Pers Ubiquit Comput (2014) 18:1705–1719
DOI 10.1007/s00779-014-0783-2

ORIGINAL ARTICLE

Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity

Oren Zuckerman · Ayelet Gal-Oz

Received: 1 November 2013 / Accepted: 16 April 2014 / Published online: 5 July 2014

Abstract Game design elements are often implemented in persuasive systems aimed to promote physical activity, a process called "gamification." Gamification is believed to motivate users to become more active, and is commonly implemented in commercial products. However, relatively few studies rigorously evaluated the effectiveness of gamification, and they yielded contradicting findings. We set out to evaluate the effectiveness of virtual rewards and social comparison—two game elements prevalent in persuasive systems. We developed a research prototype, called "StepByStep," aimed to promote routine walking. We created different versions of StepByStep, implemented as an application on Android-based mobile devices, and compared their effectiveness in two field studies. Study 1 showed that a quantified version of the application—offering continuous measurement of walking time, a daily goal, and real-time feedback on progress toward this goal—facilitated reflection on activity and significantly increased walking time over baseline level. Study 2 showed that gamified versions offering virtual rewards and social comparison were only as effective as the quantified version. Thus, we advise designers to facilitate reflection on meaningful aspects of physical activity by developing novel ubiquitous measures. Furthermore, our findings highlight the importance of systematic comparisons between quantified and gamified elements for better understanding their motivational affordances.

O. Zuckerman (✉) · A. Gal-Oz
Media Innovation Lab, Sammy Ofer School of Communications, The Interdisciplinary Center (IDC) Herzliya, Herzliya, Israel
e-mail: orenz@idc.ac.il

A. Gal-Oz
e-mail: goayelet@idc.ac.il

Keywords Persuasive technology · Behavior change · Gamification · Virtual reward · Social comparison · Physical activity

1 Introduction

In recent years, the human–computer interaction (HCI) community has shown a growing interest in persuasive technologies—technologies designed to support behavior change in everyday life [1, 2]. A special interest has been given to technologies designed to support health-related behavior change [e.g., 3–6], because it holds the potential to improve quality of life. One way for substantially improving the quality of life is performing physical activity, which improves physical and mental health, and reduces the risk of various diseases as well as overall premature mortality [7–10]. Despite this common knowledge that physical activity is healthy, many people are not regularly active, or are not active at all [11]. One of the main reasons for this is the fact that western society has become characterized by environments that promote physical inactivity [12].

Technology can assist with counteracting this trend by motivating people to maintain a more active, healthier lifestyle. In particular, mobile technology holds great promise as a vehicle for promoting physical activity, because it offers a host of sensing technologies and data visualization tools, which allow for ubiquitously sensed data to be stored, analyzed, and communicated [13]. Furthermore, mobile devices' performance is increasing, their price is decreasing, they are always connected, and people carry their device with them all day long, enabling continuous tracking [14].

Indeed, HCI researchers and commercial companies alike have been developing various systems designed to promote physical activity. Generally, such systems consist of two main components: measurement of activity, and presentation of the measured data. Some present the data in a quantified manner with concrete numerical information [e.g., 15, 16]. Notably, the quantified self-movement (http://quantifiedself.com) advocates "self-knowledge through numbers." This approach has been called "personal informatics," defined by Li et al. [17] as collecting personally relevant information for the purpose of gaining self-knowledge. Personal informatics systems facilitate collection and storage of personal information, and provide a means of exploring and reflecting on the information. Reflection leads the individual to reconsider and possibly change attitudes or behaviors [18].

In contrast, other systems present physical activity data in a game-like manner [19]. This approach is called "gamification," defined by Deterding et al. [20] as the use of game design elements in non-game contexts. The underlying assumption is that gamification would make physical activity more enjoyable, thereby motivate users to become more active.

Gamification has been gaining popularity in recent years [21, 22], and has been proposed as a design pattern for persuasive systems [23]. However, gamification has also been criticized by HCI researchers. Deterding [24] lists four accounts for criticism against gamified systems: (1) they are not systemic, in that they merely add game design elements, whereas experiences in games emerge from the dynamic interaction of users with all system components. (2) They are reward-oriented, in that they focus on motivating through external rewards instead of intrinsic motivations. (3) They are not user-centric, in that they emphasize the goals of the system owner instead of the goals of the users. (4) They are pattern-bound, in that they limit themselves to a small set of feedback interface design patterns, rather than affording the structural qualities of games that give rise to gameful experiences. Beyond this theoretical criticism, Deterding et al. [25] noted that gamified systems have a hybrid nature, being neither "pure" functional software nor a "full-fledged" game. Therefore, there are currently no established methods, let alone empirically tested ones, for the design of gamified systems. Moreover, there is no sufficient empirical evidence regarding the effectiveness of gamified systems, though more and more studies are emerging in the field. Hamari et al. [21] attempted to evaluate the effectiveness of gamification by analyzing 24 peer-reviewed empirical studies on gamified systems. The majority of the reviewed studies yielded positive results from gamification. While these findings seem promising, the authors highlighted several methodological limitations in the reviewed studies, which might qualify the results: some studies lacked control groups and relied solely on user evaluation. In other words, even though gamification was perceived positively by users, actual effects on behavior were not examined. In addition, controls between implemented elements of gamification were often lacking, so multiple elements were investigated as a whole. The authors concluded that more rigorous methodologies ought to be used in future research on gamification. Furthermore, the majority of reviewed studies were conducted in the context of education or learning, with only one study conducted in the context of physical activity. One of the main conclusions of the review was that context is an essential antecedent for engaging gamification. Therefore, it remains unclear whether gamification is effective in the context of physical activity.

In light of the criticism toward gamification and lack of rigorous studies evaluating its effectiveness, we set out to systematically evaluate the effectiveness of several gamification elements. We developed a research prototype, called "StepByStep," specifically for the purpose of conducting granulated comparisons between different elements of gamification. StepByStep is an accelerometer-based mobile application intended to motivate people to incorporate more walking into their daily routine. Consolvo et al. [26] distinguish between two types of physical activity: "opportunistic physical activity," where people incorporate activities into their everyday lives (e.g., take the stairs instead of the elevator), and "structured exercise," where people elevate their heart rate for an extended period of time (e.g., work out at a gym). We focused on opportunistic physical activity in the form of walking, because it is a relatively easy activity to incorporate into daily routines. An emerging consensus among health professionals suggests that even short bouts of physical activity confer health benefits [27, 28]. We wanted to show users that every second of walking counts, because they accumulate to a substantial amount of walking at the end of the day. Hopefully, this realization would motivate users to walk more.

We created different versions of StepByStep, each presented activity data in a different manner, and evaluated them in two field studies. Study 1 evaluated the effectiveness of a quantified version for promoting physical activity. Study 2 compared between the effectiveness of the quantified version and two gamified versions.

In this paper, we review theoretical and empirical literature regarding the effectiveness of systems aimed to promote physical activity, explain why we developed StepByStep as a research prototype for deconstructing gamification, and discuss the findings from our studies. We conclude with suggestions for both researchers and designers.

2 Prior work

Numerous persuasive systems for promoting physical activity have been introduced over the years. We review the most relevant work, beginning with quantified systems and then gamified systems.

2.1 Quantified systems for promoting physical activity

Quantified systems measure one or several parameters related to physical activity and present the data with concrete numerical information. The main goal of quantified systems is facilitating reflection on this information [17]. Schön [29] differentiated between two modes of reflection: reflection-in-action and reflection-on-action. Reflection-in-action refers to contemplation at the time of doing, whereas reflection-on-action refers to contemplation of previous activities. Quantified systems can facilitate reflection-in-action by providing real-time feedback at the time of performing physical activity, and reflection-on-action by providing information regarding previously performed activities. Both modes of reflection could potentially lead users to change their current level of physical activity [18].

Pedometers are a well-known example of a quantified system for promoting physical activity. Pedometers are small, lightweight instruments, typically worn at the waist, which record and display movement as steps taken. Some also have features to estimate energy expended and/or distance traveled, but these are considered less accurate [30]. Pedometers have been widely used in clinical interventions for increasing physical activity [31]. Such interventions are often accompanied by formal evaluations of their effectiveness, providing valuable insights into how to motivate people to become more active. In a typical pedometer-based intervention, participants are given a pedometer to wear every day, all day, as they go about their usual activities [32]. A recent meta-analysis concluded that pedometer use is associated with an average of 26.9 % increase in physical activity over baseline [30], and with clinically relevant reductions in weight and blood pressure [30, 32]. Thus, it appears that a simple quantified system is sufficient to induce behavioral changes. The effectiveness of pedometers is believed to be rooted in the heightened awareness toward physical activity [33], as well as by the flexibility to fit in walking whenever it is convenient for the individual [32].

An additional key motivational factor for increasing physical activity with pedometers is setting a step goal. Pedometer users who were given a step goal, whether fixed or personalized, were found to walk more than pedometer users who were not given a goal [31]. The importance of goal-setting is also grounded in theory [34] and in previous work within the HCI community [35].

In recent years, more advanced quantified systems have been introduced. For example, Fitbit (www.fitbit.com), BodyBugg (www.bodybugg.com), Nike + (nikeplus.nike.com/plus), Nike + FuelBand (www.nike.com/fuelband), and Jawbone UP (jawbone.com/up) are all accelerometer-based or GPS-based activity trackers, showing users activity metrics like number of steps taken, calories burned, average speed, and distance traveled. All these devices are supposed to be worn directly on the body, whether on the wrist, around the waist or inside a shoe, and are accompanied by a special Web site or mobile application where users can set goals and track their progress. In addition to the quantified information, these systems also include game elements, offering users virtual rewards based on their activity and enabling them to interact with other users. We will discuss gamification in the next chapter.

It is important to note that goal-setting and real-time feedback are regarded by some researchers as game elements [e.g., 24]. However, both goal-setting and real-time feedback are utilized in numerous non-game and non-gamified contexts. Moreover, they are regarded as general techniques for facilitating reflection [18] and behavior change [36, 37]. Therefore, in this paper, we refer to them as quantified elements in the design of persuasive systems for promoting physical activity, not as gamified elements.

2.2 Gamified systems for promoting physical activity

Gamification refers to the use of game design elements in non-game contexts [20]. The underlying assumption is that gamification would make physical activity more enjoyable, thereby motivate users to become more active. Deterding et al. [20] suggested a five-level hierarchy of game design elements. The first level refers to game interface design patterns, for example, badges, leaderboard, levels. The second level refers to game design patterns and mechanics, for example, time constraint, limited resources, and turns. The third level refers to game design principles and heuristics, for example enduring play, clear goals, and variety of game styles. The fourth level refers to game models, for example, challenge, fantasy, and curiosity. The fifth level refers to game design methods, for example, playtesting, playcentric design, and value conscious game design. Our research focuses on two first-level elements: virtual rewards and social comparison; hence, this review focuses on these elements.

2.2.1 Motivational affordances of gamification

According to the self-determination theory (SDT) of human motivation [38], three innate psychological needs determine motivation: competence, autonomy, and relatedness. When these needs are satisfied, they enhance intrinsic motivation. When they are thwarted, they diminish intrinsic motivation.

Gamification could potentially reduce intrinsic motivation, because external rewards are known to reduce intrinsic motivation [39]. As explained by Deterding [40], playing a game is voluntary and free of consequences, two characteristics that enhance perceived autonomy, which is intrinsically motivating. In contrast, using a gamified system that offers virtual rewards or public social comparison is not necessarily voluntary or free of consequence. Thus, it could be experienced as thwarting autonomy and hence intrinsic motivation. Accordingly, Nicholson [41] predicts that external gamification elements, which are artificially attached to an underlying non-game activity, would reduce motivation in the long run.

We now turn to review empirical findings regarding the effectiveness of virtual rewards and social comparison for promoting physical activity.

2.2.2 *Virtual rewards*

Virtual rewards are digital or intangible incentives given following a desired response in an attempt to reinforce the response. They can come in the form of points, badges, or extra game commodity [42]. Antin and Churchill [43] suggested five individual and social functions for badges: goal-setting, instruction, reputation, status/affirmation, and group identification. The function of a specific badge depends on the nature of the activities it rewards and its implementation in particular contexts. Badges are not expected to be universally appreciated, understood or attended to. Indeed, users of "GoalLine" [44], an application that supports setting weekly physical activity goals, journaling physical activity, reviewing past progress, and receiving rewards in the form ribbons and trophies, were indifferent to the rewards. They did, however, like setting goals and found them motivating.

"UbiFit Garden" [45, 46] uses the screen background of an individual's mobile phone to display an animated garden that represents physical activity and goal attainment—the garden blooms as the individual performs physical activities throughout the week. In addition, the system includes an interactive application with detailed information about the individual's physical activities, and a fitness device that automatically infers and transmits information about several types of activities to the glanceable display and interactive application. A 3-month field experiment evaluating the effectiveness of the various elements of the system revealed that the glanceable display contributed most to helping users remain active over time.

Similarly, "Into" [47] presents physical activity data as a virtual trip in a map-based game world. The game records steps using the accelerometer inside the mobile phone, and users "travel" on a map based on their step count. When the target destination is reached, users receive a reward—an electronic postcard from the target city. A week-long user study indicated that users perceived the application as motivating and appealing. However, its effectiveness in promoting physical activity was not formally evaluated.

A formal evaluation of the effectiveness of virtual rewards was performed by Mekler et al. [22] in the context of an image annotation task. These researchers experimentally assessed how providing points and a meaningful frame affects participants' performance. They found that the quantity of tags was determined by points, whereas quality was determined by meaningful framing. The combination of points and meaningful framing yielded the best results. It is important to note that this study examined performance in a short-term task, and was not conducted in the context of physical activity.

A similar study in the context of physical activity was conducted with Zamzee [48]—an activity meter for children that connects to a motivational game-based Web site. The Web site allows users to view their activity level, earn points, achieve goals, and select rewards, including tangible rewards in the form of gift cards. The relative effectiveness of using the activity meter with and without access to the motivational Web site (essentially a gamified vs. quantified version of the system) was evaluated over a 6-month period. Users of the gamified system showed an average increase in physical activity of 59 % compared to users of the quantified system.

In sum, prior findings regarding the effectiveness of virtual rewards for promoting physical activity were inconclusive. While some studies reported positive effects on physical activity and enthusiastic responses from users, other studies found virtual rewards to be less effective and less engaging.

2.2.3 *Social comparison*

Social comparison refers to the process of evaluating one's own abilities and opinions by comparing them to the abilities and opinions of others [49, 50]. It can also be regarded as "social traces" [18]. Social comparison is often implemented in systems aimed to promote physical activity. The underlying assumption is that users would become more physically active in order to outperform others.

For example, "Houston" [26] is an application that combines a pedometer with a mobile phone. It enables users to share their daily step count with friends, keep track of their friends' progress, and send them motivational messages. A user study revealed that sharing activity-related information resulted in social pressure to meet one's goal, beat a friend, or not have the lowest step count.

In addition, users enjoyed receiving recognition and encouragement from their friends. In contrast, users of "GoalPost" [44]—a mobile phone application that facilitates sharing physical activity information on Facebook—were hesitant to share their information with others.

"TripleBeat" [51] is a mobile phone-based system that assists runners in achieving predefined exercise goals via musical feedback and two persuasive techniques: a glanceable interface for increased personal awareness and a virtual competition. A study with runners indicated that the system's effectiveness was derived mostly from the glanceable interface rather than the virtual competition.

"Fish'n'Steps" [52] is a system designed to provide multiple levels of incentives for increasing physical activity, which could be measured by a pedometer. The incentives include, on an individual level, the growth and emotional state of the individuals' virtual pets—fish in a fish tank. Additional motivation is provided for participants in a team condition and includes competition between teams with announcements of winning teams, and comparison between the states of the fish belonging to different members of the same team. A user study revealed that when the fish avatar was not aesthetically pleasing, participants stopped looking at the tank and some even stopped using the system all together.

"The American Horsepower Challenge" [53] is a school-based competition aimed to increase physical activity in daily life. Students in the competition wear pedometers that report their daily step count information into a Web-based game. The goal of the game is to win a virtual race. Step counts from all students in the same school are aggregated to determine the school's rank in the competition. Results from a user study showed that over the course of 11 months, the average daily step count of students dropped below the baseline.

In sum, prior studies regarding the effectiveness of social comparison for promoting physical activity yielded mixed results. Some studies revealed a heightened motivation to outperform others, whereas other studies found social comparison to be less effective, sometimes even having a negative effect on physical activity. Furthermore, most systems were evaluated as a whole, when different elements of the system were confounded. As previously stated [21, 22], gamification ought to be deconstructed into individual elements for gaining a deeper understanding of its effectiveness. This is precisely what we set out to do with the development of StepByStep.

3 StepByStep

StepByStep is an accelerometer-based mobile application, intended to motivate people to incorporate more walking into their daily routine. We aimed to create a non-intrusive and effortless system, which would not require users to make any big changes to their routine or be overly engaged with the system. Therefore, StepByStep operates as a background process on Android-based mobile devices and automatically detects walking, no manual activation is required. The walking detection algorithm samples the 3-axis accelerometer 30 times per second. If the measured magnitude exceeds 30 % above the typical magnitude, a "step-up" state is defined. If the measured magnitude is <30 % below the typical magnitude, a "step-down" state is defined. A "walking" event is defined when a distinguished sequence of "step-up" and "step-down" events occur. The typical magnitude values were determined based on several field tests, in which users walked while carrying the mobile device in various positions, mainly in pocket, in hand, and in backpack. This enables the algorithm to detect walking events in all three positions. We evaluated the reliability of StepByStep by comparing the walking time measured by the application during various activities, to time measured manually using a stopwatch. The results of the reliability test are presented in Table 1. The average level of detection accuracy was 98.3 %.

Table 1 Walking time measured by StepByStep during various activities, compared to stopwatch measurements

Activity	Time measured by a stopwatch (min)	Walking time measured by StepByStep (min)	StepByStep detection accuracy level (%)
Walking	10:00	09:52	98.7
Running	10:00	09:53	98.8
Sitting	10:00	00:00	100
Driving	10:00	00:18	97.0

In addition to automatic detection of walking, StepByStep supports real-time feedback on performance and goal-setting. First, the prototype establishes the user's baseline level of walking and then automatically sets a daily walking goal reflecting a 10 % increase over the baseline level. This automatically set goal aims to keep users reasonably motivated and prevents them from setting goals that are too high and thus discouraging, or too low and thus not challenging enough. The goal-setting mechanism is explained in the "welcome" screen of the application. Real-time feedback on performance is provided in the form of a progress bar appearing on the main screen of the application, showing progress toward the goal. Furthermore, the application's icon, in the form of a walking figure, appears on the Android notifications bar at all times and changes color from yellow to green whenever walking is detected (see Fig. 1, left). When users scroll down to view their general notifications (email, text messages, etc.), current walking

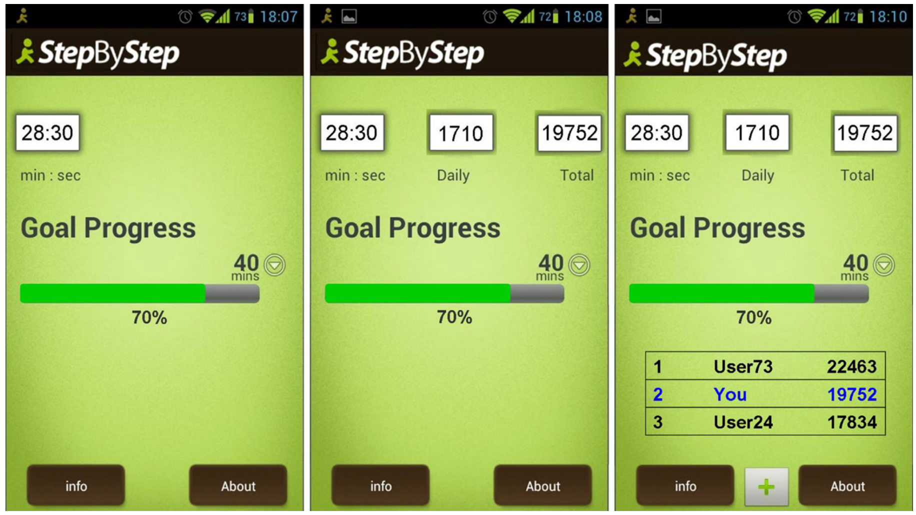

Fig. 1 Main screen of the three versions of StepByStep used in Study 2. *Left* quantified version; *middle* points version; *right* leaderboard version (presented with a "mini leaderboard," pressing the + icon opens the full leaderboard)

time is displayed as well, so there is no need to intentionally access the application in order to monitor daily progress. When users reach their daily goal, a congratulatory pop-up message appears on the screen ("Congratulations! You reached your daily goal"). The sole purpose of the message is notifying goal attainment; it completely disappears from the system once users press "OK."

StepByStep was created for research purposes, aimed to allow systematic evaluation of quantified and gamified presentation of activity-related data. Accordingly, the graphic design was a basic one, just text and a rectangular process bar over a clear background. This minimal design enabled us to evaluate the two approaches in their most basic form and ensured other variables would not become confounds.

4 Study 1: Evaluating a quantified version of StepByStep

The first study evaluated the effectiveness of a quantified version of StepByStep for promoting opportunistic physical activity, in this case walking. We tested StepByStep in the field, with users who installed the application on their personal mobile device and carried it with them throughout the day, enabling continuous tracking of their walking time. The study was conducted in 2011, over a 2-week period. During the first 3 days, the application measured walking time without engaging the users in any form of interaction ("dormant days"). This dormant mode of operation enabled us to calculate baseline walking time—the average daily walking time of each user before interacting with the system. On day 4, the application became active: setting a personalized daily goal for each user reflecting a 10 % increase from his or her baseline. Users were aware of this goal-setting mechanism. In addition, the application enabled users to monitor progress toward their goal in real time ("active days").

4.1 Participants

We recruited 40 owners of Android-based mobile devices (28 males and 12 females), between the ages of 23 and 54, to participate in the study. They were recruited through personal or professional acquaintance with the researchers and volunteered to participate. They were characterized by diverse occupations (e.g., administrator, economist, educator, engineer, human resources specialist, and real estate agent), diverse frequency of exercising, and diverse experience with Android-based mobile devices. About 93 % of participants never used an application designed to promote physical activity prior to participating in the current study.

Ten participants uninstalled the application before 2 weeks have passed. The main reason for uninstalling early was rapid battery depletion, which interrupted regular use of the mobile device. No statistically significant differences were found between those who completed the study and those who did not in regard to gender distribution, mean age, frequency of exercising, desired change in frequency of exercising, or prior experience with Android-based mobile devices. Characteristics of both groups are summarized in Table 2.

4.2 Measures

The main measure was participants' walking time during dormant days compared to active days. Walking time was automatically measured by the application using a custom algorithm, and was recorded on the mobile device in a dedicated log file. In addition, participants were asked to freely list advantages and disadvantages of the application and suggest improvements.

Table 2 Characteristics of study 1 participants who completed and did not complete the study

	Completed the study	Did not complete the study
No. of participants	30	10
Gender distribution		
Male	63.3 %	90.0 %
Female	36.7 %	10.0 %
Average age	31.8 ($SD = 6.75$)	30.6 ($SD = 4.93$)
Frequency of exercise		
Seldom	20.0 %	20.0 %
Several times a month	23.3 %	10.0 %
Once a week	26.7 %	0 %
Several times a week	20.0 %	50.0 %
Every day	10.0 %	20.0 %
Desired change in frequency of exercise		
Reduce	3.3 %	0 %
Maintain	20.0 %	40.0 %
Increase	60.0 %	50.0 %
Greatly increase	16.7 %	10.0 %
Prior experience with Android		
Novice	16.7 %	0 %
Intermediate	26.7 %	20.0 %
Advanced	30.0 %	20.0 %
Expert	26.7 %	60.0 %

4.3 Procedure

Owners of Android-based mobile devices were contacted by email and invited to participate in a study to evaluate a new application for promoting walking. The email included a link to an informed consent form and an initial background questionnaire. After answering the questionnaire, participants received an installation link for the application via email and were instructed to install it on their personal mobile device. They were asked to use the application for 2 weeks. After 2 weeks have passed, participants were contacted by email and asked to answer an online questionnaire regarding perceived advantages and disadvantages of the application. In addition, they were instructed to send their log file to the researchers. They were then thanked and debriefed.

4.4 Results and discussion

First, we report on the analysis of log files created by the StepByStep application and then on the analysis of questionnaire data.

4.4.1 Analysis of log files

In order to evaluate the effectiveness of StepByStep, we compared the average walking time during dormant and active days. During dormant days, participants walked on average 20.07 min ($SD = 14.33$), whereas during active days they walked on average 30.24 min ($SD = 14.40$). A dependent-samples t test revealed this increase was significant, $t(29) = -4.05$, $p < .001$. It seems a quantified version of StepByStep was successful in incorporating more walking into daily routines.

4.4.2 Analysis of questionnaire data

Participants were requested to list advantages and disadvantages of the application, as well as suggest improvements, in an open-ended form.

The most commonly mentioned advantage was heightened awareness toward walking throughout the day ("I always thought about walking, it was always in the back of my mind"). Additional mentioned advantages were the simplicity of the application and the continuous measurement in the background, which did not require any special effort ("It's very convenient to use, it's passive and I don't need to do anything in order for it to work").

The most commonly mentioned disadvantage was the fact that the application was perceived as ineffective ("I didn't walk more. Sure, seeing at the end of the day that I reached only half my goal made me feel bad, but the application didn't really make me walk more"). Interestingly, objective log analysis revealed the application was indeed successful in increasing daily walking time. Contradictions between users' subjective assessments and objective measurements were reported in previous studies as well [54]. An additional disadvantage was rapid battery depletion ("It depleted the battery so I had to keep a charger with me at all times"). Considering that StepByStep was created solely for research purposes, optimal battery usage was not a main concern. Furthermore, the study was conducted in 2011; since then, background processing on the Android operating system and continuous accelerometer sampling have greatly improved. As a result, battery depletion in activity monitoring applications should no longer be a critical problem.

As for improvements to the application, participants suggested the following: (1) dynamic updates to the daily goal ("It makes more sense for the goal to be constantly updated"); (2) additional metrics ("I wish it would tell me how many meters I walk and not just how much time. Perhaps also how many calories I burn"); (3) display personal history to enable reflection-on-action ("Statistics or a graph showing how much I walk everyday were missing"); and (4) inclusion of social traces, specifically

Table 3 Elements included in each version of StepByStep in study 2

Element	Version		
	Quantified	Points	Leaderboard
Continuous measurement	☑	☑	☑
Daily goal	☑	☑	☑
Real-time feedback	☑	☑	☑
Virtual rewards	☒	☑	☑
Social comparison	☒	☒	☑

competing against other users ("A table showing everyone's position would create a competition, if you want to reach a higher position you need to increase your amount of walking").

5 Study 2: Evaluating gamified versions of StepByStep

The second study evaluated the effectiveness of two game elements: virtual rewards and social comparison. We chose these elements for two reasons. First, they are considered basic elements in gamification [20], and are the two most commonly implemented elements in gamified systems [21]. Second, they are some of the most criticized elements, based on psychological research showing that many forms of rewards reduce intrinsic motivation [39, 40].

In order to empirically evaluate these elements in a granular manner, we developed two additional versions of StepByStep. The first additional version included the features of the quantified version used in study 1—continuous measurement, goal-setting, and real-time feedback, as well as receiving virtual rewards in the form of points (points version). Points were given based on walking time, so that each second of walking earned one point. If users reached their daily goal, they received a bonus—their accumulated daily points were doubled. Unlike the real-time feedback, which was reset at the end of each day, points continued to accumulate throughout the duration of the study. The second additional version included all the features of the points version, as well as a social comparison element in the form of a real-time leaderboard ranking users from first to last according to their accumulated points (leaderboard version). The elements included in each version are summarized in Table 3. The main screen of each version is displayed in Fig. 1. Unlike study 1, where goals were automatically set, in study 2, participants were prompted to define a daily goal for themselves upon the first activation of the application. Whenever the daily goal was reached 3 days in a row, a message suggesting a 10 % increase automatically appeared on the screen. This way the daily goal was constantly updated, as requested by participants in study 1.

We compared between the three versions of StepByStep in a 10-day randomized controlled study conducted during 2012. The focus of the current study was the relative effectiveness of the three versions for promoting walking, so they were all presented in "active" mode from the very first day. Furthermore, due to the risk of rapid battery depletion in some mobile devices, which interfered with regular phone use during study 1, we supplied participants in the current study with a dedicated mobile device on which the application was installed. While this required participants to carry two mobile devices at all times (their personal device and the study device), it offered several methodological advantages. First, we were no longer restricted to owners of Android-based devices, so we could recruit a larger sample. Second, all participants received the same type of device and were instructed not to install any additional applications beyond StepByStep. This ensured that everyone used StepByStep under the same conditions, which enhanced technical standardization. In order to prevent participants from forgetting the additional device, we sent two email reminders during the study, in 4-day intervals. This method of supplying participants with a second device for research purposes was successfully implemented in previous studies [e.g., 26].

5.1 Participants

Fifty-nine undergraduate communications students participated in the study (44 females and 15 males). Their age ranged from 20 to 27 ($M = 23.39$, $SD = 1.40$). They were randomly assigned to use either the quantified ($n = 18$), points ($n = 21$), or leaderboard ($n = 20$) version of StepByStep. No statistically significant differences were found between these three groups in regard to gender distribution, mean age, frequency of exercising, or desired change in frequency of exercising. All participants volunteered to participate in the study in exchange for extra course credit, which was given based solely on participation, and remained unrelated to actual walking time during the study.

5.2 Measures

The main measures reflected participants' walking behavior: daily goal, daily walking time, daily percent of goal reached, total number of days goal was reached,

Table 4 Correlation matrix between walking measures and system usability

	Daily goal	Daily walking time	Daily percent of goal reached	No. of days goal was reached	Daily no. of times accessing application	System usability
Daily goal	1.0					
Daily walking time	.63**	1.0				
Daily percent of goal reached	.15	.80**	1.0			
No. of days goal was reached	.23	.77**	.88**	1.0		
Daily no. of times accessing application	.42**	.56**	.38**	.55**	1.0	
System usability	−.11	−.09	−.04	.03	−.01	1.0

* Correlation is significant at the $p < .05$ level

** Correlation is significant at the $p < .01$ level

and daily number of times the application was accessed. All these measures were automatically calculated by the application using a custom algorithm and recorded in a local log file.

In addition, since the three versions differed in number of features, we had to verify they did not differ in ease of use. Therefore, we compared their usability using the System Usability Scale (SUS) [55], which is a ten-item scale giving a global view of subjective usability. An example item: "I felt very confident using StepByStep." All items are rated on a 5-point Likert scale ranging from (1) "strongly disagree" to (5) "strongly agree." The scale yields a single score, ranging from 0 to 100, representing a composite measure of the overall usability.

Subjective experiences while using StepByStep were evaluated with interviews. Participants were asked to express their opinion regarding the application in general, and the points and leaderboard elements in particular.

5.3 Procedure

Students were contacted by email, inviting them to participate in a study to evaluate a new application for promoting routine walking. The email included a link to an informed consent form and an initial background questionnaire. After answering the initial questionnaire, each participant met with a research assistant, who supplied him or her with an Android-based Galaxy S mobile device, on which one of the three versions of StepByStep was installed. The research assistant provided explanations regarding the specific version (determined by random assignment), and instructed participants to carry the device with them at all times. Participants then used the application for 10 days. After 10 days have passed, log files were collected, participants were asked to fill out the SUS, and a 15-min semi-structured interview was conducted with each participant. At the end of the study, participants were thanked and debriefed.

Table 5 Means and standard deviations of dependent measures as a function of StepByStep version

Measure	Quantified		Points		Leaderboard	
	M	*SD*	*M*	*SD*	*M*	*SD*
Daily goal (min)	50.58	13.34	56.30	18.68	54.21	17.04
Daily walking time (min)	43.63	20.36	49.94	24.78	40.85	18.40
Percent of daily goal reached	79.09	29.84	83.85	33.03	71.45	28.72
No. of days goal was reached	3.06	2.04	4.52	3.01	3.10	2.38
Daily no. of times accessing application	1.56	1.40	2.21	2.31	2.40	2.02
System usability (0–100)	84.72	10.00	84.29	6.90	82.63	8.49

5.4 Results and discussion

We report on the analysis of log files first, then on the usability analysis, and lastly on the analysis of user interviews.

5.4.1 Analysis of log files

First, we calculated the correlations between the various walking measures. They were mostly correlated with one another. See Table 4 for a full correlation matrix.

In order to evaluate the relative effectiveness of the three versions of StepByStep, we conducted a multivariate analysis of variance (MANOVA) with version (quantified, points, leaderboard) as the independent variable, and the various walking measures as the dependent variables. By use of Wilks' criterion, the analysis revealed no significant effect for version, F (10, 104) = 1.49, $p > .05$. Means and standard deviations are presented in Table 5. In other words, the three versions of StepByStep yielded similar walking behavior. These results were replicated when the analysis included additional control variables such as gender and frequency of exercising prior to participating in the study.

Previously, Mekler et al. [22] found that points improve performance over a control condition where no points were given. The contradiction between these findings and the findings of the current study could be explained in several ways. The study by Mekler et al. [22] was conducted in the context of image annotation, whereas the current study was conducted in the context of physical activity. Context is considered to be an essential antecedent for engaging gamification [21]. Furthermore, the study by Mekler et al. [22] examined short-term effects of points (a single sitting), whereas the current study examined the effects of points over several days. Previous studies showed that the results of gamification may be short term, potentially caused due to a novelty effect [21].

Interestingly, several differences did emerge between the three versions of StepByStep. We found a significant positive correlation between daily goal and daily walking time ($r_p = .63$, $p < .001$), indicating that a higher goal was associated with greater walking time. This is a well-documented correlation, previously established in clinical interventions involving pedometers [31]. However, this correlation existed only for the quantified ($r_p = .86$, $p < .001$) and points ($r_p = .77$, $p < .001$) versions, but not for the leaderboard version ($r_p = .24$, $p > .05$). Similarly, we found a significant positive correlation between daily number of times the application was accessed and daily walking time ($r_p = .56$, $p < .001$), indicating that a higher frequency of accessing the application was associated with greater walking time. This correlation was found in previous studies as well [e.g., 53]. Again, the correlation existed only for the quantified ($r_p = .54$, $p < .05$) and points ($r_p = .79$, $p < .001$) versions, but not for the leaderboard version ($r_p = .28$, $p > .05$). In other words, interaction with StepByStep was related to walking only among users of the quantified and points versions. Perhaps, walking leads to interaction with the application, or perhaps, interaction with the application leads to walking. In any case, level of interaction fluctuated according to the level of activity. In contrast, among users of the leaderboard version, level of interaction did not fluctuate according to the level of activity. We believe that users of the leaderboard version knew they were constantly being compared to others. Therefore, their interaction with the application was determined by the level of interest in their ranking in the leaderboard. Those who were interested, interacted with the application more than those who were not interested, regardless of how much they actually walked.

5.4.2 *Usability analysis*

In addition to actual walking behavior, we also compared the perceived usability of the three StepByStep versions. A one-way ANOVA, with version (quantified, points, leaderboard) as the independent variable, and the SUS score [55] as the dependent variable, revealed no significant effect for version, $F(2, 58) = .33$, $p > .05$. Means and standard deviations are presented in Table 5. In other words, the three versions of StepByStep were perceived as equally usable, which means the difference in number of features did not impact the ease of use. These results were replicated when the analysis included additional control variables such as gender and frequency of exercising prior to participating in the study. Furthermore, usability was unrelated to actual walking behavior (see Table 4).

5.4.3 *Interview analysis*

First, participants were asked regarding their general opinion of StepByStep. As in study 1, awareness toward walking increased ("It made me think about how much I don't walk during the day, I suddenly realized that"), and additional metrics were requested, mainly distance, speed, and number of calories burnt.

Participants who used the points version were then asked regarding their opinion of the points element, while participants who used the leaderboard version were asked regarding both the points and leaderboard elements. It appeared that participants understood the points allocation mechanism and distinguished between points and real-time feedback. Nonetheless, points were perceived by most participants as meaningless; they were more interested in their walking time than in the number of accumulated points ("So I have 33,000 points. What does it mean? It doesn't mean anything, it's just a number." "I knew about the bonus if you meet your goal, but I didn't really care how many points I had, I was more interested in how much I walked").

Interest in the leaderboard seemed to reflect interpersonal differences, as some participants expressed great interest in their current ranking ("I kept checking the application to see where I am and who is in front of me"), whereas others expressed little to no interest in their current ranking ("I didn't really care about comparing my achievement to the achievement of others"). These qualitative findings support our earlier explanation for the quantitative findings—interaction with the application was not correlated with walking time among users of the leaderboard version due to interpersonal differences in the level of interest in current ranking. Similar interpersonal differences were found in [56].

6 General discussion

Numerous persuasive systems aimed to promote physical activity have been introduced in recent years. These systems measure activity-related parameters and present the measured data to the user in various ways. Quantified systems display the data in a concrete numerical form, facilitating reflection on one's level of physical activity. Gamified systems add game-like elements on top of the numeric data. Presumably, gamification turns physical activity into a more enjoyable experience, thus motivates users to engage in physical activity. However, this assumption has yet to be empirically established, and has been challenged by several authors [24, 41].

We set out to evaluate the relative effectiveness of the quantified and gamified approaches to presenting physical activity data. We evaluated two commonly used gamification elements: virtual rewards and social comparison, in the context of opportunistic physical activity. We developed a research prototype called StepByStep, which is an accelerometer-based mobile application intended to motivate people to walk more during their daily routine. The application operates as a background process on Android-based mobile devices and automatically detects walking events. We created different versions of this application, each includes different elements, and evaluated them in two field studies.

Study 1 evaluated a quantified version of StepByStep, offering continuous measurement of walking time, a daily walking goal, and real-time feedback on progress toward this goal. Results indicated that daily walking time while interacting with the application was significantly higher than baseline walking time. Interestingly, participants noticed their awareness toward walking had increased, but they did not notice the actual increase in their walking time. They believed the application failed to motivate them and suggested adding features in order to enhance its effectiveness, mainly dynamic updates to the daily goal, additional metrics, and social comparison.

Study 2 focused on two commonly used game elements: virtual rewards and social comparison. We compared between three versions of StepByStep: a quantified version similar to the one used in study 1 (serving as a control version here), a virtual rewards version awarding users with points based on walking time and goal attainment, and a social comparison version ranking users in a leaderboard according to their accumulated points. All three versions yielded similar walking behavior as well as similar usability ratings. Interviews at the end of study 2 indicated that points were perceived as meaningless by most participants, who were more interested in actual walking time than in an arbitrary number. This finding could explain why the points version yielded similar results as the quantified version. Interviews also indicated that rank in the leaderboard was important to some participants, but less important to other participants. These interpersonal differences could explain why, on average, the leaderboard version yielded similar results as the quantified version.

Interpersonal differences in attitudes toward a leaderboard were also observed by Xu et al. [56]. Moreover, a literature review of empirical studies on gamification [21] found that, in general, different people interact with gamified systems in different manners, and for different reasons. As a result, the effects of gamification are likely to vary. Perhaps, gamified systems could be more effective for promoting physical activity if tailored to the personal characteristics and preferences of each user. This is a promising avenue for further research. For example, it might be interesting to examine the interplay between gamification and competitiveness, or the level of cognitive dissonance, as suggested by [57].

Nicholson [41] predicted that game elements that are external to the underlying activity, such as meaningless points and leaderboards, would reduce motivation to perform the activity in the long run. This was not the case in study 2. Implementation of game elements did not result in less walking compared to the quantified version. It is likely, however, that the study ended before any potential negative implications of gamification could emerge. Xu et al. [53] evaluated the effectiveness of a gamified system over the course of 11 months. As predicted by Nicholson, they found that average levels of physical activity dropped below the baseline.

According to Nicholson [41], the potential negative consequences of gamification could be avoided by creating a gamification system that is meaningful to the users. He emphasizes the importance of information—the system should provide information instead of presenting a score. This way, users could create their own games and goals. When users are given information and control over goals, they are more likely to find internal meaningful connections to the underlying activity and thus continue performing it over time.

When physical activity is concerned, information is derived from continuous measurement and monitoring of physical and physiological parameters. A large-scale review of interventions aimed to promote physical activity found that interventions prompting participants to self-monitor their activities were most effective [37]. Technology facilitates self-monitoring of physical activity, and hence reflection, both reflection-in-action and reflection-on-action. Integrating elements that facilitate reflection on physical activity with regular use of mobile devices could be particularly beneficial. In StepByStep, an icon of a walking figure constantly appears on the device's notifications bar, changing color when walking is detected. In addition, the current walking time is displayed whenever the notifications bar is scrolled down. A special widget appearing on the home screen of the device could also serve as a constant trigger for reflection.

Since measurement and monitoring of activity consistently emerge as effective techniques for promoting opportunistic physical activity, we advise designers to focus on developing novel ubiquitous measures to facilitate reflection. This recommendation is especially relevant when taking into account recent advancements in sensing technology [58]. The potential for new ubiquitous measures is vast, from already common measures such as movement and location, to physiological measures such as pulse and skin conductivity that can even imply on emotional state [59]. Moreover, sensor-based data analysis and mobile battery capacity are expected to improve in the near future, turning continuous data collection to a standard technique in everyday mobile use. Coupling this vast range of measures with goal-setting and real-time feedback holds the potential to facilitate reflection, and hopefully behavior change as well.

Development of novel ubiquitous measures could also enable "meaningful measurement." Nicholson [41] suggested involving users in the creation or customization of a gamification system, so it would correspond with their own interests, for example, allowing users to create their own tools to track different aspects of the activity, or to create their own leveling systems and achievements. Following Nicholson, we suggest involving users in the design of the measurement itself, so even the presented metrics would correspond with their interests. As one can recall, participants in both study 1 and study 2 requested that additional metrics would be provided by StepByStep. Users could benefit from the quantification of various meaningful aspects of their lives; all they need is suitable measurement and monitoring tools. For example, some users could be interested in periods of continuous walking, some users could be interested in the number of stairs they climb every day, some users could be interested in the difference between current and past performance, and some users could be interested in the difference between their own performance and the performance of a professional athlete. Enabling users to measure and thereby reflect on a parameter that is particularly meaningful to them could greatly increase physical activity, compared to standard non-personalized measurement. We plan to empirically test the effectiveness of meaningful or user-centered measurement in future work.

This idea corresponds with work by Baumer et al. [60] on open-ended social awareness. These researchers noted that many persuasive health systems tend to be prescriptive, telling the user either explicitly or implicitly what to do. In contrast, they suggest open-ended social awareness as a central design principle, making users aware of both others' and their own health decisions. Open-ended systems allow users flexibility in defining what counts as "health." On the other hand, the openness of the system has several shortcomings, namely reduced motivation to report routine activities and potentially inaccurate reports. A possible solution might be open-ended selection of measurement, offering users flexibility in defining which parameters would be measured, followed by automatic measurement and monitoring of the selected parameters.

The present research includes several limitations that should be addressed. First, we focused only on virtual rewards and social comparison, and deliberately implemented them in a very simplistic manner. However, gamification as a design approach includes additional elements. As mentioned earlier, Deterding et al. [20] classified the various game design elements into five different levels, and we focused only on the first level. Before we could draw any definitive conclusions regarding the relative effectiveness of quantified versus gamified presentation of activity data, the effects of other gamification elements should be evaluated in a systematic manner.

Moreover, when considering the effectiveness of gamification, one must consider that the motivational affordances of game elements are situated—their motivational salience is determined by their situational usage and meaning [40]. Thus, to better understand whether virtual rewards and social comparison are effective, future work should not only compare game elements, but also evaluate the usage situation. Perhaps in a different situation virtual rewards would be more meaningful, and social comparison would be more engaging. For example, it could be interesting to examine the motivational affordances of game elements implemented in communal systems, like the one described in [61].

Secondly, we focused on a specific type of physical activity, activity that can be incorporated into everyday routines, defined by Consolvo et al. [26] as opportunistic physical activity. However, some people prefer setting aside special time in their daily schedule for performing physical activity, defined by Consolvo et al. [26] as structured exercise. Context was found to be an essential antecedent for engaging gamification [21]. Therefore, future studies should examine the relative effectiveness of quantified versus gamified systems in regard to structured exercise as well.

Third, study 2 yielded statistically nonsignificant results that are hard to interpret. While we cannot completely rule out the occurrence of a type II error, the fact that other studies yielded similar results adds credence to our findings. For example, Liu et al. [62] evaluated the effectiveness of two gamified systems: a mobile crowdsourcing application designed for image-based social search across languages, and a persuasive application for motivating users to reduce CO_2 emission. In both cases, the effect of gamification incentives on behavior did not reach statistical significance. The authors concluded that the main functionalities of the system have a greater impact than the additional game elements. We too believe that the main functionality of StepByStep—the continuous measurement of walking activity—is the most important and effective component of the system.

An additional limitation stems from the relatively short-term nature of the current evaluation. Participants in study 1 used StepByStep for 14 days, and participants in study 2 used the application for 10 days. We cannot simply assume that systems offering quantified data, virtual rewards, and social comparison would remain equally effective in the long term as well. It is possible that as time passes, the gamified versions would reduce motivation for walking compared to the quantified version, as predicted by Nicholson [41] based on motivation theory. It is also possible that as time passes, the gamified versions would increase motivation for walking compared to the quantified version. This option seems less likely, because interest and engagement with new technologies are usually at their highest immediately after the introduction of a technology, gradually dissipating over time when the novelty wears off [21, 53]. Accordingly, most persuasive systems are expected to change behavior sooner rather than later, while interest and engagement are at their highest [14]. Since virtual rewards and social comparison failed to confer any additional benefits compared to a simple quantified application while excitement was at its highest, it is hard to believe such benefits would emerge once excitement has decreased.

Lastly, the participants in study 2 were all undergraduate students. While this relative homogeneity served to reduce variance in an experimental study, it should be acknowledged that this population may not be representative of the general population in regard to the level of physical activity, general health considerations, and daily routines. The findings of study 2 should be replicated with a more heterogeneous sample.

Despite the above-mentioned limitations, our work offers several contributions. First, the results of study 2 indicated that what is popular is not necessarily more effective, or that effectiveness of systems with multiple elements may stem from different reasons than those commonly presumed. Researchers and practitioners cannot simply assume that gamification would always be an effective approach for promoting opportunistic physical activity. Instead, the effectiveness of game elements must be systematically evaluated, with the various elements separated and compared. We believe that research prototypes like StepByStep, which enable comparisons between different elements of a single system, serve as a promising tool for achieving this goal.

Secondly, it seems clear that continuous measurement, especially coupled with goal-setting and real-time feedback, is an effective technique for motivating opportunistic physical activity. Therefore, we advise designers of persuasive technologies to focus on developing novel ubiquitous measures, ones that are meaningful to users, and facilitate reflection on various parameters of physical activity.

7 Conclusion

Gamification, the use of game design elements in non-game contexts, is often implemented in persuasive systems aimed to promote physical activity. Presumably, gamification makes physical activity more enjoyable, thus motivates users to become more active. However, due to contradicting findings from prior studies, and lack of systematic research in the field, this assumption cannot be supported by the existing literature. We set out to further our understanding of gamification by developing a research prototype, called "StepByStep," specifically for the purpose of deconstructing gamification into separate elements. We created three versions of StepByStep. The first version was a quantified version, offering continuous measurement of walking time, goal-setting, and feedback on progress toward the goal. The quantified version significantly increased walking time over baseline level. The second version of StepByStep was a gamified version, adding a simple virtual rewards element in the form of points, accumulated based on walking time and goal attainment. The third version was gamified as well, adding a social comparison element by ranking users in a leaderboard according to accumulated points. The two gamified versions were only as effective as the quantified version for promoting routine walking. Interviews with users revealed that the quantified elements facilitated reflection on activity and were greatly appreciated. In contrast, points were considered meaningless by most users, whereas interest in the leaderboard varied according to interpersonal differences. Based on our findings, we advise designers to focus on developing novel ubiquitous measures, intended to facilitate reflection on meaningful aspects of physical activity. We advise researchers to conduct systematic evaluations of quantified and gamified elements, for better understanding their situated motivational affordances.

Acknowledgments We would like to thank Meytal Abo, Michal Gilon-Yanai, Orad Weisberg, Dr. Yaniv Kanat-Mymon, Dr. Guy Hoffman, and Dr. Guy Doron for their assistance with conceptualizing StepByStep; Shai Yagur, Roy Ofer, Omri Baumer, and Yair Halevi for their assistance with developing the different versions of the StepByStep prototype; Dr. Ofer Bergman for his suggestions on an early draft of the manuscript, and two anonymous reviewers for their constructive comments that helped improve the final paper.

References

1. Fogg B (2002) Persuasive technology: using computers to change what we think and do. Morgan Kaufmann, San Francisco
2. Oinas-Kukkonen H, Harjumaa M (2008) A systematic framework for designing and evaluating persuasive systems. In: Proceedings of the 3rd international conference on persuasive technology (PERSUASIVE 2008). Springer Berlin Heidelberg, pp 164–176. doi:10.1007/978-3-540-68504-3_15
3. Ertin E, Stohs N, Kumar S, Raij A, al'Absi M, Shah S (2011) AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In: Proceedings of the 9th ACM conference on embedded networked sensor systems (SenSys 2011). ACM, New York, pp 274–287. doi:10.1145/2070942.2070970
4. Gay V, Leijdekkers P, Barin E (2009) A mobile rehabilitation application for the remote monitoring of cardiac patients after a heart attack or a coronary bypass surgery. In: Proceedings of the 2nd international conference on pervasive technologies related to assistive environments (PETRA 2009). ACM, New York, Article no. 21. doi:10.1145/1579114.1579135
5. Grimes A, Kantroo V, Grinter RE (2010) Let's play! mobile health games for adults. In: Proceeding of the 12th ACM international conference on ubiquitous computing (Ubicomp 2010). ACM, New York, pp 241–250. doi:10.1145/1864349.1864370
6. Tsai CC, Lee G, Raab F, Norman GJ, Sohn T, Griswold WG, Patrick K (2007) Usability and feasibility of PmEB: a mobile phone application for monitoring real time caloric balance. Mob Netw Appl 12:173–184. doi:10.1007/s11036-007-0014-4
7. Biddle SJH, Asare M (2011) Physical activity and mental health in children and adolescents: a review of reviews. Br J Sports Med 45:886–895. doi:10.1136/bjsports-2011-090185
8. Fogelholm M (2010) Physical activity, fitness and fatness: relations to mortality, morbidity and disease risk factors: a systematic review. Obes Rev 11:202–221. doi:10.1111/j.1467-789X.2009.00653.x
9. Janssen I, LeBlanc AG (2010) Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys 7:1–16. doi:10.1186/1479-5868-7-40
10. Vogel T, Brechat P-H, Leprêtre P-M, Kaltenbach G, Berthel M, Lonsdorfer J (2009) Health benefits of physical activity in older patients: a review. Int J Clin Pract 63:303–320. doi:10.1111/j.1742-1241.2008.01957.x
11. Bauman A, Ainsworth BE, Sallis JF, Hagströmer M, Craig CL, Bull FC, Pratt M, Venugopal K, Chau J, Sjöström M, the IPS Group (2011) The descriptive epidemiology of sitting. Am J Prev Med 41:228–235. doi:10.1016/j.amepre.2011.05.003
12. Lakdawalla D, Philipson T (2009) The growth of obesity and technological change. Econ Hum Biol 7:283–293. doi:10.1016/j.ehb.2009.08.001
13. Koch S, Marschollek M, Wolf KH, Plischke M, Haux R (2009) On health-enabling and ambient-assistive technologies. Methods Inf Med 48:29–37. doi:10.33414/ME9136
14. Klasnja P, Consolvo S, Pratt W (2011) How to evaluate technologies for health behavior change in HCI research. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2011). ACM, New York, pp 3063–3072. doi:10.1145/1978942.1979396
15. Ahtinen A, Mattila E, Vaatanen A, Hynninen L, Salminen J, Koskinen E, Laine K (2009) User experience of mobile wellness applications in health promotion: user study of wellness diary, mobile coach and selfrelax. In: Proceeding of the 3rd international conference on pervasive computing technologies for healthcare (Pervasive Health 2009). IEEE, pp 1–8. doi:10.4108/ICST.PERVASIVEHEALTH2009.6007
16. Dantzig S, Geleijnse G, Halteren AT (2013) Toward a persuasive mobile application to reduce sedentary behavior. Pers Ubiquit Comput 17(6):1237–1246. doi:10.1007/s00779-012-0588-0
17. Li I, Dey A, Forlizzi J (2010) A stage-based model of personal informatics systems. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2010). ACM, New York, pp 557–566. doi:10.1145/1753326.1753409
18. Ploderer B, Reitberger W, Oinas-Kukkonen H, van Gemert-Pijnen J (2014) Social interaction and reflection for behaviour change. Pers Ubiquit Comput (this issue). doi:10.1007/s00779-014-0779-y
19. Campbell T, Ngo B, Fogarty J (2008) Game design principles in everyday fitness applications. In: Proceedings of the 2008 ACM conference on computer supported cooperative work (CSCW 2008). ACM, New York, pp 249–252. doi:10.1145/1460563.1460603
20. Deterding S, Dixon D, Khaled R, Nacke L (2011) From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th international academic MindTrek conference: envisioning future media environments (MindTrek 2011). ACM, New York, pp 9–15. doi:10.1145/2181037.2181040
21. Hamari J, Koivisto J, Sarsa H (2014) Does gamification work? A literature review of empirical studies on gamification. In: Proceedings of the 47th Hawaii international conference on system sciences
22. Mekler ED, Brühlmann F, Opwis K, Tuch AN (2013) Disassembling gamification: the effects of points and meaning on user motivation and performance. In: CHI '13 extended abstracts on human factors in computing systems (CHI EA 2013). ACM, New York, pp 1137–1142. doi:10.1145/2468356.2468559
23. Oduor M, Alahäivälä T, Oinas-Kukkonen H (2014) Persuasive software design patterns for social influence. Pers Ubiquit Comput (this issue). doi:10.1007/s00779-014-0778-z
24. Deterding S (2013) Skill atoms as design lenses for user-centered gameful design. In: Proceedings of CHI '13 workshop "Designing gamification"
25. Deterding S, Björk SL, Nacke LE, Dixon D, Lawley E (2013) Designing gamification: creating gameful and playful experiences. In CHI'13 extended abstracts on human factors in computing systems (CHI EA 13). ACM, New York, pp 3263–3266. doi:10.1145/2468356.2479662
26. Consolvo S, Everitt K, Smith I, Landay JA (2006) Design requirements for technologies that encourage physical activity. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2006). ACM, New York, pp 457–466. doi:10.1145/1124772.1124840
27. Murphy MH, Blair SN, Murtagh EM (2009) Accumulated versus continuous exercise for health benefit. Sports Med 39:29–43. doi:10.2165/00007256-200939010-00003
28. Woolf-May K, Kearney EM, Owen A, Jones DW, Davison RCR, Bird SR (1999) The efficacy of accumulated short bouts versus single daily bouts of brisk walking in improving aerobic fitness and blood lipid profiles. Health Educ Res 14:803–815. doi:10.1093/her/14.6.803
29. Schön DA (1983) The reflective practitioner: how professionals think in action. Basic Books, New York
30. Tudor-Locke C (2002) Taking steps toward increased physical activity: using pedometers to measure and motivate. President's Council on Physical Fitness and Sports: Research Digest Series 3, No. 17
31. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, Stave CD, Olkin I, Sirard JR (2007) Using Pedometers to increase physical activity and improve health. JAMA 298:2296–2304. doi:10.1001/jama.298.19.2296

32. Richardson CR, Newton TL, Abraham JJ, Sen A, Jimbo M, Swartz AM (2008) A meta-analysis of pedometer-based walking interventions and weight loss. Ann Fam Med 6:69–77. doi:10.1370/afm.761
33. Tudor-Locke C, Lutes L (2009) Why do pedometers work? Sports Med 39:981–993. doi:10.2165/11319600-000000000-00000
34. Locke EA, Latham GP (2002) Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am Psychol 57:705–717. doi:10.1037/0003-066X.57.9.705
35. Consolvo S, Klasnja P, McDonald DW, Landay JA (2009) Goal-setting considerations for persuasive technologies that encourage physical activity. In: Proceedings of the 4th international conference on persuasive technology (Persuasive 2009). ACM, New York, Article no. 8. doi:10.1145/1541948.1541960
36. Ashford S, Edmunds J, French DP (2010) What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. Br J Health Psychol 15:265–288. doi:10.1348/135910709X461752
37. Michie S, Abraham C, Whittington C, McAteer J, Gupta S (2009) Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 28:690–701. doi:10.1037/a0016136
38. Deci EL, Ryan RM (1985) Intrinsic motivation and self-determination in human development. Plenum, New York
39. Deci E, Koestner R, Ryan R (2001) Extrinsic rewards and intrinsic motivation in education: reconsidered once again. Rev Educ Res 71:1–27. doi:10.3102/00346543071001001
40. Deterding S (2011) Situated motivational affordances of game elements: a conceptual model. In: Proceedings of CHI '11 workshop "Gamification: using game design elements in non-gaming contexts"
41. Nicholson S (2012) A user-centered theoretical framework for meaningful gamification. In: Proceedings of Games + Learning + Society 8.0 (GLS 8.0)
42. Lindqvist J, Cranshaw J, Wiese J, Hong J, Zimmerman J (2011) I'm the mayor of my house: examining why people use foursquare: a social-driven location sharing application. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2011). ACM, New York, pp 2409–2418. doi:10.1145/1978942.1979295
43. Antin J, Churchill EF (2011) Badges in social media: a social psychological perspective. In: Proceedings of CHI '11 workshop "Gamification: using game design elements in non-gaming contexts"
44. Munson SA, Consolvo S (2012) Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In: Proceedings of the 6th international conference on pervasive computing technologies for healthcare (PervasiveHealth 2012). IEEE, pp 25–32
45. Consolvo S, McDonald DW, Toscos T, Chen MY, Froehlich J, Harrison B, Klasnja P, LaMarca A, LeGrand L, Libby R, Smith I, Landay JA (2008) Activity sensing in the wild: a field trial of ubifit garden. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2008). ACM, New York, pp 1797–1806. doi:10.1145/1357054.1357335
46. Consolvo S, Klasnja P, McDonald DW, Avrahami D, Froehlich J, LeGrand L, Libby R, Mosher K, Landay JA (2008) Flowers or a robot army? Encouraging awareness and activity with personal, mobile displays. In: Proceedings of the 10th international conference on ubiquitous computing (UbiComp 2008). ACM, New York, pp 54–63. doi:10.1145/1409635.1409644
47. Ahtinen A, Huuskonen P, Hakkila J (2010) Let's all get up and walk to the north pole: design and evaluation of a mobile wellness application. In: Proceeding of the 6th nordic conference on human–computer interaction (NordiCHI 2010). ACM, New York, pp 3-12. doi:10.1145/1868914.1868920
48. HopeLab (2012) Study shows technology gets kids moving 59% more.https://www.zamzee.com/press/9_Research_Press_Release_120927.pdf. Accessed 31 March 2014
49. Festinger L (1954) A theory of social comparison processes. Hum Relat 108:117–140. doi:10.1177/001872675400700202
50. Kruglanski AW, Mayseless O (1990) Classic and current social comparison research: expanding the perspective. Psychol Bull 108:195–208. doi:10.1037/0033-2909.108.2.195
51. de Oliveira R, Oliver N (2008) TripleBeat: enhancing exercise performance with persuasion. In: Proceedings of the 10th international conference on human computer interaction with mobile devices and services (MobileHCI 2008). ACM, New York, pp 255–264. doi:10.1145/1409240.1409268
52. Lin JJ, Mamykina L, Lindtner S, Delajoux G, Strub HB (2006) Fish'n'Steps: encouraging physical activity with an interactive computer game. In: Proceedings of the 8th international conference on ubiquitous computing (UbiComp 2006). Springer-Verlag Berlin, Heidelberg, pp 261–278. doi:10.1007/11853565_16
53. Xu Y, Shehan Poole E, Miller AD, Eiriksdottir E, Catrambone R, Mynatt ED (2012) Designing pervasive health games for sustainability, adaptability and sociability. In: Proceedings of the international conference on the foundations of digital games (FDG 2012). ACM, New York, pp 49–56. doi:10.1145/2282338.2282352
54. Edwards HM, McDonald S, Zhao T (2011) Exploring teenagers' motivation to exercise through technology probes. Proceedings of the 25th BCS conference on human–computer interaction (BCS-HCI 2011). British Computer Society, Swinton, pp 104–113
55. Brooke J (1996) SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL (eds) Usability evaluation in industry. Taylor & Francis, London, pp 189–194
56. Xu Y, Poole ES, Miller AD, Eiriksdottir E, Kestranek D, Catrambone R, Mynatt ED (2012) This is not a one-horse race: understanding player types in multiplayer pervasive health games for youth. In: Proceedings of the ACM 2012 conference on computer supported cooperative work (CSCW 2012). ACM, New York, pp 843–852. doi:10.1145/2145204.2145330
57. Wiafe I, Nakata K, Gulliver S (2014) Categorizing users in behavior change support systems based on cognitive dissonance. Pers Ubiquit Comput. doi:10.1007/s00779-014-0782-3
58. Martin H, Bernardos AM, Iglesias J, Casar JR (2013) Activity logging using lightweight classification techniques in mobile devices. Pers Ubiquit Comput 17:675–695. doi:10.1007/s00779-012-0515-4
59. Ivonin L, Chang H-M, Chen W, Rauterberg M (2013) Unconscious emotions: quantifying and logging something we are not aware of. Pers Ubiquit Comput 17:663–673. doi:10.1007/s00779-012-0514-5
60. Baumer EPS, Katz SJ, Freeman JE, Adams P, Gonzales AL, Pollak J, Retelny D, Niederdeppe J, Olson CM, Gay GK (2012) Prescriptive persuasion and open-ended social awareness: expanding the design space of mobile health. In: Proceedings of the ACM 2012 conference on computer supported cooperative work (CSCW 2012). ACM, New York, pp 475–484. doi:10.1145/2145204.2145279
61. Parker (2014) Reflection-through-performance: personal implications of documenting health behaviors for the collective. Pers Ubiquit Comput. doi:10.1007/s00779-014-0780-5
62. Liu Y, Alexandrova T, Nakajima T (2011) Gamifying intelligent environments. In: Proceedings of the 2011 international ACM workshop on Ubiquitous Meta User interfaces (UBI-MUI 2011). ACM, New York, pp 7–12. doi:10.1145/2072652.2072655