Preprint
Not Peer-Reviewed

The Impact of a Gamified Intervention on Physical Activity in Real-Life Conditions: A Retrospective Analysis of 4800 Individuals

Alexandre Mazéas [1,2,3] *, Cyril Forestier [4], Guillaume Harel [3], Martine Duclos [2,5], Aïna Chalabaev [1]

Cite this article:

Mazéas, A., Forestier, C., Harel, G., Duclos, M., Chalabaev, A. (2022). The impact of gamification on physical activity in real-life conditions: a retrospective analysis of 4800 individuals. *MedRxiv.*

Data, code & supplemental materials:

https://osf.io/scnu7/

***Author for correspondence:**

Alexandre Mazéas
E-mail: alexandre.mazeas@univ-grenoble-alpes.fr
Twitter: @AlexMazeas

Key words

behavior change; mHealth; gamification; intervention; physical activity.

[1]Univ. Grenoble Alpes, SENS, Grenoble, France
[2]National Research Institute for Agriculture, Food and Environment (INRAE), Clermont-Ferrand, France
[3]Kiplin, Nantes, France
[4]Laboratoire Motricité, Interactions, Performance, MIP - UR4334, Nantes Université, France
[5]Department of Sport Medicine and Functional Exploration, University Hospital Clermont-Ferrand, Hospital G. Montpied, Clermont-Ferrand, France

Abstract

Background. Digital interventions integrating gamification features hold promise to promote physical activity (PA). However, results regarding the effectiveness of this type of intervention are heterogeneous. This study aimed to examine potential moderators of the effectiveness of a gamified intervention in a large-scale sample and in real-life conditions. Specifically, we tested (1) whether a gamified intervention enhanced daily steps during the intervention and follow-up periods compared to baseline, (2) whether this enhancement was higher in participants to the intervention than in nonparticipants, and (3) what participants' characteristics or intervention parameters moderated the effect of the program.
Methods. Data from 4812 individuals who registered for a Kiplin program between January 1st, 2019, and January 2nd, 2022 were analyzed. PA was assessed via the daily step count of participants. Exposure to the intervention, the intervention content, and participants' characteristics were included in multilevel models to test the study objectives.
Results. Compared with nonparticipants, participants who benefited from the intervention had a significantly greater increase in mean daily steps from baseline during the same period ($b = 0.55$, $p < .0001$). Daily steps of participants with lower baseline steps significantly improved from baseline both during the intervention and during follow-up periods, whereas participants with >7500 baseline daily steps had no improvement or significant daily step decreases during the intervention. Age ($b = 0.05$, $p < .0001$) and exposure ($b = 0.37$, $p < .0001$) positively moderated the intervention effect.
Discussion. The Kiplin gamified intervention was effective to improve PA during intervention and follow-up periods in comparison to baseline daily step count and to nonparticipants. Responses to the intervention significantly differed as a function of individuals' initial PA. The engagement with the service and the age of the participants positively moderated the intervention effect. This study confirms the effectiveness of gamified interventions to promote PA in real-life.

Introduction

Physically inactive individuals are at higher risk of developing non-communicable diseases – such as cardiovascular diseases, cancers, type 2 diabetes mellitus, or obesity – and mental health issues compared to the most active ones.[1] Yet, one-third of the world's population is insufficiently active[2,3] and the trend is downward, with adults performing on

average 1000 fewer steps than 2 decades ago[4]. Additionally, it has recently been reported that the global population step count did not return to pre-pandemic levels in the 2 years since the onset of the COVID-19 outbreak.[5] In this context, there is an urgent need to increase physical activity (PA) of individuals in primary, secondary, and tertiary prevention.

Digital behavior change interventions and more particularly gamified services are promising avenues to promote PA. Gamification refers to the use of game elements in nongame contexts[6] and allows to transform a routine activity into a more engaging one. A recent meta-analysis[7] revealed that digital gamified interventions, lasting on average 12 weeks, improved PA by 1600 daily steps on average. Importantly, the results showed that a) gamified interventions appear more effective than digital non-gamified interventions, b) seem appropriate for any type of user regardless of their age or health status, and c) the PA improvement persists in the long term. As a result, gamified interventions are emerging as high-potential behavior change tools to tackle the physical inactivity pandemic.

However, the effect sizes reported in this meta-analysis were heterogeneous, ranging from 0.00 to 2.41, and the authors found high between-study heterogeneity (e.g., I^2 = 82%). If this heterogeneity may be explained by differences in study quality or diversity of designs in the included studies, the behavior change intervention ontology proposed by Michie et al.[8] argues that heterogeneity in behavioral interventions could also be explained by different variables such as intervention characteristics (e.g., content, delivery), the context (e.g., characteristics of the population targeted such as demographics, setting such as the policy environment or physical location), exposure of participants with the program (e.g., engagement and reach), and the mechanisms of action (the processes by which interventions influence the target behavior). Considering these variables within gamification contexts could provide a useful means to better understand the conditions under which interventions are successful.

The present study investigated this question based on a retrospective analysis of real-world data collected from a large sample of participants who were proposed a mHealth gamified intervention. In this one, participants could take part in one or several games where their daily step count was tracked, allowing individuals to play with their overall activity. In addition to offering the possibility of direct intervention on people's activity habits in natural context, the capacity of this app to collect, in real-time, a large amount of objective real-world data can be useful to understand the processes and outcomes of behavioral health interventions.[9] More specifically, these data can help make explicit when, where, for whom, and in what state for the participant, the intervention will produce the expected effect, notably thanks to continuous data collection over time. The within-person evolution in daily steps obtained via the app combined with between-person individual factors and intervention parameters is of great interest in this perspective. By analyzing these data, we can therefore a) better identify for which individuals the intervention is the most effective considering their age, health status, baseline PA, or the context in which they had the intervention, b) examine the relationship between exposure and intervention effectiveness, and c) better understand which features of the app were the most effective.

Thus, the objectives of this study were to analyze the data collected in order (1) to examine within-individual evolutions of PA before, during, and after the intervention, (2) to test the effectiveness of a gamified program in real-life conditions on PA of participants versus nonparticipants, and (3) to explore the variables that could explain heterogeneity in response to the intervention. Based on previous results, we first hypothesized that PA will increase both during and after the gamified program, in comparison to initial PA (H1). Second, we hypothesized that this improvement will be greater for participants than for non-participants (i.e., participants who registered on the app but did not complete any game, H2). Finally, we expected that intervention's characteristics (i.e., type and number of games), the context within the intervention was performed (i.e., population and settings), and the exposure to the intervention (i.e., engagement of participants with the app) will moderate the intervention effect (H3).

Methods

Study design and participants

This study retrospectively analyzed data from adult participants who had registered for a Kiplin program including PA games and had given consent for their data to be collected. To be included, participants must be 18 years old or older, have registered on the app between January 1st, 2019, and January 2nd, 2022, and logged daily steps (measured via their smartphone or an activity monitor) on a time frame of at least 90 days with less than 20% of missing daily

observations.[10] Non-wear days were defined as days with fewer than 1000 steps and considered as missing observations – as previous research suggested that daily step values less than 1000 may not represent full data capture.[11,12] Days before the first day of the first game were considered as 'baseline' (*Mdn* = 14 days ± 42.9), the period between the first day of the first game and the last day of the last game as 'intervention period' (*Mdn* = 19 days ± 31.2), and the days after the last day of the last game as 'follow-up' (*Mdn* = 90 days ± 22.8). We restricted the follow-up periods to 90 days post-intervention (i.e., 3 months).

Participants could receive the Kiplin intervention a) in the context of their work (i.e., primary prevention with employees), b) in a senior program (i.e., primary prevention with volunteer retirees), or c) as part of their chronic disease care (i.e., patients mainly treated for obesity or cancer). In all the aforementioned conditions, the program was paid not by the participant but by their employer or health care center. At the beginning of the intervention, participants had to download the Kiplin app. They were given an access code by their employer or health care center, and could then create their account. Upon registration, participants agreed that their anonymized data may be stored on certified health data servers. Participants then benefited from one or several PA games (depending on the program) lasting approximately 14 days each. If several games were proposed, these games followed each other in an interval of fewer than 60 days. In programs with multiple games, there was always a break of some days between games to provide regular doses of gamification. Details on the games' content have been reported previously.[13]

Some participants registered for the program, created their account, but did not take part in the intervention (i.e., did not completed any game). These individuals were considered "nonparticipants" and were used as a control group (as proposed in previous research[14]). Similarly, the baseline period of these nonparticipants corresponds to the days prior to the date they were supposed to start the intervention period.

The study was approved by local Ethics Committee (IRB00013412, "CHU de Clermont Ferrand IRB #1", IRB number 2022-CF063) with compliance to the French policy of individual data protection.

Variables

The variables of interest were selected on the basis of the behavior change intervention ontology of Michie et al.[8] and included (1) the longitudinal evolution of daily steps, (2) the exposure of each participant to the intervention, (3) the intervention parameters, and (4) the context (participants' characteristics and settings), as these variables are likely to influence the intervention effect. Table 1 specifies the measures of interest and their operationalization.

Statistical analyses

Mixed-effects models were used to 1) analyze within-person evolution across time (i.e., changes in daily steps between baseline, intervention, and follow-up periods), and across participants and nonparticipants, and 2) examine the associations between intervention parameters, exposure to the intervention, participants' characteristics and settings, and the daily steps evolution. This statistical approach controls for the nested structure of the data (i.e., multiple observations nested within participants), does not require an equal number of observations from all participants,[15] and separates between-person from within-person variance, providing unbiased estimates of the parameters.[16,17]

First, an unconditional model (i.e., with no predictor) was estimated for each variable to calculate intra-class correlations (ICC) and estimate the amount of variance at the between and within-individual levels, which allowed us to determine whether conducting multilevel models was relevant or not. Then, a model that allowed random slope over time (i.e., model with random intercept and random slope) was compared to the null model (i.e., with only random intercept) using an ANOVA, to evaluate whether the less parsimonious model explain a significantly higher portion of the variance of the outcome, compared to the unconditional model.[18,19] Third, between-level predictors and confounding variables were added to another model (Model 1)*[1] and compared to the previous models. Finally, intervention characteristics as well as their interactions with the phases (i.e., baseline/intervention/follow-up) of the study were added in a final model excluding nonparticipants (Model

[*1] The equation for the Model was the following: $Y_{ij} = (\beta_0 + \gamma_{0i} + \theta_{0j}) + (\beta_1 + \theta_{1j})\ Time_j + \beta_2\ Phase_j + \beta_3\ Age_j + \beta_4\ Sex_j + \beta_5\ Population_j + \beta_6\ Season_j + \beta_7\ Captor_j + \beta_8\ Baseline\ PA_j + \beta_9\ Lockdown_j + \beta_{10}\ Condition_j \times Phase_j + \varepsilon_{ij}$ where β_0 to β_{10} are the fixed effect coefficients, θ_{0j} and θ_{1j} are the random effect for the participant j (one random intercept and one random slope), γ_{0i} is the random effect for the Time i (random intercept), and ε_{ij} is the error term.

2)*[2]. Model fit was assessed via the Bayesian Information Criterion (BIC) and −2-log-likehood (−2LL).[20] All models were performed using the *lmerTest* package in the R software.[21] An estimate of the effect size was reported using the marginal and conditional pseudo R^2. Models' reliability was estimated with residual analyses, performed using the *Performance* package.[22] When the interaction terms turned significant, contrasts analyses were computed using the *emmeans* package.[23]

The data and code for the statistical analyses used in the present study are available on Open Science Framework (https://osf.io/scnu7/).

Table 1. Operationalization of the variables

Outcome	Operationalization
Primary outcome – target behavior (independent variable)	
Daily step count	PA was assessed via the daily step count, measured with the smartphone or activity monitor of the participant. The daily step count is a trusted proxy for PA.[24]
Intervention (content and delivery) and mechanisms of action	
Type of game	Participants could play 4 types of games (i.e., challenge, adventure, boardgame, mission). The challenge is a competitive game where participants had to walk more than other teams to win. The three other games have been introduced elsewhere with details on embedded behavior change techniques.[13]
Exposure	
Compliance ratio	The engagement of participants with the app was computed as the compliance ratio representing the number of days with a login during the game period divided by the duration of the game periods. This variable allows measuring the frequency of the engagement with the service.[25]
Number of games played	The total number of games played during the intervention period.
Context (population and setting)	
Self-reported age and gender	Filled out by participants when they registered on the app.
Population	Employees, seniors, or patients (treated for obesity, or cancer).
Cofounders	
Season	The season (winter, spring, summer, autumn) when the data step was logged, was controlled as the season can influence PA.[26]
Type of device	The type of device used to assess daily step count (i.e., Android or iOS smartphones, Garmin, Withings, Polar, Fitbit, or Tomtom wearables) was controlled as smartphone apps and wearable devices differ in accuracy and precision.[27]
Lockdown	The study period was characterized by the COVID-19 pandemic and 3 lockdowns were set up in France to limit the spread of the outbreak. As these periods had a strong influence on PA of individuals,[28,29] we controlled the lockdown periods in our analyses.

Results

Descriptive Results

Descriptive results are presented in Table 2. The final sample included 4812 adults (mean age = 42.7 ± 11.5 years; 60% women). Sociodemographic variables and baseline physical activity were not descriptively different between participants and nonparticipants. Participants wore an activity monitor measuring their daily step count for an average of 113 days (range = 90 – 686 days). 34,922 daily steps observations were missing on a total of 544,449, which is equivalent to 6.4% of missing data on the full dataset.

*[2] The equation for the Model was the following: $Y_{ij} = (\beta_0 + \gamma_{0i} + \theta_{0j}) + (\beta_1 + \theta_{1j})\ Time_j + \beta_2\ Phase_j + \beta_3\ Age_j \times Phase_j + \beta_4\ Sex_j + \beta_5\ Population_j \times Phase_j + \beta_6\ Season_j + \beta_7\ Captor_j + \beta_8\ Baseline\ PA_j \times Phase_j + \beta_9\ Lockdown_j + \beta_{10}\ Compliance\ ratio_j \times Phase_j + \beta_{11}\ Number\ of\ games\ played_j \times Phase_j + \beta_{12}\ Type\ of\ Game_j + \varepsilon_{ij}$ where β_0 to β_{12} are the fixed effect coefficients, θ_{0j} and θ_{1j} are the random effect for the participant j (one random intercept and one random slope), γ_{0i} is the random effect for the Time i (random intercept), and ε_{ij} is the error term.

Table 2. Descriptive statistics.

	Participants	Non-participants
Participants (observations)	N=3817 (472,946)	N=995 (71,503)
Sociodemographics		
Age, mean (SD)	43.2 (11.08)	41.0 (12.81)
Female (%)	2313 (62)	510 (53)
Employees (%)	3,526 (92)	978 (98)
Patients (%)	194 (5)	17 (2)
Seniors (%)	97 (2)	-
Physical activity (daily steps)		
Baseline, mean (SD)	7,311.5 (3109.3)	6,443.8 (2854.9)
Intervention, mean (SD)	11,488.7 (5372.2)	6,528.5 (2951.6)
Follow-up, mean (SD)	7,789.9 (3000.4)	-
Exposure		
Compliance ratio, mean (SD)	0.84 (0.23)	0
Games played, mean (SD)	1.28 (0.9)	0
In-game days, mean (SD)	22.06 (16.24)	0
Observations in each type of game (in days)		
Adventure (% of all days)	21,316 (33)	-
Boardgame (% of all days)	4,093 (6)	-
Challenge (% of all days)	32,801 (50)	-
Mission (% of all days)	6,915 (11)	-
Type of device used		
Android smartphone (%)	1076 (28)	286 (29)
iOS smartphone (%)	810 (21)	533 (54)
Fitbit (%)	750 (20)	52 (5)
Garmin (%)	1,071 (28)	109 (11)
Polar (%)	5 (0.1)	-
Tomtom (%)	3 (0.08)	-
Withings (%)	90 (2)	9 (1)
Observations in each season (in days)		
Winter (% of all days)	110,517 (24)	17,451(24)
Spring (% of all days)	94,961 (20)	21,162 (30)
Summer (% of all days)	129,039 (27)	8,804 (12)
Fall (% of all days)	138,429 (29)	24,086 (34)
Observations in each lockdown (in days)		
1st lockdown – spring 2020 (% of all days)	10,872 (3)	925 (1)
2nd lockdown – fall 2020 (% of all days)	32,298 (8)	4,110 (6)
3rd lockdown – spring 2021 (% of all days)	23,435 (6)	1,757 (2)

Is the gamified program effective to promote PA? (H1)

During the intervention period, participants increased their daily steps by 4177 steps per day on average, compared to the baseline period, and by 478 steps per day on average during the follow-up period, compared to the baseline. In comparison, the daily step count of the control group remained more or less stable throughout the same timeframe with a mean increase of 84 daily steps compared to baseline.

Overall, contrast analyses of the model for the intervention participants (Model 2, Table 3) revealed a negative effect of the intervention on the daily step count during the intervention phase compared to baseline activity (b = -0.09, 95 CI [-0.14; -0.05], p <.0001) and no significant effect (b = 0.01, 95 CI [-0.05; 0.06], p = 0.79) during follow-up periods compared to baseline. However, the patterns were different when participants were stratified by baseline PA. Participants with lower baseline daily steps (<5000 steps per day or 5001-7500 steps per day) showed a significant increase of their daily steps during the intervention and the follow-up, both compared to the baseline (respectively b = 0.25, 95 CI [0.22; 0.28], p <.0001 and b = 0.12, 95 CI [0.09; 0.15], p <.0001). Participants with initial values between 7501 and 10000 steps did not have significant increase their daily steps during the intervention (b = 0.00, 95 CI [-0.05; 0.05], p = 0.99) nor during the follow-up period (b = -0.01, 95 CI [-0.04; 0.02], p = 0.44), compared to baseline. Participants who performed more than 10000 baseline steps had significant deteriorations during the intervention (b = -0.13, 95 CI [-0.19; -0.08], p <.0001) and follow-up (b = -0.06, 95 CI [-0.10; -0.03], p = 0.0001). These trends are depicted in Figure 1.

Is the intervention effect greater for participants compared to nonparticipants? (H2)

In Model 1 (Table 3), participants who received the intervention from Kiplin had a significantly greater increase in mean daily steps between baseline and the intervention period, compared with nonparticipants (b = 0.54, 95%CI [0.52; 0.58], p <.0001).

What are the moderators of the intervention effect? (H3)

The Model 2 estimates are displayed in Table 2. The variables under consideration explained 39% of the variance in daily steps. In this model, we tested the hypothesized interactions, to investigate predictors associated with the efficiency of the intervention (Table 4). Contrast analyses were conducted on significant interactions and revealed that the

Table 3. Mixed effect models' estimates.

	Unconditional Model			Model 1			Model 2		
	b [95 CI]	SE	*p*	*b* [95 CI]	SE	*p*	*b* [95 CI]	SE	*p*
Intercept	0.04 [-0.06; -0.02]	0.01	**<.001**	-0.70 [-0.78; -0.62]	0.04	**<.001**	-0.34 [-0.39; -0.29]	0.03	**<.001**
Fixed Effects									
Time	0.10 [-0.11; -0.09]	0.01	**<.001**	0.11 [0.05; 0.16]	0.03	**<.001**	-0.05 [-0.06; -0.04]	0.01	**<.001**
Phases of the study									
Baseline				Reference			Reference		
Intervention				0.01 [-0.02; 0.04]	0.01	0.441	0.31 [0.28; 0.34]	0.01	**<.001**
Follow-up				-			0.29 [0.27; 0.32]	0.01	**<.001**
Participants' profile									
Age				0.01 [-0.00; 0.02]	0.01	0.117	-0.00 [-0.02; 0.01]	0.01	0.523
Sex female				Reference			Reference		
Sex male				-0.01 [-0.03; 0.01]	0.01	0.397	0.02 [-0.00; 0.05]	0.01	0.104
Condition									
Control				Reference			-		
Kiplin				-0.03 [0.07; 0.09]	0.02	0.079	-		
Population									
Workers				Reference			Reference		
Seniors				-0.05 [-0.12; 0.03]	0.04	0.213	0.13 [0.05; 0.21]	0.04	**0.002**
Obese patients				-0.12 [-0.21; -0.02]	0.05	**0.017**	-0.01 [-0.14; 0.12]	0.06	0.864
T2DM patients				-0.37 [-0.62; -0.12]	0.13	**0.003**	-0.18 [-0.44; 0.09]	0.13	0.189
Cancer patients				0.02 [-0.05; 0.09]	0.04	0.532	0.07 [-0.01; 0.15]	0.04	0.076
Other patients				-0.11 [-0.21; -0.01]	0.05	**0.031**	0.08 [-0.05; 0.21]	0.07	0.241
Type of game									
Adventure							Reference		
Boardgame							-0.17 [-0.20; -0.14]	0.02	**<.001**
Challenge							-0.08 [-0.10; -0.06]	0.01	**<.001**
Mission							-0.17 [-0.20; -0.14]	0.01	**<.001**
Exposure									
Observance ratio							0.02 [0.00; 0.04]	0.01	**0.021**
Number of games							0.02 [-0.00; 0.04]	0.01	0.064

Season							
Winter		Reference			Reference		
Spring		0.04 [0.03; 0.06]	0.01	**<.001**	0.04 [0.02; 0.05]	0.01	**<.001**
Summer		0.11 [0.09; 0.13]	0.01	**<.001**	0.06 [0.05; 0.08]	0.01	**<.001**
Fall		0.07 [0.05; 0.09]	0.01	**<.001**	0.02 [0.01; 0.03]	0.01	**<.001**
Type of device							
Android		Reference			Reference		
iOS		-0.05 [-0.08; -0.03]	0.01	**<.001**	0.02 [-0.01; 0.05]	0.01	0.128
Fitbit		0.29 [0.26; 0.33]	0.02	**<.001**	0.50 [0.47; 0.53]	0.02	**<.001**
Garmin		0.17 [0.14; 0.20]	0.02	**<.001**	0.33 [0.30; 0.36]	0.01	**<.001**
Polar		-0.94 [-2.23; -0.35]	0.66	0.154	0.11[-0.72; 0.95]	0.43	0.788
Tomtom		0.75 [0.35; 1.15]	0.21	**<.001**	0.81 [0.45; 1.18]	0.19	**<.001**
Withings		0.05 [-0.03; 0.12]	0.04	0.199	0.11 [0.05; 0.18]	0.03	**0.001**
Lockdown							
Periods without restrictions		Reference			Reference		
1st lockdown		-0.09 [-0.21; -0.01]	0.02	**<.001**	0.18 [-0.20; -0.15]	0.01	**<.001**
2nd lockdown		0.00 [-0.02; 0.03]	0.01	0.795	0.00 [-0.02; 0.02]	0.01	0.989
3rd lockdown		0.01 [-0.01; 0.03]	0.01	0.502	0.02 [0.00; 0.03]	0.01	**0.034**
Random Effects							
Level 1 intercept variance	0.70	0.67			0.67		
Level 2 intercept variance	0.32	1.88			0.13		
-2*log (lh)	-515610.3	-236638.1			-449086		
Akaike Information Criteria	1031232.6	473334.3			898288.4		
Marginal R^2 / Conditional R^2	0.01 / 0.38	0.071 / 0.86			0.225 / 0.39		

age ($b = 0.05$, $p < .0001$) and the compliance ratio ($b = 0.37$, $p < .0001$) were positively associated with the change in daily steps between baseline and intervention. Specifically, the older the age, the more regularly the individuals played and the more effective the intervention was. On the other hand, the number of games played by participants was negatively associated with this change ($b = -0.02$, $p = 0.02$). In other words, the longer the intervention and the higher the number of games, the less effective the intervention. For categorical outcomes, contrast analyses revealed differences in the intervention effect between the different populations. Compared to employees, cancer patients ($b = -0.18$, 95 CI [-0.24; -0.12], $p < .0001$), and seniors ($b = -0.19$, 95 CI [-0.25; -0.13], $p < .0001$) observed a significantly weaker effect of the intervention in comparison to baseline PA. There was no significant difference between employees and obese patients ($b = -0.07$, 95 CI [-0.16; 0.02], $p = 0.13$). In sum, programs conducted with office workers or patients treated for obesity had better effects than programs on other populations (Figure 2). All the results of these analyses are available in supplementary materials.

Table 4. Interactions tested between the intervention phase, participants' characteristics, and intervention parameters in Model 2.

	b [95 CI]	SE	*p*
Model 1			
Intervention × Control	*Reference*		
Intervention × Kiplin	0.55 [0.52; 0.58]	0.02	**<.001**
Model 2			
Intervention × Age	0.05 [0.04; 0.06]	0.01	**<.001**
Intervention × Compliance ratio	0.37 [0.35; 0.38]	0.01	**<.001**
Intervention × Nb of games	-0.02 [-0.03; -0.00]	0.01	**0.021**
Intervention × Workers	*Reference*		
Intervention × Cancer	-0.18 [-0.24; -0.12]	0.03	**<.001**
Intervention × Obese	-0.07 [-0.16; 0.02]	0.04	0.133
Intervention × Senior	-0.19 [-0.25; -0.13]	0.03	**<.001**
Intervention × base <5000	*Reference*		
Intervention × base 5000 - 7500	-0.23 [-0.26; -0.20]	0.01	**<.001**
Intervention × base 7501 - 10000	-0.41 [-0.44; -0.38]	0.02	**<.001**
Intervention × >10000	-0.60 [-0.64; -0.57]	0.02	**<.001**

Notes base = baseline daily steps; control = nonparticipants

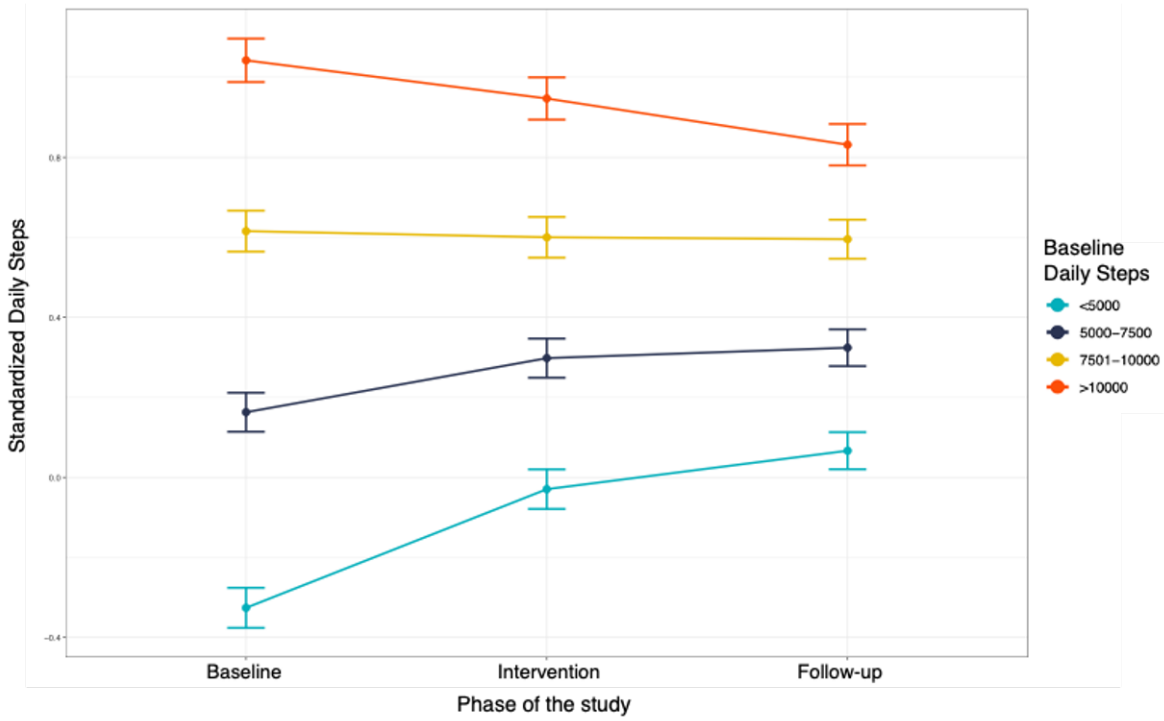

Figure 1. Changes in daily steps throughout the study phases for participants who received a Kiplin program, stratified by baseline daily steps

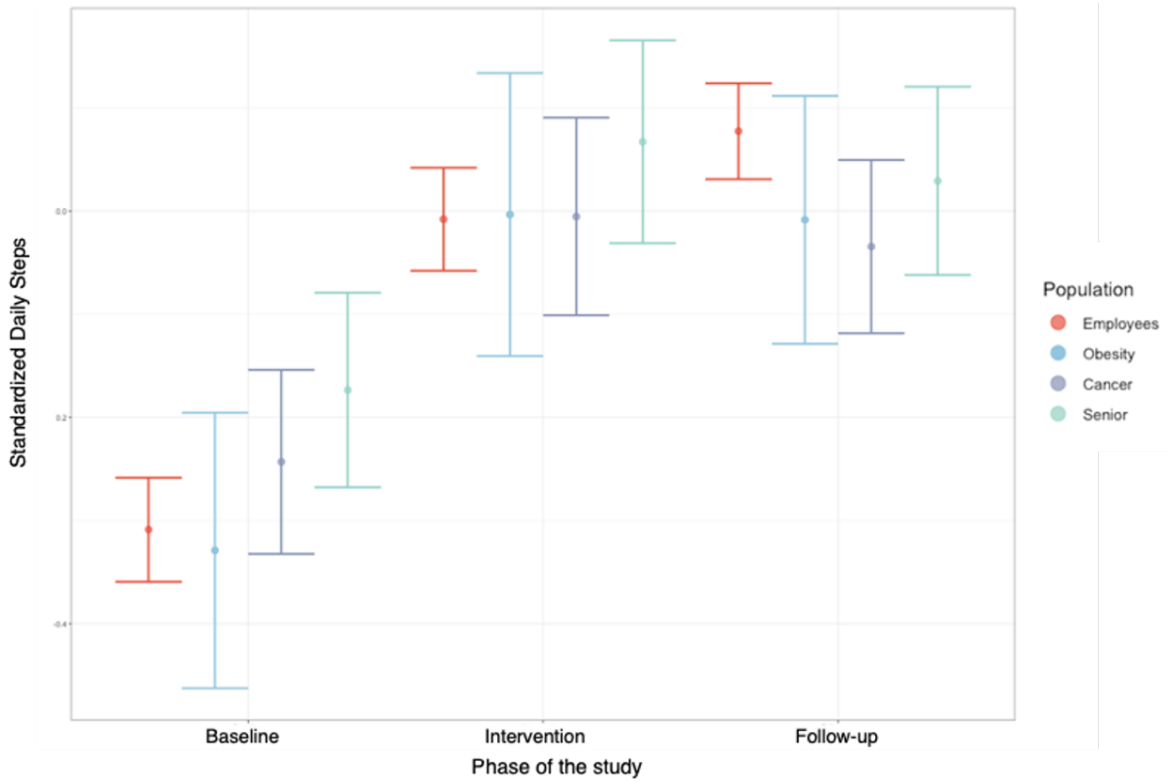

Figure 2. Changes in daily steps throughout the study phases for the different populations who received a Kiplin program.

Discussion

This observational study retrospectively analyzed the real-world data of 4800 participants who registered on the Kiplin app. We found that participants benefiting from the Kiplin games significantly increased PA compared with non-participants during the same period. We also found that the intervention effect depended on the baseline PA of individuals. Participants with lower baseline steps (<5000 steps per day or 5001-7500 steps per day) significantly improved their PA both during the intervention and follow-up periods whereas participants with more than 7500 steps had no change or significant decreases. These results suggest that a gamified program is more efficient among inactive individuals than active ones, with the existence of a plateau effect. They also confirm recent findings[7] and the ability of gamified interventions to improve PA both during and after the end of the program - at least for the more inactive individuals. This effectiveness is particularly interesting considering that current behavioral interventions struggle to change PA in the long haul.[30]

Once again, our results stressed that older age may not be incompatible with gamified interventions. Indeed, we found that intervention effectiveness was moderated by the age of the individual and that gamification was more efficient among older individuals, compared to younger ones. Whether in a global way the literature on gamification shows cautiously positive results in the use of gamification for older people,[31] the present results are in line with a previous study[32] which found that older users had a greater degree of use of the gamification features. The authors proposed the explanation that older adults pay generally more attention to their health and thus have a stronger intention to engage in a health program. From another perspective and in light of the Kiplin games' characteristics, these results could also be explained by the fact that these games are accessible – inspired by traditional board game rules and mechanics widely known in the general population (e.g., Cluedo, snakes and ladders) – and thus may be more attractive for older populations. Indeed, the most engaging game mechanics may diverge between youth and other populations,[33] and we can expect that younger populations may prefer more complex game mechanics and need more novelty during the intervention to stay interested by the service.

Regarding the effects of our gamified intervention according to the characteristics of the population, we found a stronger effect on office workers and obese patients. If these results do not allow to draw conclusions that would be too hasty considering the high variability observed in patients or senior participants, these findings highlight that, beyond the attributes and health status of the participants,

the setting of the intervention can be important. For example, interventions proposed with employees were conducted within their company. Participants thus know each other which can enhance the motivational impact of some gamification features (such as social comparison with leaderboards and social connectedness with teams) whereas interventions in healthcare settings usually involve patients who do not know each other.

Our findings also revealed several insights that could help to improve future intervention design. First, exposure to the content is essential for the gamified intervention to be effective. It is interesting as gamification has often been assimilated into a self-fulfilling process permitting automatic engagement of participants. These results are consistent with previous findings demonstrating that higher use of gamification features was associated with greater intervention effectiveness.[32,34] If gamification can ultimately increase program engagement, developers need first to design their apps to be as attractive as possible and optimize retention.

Second, the total number of games played was negatively associated with the intervention effect, suggesting that a shorter intervention could be more beneficial for behavior change. These results are in line with previous research[7,35] suggesting that users benefit more from digital interventions shorter than 3 months. It also suggests a «dose-response» relationship in inverted U shape, with an optimal "middle" to find. Nevertheless, it is necessary to take into consideration the fact that Kiplin programs of more than one game are built in such a way as to decimate several doses at regular intervals. Periods without games were therefore considered in the intervention phases and could explain why, overall, the shorter games were more efficient. More refined analyses of the intervention effect over time will be necessary in the future.

Third, the daily step count of participants was significantly higher in the adventure and the challenge. These two games share the characteristic of being more competitive with a stronger emphasis on leaderboards than the two other games more focused on collaboration. In that idea, Patel et al.[36] observed that the competitive version of their gamified intervention outperformed the collaborative and supportive arms. Moreover, various studies demonstrated that leaderboards are a particularly successful gamification mechanic.[32,37]

Strengths and limitations

This study has several strengths, including the large number of participants included, the intensive objective PA measurement in real-life conditions, and the longer baseline and follow-up duration compared with most trials on gamification that typically incorporate measurement bursts dispersed across time.[7] However, several limitations should be considered. First, this study was observational and not a randomized controlled trial. Thus, we cannot establish the causality of the intervention's effect on outcome improvement. The non-participants are not a true control group. If they did not receive the intervention, it may be for underlying motivational reasons that could impact their PA. Second, intervention lengths differed between participants. Third, if multilevel models are useful for describing trends in PA behavior change over time, they are limited in their capacity to assess precise fluctuations patterns of non-stationarity behavior such as PA[38] across time. Slightly more complex options are available to precisely describe time changes and patterns (e.g., time series analyses) and could be used in future longitudinal studies. Finally, the compliance ratio used in this study as a proxy for engagement tends to oversimplify the exposure of participants to the service. Complementary measures of engagement (e.g., using the number of logins, time spent per login, and the number of components accessed) will need to be conducted to draw the longitudinal impact of the engagement of the participants on the intervention effect.

Conclusion

In this study in which we retrospectively analyzed the daily step count of 4800 individuals in real-life conditions, participants who benefited from the Kiplin gamified intervention had a significantly greater increase in mean daily steps from baseline than nonparticipants. Responses to the intervention were significantly different as a function of individuals' initial PA. Whereas participants with less than 7500 baseline daily steps had significant improvements both during the intervention and follow-up periods, the intervention had no effect on participants with initial values >7500. The age of participants and the engagement with the app were positively and significantly associated with the intervention effect while the number of games played was negatively associated with it. The results of this study suggest that gamification is effective to promote PA of inactive populations at short and medium-term effects. To our knowledge, this study is the first to examine the longitudinal effect of a gamified program outside the context of a trial, with real-world

data. The results of this study are therefore highly generalizable and confirm the interest of gamification in both primary and tertiary prevention.

Additional Information

Contributions

Conceptualization of the study: AM; data curation: AM, GH; investigation: AM; methodology and statistics: AM, CF; software: AM; writing – original draft: AM; writing – review & editing: all the authors.

Declaration of interests

AC, CF, and MD declare that they have no competing interests. AM's PhD grant is funded by the French National Association for Research and Technology (ANRT) and Kiplin. GH is employed by Kiplin.

Funding

The work of AM is supported by an ANRT grant (Cifre PhD Thesis) and by the company Kiplin. The funders had no input in the design of the study and no influence on the interpretation or publication of the study results.

Data & Code Accessibility

The anonymized data used in this study and the R code are available on the Open Science Framework (https://osf.io/scnu7/).

References

1 Lee IM, Shiroma EJ, Lobelo F, *et al.* Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. *Lancet* 2012. DOI:10.1016/S0140-6736(12)61031-9.

2 Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. *Lancet Glob Heal* 2018; **6**: e1077–86.

3 Hallal PC, Andersen LB, Bull FC, *et al.* Global physical activity levels: Surveillance progress, pitfalls, and prospects. Lancet. 2012. DOI:10.1016/S0140-6736(12)60646-1.

4 Conger SA, Toth LP, Cretsinger C, *et al.* Time Trends in Physical Activity Using Wearable Devices: A Systematic Review and Meta-analysis of Studies from 1995 to 2017. *Med Sci Sports Exerc* 2022; **54**: 288–98.

5 Tison GH, Barrios J, Avram R, *et al.* Worldwide physical activity trends since COVID-19 onset. *Lancet Glob Heal* 2022; **10**: e1381–2.

6 Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: Defining 'gamification'. *Proc 15th Int Acad MindTrek Conf Envisioning Futur Media Environ MindTrek 2011* 2011; : 9–15.

7 Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials. *J Med Internet Res* 2022; **24**: e26779. doi: 10.2196/26779.

8 Michie S, Thomas J, Johnston M, *et al.* The Human Behaviour-Change Project: Harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implement Sci* 2017; **12**: 1–12.

9 Hekler EB, Michie S, Pavel M, *et al.* Advancing Models and Theories for Digital Behavior Change Interventions. *Am J Prev Med* 2016; **51**: 825–32.

10 Chevance G, Baretta D, Heino M, *et al.* Characterizing and predicting person-specific, day-to-day, fluctuations in walking behavior. *PLoS One* 2021; **16**: e0251659.

11 Bassett DR, Wyatt HR, Thompson H, Peters JC, Hill JO. Pedometer-measured physical activity and health behaviors in U.S. adults. *Med Sci Sports Exerc* 2010; **42**: 1819–25.

12 Kang M, Rowe DA, Barreira T V., Robinson TS, Mahar MT. Individual Information-Centered Approach for Handling Physical Activity Missing Data. *https://doi.org/101080/02701367200910599546* 2013; **80**: 131–7.

13 Mazéas A, Chalabaev A, Blond M, Pereira B, Duclos M. Digital intervention promoting physical activity among obese people (DIPPAO) randomised controlled trial: study protocol. *BMJ Open* 2022; **12**: e058015. doi:10.1136/ bmjopen-2021-058015.

14 Wang G, Yang M, Hong M, Krauss J, Bailey JF. Clinical outcomes one year after a digital musculoskeletal (MSK) program: an observational, longitudinal study with nonparticipant comparison group. *BMC Musculoskelet Disord* 2022; **23**: 1–14.

15 Sonnentag S. Dynamics of Well-Being. *Annu Rev Organ Psychol Organ Behav* 2015; **2**: 261–93.

16 Boisgontier MP, Cheval B. The anova to mixed model transition. *Neurosci Biobehav Rev* 2016; **68**: 1004–5.

17 Judd CM, Westfall J, Kenny DA. Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annu Rev Psychol* 2017; **68**: 601–25.

18 Bliese P. Multilevel Modeling in R (2.7) A Brief Introduction to R, the multilevel package and the nlme package. 2022.

19 Finch WH, Bolin JE. Multilevel modeling using Mplus. *Multilevel Model Using Mplus* 2017; : 1–321.

20 Bollen KA, Harden JJ, Ray S, Zavisca J. BIC and Alternative Bayesian Information Criteria in the Selection of Structural Equation Models. *Struct Equ Model a Multidiscip J* 2014; **21**: 1–19.

21 Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *J Stat Softw* 2017; **82**: 1–26.

22 Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *J Open Source Softw* 2021; **6**. DOI:10.21105/joss.03139.

23 Lenth R, Singmann H, Love J, Buerkner P, Herve M. Emmeans: Estimated marginal means, aka least-squares means. *R Packag version* 2018; **1**: 3.

24 Saint-Maurice PF, Troiano RP, Bassett DR, *et al.* Association of Daily Step Count and Step Intensity with Mortality among US Adults. *JAMA - J Am Med Assoc* 2020; **323**: 1151–60.

25 Short CE, DeSmet A, Woods C, *et al.* Measuring Engagement in eHealth and mHealth Behavior Change Interventions: Viewpoint of Methodologies. *J Med Internet Res 2018;20(11)e292 https//www.jmir.org/2018/11/e292* 2018; **20**: e9397.

26 Turrisi TB, Bittel KM, West AB, *et al.* movement behaviors : a scoping review from 2006 to 2020. 2021; : 1–26.

27 Piccinini F, Martinelli G, Carbonaro A. Accuracy of Mobile Applications versus Wearable Devices in Long-Term Step Measurements. *Sensors 2020, Vol 20, Page 6293* 2020; **20**: 6293.

28 Maltagliati S, Rebar A, Fessler L, *et al.* Evolution of physical activity habits after a context change: The case of COVID-19 lockdown. *Br J Health Psychol* 2021; **26**: 1135.

29 Genin PM, Lambert C, Larras B, *et al.* How Did the COVID-19 Confinement Period Affect Our Physical Activity Level and Sedentary Behaviors? Methodology and First Results From the French National ONAPS Survey. *J Phys Act Heal* 2021; **18**: 296–303.

30 Dunton GF, Rothman AJ, Leventhal AM, Intille SS. How intensive longitudinal data can stimulate advances in health behavior maintenance theories and interventions. *Transl Behav Med* 2021; **11**: 281–6.

31 White BK, Martin A, White J. Gamification and older adults: Opportunities for gamification to support health promotion initiatives for older adults in the context of COVID-19. *Lancet Reg Heal - West Pacific* 2022; : 100528.

32 Maher CA, Olds T, Vandelanotte C, *et al.* Gamification in a Physical Activity App: What Gamification Features Are Being Used, by Whom, and Does It Make a Difference? *Games Health J* 2022; **11**: 193–9.

33 Schwarz AF, Huertas-Delgado FJ, Cardon G, Desmet A. Design Features Associated with User Engagement in Digital Games for Healthy Lifestyle Promotion in Youth: A Systematic Review of Qualitative and Quantitative Studies. *Games Health J* 2020; **9**: 150–63.

34 Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. *PLoS One* 2017; **12**: e0173403.

35 Romeo A, Edney S, Plotnikoff R, *et al.* Can Smartphone Apps Increase Physical Activity? Systematic Review and Meta-Analysis. *J Med Internet Res* 2019; **21**: e12053.

36 Patel MS, Small DS, Harrison JD, *et al.* 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* 2019. DOI:10.1001/jamainternmed.2019.3505.

37 De Croon R, Geuens J, Verbert K, Vanden Abeele V. A Systematic Review of the Effect of Gamification on Adherence Across Disciplines. *Int Conf Human-Computer Interact* 2021; **12789 LNCS**: 168–84.

38 Heino MTJ, Knittle K, Noone C, Hasselman F, Hankonen N. Studying Behaviour Change Mechanisms under Complexity. *Behav Sci 2021, Vol 11, Page 77* 2021; **11**: 77.