weeks 1 and 6, and 2 (Slightly Motivated) for weeks 2-5. For the strangers group, the average score of the perceived motivation was 3 (moderate motivated) for weeks 1 and 2, and 2 (Slightly Motivated) for weeks 3 to 6. Same as our initial evaluation, both groups perceived that *StepQuest* motivates them in a slight degree to perform PA over six weeks.

*RQ2: Gameplay. Captured Spaces.* The number of captured spaces during week 1 and week 4 was significantly higher in the friends group than the strangers group ($p$ <0.05). There was no significant difference between the captured spaces for the weeks 2, 3, 5 and 6 (Figure 3–(b)). The intra-group analysis showed that the friends group increased significantly the number of captured spaces from week 1 to week 4 (p-value = 0.026) and decreased significantly from week 4 to week 6 (p -value = 0.046). For the strangers group, the captured spaces increased significantly from week 1 to week 2 (p-value = 0.0047), but decreased from week 2 to week 3 (p-value = 0.0107) (Figure 3–(b)).

*Power-ups.* The number of power-ups was significantly higher in all weeks for the friends group ($p$ <0.05) than the strangers group (Figure 3–(c)). The intra-group analysis showed that the number of power-ups used by the friends group decreased significantly from week 3 to week 4 (p-value = 0.0016). Then the number was maintained until week 6. Similarly, in the strangers group, the number of power-ups decreased significantly from week 3 to week 4 (p-value = 0.018). Then power-ups were nearly never used for the remaining weeks (Figure 3–(c)). In addition, no correlation was found between the number of steps and the number of in-game actions including captured spaces and power-ups in each group.

*GEQ.* Both groups perceived a similar degree of game experience, as no significant difference was found in the scores of each GEQ dimension between the groups (Figure 4). The score for *positive affect* was the higher comparing to other dimensions, that is if players of both groups felt content and good when played *StepQuest*: Friends' score = 3 (fairly), strangers' score = 2 (moderately). On the other hand, scores for *challenge*, regarding if players felt challenged by the game or felt they had to put a lot of effort into it were also acceptable in both groups. However, the friends group perceived most challenge (score = 2.5 [between moderately and fairly]), comparing to strangers' (score 1.6 [between slightly and moderately]). We attribute this result to the fact that in-game actions could have impacted the perceived challenge of both groups, as friends group used more power-ups in all weeks and captured more spaces in two weeks.

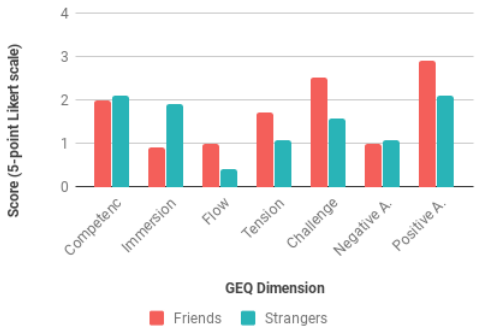



**Figure 4: GEQ scores by dimension.**

## Discussion

In summary, our results show positive effects of a pre-existing social relationship (i.e., friendship) on PA levels (steps) and

in-game actions (captured spaces and power-ups) when participants played *StepQuest*. Extending the results of our initial evaluation, we found a between-group difference in the first five weeks in terms of steps, comparing to the difference in the first three weeks in the initial evaluation. We hypothesize that this result is related to the addition of the variety of in-game actions (i.e., power-ups). This feature strengthened the pre-existing social relationship, and made friends interact more with the MGH:

> *"[I remembered one time] my brothers [friends] were leaving [and I heard someone] 'he's trying to get everything around [using a power-up]' and sudden all the three of us were like in the corner, strategizing our game... "* P4, friends group.

On the other hand, in the strangers group, the power-ups were not so engaging and players were not moving constantly on the board. They expressed that playing with friends, would better enhance the game experience:

> *"Playing with friends would be a little more fun, because you would like mess up other people [using power-ups]... cause for us, it was like the red team just not the ones moving. I was like, well, I could have put bombs [power-ups] in front of the blue person that never moved, but this is not fun"* P1, strangers group.

The power-ups added to *StepQuest* widened the effects of the pre-existing social relationship between both groups. Compared to the initial evaluation, the PA levels in the friends group were significantly greater in all weeks except week 2. Regarding power-ups, the difference was observed in all six weeks. That is, the number of power-ups used was greater in all weeks in friends group, compared to strangers. Overall, the friends group perceived power-ups engaging and a potential way to balance gameplay.

> *"Everything was really simplified [in the game] and the power-ups give you that –It does balance the game, give people that are behind a chance at catching a whole bunch of blocks."* P3, friends group.

Also, they attributed power-ups to motivate them to take an extra walk to obtain steps.

> *"When you're in the center of the map, you're putting a cell around you and I thought that was hilarious because like now it's one spot for you to get out so I just walked over [to get steps] and then I put a bomb [power-up] there."* P2, friends group.

However, we found that *StepQuest* sustained PA levels until week 4 in both groups. Comparing to the initial evaluation, the main difference over time was that strangers sustained the number of steps for more one week. Even so, the new variety of the in-game actions of our MGH was not enough to sustain PA levels and extended play for more than 4 weeks. Our qualitative results indicate that in both groups, the fact that one or more players stopped to play the game affected others motivation to play.

> *"I stopped [playing] because I was like, 'I want them to actually move before I do something' [laughs] That was what I was doing most of the time... like, 'I'm not gonna move this if they're not gonna move' "* P1, friends group.

> *"The first weeks, I was looking on the other players like 'you had more steps than me' and I had already gotten to the star and then, like in the middle [of the study], was just like they haven't moved at all and I was like 'okay'... "* P5, strangers group.

More research is needed to iterate our game mechanics and social interaction in *StepQuest* to sustain PA and extended play for more than 4 weeks. Key considerations has to be made when designing social MGHs to sustain PA levels for an extended period of time, specially when designing for a group a friends or strangers.

## 6 DISCUSSION OF BOTH EVALUATIONS

We discuss the implications of results from both evaluations.

### Considerations for Designing Social MGHs

*Gameplay Complexity and Variations. Incorporating reasonable gamplay complexity*. Participants in the first evaluation identified a lack of variety in gameplay as a major component in their loss of motivation over time, while participants in the second evaluation, especially the friends group, reported higher motivation to play with the power-ups. This indicates that a reasonable level of gameplay complexity that go beyond the simple dynamic of *"whoever walks more wins"* is crucial to a social MGH aimed at sustain PA over time, which resonates the importance to incorporate various or even personalized persuasive strategies in MGHs[26]. However, two participants from the second evaluation expressed initial confusion about the functions of different power-ups in the focus group interviews. This alerts us that a social MGH like *StepQuest* should stay intuitive to players when more complex features are added, which requires effective and clear user experience designs to explain new features and make players aware of their actions.

*Engaging players through game feature updates*. In both studies, participants noted that they were less motivated to gain extra steps as the novelty of the *Fitbit* device and game itself wore off. Even with an increased gameplay complexity in the second evaluation study, it was still challenging for

*StepQuest* to sustain players' motivation beyond 4 weeks. This calls for comprehensive design considerations to support PA over time. According to our qualitative analysis from both evaluation studies, a promising approach is to add new game features, including various in-game actions, gradually over time. Contemporary commercial multiplayer games thrive on constant updates over time, from *League of Legends* to *Pokemon GO*. Players value game updates and changes in gameplay experiences as they give them new excitement to revisit the game.

*Balancing Gameplay.* One emergent theme from the qualitative analysis was game balance. Participants of both studies identified balance problems that negatively affect their gameplay experience. Although a well-known step in the game design process [13], game balancing was rarely discussed in MGHs literature. One possible reason is that existing research often examine short sessions of gameplay, where the issues of imbalance have not yet surfaced. We identified the following factors of game balancing in MGHs that requires further investigation.

*Balancing different players' PA levels.* In a social MGH such as *StepQuest*, players may have different PA levels. In both evaluation studies, some participants experienced the snowballing effect when they realized that others are more active and therefore much more likely to win. One potential future direction is to better match players with similar PA levels. We believe that different and more complex game mechanics also need to be investigated.

*Balancing individual and collective agency.* Participants from both evaluation studies, regardless the condition (friends or strangers), reported decreased motivation when their team members were inactive in the game. This raised an interesting issue of balancing. If an individual perceives that her action alone is not enough to make a meaningful impact on the game, her engagement drops. However, if an individual is too powerful, it renders other players' action and social interaction inconsequential. A possible solution may be to provide both individual-based and team-based incentives and win conditions. For example, our participants recommended an avatar customization system where a player can earn different looks based on her performance. This would be evidence of an in-game achievement visible to the whole group. Finally, another potential solution might include the incorporation of AI agents, whose can help to take actions in the game, when detect players are inactive.

### Considerations for Designing Health interventions

*Taking Advantage of Pre-existing Social Relationships.* Our results from both studies suggest that social MGHs for supporting PA over time can benefit from including people with pre-existing social relationships. Additionally, our results also confirm that these pre-existing relationships can enhance participants' interactions inside the gameplay of the MGHs. A broader implication of our research is that taking advantage of pre-exisitng social relationships could lead to more effective health interventions for sustaining healthy PA levels.

*Including Variations and Updates in Health Interventions.* In both studies, we found that participants' overall PA levels started to decrease in week 3 or 4 in all groups. As discussed above, yet increasing game complexity and variations helped keep players engaged with the MGH longer. Our results indicate that when designing health interventions for promoting PA over an extended period of time, it would be beneficial to avoid repetitive tasks and include a range of variations and updates through the course of the intervention. However, in our particular case, having internal games of two or three weeks long, as well as, more variety of in-game actions, was still not enough to sustain PA for more than 4 weeks. Thus, further research is needed in other aspects to promote PA through health interventions, such as personalized motivation mechanisms based on individual preferences or personality characteristics [22, 26].

## 7 LIMITATIONS

This research was conducted only with undergraduate students from a local university. In addition, the population of both user study evaluations is small ($n$ 24). To generalize the results for a larger group, our studies should involve more diverse participants. The levels of PA were measured using an external device (i.e., *Fitbit*), we are aware that the steps counts calculations might present an error rate, which is beyond of our control.

## 8 CONCLUSION AND FUTURE WORK

In conclusion, we presented the iterative design of *StepQuest*, a *Fitbit*-based social MGH to sustain PA and support extended play. We conducted two 6-week user studies ($n$ 24) to evaluate the effectiveness of the game to promote PA for an extended period of time as well as the impact of existing social relationships on player engagement. Our findings indicate that a pre-existing positive relationship (e.g., friendship) has a positive impact on players' PA levels when they play a social MGH, compared to strangers, and that this effect was amplified when more gameplay actions were available. However, our results also show that overall PA levels declined for both groups in week 4, and that pre-existing social relationship and a variety of gameplay actions are not enough to sustain long-term motivation for PA. Based an these results and drawing from game design literature, we presented a list of design implications including less-discussed key topics such as game balancing.

For future work, we plan to expand our evaluations with a larger group of users. We also plan to explore different in-game actions such as weekly game updates to continue investigating how can we extend our sustained PA levels beyond 4 weeks. Finally, we plan to explore more about game balancing in the context of social MGHs intended for extended play.

## ACKNOWLEDGMENTS

The authors would like to thank all members of the project.

## REFERENCES

[1] Aino Ahtinen and Pertti Huuskonen. 2010. Let ' s All Get Up and Walk to the North Pole: Design and Evaluation of a Mobile Wellness Application. In *NordiCHI 2010*. 3–12.

[2] Tim Althoff, W. Ryen White, and Eric Horvitz. 2016. Influence of Pokémon Go on Physical Activity: Study and Implications. *J Med Internet Res* (2016).

[3] Ian Anderson, Julie Maitland, Scott Sherwood, Louise Barkhuus, Matthew Chalmers, Malcolm Hall, Barry Brown, and Henk Muller. 2007. Shakra: Tracking and Sharing Daily Activity Levels with Unaugmented Mobile Phones. *Mobile Networks and Applications* 12, 2-3 (2007), 185–199.

[4] Jacob E Barkley, Andrew Lepp, and Ellen L Glickman. 2017. "Pokémon Go!" May promote walking, discourage sedentary behavior in college students. *Games for health journal* 6, 3 (2017), 165–170.

[5] Taj Campbell, Brian Ngo, and James Fogarty. 2008. Game Design Principles in Everyday Fitness Applications. In *ACM conference on Computer supported cooperative work*. ACM, San Diego, CA, USA, 249–252.

[6] Karina Caro, Monica Tentori, Ana I Martinez-Garcia, and Ivan Zavala-Ibarra. 2017. FroggyBobby: An Exergame to Support Children with Motor Problems Practicing Motor Coordination Exercises during Therapeutic Interventions. *Computers in Human Behavior* 71 (2017), 479–498.

[7] Gerry Chan, Anthony Whitehead, and Avi Parush. 2017. Dynamic Player Pairing: Quantifying the Effects of Competitive Versus Cooperative Attitudes. In *Game Dynamics*. Springer, 71–93.

[8] Yu Chen and Pearl Pu. 2014. HealthyTogether: Exploring Social Incentives for Mobile Fitness Applications. In *Second International Symposium of Chinese CHI (Chinese CHI '14)*. ACM, New York, NY, USA, 25–34.

[9] Mihaly Csikszentmihalyi. 1996. *Flow and the psychology of Discovery and Invention*. New York: Harper Collins.

[10] Amanda J Daley. 2009. Can exergaming contribute to improving physical activity levels and health outcomes in children? *Pediatrics* 124, 2 (2009), 763–771.

[11] Yvonne A. W. De Kort and Wijnand A. Ijsselsteijn. 2008. People, Places, and Play: Player Experience in a Socio-spatial Context. *Comput. Entertain.* 6, 2 (2008), 1–11.

[12] Samuel T Forlenza, Norbert L Kerr, Brandon C Irwin, and Deborah L Feltz. 2012. Is My Exercise Partner Similar Enough? Partner Characteristics as a Moderator of the Kohler Effect in Exergames. *Games for health journal* 1, 6 (2012), 436–441.

[13] Tracy Fullerton. 2014. *Game Design Workshop: a Playcentric Approach to Creating Innovative Games*. CRC press.

[14] Kathrin M Gerling, Regan L Mandryk, and Conor Linehan. 2015. Long-Term Use of Motion-Based Video Games in Care Home Settings. In *CHI 2015*. 1573–1582.

[15] Stefan Göbel, Sandro Hardy, Viktor Wendel, Florian Mehm, and Ralf Steinmetz. 2010. Serious games for health: personalized exergames. In *18th ACM international conference on Multimedia*. ACM, 1663–1666.

[16] Diana L Graf, Lauren V Pratt, Casey N Hester, and Kevin R Short. 2009. Playing Active Video Games Increases Energy Expenditure in Children. *Pediatrics* 124, 2 (2009), 534–540.

[17] Juho Hamari and Jonna Koivisto. 2015. "Working Out for Likes": An Empirical Study on Social Influence in Exercise Gamification. *Computers in Human Behavior* 50 (2015), 333–347.

[18] Hamilton A Hernandez, Mallory Ketcheson, Adrian Schneider, Zi Ye, Darcy Fehlings, Lauren Switzer, Virginia Wright, Shelly K Bursick, Chad Richards, and T C Nicholas Graham. 2014. Design and Evaluation of a Networked Game to Support Social Connection of Youth with Cerebral Palsy. In *ASSETS 2014*. 161–168.

[19] Katherine B Howe, Christian Suharlim, Peter Ueda, Daniel Howe, Ichiro Kawachi, and Eric B Rimm. 2016. Gotta catch'em all! Pokémon GO and physical activity among young adults: difference in differences study. *bmj* 355 (2016), i6270.

[20] Wijnand IJsselsteijn, Karolien Poels, and Yvonne AW de Kort. 2008. The Game Experience Questionnaire: Development of a self-report measure to assess player experiences of digital games. *TU Eindhoven, Eindhoven, The Netherlands* (2008).

[21] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-based Model of Personal Informatics Systems. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 557–566.

[22] Elke Mattheiss, Christina Hochleitner, Marc Busch, Rita Orji, and Manfred Tscheligi. 2017. Deconstructing Pokémon Go – An Empirical Study on Player Personality Characteristics. In *Persuasive Technology: Development and Implementation of Personalized Technologies to Change Attitudes and Behaviors*, Peter W. de Vries, Harri Oinas-Kukkonen, Liseth Siemons, Nienke Beerlage-de Jong, and Lisette van Gemert-Pijnen (Eds.). Springer International Publishing, Cham, 83–94.

[23] Florian Floyd Mueller, Rohit Ashok K, Kathrin M. Gerling, and Regan Mandryk. 2016. Exertion Games. *Foundations and Trends in Human-Computer Interaction* 10, 1 (2016), 1–86.

[24] Florian Floyd Mueller, Martin R Gibbs, Frank Vetere, and Darren Edge. 2017. Designing for Bodily Interplay in Social Exertion Games. *ACM Transactions on Computer-Human Interaction* 24, 3 (2017), 1–41.

[25] Cynthia L Ogden, Margaret D Carroll, Brian K Kit, and Katherine M Flegal. 2012. Prevalence of Obesity in the United States, 2009-2010. (2012).

[26] Rita Orji, Regan L. Mandryk, and Julita Vassileva. 2017. Improving the Efficacy of Games for Change Using Personalization Models. *ACM Trans. Comput.-Hum. Interact.* 24, 5, Article 32 (Oct. 2017), 22 pages. https://doi.org/10.1145/3119929

[27] Chad Richards and TC Graham. 2016. Developing Compelling Repetitive-motion Exergames by Balancing Player Agency with the Constraints of Exercise. In *ACM Conference on Designing Interactive Systems*. ACM, 911–923.

[28] Sarah-Jeanne Salvy, James N Roemmich, Julie C Bowker, Natalie D Romero, Phillip J Stadler, and Leonard H Epstein. 2008. Effect of peers and friends on youth physical activity and motivation to be physically active. *Journal of pediatric psychology* 34, 2 (2008), 217–225.

[29] Sarah Justine Guy Skriloff, Dario C Gonzalez, Kurtis C Christensen, Logan J Bentley, and Cody V Mortensen. 2016. FitPlay Games: Increasing Exercise Motivation Through Asynchronous Social Gaming. In *CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 164–167.

[30] A. Strauss and J. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for developing grounded theory.* (second ed.). SAGE Publications.

[31] G. F. Tondello, D. L. Kappen, E. D. Mekler, M. Ganaba, and L. E. Nacke. 2016. Heuristic Evaluation for Gameful Design. In *Annual Symposium on Computer-Human Interaction in Play extended abstracts - CHI PLAY EA '16*. ACM, Austin, TX, USA.

[32] Richard P Troiano, David Berrigan, Kevin W Dodd, Louise C Masse, Timothy Tilert, Margaret McDowell, et al. 2008. Physical Activity in the United States Measured by Accelerometer. *Medicine and science in sports and exercise* 40, 1 (2008), 181.