Keeping Users Engaged through Feature Updates: A Long-Term Study of Using Wearable-Based Exergames

Zhao Zhao, Ali Arya, Anthony Whitehead, Gerry Chan, S. Ali Etemad
Carleton University
Ottawa, Canada
{zhao.zhao, ali.arya, anthony.whitehead, gerry.chan, ali.etemad}@carleton.ca

ABSTRACT
Gamification and exergames in particular have been broadly employed in health and fitness as an attempt to promote exercise and more active life styles. Motivated by popularity and availability of wearable activity trackers, we present the design and findings of a study on the motivational effects of using activity tracker-based games to promote daily exercise. Furthermore, we have investigated user behaviors, usage patterns, engagement, and parameters that affect them. An exergame was developed with an accompanying wearable device, for which different variations of application updates were pushed out periodically over a 70-day period. The results of this long-term study show that the usage of wearable activity trackers during exercise, even when gamified for increased entertainment, sees a consistent decline over time. This decline, however, is observed to be reversible with periodic updates to the game. This work, we believe, can make a significant contribution to solving the user retention problem of wearable-based exergames.

Author Keywords
Gamification; fitness; exercise; wearable device; motivation; game sustainability; exergames.

ACM Classification Keywords
H.5.2 [User Interfaces]: Input Devices and Strategies; K.8.0 [Personal Computing]: Games

INTRODUCTION
Gamification (utilizing game thinking and game mechanics in non-game context [7]) in exercise and fitness has rapidly emerged over the past years as a tool to promote health and wellness [2, 6, 12, 24, 29, 33, 35]. It helps users to achieve certain fitness goals and increase engagement by adding game features to physical activity.

CHI 2017, May 06-11, 2017, Denver, CO, USA
 ISBN 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025982
This work is supported by Social Sciences and Humanities Research Council of Canada through IMMERSE Network.

In recent years, there has been a proliferation of physical activity tracking devices, such as Fitbit, Jawbone, Garmin, etc., and these wearable devices, along with their associated applications have proven to be popular platforms for promoting health and fitness due to their wide adaption, ease of use, and continuous proximity to the users [25].

Consequently, previous research [36] suggests that off-the-shelf wearable devices have considerable potential to be utilized in gamification of exercise and fitness. In our previous work, we utilized this concept through the design and implementation of a smartphone game application, which used wearable devices as input systems [36]. Short-term studies were conducted with combinations of different exercises and types of wearable devices to evaluate the usability and motivations of our system. The results showed that based on existing technologies and user needs, the idea of employing wearable activity trackers for gamification of exercise and fitness is feasible, motivating, and engaging.

In this paper, we extend the application with more diverse features and scenarios, followed by a 70-day long-term user study, to further investigate the motivational effects of using our system for promoting and sustaining exercise engagement. In order to have people maintain a constant habit or routine, applications should be designed with not only desirable usability but also sustainability, to provide users with sustainable motivation and engagement over time [4]. We designed the application to update regularly by adding new features to enable us to evaluate the effect of such time-based updates. Our research question is whether game features and their release system can improve engagement over time. We aimed to verify the hypothesis that gradually releasing game features will improve user retention over a longer period. Therefore, we divided our participants into three groups where the first group received the game with only very basic features, the second group received a more comprehensive set of various features such as customization and multiplayer mode. The third group started from only basic features but had different features gradually unlocked every 10 days. The behavioral patterns of all three groups were monitored by tracking users' activities as well as collecting their objective experiences through questionnaires both before and after the study.

This paper continues by covering related work, followed by the application design and implementation. It continues with the description of user study design, followed by the study results and analysis. Finally, some conclusions will be drawn to summarize our findings, as well as introduce some potential directions for future work.

RELATED WORK

Studies have shown that gamification in the area of health and fitness could motivate users to become more active [23]. The advantage of gamified exercise over traditional forms of exercise is that it is not just a physical activity, but also a form of entertainment. For example, exergames such as Just Dance (and other Wii or Kinect based exergames) offer participants a much higher level of engagement compared to regular exercise, which could reduce the level of perceived exertion. As a result, the level of motivation to stick with the game is also much higher than with traditional exercise [1]. Exergames can not only make exercise seem more enjoyable [35], but can also help improve overall well-being of older adults [17]. However, exergames may not be able to maintain long term interest in exercise [30] and some exergames are not intense enough to contribute toward the recommended daily amount of exercise for children because these games do not require as much energy expenditure as the actual sport [12]. To ensure the effectiveness of exergames, designers must consider two core issues for promoting adequate energy expenditure: 1) rewards for encouraging long term motivation and 2) better physical benefits which involves full body movements [33].

The smart phone game Zombies Run (https://zombiesrungame.com) was a successful example of gamified exercise where headphones were used to interact with users during running. This approach proved to have a positive and motivational effect [26] in promoting people to be more active. Göbel et al. [9] proposed a set of personalized exergames which combine the concepts of serious games and sensor technologies. It added customized features to the gameplay and result show that it enhances users' motivation toward the gameplay. Campbell et al. [4] discussed the concept of everyday fitness games and suggested that for applications that people frequently use in their everyday lives, designs need to be fun and sustainable, as well as adapt to behavioral changes. Wylie [34] indicated that with the use of smart phones, gamified components such as a leaderboard, achievements and challenge amongst friends are effective ways of motivating and encouraging users to reach their personal goals and track their physical fitness activities. Hamarji and Kovisto [13] investigated how social factors affect users' intention to continue using the gamified system, and their results also indicated that social factors are strong predictors for game motivation. Schwenk et al. [27] implemented a virtual reality (VR) game–based system prototype for balance rehabilitation, in which wearable sensors and interactive user interface were involved for real-time visual feedback. Findings support that real-time feedback is an important factor that could affect user's in-game experience.

In addition, Smeddinck et al. [28] conducted a 5-week study comparing exergames with traditional therapy. Results showed that a significant increase on the measure of functional reach when training with exergames. Macvean and Robertson [21] reported a 7-week study on user's physical activity, motivation and behavioral patterns on using exergames, and suggested that longitudinal studies are necessary for evaluating motivational effects, since it ensures that the intensity of a user's behavior is appropriate and sustained.

According to our literature review, game elements such as customization, social factors, as well as real-time feedback have been found to be important features of a gamified fitness application, and thus, we make use of these ideas in the design and implementation of our gamified exercise system to be evaluated over a long-term study.

To the best of our knowledge, no previous work has specific parameters of our study, i.e.: (1) combining a game with an activity tracker, (2) long-term continuous observation, (3) comparing 3 feature update methods. Gouveia et al. [11] conducted a long-term study, but users were randomly joining throughout the period; a game was not used; and updates were limited to different feedback messages/goals., In contrast, we invited those who are not very active but want to be more active, updated app features periodically, and used a game. We have followed up and improved on their suggestions of focusing on "intermediate stage", playful goals, and updated feedback. Other researchers [e.g.: 14, 20, 22] have also suggested similar strategies with no study, or have done studies with no app/game or update process.

APPLICATION DESIGN AND IMPLEMENTATION

An iOS game created by the authors for a short-term study [36] is used as the basis for this work. However, we have added a series of new features to the original single-user single-level game, as shown in Table 1. These added game features were designed to provide users with choices of customization, to enhance real-time game experience, as well as to incorporate social factors.

Customization features	Customized bird color
	Customized background color
Real-time game features	Bird nets (in level 2)
	Thunderstorm (in level 3)
Social features	Leaderboard
	Achievement
	Challenge box

Table 1. List of features added in the new application

The iOS application communicates with wearable activity trackers (TI Sensortag [32]) in real-time and utilizes the data received from its inertial measurement units (IMU) as inputs to the game. Players can play the games with three exercises: running, cycling or rope skipping to control the movements of the game character in real-time. For running and cycling we have a phone mounted in front of the user on the treadmill/bicycle while they exercising, and for rope skipping we casted the phone screen to a TV to make sure users can easily see the screen while jumping. (see Figure 1) The story of the game is based on a bird that has fallen behind its flock. The objective is to regroup with the flock before time runs out and before the flock reaches its destination. There are also physical and environmental parameters that may affect the movement behavior of the bird and flock, for example a bird net may appear to catch the bird and the player has to reach a certain speed in order to break the net. Another example is a thunderstorm that may occur and threaten the flock. The game features a mission-based structure and point-based system. Higher levels can be unlocked when missions have been achieved. The achievement system is also designed to encourage players during the gameplay. Multiplayer mode brings in leaderboard and challenge boxes that allows players to compete with each other. Figure 2 presents several screens from the application. Additionally, the ability to customize character and background colors was implemented in the application (see Figure 3). We used HockeyApp [14] to distribute different versions of our application to participants, and to receive user feedback.

Figure 1. The three types of exercises in our user study: running, cycling, and rope skipping.

Figure 2. Screen shots from our original application. From left to right, row one: choice of player mode, choice of exercises, result screen, achievement screen; row two: landing screen, level 1 view, level 2 view, and level 3 view.

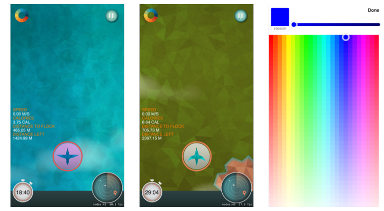

Figure 3. Screen shots from our updated application. From left to right: customized bird color for level1, customized background color for level 2, color picker panel view.

USER STUDY DESIGN

Participants and Updating Schedule

Our user study was designed to verify the hypothesis that adding different game features and gradually releasing them can positively affect the user engagement and retention. 36 participants were recruited from Carleton University by posters for our study. 30 were recruited as main subjects and 6 as backups who would replace any of the original 30 participants if they quit the study during the 70-day duration of the experiment. The backup participants started and participated in the experiments as others. Each participant was provided with a TI Sensortag [32] wearable device (see Figure 4) to take home. The wearable could be worn on the wrist or ankle (depending on the exercise they choose; for example, for cycling the sensor would be worn on the ankle). The wearable then connects to the game application that we distributed to the users. No particular set of instructions was provided, and users were allowed to explore the system on their own.

Figure 4. The wearable device used in this study (TI Sensortag).

Basic Group	Basic features: single level, single player.
Full Group	Basic feature plus all the features introduced in Table 1.
Updating Group	Start with basic feature (same as the Basic Group), with other features gradually added (as described in Table 3).

Table 2. List of different features that each group received.

The participants were then randomly divided into three equal groups of 12 (10 as main and 2 as backup). To users in the first group (referred to as Basic Group), we only provided very basic features of the game such as a single level (level 1) version of the game, as well as single player mode. Users in the second group (referred to as Full Group) were provided the full version of the application, including the ability to customize colors, as well as all 3 levels of the game, and multiplayer mode with leaderboard and challenge box, etc. Users in the third group (referred to as Updating Group) started with the same application as those in the Basic Group, but periodically received updates (once every 10 days), until after 70 days they ended up with an application that was identical to those in the Full Group. Users were told in Updating group about receiving updates every 10 days. On each update, we used push notification to let them know about new feature. The different features each group had received are summarized in Table 2 below, and the update schedule for the Updating Group is shown in Table 3.

Day	Updates
Day 1	Single level, single player
Day 11	Add customized bird color
Day 21	Add multiplayer leaderboard and challenge-box
Day 31	Add level 2 (with bird net)
Day 41	Add customized background color
Day 51	Add achievements
Day 61	Add level 3 (with thunderstorm)

Table 3. Automatic updates schedule used for the Updating Group.

Procedure and Data Collection

Pre-study

We asked participants to answer questionnaires upon receiving and returning the wearable device. The pre-study questionnaire asked demographic questions including age, gender, height, weight, the hours they spend per week in exercising, any wearables owned, and the types and duration of video games (PC, console, and mobile) played.

In-game data

For in-game data collection, we used Google Analytics API [10] to track users' comprehensive behavior data, including screen views and tapped events with associated timestamp, exercise sessions, performance, scores, and feedback and comments. We also asked users about their level of encouragement through a 5-point Likert question (from 1 = highly discouraged to 5 = highly encouraged) after each completed session they have exercised.

Post-study

A post-study questionnaire was conducted to comprehensively evaluate participants' experience during the study. We firstly asked three general questions to measure participants' overall motivation, satisfaction and preference based on their experience with the game:

1) I find this kind of application motivation for exercise;
2) I was overall satisfied with this application;
3) I prefer using this type of application for exercise over regular exercises.

We included parts of Intrinsic Motivation Inventory (IMI) [15] measurement to assess participants' subjective experience related to the game experience. The detail questions are attached in Appendix for reference. Since the goal of this research project is to encourage continuous participation in exercise, IMI was selected as a measure of motivation because it evaluates a person's experience while performing an activity based on criteria such as enjoyment, effort, and competence [15]. In exergames, an enjoyable game experience is associated with greater amount of energy expenditure [20] and player achievement predicts feelings of autonomy, competence, presence, enjoyment and continued motivation to play [19]. If participants report high levels on these criteria, it is likely they are intrinsically motivated and enjoy what they were doing, and will continue doing it.

In the post-study questionnaire, participants used a 7-point Likert scale (from 1 = strongly disagree to 7 = strongly agree) to rate the game-play experience, followed by some open-ended questions. Note that compared to the question regarding encouragement we asked in-game after each complete session, here we switched from 5-point to 7-point because for the in-game question (encouragement) we kept a consistent scale with our previous work [36] with a 5-point Likert scale, but because of our interest in the intrinsic motivation in this current study, we employed the IMI [15] which recommends the use of a 7-point scale.

We also included some open-ended questions regarding their preferred exercise and game mode, suggestions on both software and hardware of the system, as well as their opinion on how the game could be more engaging based on their game experience.

Ethics approval was received from the Carleton University Ethics committee. Participants received a $10 honorarium gift card to thank them for their participation in the study. During the 70 days of the user study, 4 out of the 30 participants withdrew for different reasons, and 4 participants from the back-ups of the same study group were randomly selected to replace the original participants.

RESULTS AND ANALYSIS

Demographic information

Out of the 30 volunteers who participated in this study, 19 were males and 11 were females. Their average age was 24.71 years with the standard deviation of 4.23 years. Their self-reported average hours of exercise per week was 3.81 hours with a standard deviation of 2.83 hours, while the self-reported average hours per week spent on playing games was 6.33 for pc/console games (with a standard deviation of 9.51) along with a 4.73 hours per week spent on playing smart-phone games (with a standard deviation of 6.33). 11 out of 30 participants (36.7%) previously (or currently) owned an activity tracking wearable device.

In-game Data

Figure 5 shows participants' choice of exercise among the three classes enabled during our study (from in-game data monitoring), where running is the most popular, followed by cycling. This result matches the subjective results we collected from the post-study questionnaire, where the vast majority preferred running as well. Figure 6 shows the feedback that we collected during the study (after each completed session) using a 5-point Likert scale (from 1 = highly discouraged to 5 = highly encouraged) for regarding the level of encouragement by the system towards more exercise. A total of 322 answers were collected, 47 from the Basic Group, 105 from the Full Group and 170 from the Updating Group. We can see that most of the users felt positively regarding this gamified exercise experience. The Full Group felt most encouraged (with 82% of participants choosing options higher than "no difference", followed closely by the Updating (80%) and Basic Groups (70%).

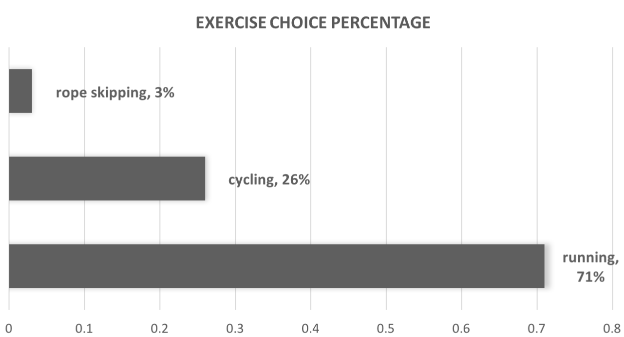

Figure 5. Choice of exercise (from in-game data monitoring).

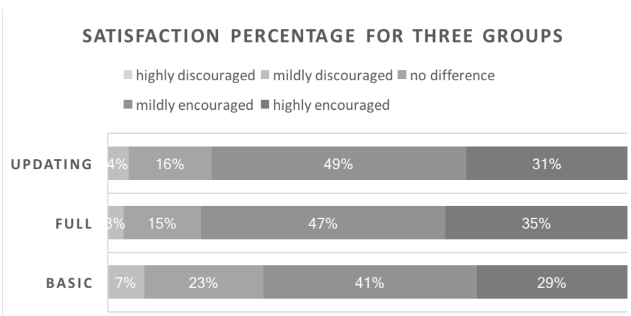

Figure 6. Level of encouragement for three groups (asked after each completed workout session).

Overall Motivation, Satisfaction and Preference

A one-way between-groups analysis of variances (ANOVA) and Post-Hoc Tukey-Kramer HSD test [16] was conducted to analyze the main effects between three groups. The alpha was set at 0.05 for all statistical tests.

Figure 7 shows the average and standard deviations of the scores for question 1 to 3. Please note that the subscript B represents the Basic Group, F represents the Full Group and U represents the Updating Group. For motivation, M_B = 3.80 and SD_B = 0.79, M_F = 5.30 and SD_F = 0.95, M_U = 5.50 and SD_U = 0.92. There was a significant effect at the $p < 0.05$ level for the three groups [$F(2, 27) = 10.50$, $p = 0.0042$]. For preference, M_B = 3.90 and SD_B = 1.49, M_F = 5.7 and SD_F = 0.82, M_U = 5.6 and SD_U = 0.66. There was a significant effect at the $p < 0.05$ level for the three groups [$F(2, 27) = 12.42$, p = 0.00014]. For preference over regular exercise, M_B = 3.90 and SD_B = 1.49, M_F = 5.70 and SD_F = 0.82, M_U = 5.60 and SD_U = 0.66. ANOVA shows no significant effect between groups [$F(2, 27) = 3.17$, $p = 0.0581$]. The Tukey-Kramer HSD post-hoc test results for significant factors are shown in Table 4 below and the pairwise comparison significance are also marked in Figure 7. The output of the Tukey test shows the average difference, a confidence interval as well as whether we should reject the null hypothesis for each pair of groups at the given significance level, in which "True" indicates there is a significance between the two groups.

	group1	group2	meandiff	lower	upper	reject
Motivation	B	F	1.5	0.4947	2.5053	True
	B	U	1.7	0.6947	2.7053	True
	F	U	0.2	-0.805	1.2053	False
Satisfaction	B	F	1.9	0.8613	2.9387	True
	B	U	1.7	0.6613	2.7387	True
	F	U	-0.2	-1.239	0.8387	False

Table 4. Multiple Comparison of Means (general question) - Tukey HSD, FWER=0.05

From the results, we can see that the Full Group and the Updating Group show significantly higher motivation and satisfaction compared to the Basic Group. This suggests that the number and types of features are the main factor affecting users' overall motivation and satisfaction. No significant effect in preference over regular exercise, suggests that all participants, even in Basic group, still prefer using the gamified exercise experience over regular exercises.

Feature Importance Level

Only the Full Group and the Updating Group were asked to answer the second section of the questionnaire regarding the different game features (the Basic Group had only received a basic version of the game without any features for evaluation). We used statements such as: "Leaderboard was important in my game experience." to evaluate the importance level of each separate game feature. The results are shown in Figure 8, in which feature 1 to 7 represent:

1) Customized bird color,
2) Customized background color,
3) Bird nets (in level 2),
4) Thunderstorm (in level 3),
5) Leaderboard,
6) Achievements,
7) Challenge box.

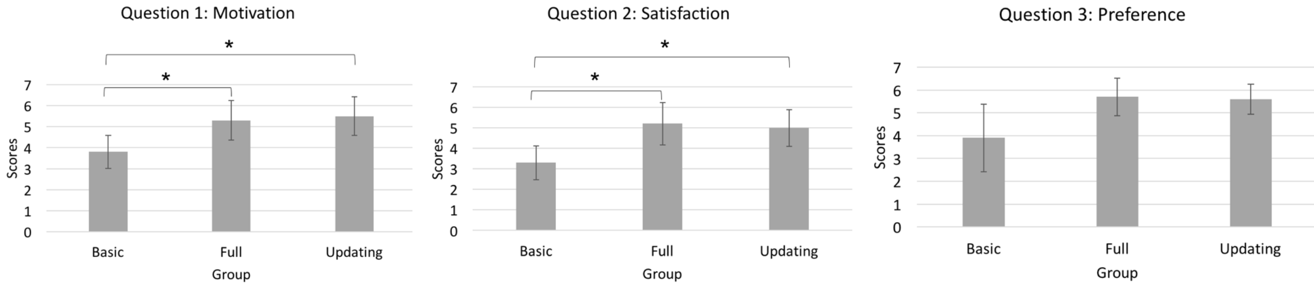

Figure 7. Result for post-study question 1, 2 and 3. From left to right: overall motivation, overall satisfaction, and overall preference over regular exercise.

Results on the importance level of different game features show that social factors including Leaderboard ($M_F = 5.71$ and $SD_F = 0.82$, $M_U = 5.04$ and $SD_U = 1.05$) and Achievements ($M_F = 5.78$ and $SD_F = 0.79$, $M_U = 5.34$ and $SD_U = 1.16$) are playing important roles in game experience. However, also being one of the social factors, challenge box is considered less important. A possible reason leading to this result could be the lack of real-life relationships between our participants. We believe a closer relationship between users could strengthen the importance level of challenge box. We did not add invitation feature to this study since we need to keep a consistent user group, but adding such ability in future can potentially add to the social value. Also, we found the scores of Full Group are relatively higher than the Updating group in most of the game features except for customized bird color. This might because the customized bird color was the first updating feature that the Updating Group had received, so that it became more impressive compared to other features, although this is not conclusive.

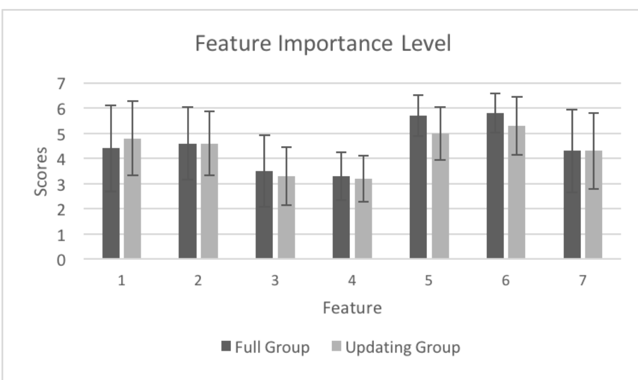

Figure 8. Impact of different features for the Full Group and the Updating Group

IMI sub-scales

A reliability analysis is conducted because we modified the original IMI scale to better fit the concept of game. A Cronbach's alpha value of 0.878 ($\alpha = 0.878$) is obtained (a commonly accepted rule is that an alpha of 0.7 indicates acceptable reliability and 0.8 or higher indicates good reliability. [31])

As shown in Figure 9, for interest and enjoyment ($\alpha = 0.862$), $M_B = 3.83$ and $SD_B = 1.34$, $M_F = 5.17$ and $SD_F = 1.049$, $M_U = 5.56$ and $SD_U = 0.80$. There was a significant effect at the $p < 0.05$ level for the three groups [$F(2, 27) = 29.43$, $p = 1.65e\text{-}07$].

For perceived competence ($\alpha = 0.887$), $M_B = 4.05$ and $SD_B = 1.32$, $M_F = 5.18$ and $SD_F = 1.03$, $M_U = 5.36$ and $SD_U = 0.80$. There was a significant effect at the $p < 0.05$ level for the three groups [$F(2, 27) = 25.15$, $p = 6.82e\text{-}07$].

For effort/importance ($\alpha = 0.848$), $M_B = 2.70$ and $SD_B = 0.76$, $M_F = 3.44$ and $SD_F = 0.84$, $M_U = 3.80$ and $SD_U = 0.78$. There was a significant effect at the $p < 0.05$ level for the three groups [$F(2, 27) = 12.94$, $p = 0.0001$].

For pressure/tension ($\alpha = 0.904$), $M_B = 5.64$ and $SD_B = 0.63$, $M_F = 6.27$ and $SD_F = 0.65$, $M_U = 5.98$ and $SD_U = 0.65$. There was no significant effect at the $p < 0.05$ level for the three groups [$F(2, 27) = 2.69$, $p = 0.086$].

Finally, for value/usefulness ($\alpha = 0.925$), $M_B = 3.56$ and $SD_B = 0.86$, $M_F = 5.15$ and $SD_F = 0.97$, $M_U = 5.13$ and $SD_U = 0.88$. There was a significant effect at the $p < 0.05$ level for the three groups [$F(2, 27) = 28.98$, $p = 1.90e\text{-}07$].

The Tukey-Kramer HSD test results for those sub-scales with significance factors are shown in table 5 below and the pairwise comparison significance are also marked in Figure 9.

	group1	group2	meandiff	lower	upper	reject
Interest/ enjoyment	B	F	1.3429	0.7565	1.9292	True
	B	U	1.7286	1.1422	2.3149	True
	F	U	0.3857	-0.201	0.9721	False
Perceived competence	B	F	1.1333	0.6347	1.6319	True
	B	U	1.3167	0.8181	1.8153	True
	F	U	0.1833	-0.315	0.6819	False
Effort/ importance	B	F	0.74	0.1934	1.2866	True
	B	U	1.1	0.5534	1.6466	True
	F	U	0.36	-0.187	0.9066	False
Value/ usefulness	B	F	1.6	1.0037	2.1963	True
	B	U	1.5714	0.9752	2.1677	True
	F	U	-0.029	-0.063	0.5677	False

Table 5. Multiple Comparison of Means (IMI sub-scales) - Tukey HSD, FWER=0.05

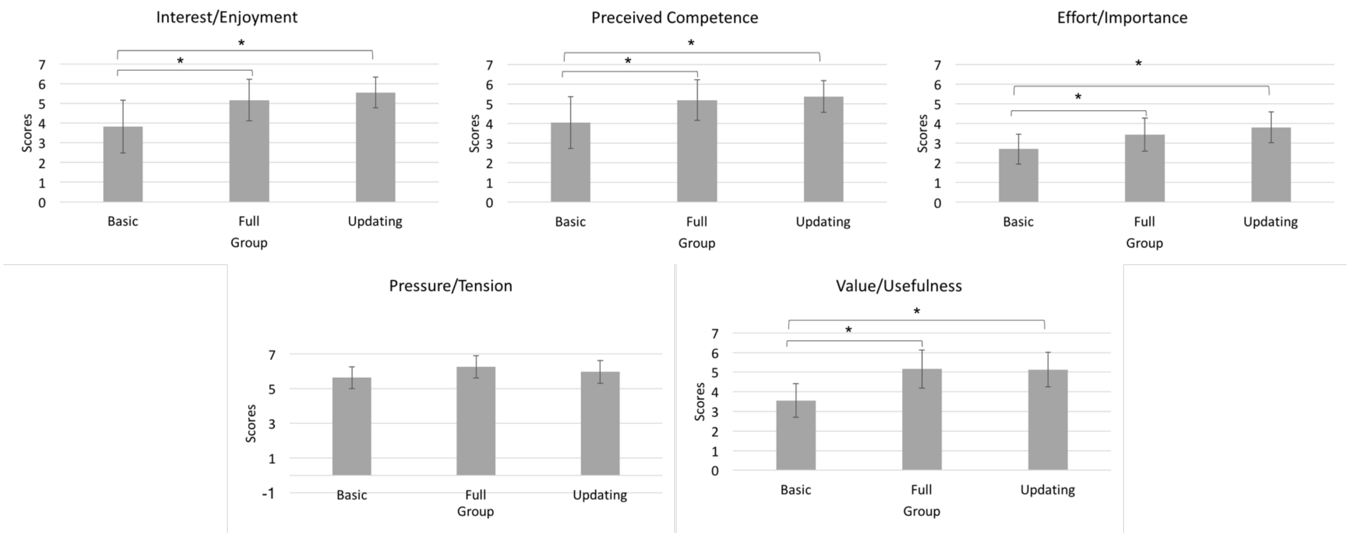

Figure 9. Average and Standard deviation of IMI: Interest/Enjoyment, Perceived Competence, Effort/Importance, Pressure/Tension and Value/Usefulness.

From the result, we can see that the number and diversity of game features could have impacts on user's motivation towards the gamified exercise system, because there are significant effects between the Basic Group and the Full Group, as well as between the Basic Group and the Updating Group. On the other hand, gradual updating does not take away from motivation, since there is no significant effect between the Full Group and the Updating Group. This indicates that by pushing out gradual updates, we can benefit from the advantages of the approach, namely increased engagement and retention, while not negatively impact motivation.

Daily Active Users and Workout Sessions

Figure 10 shows the change of active users of all three groups during the 70 days. We can see for active users, there is an overall descent trend appears as time grows for all three groups, while the Basic Group declines the fastest. The waves of the Full Group were basically caused by weekends. The large variance in the Updating Group was caused by the regular updating plus the weekend. Figure 11 shows the accumulated activate users for three groups with regression line, in which the slope of the Basic Group is 1.75, for the Full Group is 3.23, and for the Updating Group is 4.30. Figure 12 shows the total started workout sessions and complete sessions of all three groups, from which we noticed a basically same trend as the active users. The decline of Basic Group is the fastest, and the Full Group is mostly affected by weekends when people have more time for exercise, as in Figure 14 shows the participants' overall completed sessions by time of day. The Updating Group maintains a relatively higher value both in started and completed sessions, which indicates the gradually updating not only brings back users to check the updates, but actually promotes engagement towards the actual behavior of exercise. Figure 13 shows accumulated completed sessions with regression line, in which the slope of the Basic Group is 1.18, for the Full Group is 2.33, and for the Updating Group is 3.26. Compared to Figure 11, we found the accumulated complete sessions are overall rise slower compared to active users. This indicates there are circumstances that participants only either open the application to check updates or play around with it, without actually accomplishing any exercise session. This made us realize that despite the elements we considered important regarding exercise experience, other features that not directly linked to gameplay may also affect user retention and the application's life cycle.

Users received notifications (through Apple Push Notification service) when updates were available for the application. It was observed that many users returned to the application to check the updates. This contributed to the number of completed workout sessions being increased. Since the Updating Group contains the most active users and completed sessions, we can conclude that the gradual releasing of updates can not only attract users to interact more with the application, but also motivate them to complete gamified exercise sessions. The total in-game workout time of the Basic Group was 439 minutes (4.3 minutes per session on average), the total time for the Full Group was 1073 minutes (an average 6.1 minutes per session), and for the Updating Group the total was 1322 minutes (an average of 5.7 minutes per session). The reason that average time for the Updating Group is slightly lower than the Full Group is that we released the second and third level at the beginning to the Full Group but in the middle of the study to the Updating Group, and higher levels are longer in duration. We can see that both in Figure 8 and Figure 10, at the start of the study, the Full Group accumulated more active users and completed sessions, but roughly after the second update, the accumulated numbers of the Updating Group exceeded the Full Group and kept a higher slope till the end of the study. We can conclude that user's curiosity towards new features can make the game more sustainable over the long run, resulting in relatively higher user retention rate. When we asked participants which exercise (running, cycling or rope skipping) they preferred, 69% chose running, 24% chose cycling, and 7% chose rope skipping. This answer was consistent with the data acquired for gameplay. For the question of which game mode (single or multi-player) they preferred, 73% chose multiplayer while 27% chose single player. We collected 122 pieces of feedback from participants by in-game pop-up question asked after each complete workout session (23.9% of the total completed sessions). The most popular comment (56 of 122) suggested they would like this kind of application to support outdoor activities, and some of them (25 of 122) stated they would prefer real-time competition mode. Additional comments were received regarding including more in-game interactions, more rewards, and more social factors (such as allowing to share to social networks).

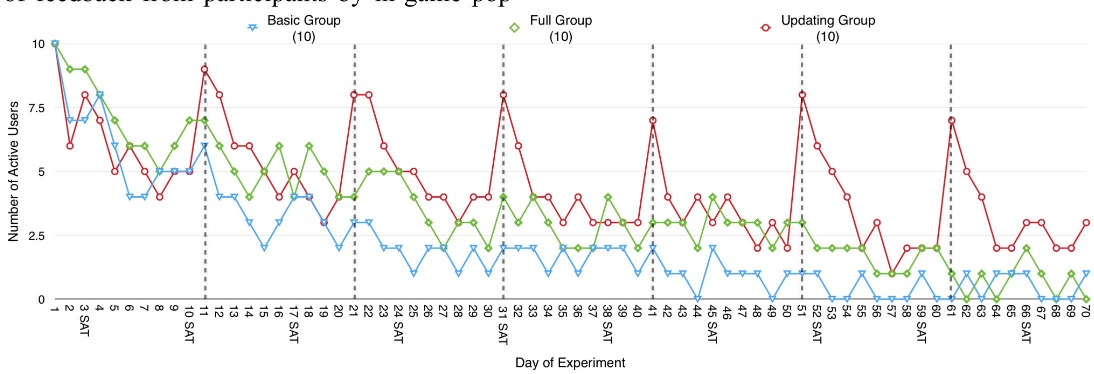

Figure 10: Number of active users for each group is depicted per day (dotted vertical lines indicate the application update days).

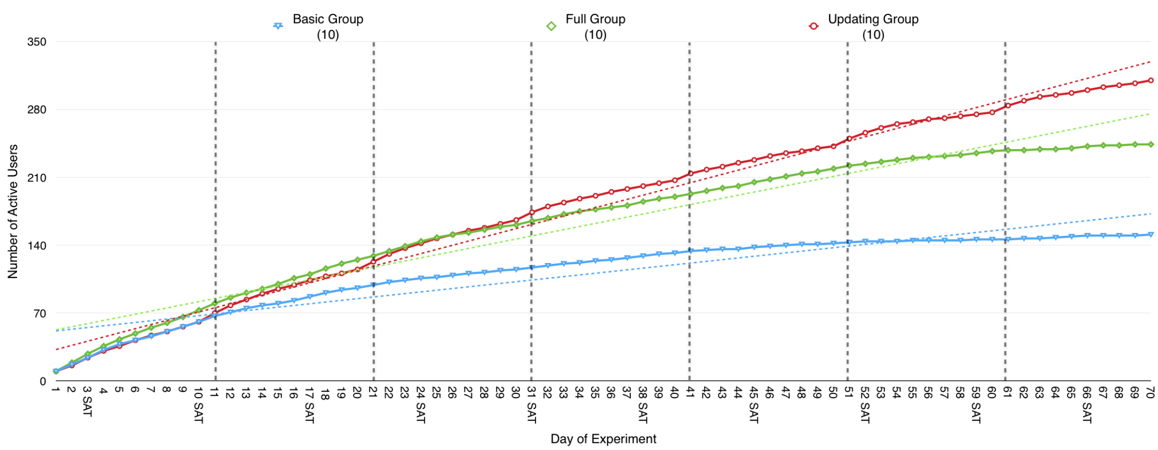

Figure 11. Plot of accumulated active users for all 3 groups with regression line

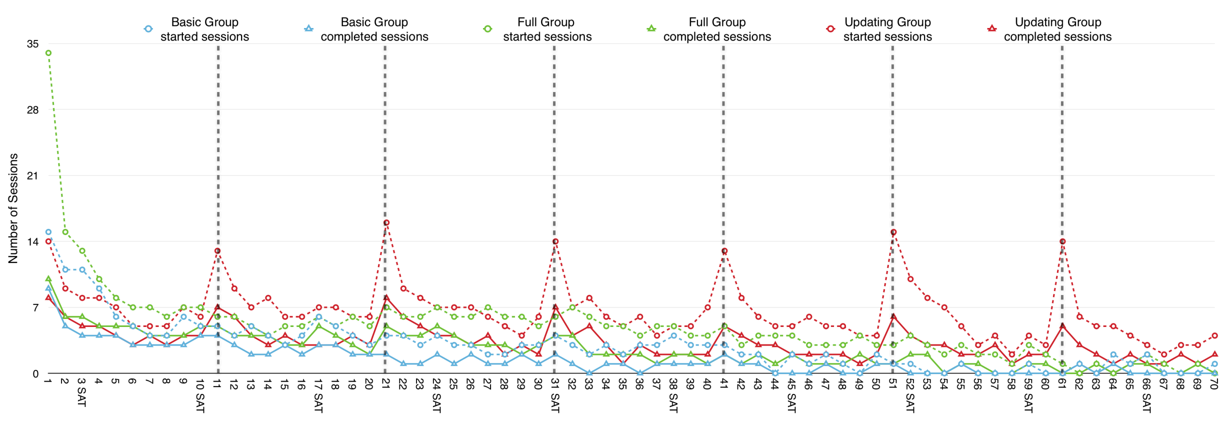

Figure 12. Number of started and completed sessions per day (dotted vertical lines indicate the application update days)

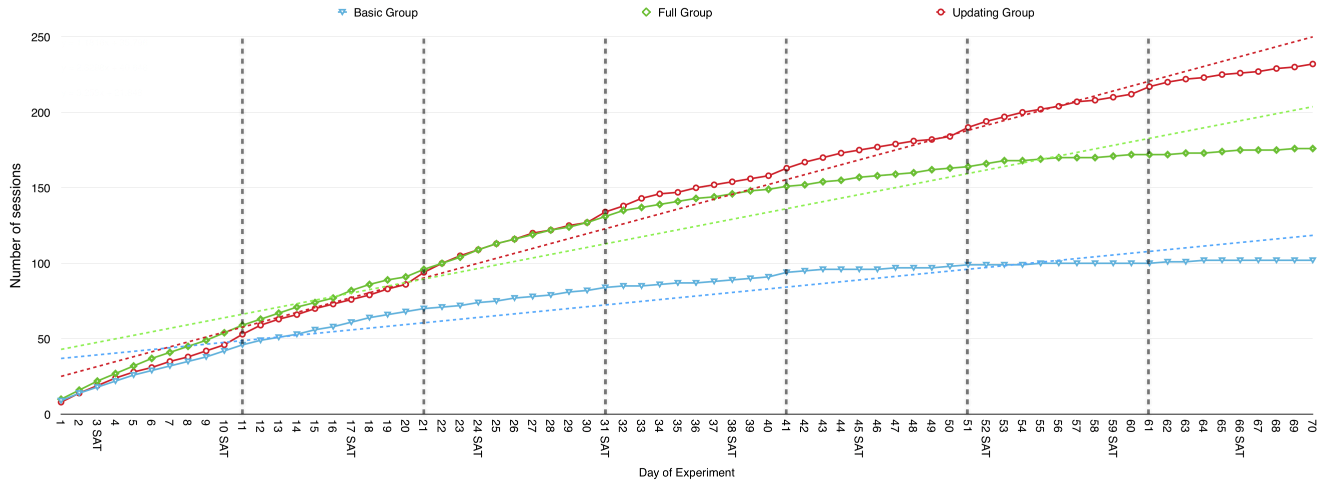

Figure 13. Accumulated completed sessions for all 3 groups with regression line

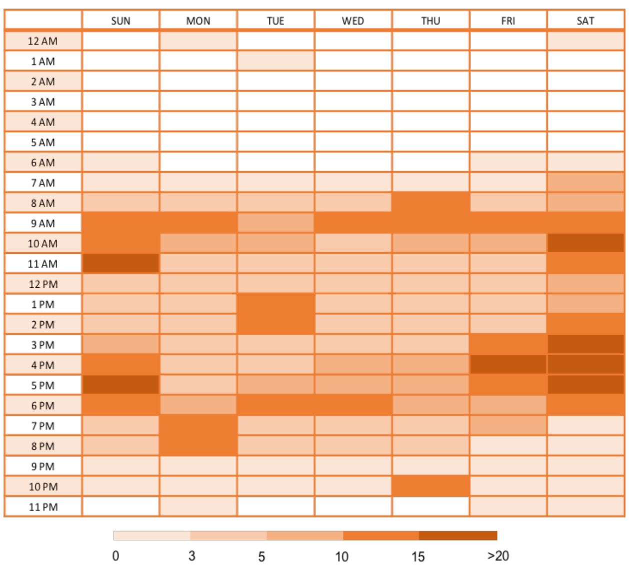

Figure 14. Total sessions by time of day.

DISCUSSION

Overall in this 70-day user study, the evaluation on motivation showed that more and diverse features could increase user motivation toward the gamified exercise system, and the gradual release of game features does not negatively affect user motivation. The analysis on game sustainability showed that while there was a decline in activity level in all groups, the gradual release of features improved the sustainability to a certain extent which verifies our hypothesis that various game features and the way they are released will influence user retention.

In this study we invited intermediate stage subjects and selected them randomly. However, they had a slightly higher initial exercise rate. Looking at the results from the 5 subjects with the least self-reported exercise rates, we see that the pattern reported in the paper still exists, even stronger (higher performance in Updating group). Among them, 2 were from the Basic (total in-game workout time 1.9 and 3.0; group average 2.8), 2 from the Updating (4.2 and 6.4; group average 5.3), and 1 from the Full (3.9 vs. average 3.9).

While the Basic and Full feature groups started to asymptotically level off within the 10-week study (in weeks 5 and 9 respectively), this was not apparent in the Updating group (see Figure 13). This suggests that minor feature updates can have the effect of increasing sustainability over the long term. We further believe that a combination of minor and major updates will likely have an even longer effect on sustainability.

It can be seen in Figure 12 that sometimes the notification of new features only brings people back to check updates but without completing workouts. This increases the number of minutes of engagement with the game, not the number of actual exercises. One of the participants indicated that changes such as customized color did not create enough interest to complete a session. However, addition of the leaderboard significantly increased the motivation to complete sessions with improved results. Accordingly, we can see that different users were motivated by different features, and this can be considered as the main cause of some incomplete sessions. Player modeling techniques can determine the types of features and updates that can motivate different users. Therefore, customized feature updates may help improve workout completion rates, which is part of our future work.

The results of this study also show that participants in all groups preferred the gamified exercise experience over regular exercises; features related to social factors played a relatively more important role in this game experience, followed by customization features; and participants spent more time interacting with the system on weekends, especially during 9:00 AM to 11:00 AM, and 3:00 PM to 6:00 PM.

There are some limitations regarding this work. One limitation is with respect to participant characteristics. The participants consisted only of university students, which limits the generalizability of the results. Future research may replicate this study by recruiting participants of various age groups, gamer types (e.g.: gamer vs. non-gamers) and people with different exercise goals (e.g.: leisure vs. fitness). People of different age or who have different exercise goals could have different motivations for using the system. It is important to understand what these goals are and continue to redesign and redevelop game features to keep players engaged in the game and exercise. Another limitation of this study was data collection. We used Google Analytics in this work, which contains less than 1.5% missing data. While this is considered acceptable by most standards [3], the accuracy of the results is reduced nonetheless. Thus, a future study could employ more accurate analytical tools/software for obtaining better results.

CONCLUSION

In this study, we extended an existing gamified exercise application with a set of more diverse features to evaluate the motivational impacts and game sustainability of a wearable-based gamified exercise and fitness system. The results of this long-term (70-day) user study shows that engagement and game-play are highly linked, and the number of game features could have impacts on user motivation towards the gamified exercise system. Moreover, consistent updates (gradual addition of new features) not only resulted in an increased usage of the application, but also had positive impacts on the actual amount of workout activity.

Finally, the feedback collected from the participants indicated a few potential paths for future work, such as the addition of support for outdoor activities, addition of real-time multiplayer modes, and others. In our future work, we will also focus on utilizing player modelling techniques to better understand users from different perspectives, and to make adaptive adjustments to the game. A comprehensive gamified fitness advisor and recommendation system can be further developed. A key follow-up to this study may be a longer study (over many months) to see how sustainability can be better understood and impacted.

APPENDIX

The Intrinsic Motivation Inventory [15]:

Interest/Enjoyment was measured by seven sentences:

1) I enjoyed playing this game very much;
2) This game was fun to play;
3) I thought this was a boring game (reverse coded);
4) This game did not hold my attention at all (reverse coded);
5) I would describe this game as very interesting;
6) I thought this game was quite enjoyable;
7) While I was playing this game, I was thinking about how much I enjoyed it.

Perceived Competence was measured by six sentences:

1) I think I am pretty good at this game;
2) I think I did pretty well at this game, compared to other players;
3) After playing this game for a while, I felt pretty competent;
4) I am satisfied with my performance at this time;
5) I was pretty skilled at this game;
6) This was a game that I couldn't do very well (reverse coded).

Effort/Importance was measured by five sentences:

1) I put a lot of effort into this game;
2) I didn't try very hard to do well at this game (reverse coded);
3) I tried very hard at this game;
4) It was important to me to do well at this game.
5) I didn't put much energy into this game (reverse coded).

Pressure/Tension was measured by five sentences:

1) I did not feel nervous at all while doing this (reverse coded);
2) I felt very tense while doing this activity;
3) I was very relaxed in doing these (reverse coded);
4) I was anxious while working on this task;
5) I felt pressured while doing these.

Value/Usefulness was measured by 7 statements;

1) I believe this game could be of some value to me;
2) I think that playing this game is useful for keeping me active;
3) I think this is important to do because it can motivate me to engage in exercise;
4) I would be willing to lay it again because it has some value to me;
5) I think playing this game could help me to exercise more often;
6) I think playing this game could be beneficial to me;
7) I think this is an important game.

REFERENCES

1. Riikka Ahola, et al. 2013. Gamified physical activation of young men: A multidisciplinary population-based randomized controlled trial (MOPO study). *BMC Public Health* 13, 32: 1–8.
2. Marie-Louise Bird, Brodie Clark, Johanna Millar, Sue Whetton, and Stuart Smith. 2015. Exposure to "exergames" increases older adults' perception of the usefulness of technology for improving health and physical activity: A pilot study. *JMIR Serious Games* 3, 2: e8. doi: 10.2196/games.4275.
3. Blast Analytics & Marketing. 2013. *Can You Trust Your Google Analytics Data?* Retrieved September 5

2016 from
http://www.blastam.com/blog/can-you-trust-your-google-analytics-data

4. Taj Campbell, Brian Ngo, and James Fogarty. 2008. Game design principles in everyday fitness applications. *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, ACM2008*, 249–252.
5. James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 647–658.
6. Amanda J. Daley. 2009. Can exergaming contribute to improving physical activity levels and health outcomes in children? *Pediatrics,* 124, 2: 763–771.
7. Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining gamification. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. ACM2011.
8. Daniel A. Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A. Munson. 2016. Beyond abandonment to next steps: Understanding and designing for life after personal informatics tool use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1109–1113.
9. Stefan Göbel, Sandro Hardy, Viktor Wendel, Florian Mehm, and Ralf Steinmetz. 2010. Serious games for health: Personalized exergames. *Proceedings of the 18th ACM international conference on Multimedia*, ACM2010, 1663–1666.
10. Google.com. 2016. *Google Analytics - Mobile, Premium and Free Website Analytics – Google*. Retrieved September 5 2016 from https://www.google.com/analytics
11. Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2015. How do we engage with activity trackers? A longitudinal study of Habito. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1305–1316.
12. Lee Graves, Gareth Stratton, N. D. Ridgers, and N. T. Cable. 2008. Energy expenditure in adolescents playing new generation computer games. *British journal of sports medicine* 42, 7: 592–594.
13. Juho Hamari and Jonna Koivisto. 2013. Social motivations to use gamification: An empirical study of gamifying exercise. In *ECIS* 2013, 105.
14. Hockeyapp.net. 2016. *HockeyApp - The Platform for Your Apps*. Retrieved August 5, 2016 from https://www.hockeyapp.net
15. Intrinsic Motivation Inventory. 2016. Self-Determination Theory. http://selfdeterminationtheory.org/intrinsic-motivation-inventory/
16. James Jaccard, Michael A. Becker, and Gregory Wood. 1984. Pairwise multiple comparison procedures: a review. *Psychological Bulletin,* 96, 3: 589–596.
17. Younbo Jung, et al. 2009. Games for a better life: effects of playing Wii games on the well-being of seniors in a long-term care facility. *Proceedings of the Sixth Australasian Conference on Interactive Entertainment*. ACM2009.
18. Evangelos Karapanos. 2015. Sustaining user engagement with behavior change tools. *Interactions,* 22, 4: 48–52.
19. Anthony M. Limperos and Mike Schmierbach. 2016. Understanding the relationship between exergame play experiences, enjoyment, and intentions for continued play. *Games for Health Journal,* 5, 2: 100–107.
20. Elizabeth J. Lyons, Deborah F. Tate, Dianne S. Ward, Kurt M. Ribisl, J. Michael Bowling, and Sriran Kalyanaraman. 2014. Engagement, enjoyment, and energy expenditure during active video game play. *Health Psychology,* 33, 2: 174–181.
21. Andrew Macvean and Judy Robertson. 2013. Understanding exergame users' physical activity, motivation and behavior over time. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM.
22. Edward McAuley, Terry Duncan, and Vance V. Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*, 1: 48–58.
23. Florian 'Floyd' Mueller, Firaz Peer, Stefan Agamanolis, and Jennifer Sheridan. 2011. Gamification and exertion. *Proceedings of the Workshop on Gamification at the Conference on Human Factors in Computing Systems*. ACM.
24. Huynh Van Nguyen, et al. 2016. Double-edged sword: The effect of exergaming on other forms of exercise; a randomized controlled trial using the self-categorization theory. *Computers in Human Behavior,* 62: 590–593.
25. Lukasz Piwek, et al. 2016. The rise of consumer health wearables: promises and barriers. *PLoS Med.,* 13, 2: e1001953.
26. Positive Gaming. 2016. The Benefits of Exergaming. Retrieved September 5, 2016 from

http://www.positivegaming.com/positivegaming/benefits/exergaming-benefits

27. Michael Schwenk, Gurtej S. Grewal, Bahareh Honarvar, Stefanie Schwenk, Jane Mohler, Dharma S. Khalsa, and Bijan Najafi. 2014. Interactive balance training integrating sensor-based visual feedback of movement performance: a pilot study in older adults. *Journal of Neuroengineering and Rehabilitation*, 11, 1: 1.
28. Jan David Smeddinck, Marc Herrlich, and Rainer Malaka. 2015. Exergames for physiotherapy and rehabilitation: A medium-term situated study of motivational aspects and impact on functional reach. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM.
29. Amanda E. Staiano and Sandra L. Calvert. 2011. Exergames for physical education courses: Physical, social, and cognitive benefits. *Child development perspectives,* 5, 2: 93–98.
30. Haichun Sun. 2013. Impact of exergames on physical activity and motivation in elementary school students: A follow-up study. *Journal of Sport and Health Science,* 2, 3: 138–145.
31. Mohsen Tavakol and Reg Dennick. 2011. Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2: 53–55.
32. Ti.com. 2016. *Simplelink SensorTag - TI.com*. Retrieved September 5, 2016 from http://www.ti.com/ww/en/wireless_connectivity/sensortag2015/?INTC=SensorTag&HQS=sensortag
33. Anthony Whitehead, Hannah Johnston, Nicole Nixon, and Jo Welch. 2010. Exergame effectiveness: what the numbers can tell us. *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games* ACM. 55–62.
34. Justin Wylie. 2010. Fitness gamification: concepts, characteristics, and applications. Retrieved from: http://www.justintwylie.com/wp-content/uploads/2014/12/JUSTIN-WYLIE-Fitness-Gamification.pdf
35. Jeffrey Yim and T. C. Nicholas Graham. 2007. Using games to increase exercise motivation. *Proceedings of the 2007 conference on Future Play*. ACM2007.
36. Zhao Zhao, S. Ali Etemad, Ali Arya, and Anthony Whitehead. 2016. Usability and motivational effects of a gamified exercise and fitness system based on wearable devices. In *International Conference of Design, User Experience, and Usability*, Springer International Publishing, 333–344.