MethodsX 12 (2024) 102551

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

Protocol for a randomised controlled field experiment on the effect of different gamification designs of physical activity ☆

Elaine Marie Grech*, Marie Briguglio, Emanuel Said

University of Malta, Msida MSD2080, Malta

ARTICLE INFO

Method name:
Protocol for a randomised controlled field experiment on the effect of different gamification designs of physical activity

Keywords:
Gamification
Randomised controlled field experiment
Physical activity
Behaviour change
Psychological outcomes
Motivation
Steps
Competition
Cooperation
Coopetition

ABSTRACT

Gamification is finding growing application in the field of physical activity, promising engaging and motivating experiences that foster behavioural change. However, existing empirical work has insufficiently scrutinised whether the reported positive outcomes emerge because of gamification and what type of gamification design leads to optimal results. This protocol for a parallel four-arm randomised controlled field experiment was purposely designed to investigate the effect of different gamification designs on motivation, perceived usefulness, and the intended behavioural change in physical activity (increase in step counts). Participants were randomly assigned to either: 1) a competitive gamified group; 2) a cooperative gamified group; 3) a hybrid (competitive-cooperative) gamified group; or 4) a control group. The design of the gamified interventions was guided by gamification design frameworks identified in literature. The data gathered includes: 1) a longitudinal panel dataset of step counts to investigate the causal effect of gamification on physical activity behaviour; and 2) self-reported data to examine the effect of gamification on the users' intrinsic motivation and perceived usefulness of the experience. This protocol outlines the procedure and processes followed during this experiment to facilitate replicability for future studies.

Specifications table

Subject area:	Psychology
More specific subject area:	Motivation and Behaviour Change
Name of your protocol:	Protocol for a randomised controlled field experiment on the effect of different gamification designs of physical activity
Reagents/tools:	Wearable physical activity trackers Gamification platform - Pointagram.com STATA™ (version 16.1, StataCorp)
Experimental design:	A parallel four-arm randomised controlled field experiment, examining the effect of three gamified interventions (1 - competition; 2 - cooperation; 3 - hybrid involving an inter-team competition) versus an active control group.
Trial registration:	n/a
Ethics:	Research ethics application submitted to the Faculty Research Ethics Committee at the University of Malta (Reference number: 3829_10122019). Informed consent was obtained from participants.
Value of the Protocol:	• Addresses pertinent questions in the field of gamification research to determine whether gamification works in the context of physical activity and the optimal choice of gamification design. • Explains the design process of three different gamified interventions of physical activity. • Outlines the guidelines, procedure and processes followed to facilitate replicability for future studies.

☆ **Related research article:** Grech, E. M., Briguglio, M., & Said, E. (2024). A field experiment on gamification of physical activity – Effects on motivation and steps. International Journal of Human-Computer Studies, 184, 103205.

* Corresponding author.
E-mail address: elaine.m.grech@um.edu.mt (E.M. Grech).

https://doi.org/10.1016/j.mex.2024.102551
Received 5 June 2023; Accepted 4 January 2024
Available online 11 January 2024

Description of protocol

Background

Gamification is finding growing application in the field of physical activity, promising engaging and motivating experiences that foster behavioural change [10,11]. Extant empirical evidence suggests that the effect of gamification of physical activity is predominantly positively oriented [10,15]. However, literature has insufficiently scrutinised whether the reported positive outcomes emerge because of gamification and what type of gamification design leads to optimal results [15].

Objectives

This protocol for a randomised controlled field experiment was purposely designed to investigate the effect of different gamification designs on the intended behavioural change (physical activity) and the users' psychological (emotional and cognitive) responses. Specifically, this study investigates:

i. whether the use of gamification stimulates the desired behavioural change in physical activity, in terms of an increase in step counts;
ii. how the choice of gamification design affects the behavioural change in physical activity; and
iii. how the use of gamification in the context of physical activity influences the users' intrinsic motivation and perceived usefulness.

Physical activity is tracked and measured objectively through the use of wearable activity trackers that incorporate sensor-based technologies. The data gathered includes a longitudinal panel dataset of step counts to investigate the causal effect of gamification on physical activity, as well as self-reported data to examine the effect of gamification on psychological outcomes.

Hypotheses

The data gathered through this field experiment tests the following hypotheses:

Hypothesis 1. Gamification improves physical activity - Gamified groups will report higher step counts than the control group during the intervention period.

Hypothesis 2. Hybrid (competitive-cooperative) design will facilitate the strongest effect on step counts.

Hypothesis 3a. Gamified groups will report higher intrinsic motivation than the control group.

Hypothesis 3b. Gamified groups will report higher perceived usefulness than the control group.

Material and Methods

Study design

A four-arm randomised controlled field experiment examining the effect of three gamified interventions (1 - competition; 2 - cooperation; 3 - hybrid design involving an-inter team competition) versus an active control group. This study involves a parallel group design, where each participant was allocated to one group throughout the experimental period. The four-week experimental period consists of a one-week baseline period, followed by a three-week intervention period. The overall timeline of this study is illustrated in Fig. 1.

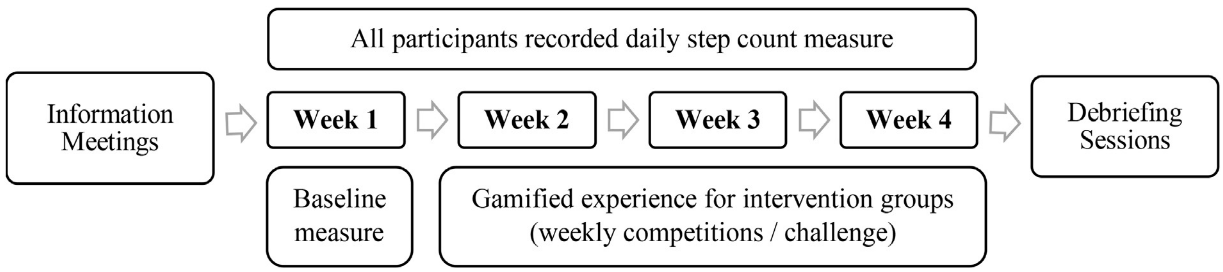

Fig. 1. Study timeline.

Fig. 2. Wearable activity trackers prepared for participants.

Eligibility criteria

Key inclusion criteria for participation in this study:

- over 18 years of age
- did not use a smartwatch or a wearable device to monitor their physical activity during the previous 12-month period
- had no health issues (such as heart condition, chest pain, bone or joint pain, or dizziness) that they are aware of, which could prevent them from engaging in physical activity

Key exclusion criteria:

- ineligible to participate if the participant is currently pregnant or have been told by their doctor not to engage in physical exercise

Procedures and interventions

Eligible participants who provided informed consent attended a group information meeting (see Fig. 1). Separate information meetings were held for each group of participants to avoid cross-contamination between groups. During the information meetings, all participants were given a smartwatch (Xiaomi Mi Band) to monitor their physical activity (see Fig. 2). Earlier studies [26,30] show that these wearable devices are adequately reliable in measuring step counts, and hence these were preferred against other brands of pedometers due to their cost and battery lifespan (lasting approximately two weeks). The use of wearable devices permits the collection of objective data. Literature indicates that objective data is a more reliable measure of physical activity than self-reported data based on recall [5,21].

During the briefing session, the researcher assisted all participants to set up the wearable physical activity tracker and install a corresponding mobile application (MiFit app[1]) that synchronises with the device. The installation of the MiFit application on all the participants' smartphones was done as a validation check to retrieve and confirm the panel dataset of step counts utilised for this study. A pilot study highlighted the importance of synchronising the wearables with the smartphone application. As a result, participants were asked to wear the device at all times, keep the Bluetooth feature on their smartphones on, and synchronise the wearables daily every evening to ensure the correct estimate of step counts was recorded. Participants had to charge their wearable device only once during the experiment and were advised to do so during the night to ensure that the device recorded all the step counts undertaken by participants during the day.

During the set-up of the wearables and the corresponding application installed on their smartphones, all participants were allowed to choose a personalised daily step target. Goal-setting is a commonly used feature in self-tracking motivational technologies [1] that supports users' intrinsic motivation and self-regulation [13].

During the information meetings, participants were asked to complete a pen-and-paper questionnaire to gather the participants' demographic information and lifestyle characteristics. Participants' weight and height evaluations were done using a weighing scale and a tape measure. Based on this information, the body mass index was calculated as weight divided by height in meters squared.

At the end of the study, all the participants were asked to return their wearable devices. As a validation check, two researchers retrieved and independently validated the daily step count data from the MiFit app that was installed on the smartphone of each participant. This process was done to confirm the panel dataset of step counts that was analysed for this study, ensure that no data was lost and minimise human error. During this meeting, all participants were also asked to complete a pen-and-paper questionnaire to measure the participants' interest and enjoyment during this experience and their perceived usefulness of the experience.

Control group

Participants in the control group could only monitor whether they achieved their personal daily step target set on their smartwatch. Empirical evidence [15] concluded that the positive effects reported for gamified interventions are considerably higher when gamification was compared to inactive control groups such as individuals on waiting lists, rather than active control groups where

[1] After the experiment period, Xiaomi's MiFit app was renamed to Zepp life app.

Table 1
Gamification design of the interventions.

Gamification Design	
Gamification design principles, elements and mechanics	*Applicable to ALL gamified interventions:* ▪ *Points:* users earn individual points for step counts recorded (one step = one point). ▪ *Badges:* users earn individual badges for achieving higher daily step counts. ▪ *Progression status*: progression bar indicating the progress and remaining effort required to achieve the next badge. ▪ *Opportunities for social interaction and support:* users can post comments, send likes to each other comments and notifications on the newsfeed section of the gamified application. ▪ *Opportunities for social comparison:* users can see others' performance, progress and achievements. ▪ *User identity*: users are anonymised and represented by a URN code. ▪ *Feedback:* users are notified when they earn points and badges through a notification on the gamified application. ▪ *Episodical:* competitions and challenges/quests run from Monday to Sunday, users' progress in the competition / challenge resets every week. *Applicable to the Competitive Gamified Design (Player vs. Player competition):* ▪ *Leaderboard:* showing the ranking of all the players ▪ *Virtual trophies:* awarded to the top three players with the highest step counts *Applicable to the Hybrid Gamified Design (Team vs. Team competition):* ▪ *Teams:* players were randomised to teams of 4 players each ▪ *Leaderboard:* showing the ranking of all the players ▪ *Virtual trophies:* awarded to the top three teams with the highest step counts *Applicable to the Cooperative Gamified Design (Shared group challenge/quest):* ▪ *Visualisation /Plot:* a map showing a pirate making his way to reach the treasure chest with a countdown timer indicating the time left for the participants to complete the challenge. Users' step counts are reflected in the progress that the pirate towards the treasure chest. ▪ *Challenge:* Quest to reach a target step count (shared goal) which automatically opens the pirate treasure chest by the end of the week.
Desirable outcomes	Behaviour change - Increase in step counts. Positive psychological responses in terms of perceived enjoyment and interest (intrinsic motivation) and perceived usefulness of the experience.

participants utilised a wearable device to track their physical activity.[2] Thus, an active control group was adopted in this study to ensure that the observed effect on physical activity is not the result of having a wearable device to monitor physical activity.

Interventions

A gamified platform (pointagram.com) was used to design a gamified experience for each treatment group. Each treatment group had a separate gamified interface on the platform, so the participants in the gamified groups would not become aware of the other groups. All participants could access the gamified platform through an application that was installed on their smartphone or through a web browser.

The game elements and principles implemented in the interventions, and the desirable outcomes are detailed in Table 1. The design of the gamified interventions was guided by the taxonomy of gamification concepts utilised in health applications [24] and gamification design frameworks [2,14,17] identified in literature. Following consideration of the psychological model of the self-determination theory [23], all the gamified interventions incorporated a common set of game elements related to the motivational constructs of the self-determination theory to afford an appealing and motivating experience that supports the users' intrinsic motivation [2]. These included points, badges, progress feedback and opportunities for social support, social comparison and interaction on the newsfeed section of the gamified application. Participants earned points for their daily step counts (one point for each step count recorded). Based on their daily step count, individual badges were awarded at increments of 2K step counts, starting from a 2K badge going up to a 20K badge. Participants could see others' performance (social comparison) and interact with other participants in their respective group through posts, comments, and likes (social interaction and support). Screenshots from the gamified application are presented in Figs. 3 to 5.

Furthermore, based on the classification of gamification features proposed by Morschheuser et al. [17], each gamified intervention incorporated specific game elements to create 1) a competitive gamified design; 2) a cooperative gamified design; and 3) a hybrid (competitive-cooperative) gamified design.[3] Participants in the **competitive treatment group** had a weekly individual competition,

[2] Effect size of 0.58 in the case of gamified intervention versus inactive control group; and an effect size of 0.23 in the case of gamified intervention versus active control group.

[3] Drawing on the same classification of gamification features utilised for this study, Morschheuser et al. (2019) investigated the effect of different gamified designs in the crowdsourcing domain.

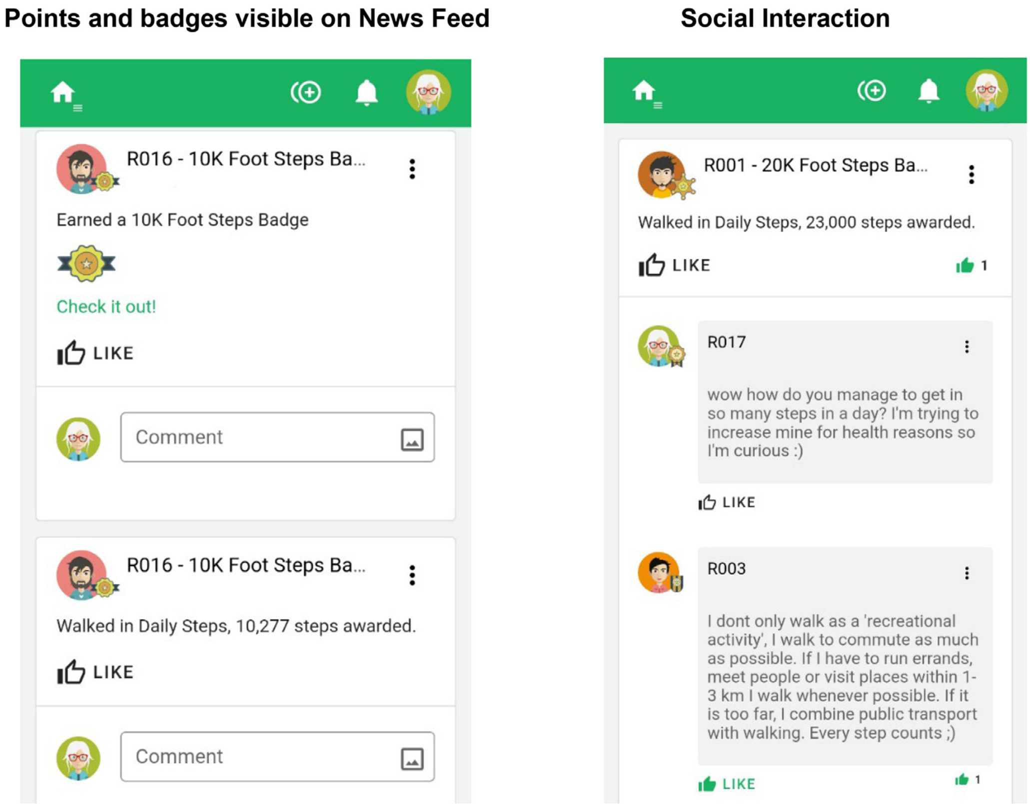

Fig. 3. Screenshots from the gamified application.

where the accumulated points were visible on a leaderboard (Fig. 4) and the top three players were awarded a virtual trophy. By contrast, participants in the **hybrid (competitive-cooperative) treatment group** had a weekly inter-team competition (participants were randomly assigned in teams of four participants each) where the accumulated points of each team were visible on a leaderboard (Fig. 4) and the top three teams were also awarded virtual trophies. Finally, the **cooperation treatment group** had a weekly group challenge (quest) to reach a target step count (shared goal) by the end of the week. Their steps were accumulated and depicted on a visualisation of a pirate making his way to reach the treasure chest on an island, with a countdown timer indicating the time left for the participants to complete the challenge (Fig. 5). The group target step count was 700K steps for the first week (based on approximately 5K daily step count per participant) and then increased every week based on the equivalent of 7.5K and 10K daily step counts per participant as a group target. All the challenges and competitions were scheduled to run on a weekly basis from Monday to Sunday.

The game elements and design principles implemented for this study are amongst the most commonly adopted gamification design strategies in gamified fitness applications in industry practice [18].

Outcome measures

The behaviour change in physical activity was measured in terms of the change in step counts. To ensure accuracy and avoid human error, this study relies solely on the step count data gathered from the wearable devices. Step counts were recorded daily as a continuous variable. The panel dataset considered for this study included four weeks of step count data, each week starting on Monday. Earlier studies suggested that the most reliable measures are achieved when monitoring of step count data starts on Monday [25]. The first week of step count data was considered as the baseline measure, during which no treatment was administered, whilst the following three weeks of step count data were during the intervention period.

The psychological outcomes were measured in terms of the participants' intrinsic motivation (based on the users' interest and enjoyment) and the users' perceived value of the experience. Self-reported data on these experiential outcomes was gathered using adaptations of the Intrinsic Motivation Inventory (IMI), which is a validated instrument based on the self-determination theory [16,22]. The Interest/Enjoyment sub-scale of the IMI is considered as a validated self-reported measure of intrinsic motivation [16,22]. The Perceived Usefulness sub-scale of the IMI refers to the perceived value of an activity which facilitates internalisation and self-regulation of activities that are found to be useful [4]. Table 2 presents the details on the scale items for each construct.

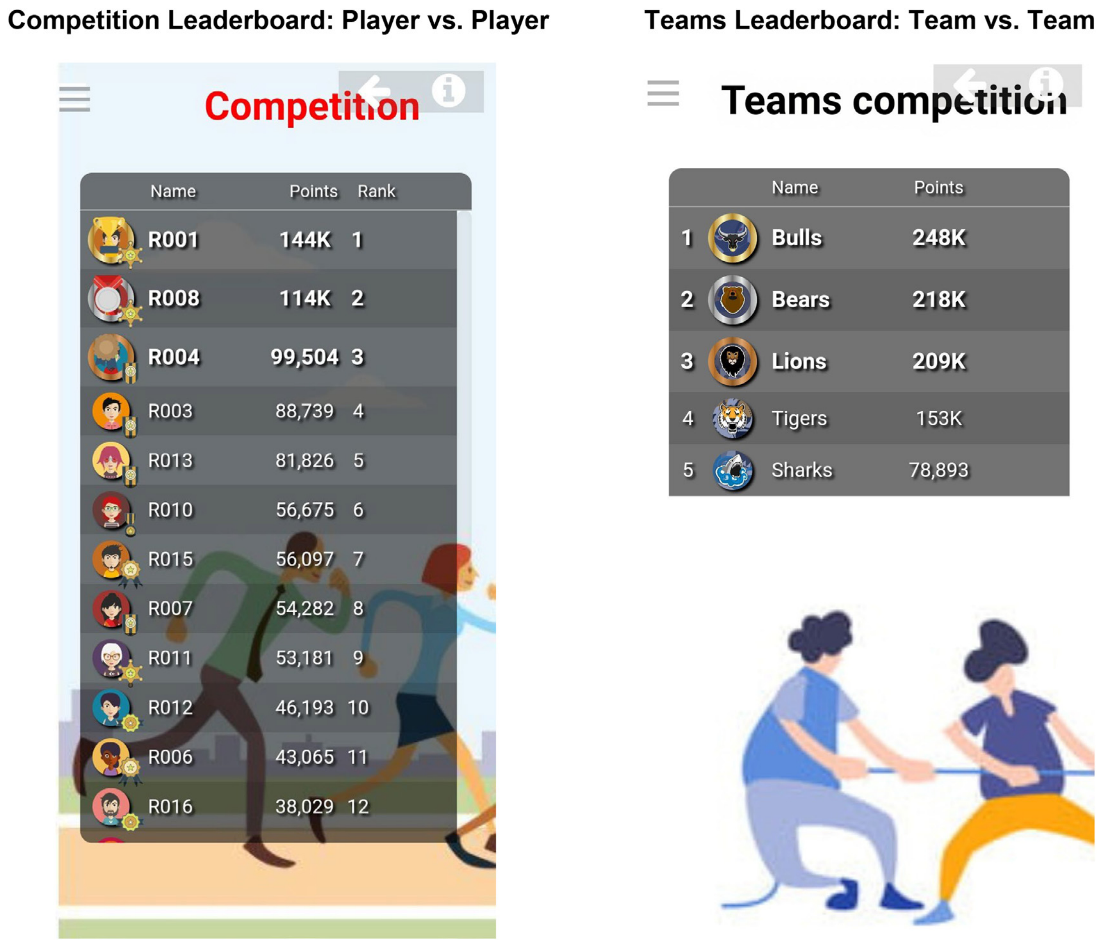

Fig. 4. Screenshots showing the leaderboards used in the competition and hybrid gamified groups.

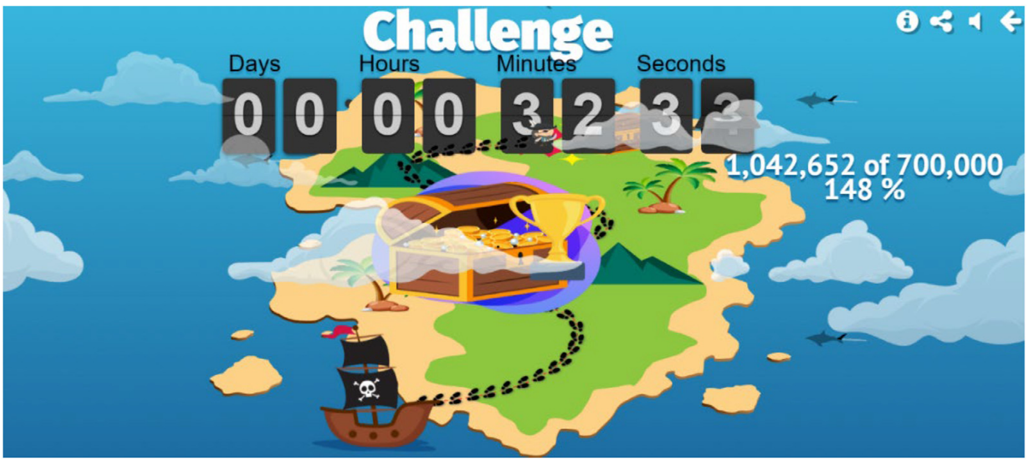

Fig. 5. Screenshot of the cooperative group challenge.

Table 2
Measure, items and scales.

Measure	Source	Item wording	Notes
Intrinsic Motivation *(Emotional response)*	Adapted from the Interest / Enjoyment sub-scale of the IMI	I enjoyed doing this experience very much This experience was fun to do I thought this was a boring experience (R)	7-point Likert scale anchored 'Not at all true' to Very true'
Perceived Usefulness *(Cognitive response)*	Adapted from the Value / Usefulness sub-scale of the IMI	I believe this experience was of some value to me I think that doing this experience was useful to increase my physical activity I think doing this experience helped me to increase my physical activity	7-point Likert scale anchored 'Not at all true' to Very true'

Sample size calculation

The sample size required to establish superiority of the gamified interventions compared to the control group was based on a targeted power of 80 % (1 - β = 0.8) at 5 % significance level (α = 0.05) with equal allocation between the groups (k = 1). Calculation of the sample size was based on the recommended guidelines on sample size estimation for randomised controlled trials suggested by Chow et al. [3]. On the basis of previous literature [6], the expected difference in daily steps between the gamified intervention groups and those in an active control group using a wearable device was around 2000 steps per day. The standard deviation was assumed to be about 2500 steps per day. The sample size calculation indicated that 20 participants were required for each treatment group and control group respectively. Thus, for a four-arm randomised controlled experiment a total sample size of 80 participants was required to detect between-group differences on the daily step count.

Recruitment

The invitation to participate in this study was sent as an email invitation through the University of Malta and a post on social media. Recruitment was carried out over a two-month period. Interested participants were invited to review the information about the study (including its objectives, duration, and eligibility criteria) and provide informed consent through the link provided. Participants were informed that of all the data generated through the physical activity tracker, only step count data would be collected for this study.

Sampling method, randomisation and blinding

A non-probabilistic convenience sampling method was utilised. Following the eligibility screening criteria, a Unique Reference Number (URN) was assigned to all participants to ensure anonymity all throughout the study. Using an online random sequence generator (random.org), eligible participants (n = 80) who provided informed consent were randomly allocated to the control or one of the treatment groups using a 1:1:1:1 ratio. Participants were blinded to group allocation and groups were colour-coded to hide the identity of each group from participants. The participant flow diagram based on CONSORT guidelines for transparent reporting of randomised trials is presented in Fig. 6.

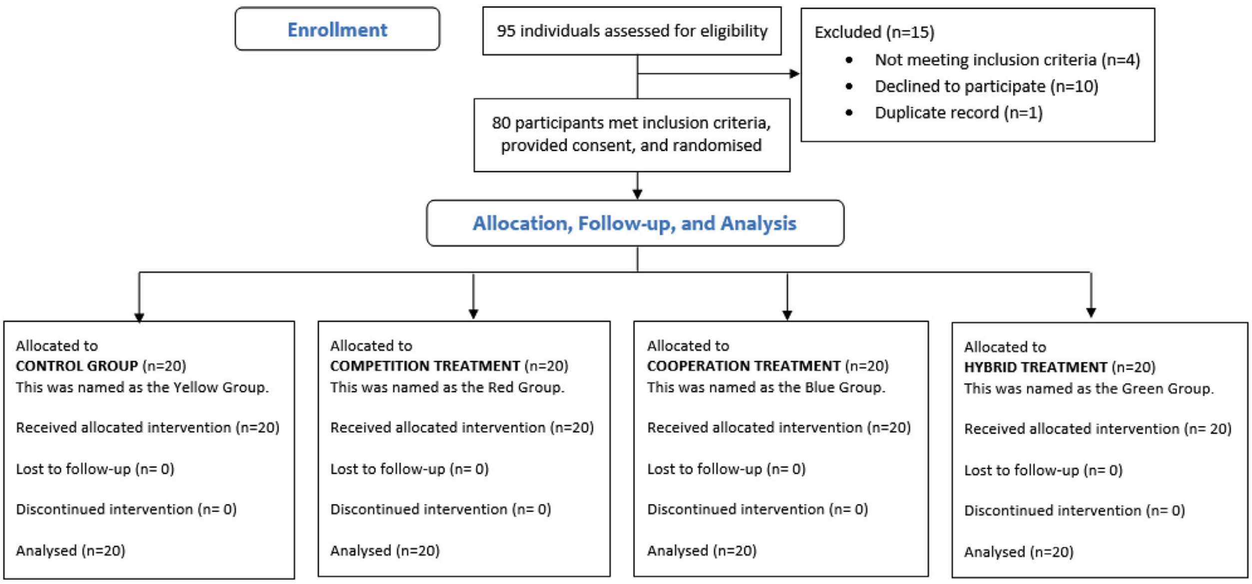

Fig. 6. Participant flow diagram.

Table 3
Data structure for the longitudinal analysis of covariance.

Respondent	Treatment	Group	Baseline Steps	Intervention Steps	Time
1	1	2	Yt_0	Yt_1	1
1	1	2	Yt_0	Yt_2	2
1	1	2	Yt_0	Yt_3	3

Statistical data analysis

Data cleaning

All randomly assigned participants were included with the intention-to-treat principle, and thus all randomised participants were included in the analysis. Step count data on the days when smartwatches were given to participants during the initial information meetings was discarded since this did not capture full-day data of the physical activity of participants. The following two days of initial wearable use which were on weekend days were not included in the baseline estimate due to potential higher activity during initial wearable use. This approach is similar to that adopted in previous studies [19,20].

Previous literature suggests that daily step count values that are less than 1000 steps do not reflect full day data activity and should thus be excluded and imputed [12,19]. In this study, step counts values less than 1000 were imputed at the weekly mean step count. The mean daily step count for each week was derived by summing up the daily step count for each respondent and dividing it by the number of days on which step counts were recorded. Days with unrecorded steps could result if a participant did not wear the device, or the device did not synchronise with the smartphone application. Missing step counts were imputed with the weekly mean step count. Research on pedometer monitoring indicates that three days of step count data within a week can provide a sufficient reliable estimate of physical activity [28]. If less than three days of step count data within a week were recorded, these remained as missing observations in the dataset.

Data analysis of behavioural outcome

Data was restructured into the long data format (structure of the data is presented in Table 3) and analysed using Generalized Linear Mixed-Effect Models (GLMM) in STATA™ (version 16.1, StataCorp). Statistical methods of analysis which assume independence of observations and ignore correlations between repeated measures are not appropriate. By contrast, GLMM fit statistical models to data with multi-level or repeated data, where the dependent variable is not necessarily normally distributed and are able to handle missing observations in the dataset.

The effect of gamification (*Treatment*) was estimated using the longitudinal analysis of covariance adjusting for the baseline values of the outcome variable, even though the differences at baseline are attributed to chance and random fluctuations [29]. When the data is analysed using longitudinal data analysis techniques, an adjustment for the baseline differences of the outcome variable is recommended (even though the differences at baseline are not significant), to provide a precise estimate of the treatment effect [29]. If no adjustment is made for the baseline differences in the outcome variable, an artificial intervention effect may be estimated due to the regression to the mean. By including the baseline daily step count as a covariate in the analysis and taking repeated measures of the outcome measure, error variability is reduced, thus leading to a possible increase in the attained power. The generalized mixed-effects model analysis also included a random intercept to adjust for the repeated observations over time at individual level and was estimated using a robust estimator of variance.

For the first hypothesis *(H1: Gamification improves physical activity),* gamified groups were expected to report higher step counts than control group during the intervention period. The overall effect of gamification *(Treatment)* on the mean daily step count was estimated by analysing the change in the mean daily step count for the gamified group in comparison to the control group from the baseline period to the intervention period. The mean daily step count before the intervention started was taken as the baseline measure (*Baseline Steps*). The dependent variable (*Intervention Steps*) was the mean daily step count during the intervention period. The effect of gamification was estimated using the longitudinal analysis of covariance model as recommended by Twisk et al. [29] (see Eq. (1)):

Y_t➔Intervention Steps: the mean daily step count during the intervention

X➔Treatment (0 for control and 1 for gamified treatment)

Yt_0➔Baseline Steps: the mean daily step count at baseline

$$Y_t = \beta_0 + \beta_1 X + \beta_2 Y t_0 \tag{1}$$

To estimate the effect size, Hedge's *g* (also known as the corrected effect size) was computed. Hedge's *g* is preferable to Cohen's *d* in the case of small sample sizes [8,9].

Earlier studies suggest that physical activity levels of participants decline over time during the intervention period [6,20,27]. To examine whether a declining trend was present in this study, the effect of gamification at different timepoints during the intervention period was examined by extending the longitudinal analysis of covariance model to include Time as a main effect (fixed variable) and the interaction between Time and Treatment Group as shown in Eq. (2):

Y_t➔Intervention Steps: the mean daily step count during the intervention

X➔Treatment (0 for control and 1 for gamified treatment)

Yt_0➔Baseline Steps: the mean daily step count at baseline

Time➔Timepoints during the intervention period

$$Y_t = \beta_0 + \beta_1 X + \beta_2 Y t_0 + \beta_3 Time + \beta_4 X x Time \quad (2)$$

For the second hypothesis *(H2: Hybrid design expected to facilitate the strongest effect on step counts*), the effect of each respective treatment group (competition, cooperation, and hybrid) during the intervention was estimated by including *Group* as the treatment variable as shown in Eq. (3):

Y_t➔Intervention Steps: the mean daily step count during the intervention

X➔Treatment Group (0 for control,1 for cooperation, 2 for hybrid, and 3 for competition)

Yt_0➔Baseline Steps: the mean daily step count at baseline

$$Y_t = \beta_0 + \beta_1 X + \beta_2 Y t_0 \quad (3)$$

Data analysis of psychological outcomes

The constructs' reliability for *Intrinsic Motivation* and *Perceived Value / Usefulness* were measured using Cronbach's alpha (α), composite reliability (CR), and average variance extracted (AVE). All the convergent validity metrics obtained were checked against the thresholds (Cronbach's $\alpha > 0.7$, CR > 0.7, and AVE > 0.5) suggested in literature [7].

Descriptive statistics including means and standard deviation were computed for each experiential outcome for the control and gamified groups. Following that, a Shapiro-Wilk test was conducted to determine whether the score distribution of each construct follows a normal distribution. In the case of a normal distribution, parametric statistical tests were utilised, otherwise when the normality assumption was not satisfied, the non-parametric equivalent tests were performed.

To test the third hypothesis set out for the study (*H3: Gamified groups expected to report higher scores for intrinsic motivation and perceived usefulness than the control group*), an analysis of the differences in the means between the groups was carried out for each construct. A Mann Whitney test (the non-parametric equivalent of the independent t-test) was carried out for each construct to test whether there were significant differences in the means reported between the control and the gamified groups. Furthermore, a Kruskal-Wallis test (the non-parametric equivalent of the one-way ANOVA test) was conducted for each construct to check for any significant differences between the means of each treatment group.

Strengths and limitations

This randomised controlled field experiment was purposely designed to address pertinent questions in the field of gamification research, specifically to determine whether gamification works in the context of physical activity and the optimal choice of gamification design. To our knowledge this is the first randomised controlled field experiment investigating competitive, cooperative, and hybrid gamified designs in the context of physical activity. The behavioural effect of gamification was examined on step count data gathered through physical activity trackers, rather than relying on self-reported measures. During the pilot study we considered the possibility of including the short version of the International Physical Activity Questionnaire to provide a self-reported measure of physical activity. However, estimating the amount of physical activity based on recall proved to be challenging for participants. Nonetheless, physical activity is more accurately measured having a wearable tracker [5]. Introducing an active control group ensured an unbiased estimate of the treatment causal effect. Randomisation was implemented to prevent selection bias and control for extraneous variables. Furthermore, while the majority of existing studies focus on behavioural outcomes, this study investigated both psychological and behavioural outcome measures resulting from gamification of physical activity.

The challenges encountered during the study and the limitations are discussed hereunder as these provide avenues that could be explored in future research. First, this study was conducted amongst academic members and post-graduate students. Future studies should widen the applicability and generalisability of the results by examining the effects of gamification amongst other segments of the population and in different settings. If the use gamification stimulates the desired favourable outcomes, it could have positive effects for society in general when adopted on a wide scale. Furthermore, our study was conducted amongst participants who do not have pre-existing social connections and the identity of participants was anonymised. Future studies could investigate the effect of the three gamification designs outlined in this protocol in a setting where the participants' identities are disclosed (provided that ethical issues are complied with) and/or amongst participants who are related through existing social bonds. Having identities disclosed helps users to establish social connections. Moreover, providing opportunities for social interaction may facilitate group cohesion and the development of a social network which provides social support. Consistent with the self-determination theory, motivation is more likely to thrive in contexts characterised by a sense of relatedness and social support [23].

Second, the sample size limited the potential of further sub-groups to test specific game elements and further mechanisms. The difference in step counts between the gamified conditions was expected to be much lower than the difference between the gamified and the control group. The sample size utilised for this study does not yield adequate power to detect differences between the gamified conditions. Having a larger sample size would increase the statistical power, support more covariates in the analysis and enhance the generalisability of the results. Understandably, there are challenges to conduct randomised controlled trials using wearable devices (to achieve objective data) with large sample sizes and longer timeframes. However, the accumulation of knowledge from rigorous empirical studies on the effect of gamified interventions on health-related behaviours would have practical relevance. Some ideas that could be tested in future studies include the possibility of introducing new game elements during the intervention, offering tangible rewards rather than virtual rewards when targets are achieved, and including more opportunities for social interaction amongst participants. Moreover, although this study is conducted over a period of four weeks, it is still considered as a relatively short timeframe and longer interventions are encouraged in future studies.

Third, at the time of planning this experimental study, an off-the-shelf fitness application that catered for the scope of the study with three distinct gamified experiences (competition, cooperation, hybrid) could not be identified. It was important to ensure that the gamified interventions remain distinct from each other to ensure that participants were only exposed to one gamified condition. In view of this, a gamified application (Pointagram) was utilised to design three distinct gamified experiences (with a separate interface) for the purpose of this study. Advances in technological developments including data integration, data analytics, artificial intelligence, chatbots, and internet of things could facilitate the design of gamified systems and the data insights that could be generated. For instance, a fitness application could integrate various gamification design features that may be customised to the individuals' preferences and needs. Data generated from wearable devices could be integrated automatically with the gamified application and real-time customer insights on the users' experiences could be generated through in-built features of the application.

Fourth, having a daily step goal (also in the control group) could arguably be considered as a gamification feature. However, goal setting is a standard feature even in the case of basic smartwatches. Thus, participants in the active control group would inevitably be exposed to setting daily step targets as a goal. Having an active control group (using a wearable device) was an important consideration in this study to ensure that the observed effect on step counts is not the result of having a wearable device to monitor physical activity. This was especially important since recent literature [15] suggests that positive effects reported for gamified interventions are considerably higher when compared to inactive control groups (such as individuals on waiting lists), rather than active control groups (where participants utilised a wearable device to track their physical activity or a non-gamified version of the app). In the domain of physical activity, goals are commonly integrated along with several other game elements that leverage social influences, such as competition and challenges [31,32]. Indeed, the three gamified conditions included in this protocol are social-oriented, include several game elements that are commonly adopted in industry practice [18], and are designed on gamification design frameworks established in literature [2,17].

Ethics statements

The work described has been carried out in accordance with the guidelines of the Research Ethics Committee of the University of Malta. During the design of this study, care was taken to ensure that participants would not be exposed to any physical harm or discomfort. To prevent harm to participants, all participants were required to complete pre-screening related to any prevailing health conditions where doctors would have recommended avoidance of physical activity. In such cases, participants would have automatically become ineligible to participate in this study.

Participation was voluntary and involved no compensation or incentives to participants. Participants were informed about the right to decline or withdraw from the study anytime. The contact information of the researcher was made available to all participants in case they encounter any issue with their smartwatch device, or if they had any query or difficulty that required clarification. Written informed consent was obtained from all participants. To protect participants' privacy and guarantee confidentiality at all stages during this study, a unique reference number was assigned to each participant. This unique reference number was linked to all the data collected from the participants' wearable devices and surveys, rather than using the participants' name. Of all the data generated through the physical activity tracker, only step count data was collected. All the primary data collected was stored in an encrypted folder and data back-ups were stored on a separate hard drive.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Elaine Marie Grech: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Resources, Data curation, Writing – original draft, Visualization, Project administration, Writing – review & editing. **Marie Briguglio:** Supervision, Resources, Writing – review & editing. **Emanuel Said:** Supervision, Writing – review & editing.

Data availability

The authors do not have permission to share data.

References

[1] N. Aldenaini, O. Oyebode, R. Orji, S. Sampalli, Mobile phone-based persuasive technology for physical activity and sedentary behavior: a systematic review, Front. Comput. Sci. 2 (July) (2020) 1–17, doi:10.3389/fcomp.2020.00019.
[2] J. Buckley, T. DeWille, C. Exton, G. Exton, L. Murray, A Gamification–motivation design framework for educational software developers, J. Educ. Technol. Syst. 47 (1) (2018) 101–127, doi:10.1177/0047239518783153.
[3] S.C. Chow, J. Shao, H. Wang, Y. Lokhnygina, S.C. Chow, J. Shao, H. Wang, Y. Lokhnygina, Sample Size Calculations in Clinical Research, Third Ed., Chapman and Hall/CRC, 2017, doi:10.1201/9781315183084.
[4] E.L. Deci, H. Eghrari, B.C. Patrick, D.R. Leone, Facilitating internalization: the self-determination theory perspective, J. Pers 62 (1) (1994) 119–142, doi:10.1111/j.1467-6494.1994.tb00797.x.
[5] J. Fiedler, T. Eckert, A. Burchartz, A. Woll, K. Wunsch, Comparison of Self-Reported and Device-Based Measured Physical Activity Using Measures of Stability, Reliability, and Validity in Adults and Children, Sensors 21 (8) (2021) 2672, doi:10.3390/s21082672.
[6] A.L. Gremaud, L.J. Carr, J.E. Simmering, N.J. Evans, J.F. Cremer, A.M. Segre, L.A. Polgreen, P.M. Polgreen, Gamifying accelerometer use increases physical activity levels of sedentary office workers, J. Am. Heart Assoc. 7 (13) (2018) 1–12, doi:10.1161/JAHA.117.007735.
[7] J. Hair, W. Black, B. Babin, R. Anderson, Multivariate Data Analysis, Seventh Ed, Pearson Education, 2010.
[8] L.V. Hedges, Distribution theory for glass's estimator of effect size and related estimators, J. Educ. Stat. 6 (2) (1981) 107–128, doi:10.3102/10769986006002107.
[9] L.V. Hedges, I. Olkin, Statistical Methods for Meta-Analysis, Elsevier, 1985, doi:10.1016/C2009-0-03396-0.
[10] J. Koivisto, J. Hamari, Gamification of physical activity: A systematic literature review of comparison studies, in: Proceedings of the 3rd International GamiFIN Conference, GamiFIN 2019. CEUR-WS, 2019, pp. 106–117.
[11] J. Koivisto, J. Hamari, The rise of motivational information systems: A review of gamification research, Int. J. Inf. Manag. 45 (December 2018) (2019) 191–210, doi:10.1016/j.ijinfomgt.2018.10.013.
[12] G.W. Kurtzman, S.C. Day, D.S. Small, M. Lynch, J. Zhu, W. Wang, C.A.L. Rareshide, M.S. Patel, Social incentives and gamification to promote weight loss: the LOSE IT randomized, controlled trial, J. Gen. Intern. Med. 33 (10) (2018) 1669–1675, doi:10.1007/s11606-018-4552-1.
[13] G.P. Latham, E.A. Locke, Self-regulation through goal setting, Organ. Behav. Hum. Decis. Process. 50 (2) (1991) 212–247, doi:10.1016/0749-5978(91)90021-K.
[14] D. Liu, R. Santhanam, J. Webster, Toward meaningful engagement: a framework for design & research of gamified information systems, MIS Q. 41 (4) (2017) 1011–1034.
[15] A. Mazéas, M. Duclos, B. Pereira, A. Chalabaev, A. Mazeas, M. Duclos, B. Pereira, A. Chalabaev, Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials, J. Med. Internet Res. 24 (1) (2022) e26779, doi:10.2196/26779.
[16] E. McAuley, T. Duncan, V.V. Tammen, Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: a confirmatory factor analysis, Res. Q. Exerc. Sport 60 (1) (1989) 48–58, doi:10.1080/02701367.1989.10607413.
[17] B. Morschheuser, A. Maedche, D. Walter, Designing cooperative gamification: conceptualization and prototypical implementation, in: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 2017, pp. 2410–2421, doi:10.1145/2998181.2998272.
[18] A. Neupane, D. Hansen, A. Sharma, J.A. Fails, B. Neupane, J. Beutler, A review of gamified fitness tracker apps and future directions, in: Proceedings of the Annual Symposium on Computer-Human Interaction in Play, 2020, pp. 522–533, doi:10.1145/3410404.3414258.
[19] M.S. Patel, E.J. Benjamin, K.G. Volpp, C.S. Fox, D.S. Small, J.M. Massaro, J.J. Lee, V. Hilbert, M. Valentino, D.H. Taylor, E.S. Manders, K. Mutalik, J. Zhu, W. Wang, J.M Murabito, Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial, JAMA Intern. Med. 177 (11) (2017) 1586–1593, doi:10.1001/jamainternmed.2017.3458.
[20] M.S. Patel, D.S. Small, J.D. Harrison, M.P. Fortunato, A.L. Oon, C.A.L. Rareshide, G. Reh, G. Szwartz, J. Guszcza, D. Steier, P. Kalra, V. Hilbert, Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States: the STEP UP randomized clinical trial, JAMA Intern Med 179 (12) (2019) 1624–1632, doi:10.1001/jamainternmed.2019.3505.
[21] S.A. Prince, K.B. Adamo, M. Hamel, J. Hardt, S. Connor Gorber, M. Tremblay, A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review, Int. J. Behav. Nutr. Phys. Act. 5 (1) (2008) 56, doi:10.1186/1479-5868-5-56.
[22] R.M. Ryan, Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory, J. Pers. Soc. Psychol. 43 (3) (1982) 450–461, doi:10.1037/0022-3514.43.3.450.
[23] R.M. Ryan, E.L. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, Am. Psychol. 55 (1) (2000) 68–78, doi:10.1037/0003-066X.55.1.68.
[24] M. Schmidt-Kraepelin, S. Thiebes, M.C. Tran, A. Sunyaev, What's in the game? Developing a taxonomy of gamification concepts for health apps, in: Proceedings of the 51st Hawaii International Conference on System Sciences, October 2017, 2018, doi:10.24251/hicss.2018.150.
[25] D. Sigmundová, J. Vašícková, J. Stelzer, E. Řepka, The influence of monitoring interval on data measurement: An analysis of step counts of university students, Int. J. Environ. Res. Public Health 10 (2) (2013) 515–527, doi:10.3390/ijerph10020515.
[26] M.K. Tam, S.Y. Cheung, Validation of consumer wearable activity tracker as step measurement in free-living conditions, Finn. J. EHealth EWelfare 11 (1–2) (2019) 68–75, doi:10.23996/fjhw.76673.
[27] K. Thorsteinsen, J. Vittersø, G.B. Svendsen, Increasing physical activity efficiently: An experimental pilot study of a website and mobile phone intervention, Int. J. Telemed. Appl. (2014), doi:10.1155/2014/746232.
[28] C. Tudor-Locke, L. Burkett, J. Reis, B. Ainsworth, C. Macera, D. Wilsonilson, How many days of pedometer monitoring predict weekly physical activity in adults? Prev. Med. 40 (3) (2005) 293–298, doi:10.1016/j.ypmed.2004.06.003.
[29] J. Twisk, L. Bosman, T. Hoekstra, J. Rijnhart, M. Welten, M. Heymans, Different ways to estimate treatment effects in randomised controlled trials, Contemp. Clin. Trials Commun. 10 (November 2017) (2018) 80–85, doi:10.1016/j.conctc.2018.03.008.
[30] J. Xie, D. Wen, L. Liang, Y. Jia, L. Gao, J. Lei, Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study, JMIR Mhealth Uhealth 6 (4) (2018), doi:10.2196/mhealth.9754.
[31] V. Cotton, M.S. Patel, Gamification use and design in popular health and fitness mobile applications, Am. J. Health Promot. 33 (3) (2019) 448–451, doi:10.1177/0890117118790394.
[32] A. Neupane, D. Hansen, J.A. Fails, A. Sharma, The role of steps and game elements in gamified fitness tracker apps: a systematic review, Multimodal. Technol. Interact. 5 (2021) 5, doi:10.3390/mti5020005.