





Article

Health Promotion through Monetary Incentives: Evaluating the Impact of Different Reinforcement Schedules on Engagement Levels with a mHealth App

Raoul Nuijten ^{1,*}, Pieter Van Gorp ¹, Alireza Khanshan ², Pascale Le Blanc ¹, Astrid Kemperman ³,
Pauline van den Berg ³ and Monique Simons ⁴

- ¹ Department of Industrial Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands; p.m.e.v.gorp@tue.nl (P.V.G.); p.m.le.blanc@tue.nl (P.L.B.)
- ² Department of Industrial Design, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands; a.khanshan@tue.nl
- ³ Department of the Built Environment, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands; a.d.a.m.kemperman@tue.nl (A.K.); p.e.w.v.d.berg@tue.nl (P.v.d.B.)
- ⁴ Department of Social Sciences, Wageningen University & Research, 6708 PB Wageningen, The Netherlands; monique.simons@wur.nl
- * Correspondence: r.c.y.nuijten@tue.nl



check for updates

Citation: Nuijten, R.; Van Gorp, P.; Khanshan, A.; Le Blanc, P.; Kemperman, A.; van den Berg, P.; Simons, M. Health Promotion through Monetary Incentives: Evaluating the Impact of Different Reinforcement Schedules on Engagement Levels with a mHealth App. *Electronics* **2021**, *10*, 2935. <https://doi.org/10.3390/electronics10232935>

Academic Editors: José García Moros and Christos J. Bouras

Received: 15 October 2021

Accepted: 23 November 2021

Published: 26 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Financial rewards can be employed in mHealth apps to effectively promote health behaviors. However, the optimal reinforcement schedule—with a high impact, but relatively low costs—remains unclear. Methods: We evaluated the impact of different reinforcement schedules on engagement levels with a mHealth app in a six-week, three-arm randomized intervention trial, while taking into account personality differences. Participants (i.e., university staff and students, N = 61) were awarded virtual points for performing health-related activities. Their performance was displayed via a dashboard, leaderboard, and newsfeed. Additionally, participants could win financial rewards. These rewards were distributed using a fixed schedule in the first study arm, and a variable schedule in the other arms. Furthermore, payouts were immediate in the first two arms, whereas payouts in the third arm were delayed. Results: All three reinforcement schedules had a similar impact on user engagement, although the variable schedule with immediate payouts was reported to have the lowest cost per participant. Additionally, the impact of financial rewards was affected by personal characteristics. Especially, individuals that were triggered by the rewards had a greater ability to defer gratification. Conclusion: When employing financial rewards in mHealth apps, variable reinforcement schedules with immediate payouts are preferred from the perspective of cost and impact.

Keywords: financial reward; monetary incentive; reinforcement schedule; gamification; lootbox; health promotion; physical activity; dietary intake

1. Introduction

1.1. Research Context

Research suggests that mHealth apps can be employed to effectively promote physical activity and improve dietary intake, particularly if these tools employ evidence-based intervention strategies [1]. Promising results have been obtained by using gamification techniques as intervention strategies [1–3]. Gamification is a set of motivational techniques that employ game mechanics outside of game contexts, in order to promote participation, engagement, and loyalty [4,5]. Rewards are fundamental components of gamified applications [6]. Typically, rewards are subdivided into two categories: (1) non-financial (e.g., virtual) rewards, and (2) financial (i.e., monetary) rewards [3,6]. Monetary incentives have proven to foster health behavior change. In a meta-analysis, monetary incentives were reported to increase exercise frequency in eight of the eleven included studies [7].

Furthermore, a review of randomized controlled trials found that monetary incentives were reported to have a positive effect on food purchases, food consumption, and weight loss, in all four included studies [8].

However, for an intervention organizer, prolonged provision of monetary incentives may not be sustainable in the long run. By designing alternative schedules of reinforcement (e.g., variable schedules, such as lotteries), practitioners can determine whether every response is followed by reinforcement, or whether only some responses are followed by reinforcement [9]. Additionally, schedules that do not reinforce every user response (e.g., lotteries) are effective to promote health behaviors. For example, a study (N = 1299) even found that, to promote the completion of a health risk assessment, it is more effective to distribute financial incentives through a lottery (i.e., prize value of USD 100, with an expected value of USD 25), than to pay end users a fixed fee of USD 25 for completing the assessment [10]. Furthermore, it was found that financial rewards (i.e., EUR 100, or a voucher for a luxury family-vacation) distributed through lotteries are especially promising to increase physical activity levels and gym attendance in overweight adults (N = 163) [11]. Nevertheless, the relationship between different reinforcement schedules and their implications for the overall mHealth campaign costs remain unclear.

1.2. Theoretical Background

When designing a reinforcement schedule, there are some important dimensions to consider. For example, the schedule can either be an interval-based schedule, or a ratio-based schedule. In an interval schedule, reinforcement is based on a number of units of time, whereas, in a ratio schedule, reinforcement is given for a number of behaviors [9]. Besides, the schedule can either be fixed, or variable. A fixed schedule, executes every x time units (i.e., interval-based) or n behaviors (i.e., ratio-based), whereas in a variable schedule, reinforcement is stochastically related to the amount of time units that have elapsed, or the amount of behaviors that have been performed [9]. Summarizing, these two dimensions can be combined into a total of four (i.e., a two-by-two matrix) reinforcement schedules: (1) a fixed interval schedule that reinforces every x amount of time, (2) a variable interval schedule that reinforces at varying amounts of time, (3) a fixed ratio schedule that reinforces every n th behavior, and (4) a variable ratio schedule that reinforces at a varying number of behaviors [9].

It was found that particularly variable ratio schedules tend to produce very rapid and steady responding rates, in contrast with fixed ratio schedules, where the frequency of response usually drops after the reinforcement occurs [12,13]. Furthermore, the effectiveness of a reinforcement schedule is influenced by a number of other factors. For example, the time between the occurrence of a behavior and the reinforcing consequence is important: For a consequence to be most effective as a reinforcer, it should occur immediately after the behavior occurs [9]. In a 12-month clinical trial (N = 512), it was found that immediate reinforcement schedules are more effective at promoting moderate-to-vigorous physical activity than delayed reinforcement schedules [13]. In that study, participants that were assigned an immediate reinforcement schedules obtained a variable number of points (i.e., varying from 0 to 500 points) every day for reaching their moderate-to-vigorous physical activity goals, and wearing their wearable devices. Subsequently, they could exchange 500 points for gift cards with a value of USD 5.00. On the other hand, in that study, participants that were assigned a delayed reinforcement schedule received a financial incentive only at a 60-day interval, for wearing a wearable device, with a total value of up to USD 265. After only 5 days, participants assigned to the immediate reinforcement schedule were already engaging in more moderate-to-vigorous physical activity, on average: A trend that continued throughout the remainder of the intervention period [13]. Lastly, the effectiveness of a reinforcement schedule is influenced by individual differences [9]. The likelihood of a consequence being a reinforcer varies from person to person, so it is important to determine whether a financial reward is actually a reinforcer for a particular study participant. It is important not to assume that a particular stimulus will be a reinforcer

for a person just because it appears to be a reinforcer for most people [9]. For example, it was found that people with more extravert personalities are typically more sensitive to monetary incentives [14].

1.3. Research Case

In this study, we evaluated the impact of three reinforcement schedules on engagement levels with a mHealth app and overall campaign costs. Especially, we focused at reinforcement schedules that reinforce an ‘amount of behavior’ (i.e., ratio-based). The first schedule we evaluated was a fixed schedule that reinforced every n health behaviors with EUR 5. The second schedule was a variable schedule that reinforced every n health behaviors with EUR 5, if the end user was lucky enough to win a virtual lottery. In both these schedules, reinforcement followed the behavior immediately (although not deterministic in the latter case). The third reinforcement schedule was similar to the second schedule, but in this schedule, reinforcement occurred only at the end of the program (i.e., delayed), and not immediately after the behavior was performed. We hypothesized that participants receiving the variable ratio reward would be more engaged (**H1**). Particularly, because it was found that variable ratio schedules tend to produce long-lasting responding rates, as opposed to fixed ratio schedules, where the frequency of response usually drops after the reinforcement occurs [12,13]. Furthermore, we have evaluated the difference between immediate reinforcement and delayed reinforcement in terms of impact on engagement and cost. We hypothesized that respondents are particularly engaged through schedules that trigger reinforcement immediately after a behavior, probably at lower, or similar costs (**H2**), because it was found that immediate reinforcement is more effective than delayed reinforcement [9,13]. Finally, we hypothesized that subjects who are triggered by monetary incentives (e.g., subjects that strive to obtain a financial reward), have more extraverted personalities (**H3**), as people with more extravert personalities were found to be more sensitive to financial rewards [14].

2. Methods

2.1. Recruitment

Participants were recruited among staff members and students at Eindhoven University of Technology in the Netherlands, in February 2021. The entire population comprised roughly 3000 staff members and 12,000 students, distributed over 9 different departments. The study was advertised as a health promotion program and conducted only after explicit consent of the participants. The explicit consent of participants was collected upon registration for the program. All (operational) procedures were also approved by the ethical committee of Eindhoven University (experiment ID: ERB2021IEIS7).

2.2. Intervention Context

To evaluate our reinforcement schedules in a health promotion context, we have designed our intervention using the mHealth platform GameBus (see, e.g., www.gamebus.eu, accessed on 25 November 2021). This platform was especially designed for health promotion and provides a highly configurable gamification engine that can be used to host multiple experimental designs on a single platform. At the same time, GameBus enables researchers to gather health data in a manner compliant to European (privacy) legislations. Lastly, since GameBus is built of modular components, a web interface with just the components that are relevant for a study can relatively easily be assembled.

For this study, a customized variant of GameBus was employed. This custom web app was used to promote a set of health-related activities amongst participants. Participants themselves selected a total of four activities from a list of 18 predefined health-related activities. The list was designed to include activities that have a potential health benefit. The activities fell in any one of three health-related categories [15]: physical activity (e.g., “Make a total of 6000 steps per day”), healthy dietary intake (e.g., “Eat a piece of fruit”), or meaning and happiness (e.g., “Write down three things you are grateful for”), see

Appendix A for a complete overview of suggested activities. Participants could update their selection of activities at any moment in time throughout the program. The entire program lasted six weeks.

Participants were awarded virtual points if they performed a health-related activity, again see Appendix A for an overview of the number of points that were awarded per activity. Subjects could prove their engagement in an activity by uploading to the mHealth app a recent photo or video of themselves performing the activity. Additionally, participants were equipped with a smartwatch (i.e., a Samsung Galaxy Watch Active 2) to automatically track their daily number of steps (i.e., for a detailed description of the smartwatch application, see [16]).

Virtual points were displayed on an individual dashboard and social leaderboard. The individual dashboard displayed charts (e.g., heatmaps and trendlines) of a user's performance in different categories (e.g., walk performance in number of steps, or run performance as the frequency of engaging in a run), see Figure 1a. Users could customize the dashboard to include the widgets they found relevant. The social leaderboard displayed the total number of points per participant, and the average number of points per department, see Figure 1b. Users could customize the leaderboard to display only the individuals they wanted to compare themselves with. Moreover, all health-related activities that were performed by the study participants were displayed in a newsfeed, see Figure 1c. Participants could 'like' or commented upon each other's activities in a manner similar to mainstream social media platforms such as Facebook and Instagram.

Finally, every week, participants were notified via e-mail. These e-mails included instructions (e.g., on using the smartwatch, or on calling for support). Additionally, to inspire (passive) participants, these e-mails included statistics on the overall performance of all participants. All e-mail templates can be retrieved from Figshare [17].

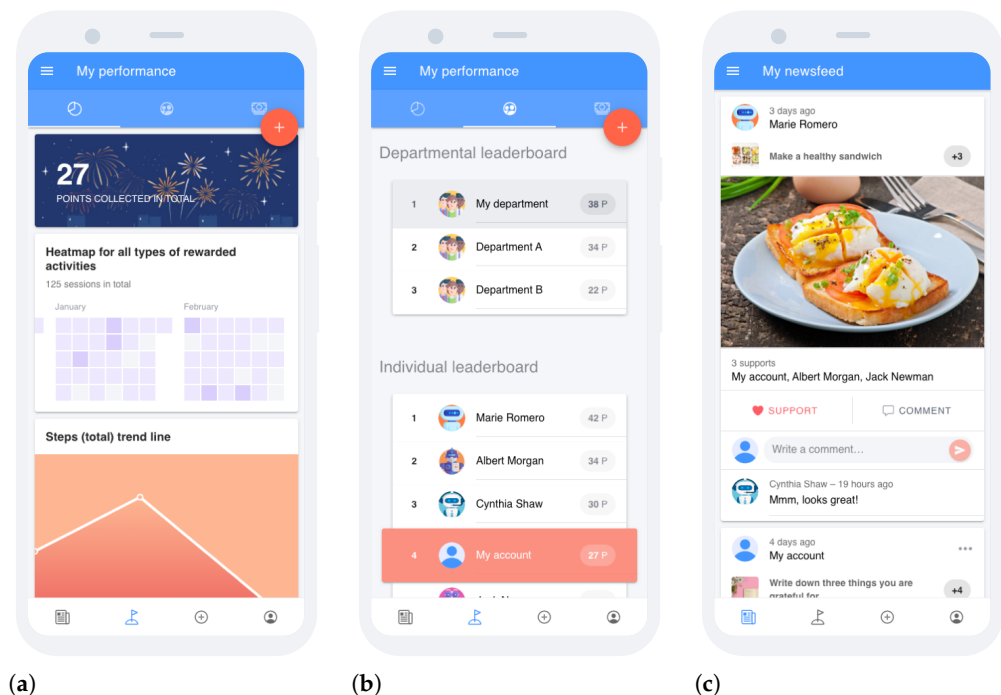


Figure 1. Overview of different performance visualizations that were available to all participants: (a) displays the individual dashboard, (b) displays the social leaderboard, and (c) displays the social newsfeed.

2.3. Study Design

In this study, three study arms were evaluated. Each arm got assigned a unique reinforcement schedule. In all study arms, participants had the opportunity to collect vouchers that could be cashed-in at a well-known Dutch online shopping platform. In the

first study arm (*FR-now*), participants received a voucher worth EUR 5 for every 80 points they scored (i.e., fixed ratio), with a maximum of five vouchers, with a total value of EUR 25, see Figure 2a. Reinforcement was immediate in this schedule, as the payout followed immediately from the behavior (i.e., receiving 80 points).

In the second study arm (*VR-now*), participants could exchange 10 points to open a virtual lootbox, and have a 10% chance of winning a voucher worth EUR 5, see Figure 2b. Again, subjects could win at most 5 vouchers, with a maximum value of EUR 25. Participants were not informed on the 10% odd ratio. This treatment had a variable reinforcement schedule, as payout was not guaranteed when a participant exchanged their points to open the lootbox. Additionally, in this schedule, reinforcement was immediate, as the payout followed immediately from the behavior (i.e., opening a lootbox).

In the third study arm (*VR-delay*), participants could also exchange 10 points to open a virtual lootbox. However, this time they did not have a 10% chance of winning an actual voucher, but rather did they have a 10% chance of winning a lottery ticket for a prize lottery that was hosted at the end of the program, see Figure 2c. In this prize lottery, vouchers worth EUR 5 (10×), EUR 25 (5×) and EUR 125 (1×) were distributed randomly over the lottery tickets, with a maximum of one voucher per lottery ticket. Hence, obtaining more lottery tickets eventually increased a participant's chance to win a voucher. Again, this treatment had a variable reinforcement schedule. However, reinforcement in this schedule was delayed, as the payout only followed at the end of the program.

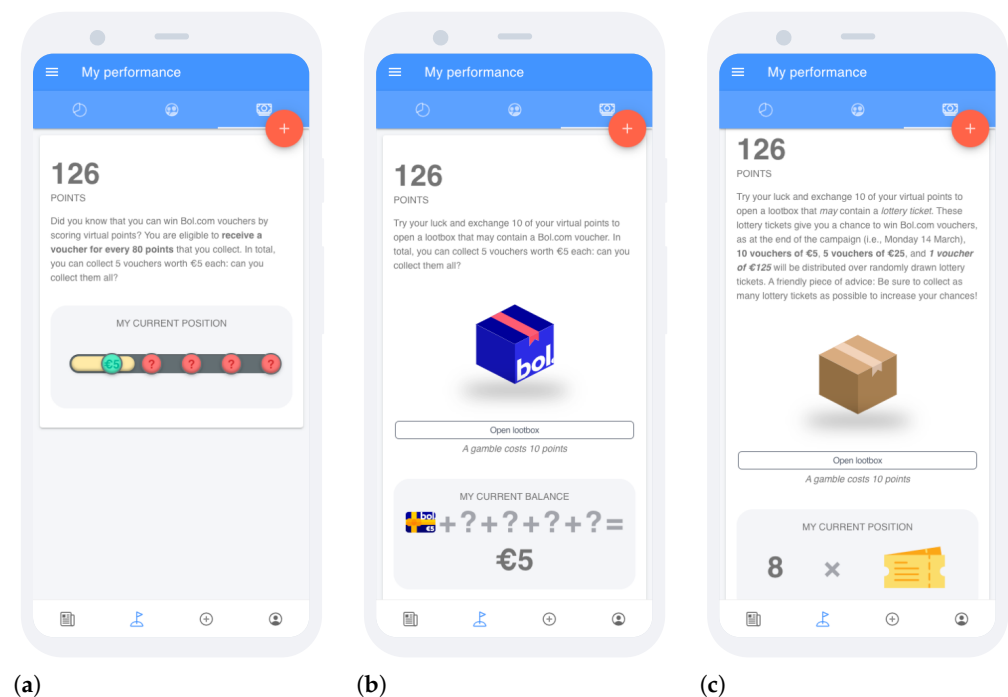


Figure 2. Implementation of different reinforcement schedules that were unique per study arm: (a) displays *FR-now* (i.e., an immediate, fixed ratio schedule), (b) displays *VR-now* (i.e., an immediate, variable ratio schedule), and (c) displays *VR-delay* (i.e., a delayed, variable ratio schedule)

2.4. Study Procedures

2.4.1. Estimated Study Costs

From the start of the program, the maximum expense of *VR-delay* was already known, as in the prize lottery at the end of the program, a maximum of 16 vouchers with a total value of EUR 300 would be raffled. However, the expense of *FR-now* and *VR-now* were unknown at the start of the program, as the monetary value of these treatments depended directly on user actions that were unknown at the start. Hence, to estimate the potential expenses of these treatments, we have performed a stimulation study, see Appendix B. After 10,000 simulations, we found that *FR-now* had an expected cost of EUR 450.07,

compared to EUR 397.91 for VR-now. Furthermore, the simulation suggested a minimum cost of EUR 380 and EUR 255 for FR-now and VR-now, respectively, and a maximum cost of, respectively, EUR 530 and EUR 560. In summary, the total expected cost of this study were estimated to be $\text{EUR } 450.07 + \text{EUR } 397.91 + \text{EUR } 300 = \text{EUR } 1147.98$.

2.4.2. Fraud Detection

The experimental setup was vulnerable to fraudulent usage. In theory, participants could upload photos and videos of activities that were not performed by themselves, or that they had already used to claim points. The research team has manually validated the registered activities. When any form of fraud was detected in an activity registration, all the potential virtual points and monetary rewards that were obtained from that activity were withdrawn. Participants who committed fraud were also alerted by a pop-up stating that their user account could be suspended, whenever they continued cheating. Throughout the study, no accounts were suspended, and no points were withdrawn, but one user received a warning once to upload a more recent photo as valid proof.

2.5. Measurements

2.5.1. Objective Exposure Data

In mHealth research, engagement is most commonly captured via measures of app usage [18,19]. Using our mHealth app, engagement of participants was repeatedly measured as: (1) the number of days a participant had been online (i.e., the number of distinct days the participant opened the web app), and (2) the number of activities a respondent performed (i.e., and registered in our app). These variables complement each other since the former may be limited to passive engagement, while the latter requires active participation (i.e., performing the suggested activities). Both measurements were recorded per participant, per week. Lastly, users were categorized in one of two categories: (1) subjects that enrolled in the program, but did not engage in the suggested activities and instead only watched the mHealth app were labeled *passive users*, and (2) subjects that enrolled in the program and did perform at least one suggested activity were labeled *active users*.

2.5.2. Subjective Survey Data

In a pre-test survey (also see Appendix C), Big Five personality profiles were estimated using the mini-IPIP scales [20], that measured, using 5-point Likert scales (i.e., measured between -2 and $+2$), respondents' degrees of: (1) openness to experience (i.e., "the breadth, depth, originality, and complexity of an individual's mental and experiential life", as opposed to closed-mindedness); (2) conscientiousness (i.e., "socially prescribed impulse control that facilitates task- and goal-directed behavior"); (3) extraversion (i.e., "an energetic approach toward the social and material world and includes traits such as sociability, activity, assertiveness, and positive emotionality"); (4) agreeableness (i.e., "prosocial and communal orientation towards others"); and (5) neuroticism (i.e., "negative emotionality, such as feeling anxious, nervous, sad, and tense") [21].

Additionally, a post-test survey (also see Appendix C) that participants completed after the six-week program, was used to collect demographic data from our sample. First, the post-test survey was used to record participants' gender, age group, and their affiliation with the host university. Second, five custom survey items were included to measure the perceived health impact of the program on: (1) walking frequency, (2) physical activity frequency, (3) improvement of dietary intake, and (4) frequency of contact with peers. These items were measured on 5-point Likert scales (i.e., coded between -2 and $+2$). Third, overall interest/enjoyment with the program was measured using 7 items from the Intrinsic Motivation Inventory (IMI) [22], measured using 5-point Likert scales (i.e., coded between -2 and $+2$). This scale is considered a self-report measure of intrinsic motivation [22]. Fourth, a subject's willingness to postpone receiving an immediate reward in order to gain additional benefits in the future was measured using 12 items from the deferment of gratification survey [23], using 5-point Likert scales (i.e., coded between -2 and $+2$). Delay

discounting has been found to significantly improve the prediction of health behaviors such as exercise frequency, eating breakfast, and estimated longevity [24]. We included this measure, because its predictive power for these behaviors had been stronger than the Big Five personality traits [24].

2.6. Statistical Analysis

To test our hypotheses a total of four sets of statistical analyses were performed. All study data and statistical procedures can be retrieved from Figshare [17]. The first set of statistical analyses included an exploration of user statistics, including descriptive statistics of demographic characteristics. Additionally an exploratory analysis was performed to evaluate dropout rates of participants. Furthermore, in this first set of analyses, a linear model was fit to determine whether the number of dropouts changed over time and were different per study arm. Lastly, we have performed exploratory analyses and statistical tests to evaluate the impact of the program on subject's self-perceived health status (e.g., frequency of walking, improvement of dietary intake, etc.). Particularly, we have performed a Wilcoxon rank sum test for each health dimension to test if the impact on a specific health dimension was significant (i.e., larger than zero).

The second set of statistical analyses was executed to evaluate the impact of the program on user engagement levels. First, our statistical analyses focused on evaluating passive user engagement levels (i.e., based on the number of days participants visited the web app), second our analyses focused on evaluating active user engagement levels (i.e., based on the number of health-related activities participants performed). For both measures, exploratory analyses were performed using mean plots to detect potential differences between study arms. Subsequently, several hierarchical linear models were estimated for: (1) the number of days a participant had been online, and (2) the number of activities a respondent had performed, respectively. For both measures, time (i.e., the ordinal week number), and study arm (i.e., the assigned reinforcement schedule) were used as predictors. It was tested whether significant second-order interaction effects existed amongst these variables. In all models we have allowed random intercepts for individuals. The final models were selected based on Akaike information criterion [25]. The Akaike information criterion estimates the relative quality of statistical models for a given set of data. The measure rewards goodness of fit, and includes a penalty for increasing the number of predictors (i.e., to prevent overfitting, as increasing the number of predictors generally improves goodness of the fit). For the final models, a set of post hoc comparisons were performed through estimated marginal means contrasts [26], in order to test more focused hypothesis than the overall omnibus F-test of a linear model. Particularly, post hoc comparisons were performed for all three contrasts of reinforcement schedules, as well as for generating all pair-wise contrasts for each reinforcement schedule, comparing the outcome variable for all weeks to the first week.

The third set of statistical analyses was similar to the second set, although this set focused on distinguishing engagement levels between subjects that won a voucher, and subjects that did not win a voucher. Again, hierarchical linear models were estimated for (1) the the number of days a participant had been online, and (2) the number of activities a respondent had performed, but now using the ordinal week number and user performance classification (i.e., whether a user won a voucher, or not) as predictors. The Akaike information criterion [25] was again used to select the final models, and post hoc comparisons were again performed using estimated marginal means contrasts [26].

Finally, a set of statistical analyses was executed to evaluate individual differences among participants that either won a voucher, or did not win a voucher. Particularly, we have evaluated whether significant differences in personality scores existed between these two groups. Again, an exploratory analysis was performed using mean plots to detect potential differences between these groups. Subsequently, several linear models were estimated for all five personality traits that were measured. Additionally, we have evaluated whether these groups differed significantly in terms of their self-reported capability to defer

gratification, and self-reported intrinsic motivation. Again we have performed exploratory analysis and fitted several linear models to detect potential differences between subjects that won a voucher, and subjects that did not win a voucher, on these dimensions, using the user performance classification (i.e., whether a user won a voucher, or not) as a predictor.

3. Results

3.1. User Statistics

In total, 61 unique participants joined the study. The participants were randomly assigned to the different study arms: 14 subjects were assigned to SA1, 26 to SA2, and 21 to SA3. Of the 61 participants, 39 subjects performed at least one activity (i.e., and were labeled active users). The other 22 subjects have only been checking the app (i.e., and were labeled passive users). Figure 3 displays a cohort diagram which details the number of subjects enrolled in different study phases. The pre-test survey was completed by 41 participants, of which 41 responses were complete. The post-test survey was completed by 26 participants, of which 26 responses were complete. A total of 22 subjects completed both the pre-test survey and the post-test survey, also see Figure 3.

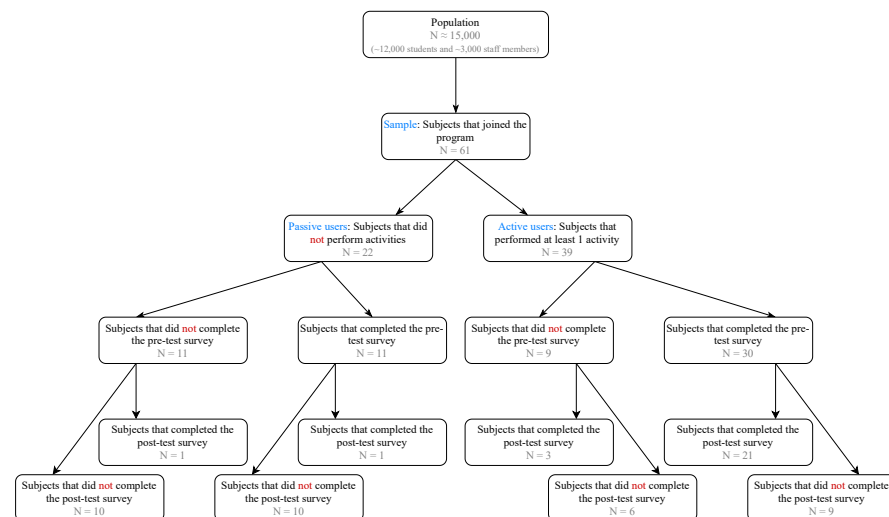


Figure 3. Cohort diagram that details the number of subjects engaged in different study phases.

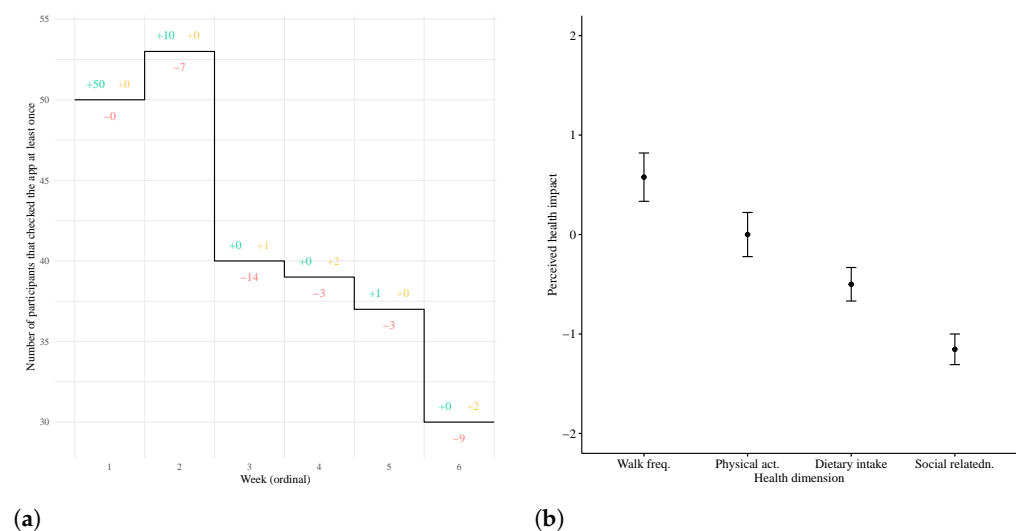
Sample demographics based on pre-test and post-test survey responses are displayed in Table 1. Note that the breakdown of gender, age group, and affiliation in Table 1 is based on 26 responses (i.e., as 26 participants took and completed the post-test survey), whereas the breakdowns of personality traits are based on 41 responses (i.e., the number of respondents that completed the pre-test survey).

Evaluation of Dropouts and Health Impact

Figure 4a displays the decrease in the number of participants that checked the mobile app during a given week. The number of participants that joined the program for the first time in a given week are displayed in green. The number of participants that dropped out in a specific week are displayed in red. The number of participants that checked the mobile app in a specific week, although they dropped out in an earlier week are displayed in yellow. No significant differences in dropout rates between study arms could be detected. Hence, it is assumed that dropouts are spread equally over study arms. Figure 4b displays participants' perceived health impact. From a Wilcoxon rank sum test, it was found that subjects generally reported to have increased their walk frequency (i.e., walk frequency is significantly higher than zero at $p = 0.0178$). Significant improvements were not reported for any other health dimension.

Table 1. Sample demographics based on pre-test and post-test survey responses.

Characteristic	Sample	FR-Now	VR-Now	VR-Delay
Gender (N; %)	N = 26	N = 6	N = 10	N = 10
Female	10; 39%	1; 17%	5; 50%	4; 40%
Male	16; 61%	5; 83%	5; 50%	6; 60%
Age group (N; %)	N = 26	N = 6	N = 10	N = 10
18–24	9; 34%	2; 34%	3; 30%	4; 40%
25–34	14; 54%	4; 66%	6; 60%	4; 40%
35–44	1; 4%	0; 0%	1; 10%	0; 0%
45–54	2; 8%	0; 0%	0; 0%	2; 20%
Affiliation (N; %)	N = 26	N = 6	N = 10	N = 10
Student	14; 54%	4; 66%	4; 40%	6; 60%
Ph.D.	3; 11%	0; 0%	2; 20%	1; 10%
Staff	6; 23%	1; 17%	2; 20%	3; 30%
Alumnus	2; 8%	1; 17%	1; 10%	0; 0%
Other	1; 4%	0; 0%	1; 10%	0; 0%
Personality (μ; σ)	N = 41	N = 9	N = 16	N = 16
Openness	−0.780; 0.633	−0.694; 0.950	−0.781; 0.625	−0.828; 0.435
Conscientiousness	−0.396; 0.768	−0.722; 0.678	−0.375; 0.816	−0.234; 0.755
Extraversion	0.146; 0.716	−0.167; 0.781	0.172; 0.568	−0.297; 0.797
Agreeableness	−0.846; 0.671	−0.444; 0.601	−0.792; 0.842	−1.125; 0.342
Neuroticism	0.250; 0.831	0.500; 0.545	0.078; 0.999	0.281; 0.785

**Figure 4.** Overview of dropout rates and health impact: (a) displays the number of subjects that visited the app at least once, per week, among all participants, N = 61, and (b) displays a mean plot of perceived health impact among post-tested participants, N = 26.

3.2. Evaluation of the Impact of Different Reinforcement Schedules

Engagement Analyses

Passive engagement was measured as the number of days participants visited the web app within a given week. Figure 5a displays the average number of days participants were online (i.e., visited the mHealth app) per week, per reinforcement schedule. From the second set of statistical analyses, performed on all enrolled subjects (N = 61), no significant differences were detected in the number of days participants were online in different reinforcement schedules. However, every week had a significant impact on the number of days participants were online of -0.226 days at $p < 0.0001$; -0.256 days at $p < 0.0001$; -0.319 days at $p < 0.0001$; and -0.934 days at $p < 0.0001$, for the third, fourth, fifth and sixth week, respectively. No significant (interaction) effects (between reinforcement schedule and week number) were detected.

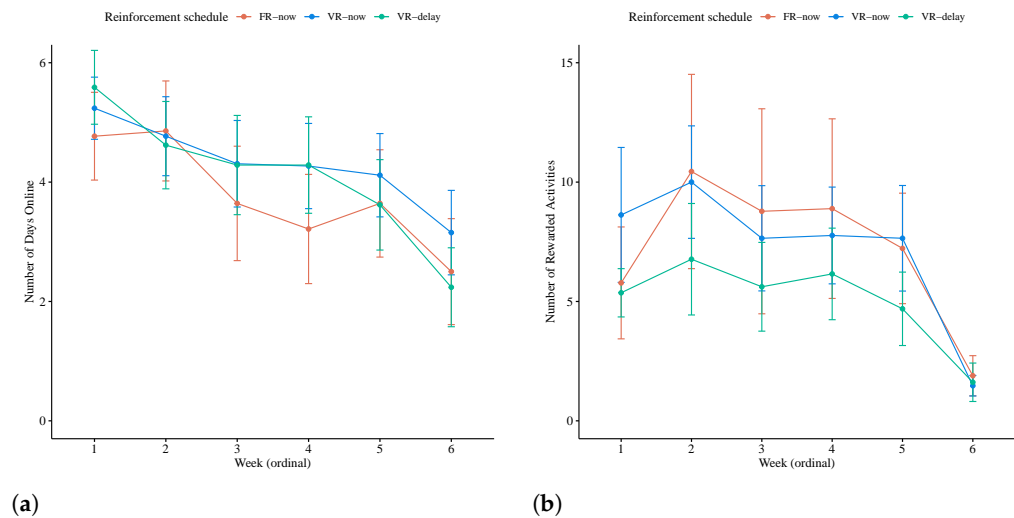


Figure 5. Mean plots per reinforcement schedule of: (a) the number of days participants visited the web app (all participants, $N = 61$), and (b) the number of activities participants registered (active participants, $N = 39$).

Additionally, we have evaluated the active engagement of participants in each reinforcement schedule. Figure 5b displays the average number of activities participants have performed per week, per reinforcement schedule. From our statistical analyses, performed on all active subjects ($N = 39$), again no significant differences were detected in the number of activities participants performed in different reinforcement schedules. Furthermore, the number of activities participants performed did not change significantly per week. However, in the last week, participants performed significantly less activities (i.e., -2.013 activities at $p < 0.0001$). Again, no significant (interaction) effects (between reinforcement schedule and week number) were detected.

Finally, the total costs of each reinforcement schedule are displayed in Table 2. Eventually, 16 participants have won a total value of 24 vouchers, worth EUR 440.00. In the first study arm (FR-now) a total of 13 vouchers, worth EUR 65.00 were distributed over four participants, out of 14 participants that were assigned to this arm; in the second study arm (VR-now) a total of 24 vouchers, worth EUR 120.00 were distributed over nine participants, out of 26 participants that were assigned to this arm; in the third study arm (VR-delay) a total of 11 vouchers, worth EUR 255.00 were distributed over three participants, out of 21 participants that were assigned to this arm. These results translate to costs of EUR 4.64, EUR 4.62, and EUR 12.14 per individual, for the first, second, and third arm, respectively.

Table 2. Total costs of reinforcement schedules.

Arm	N	Expected Value	Lootboxes	Winners	Total Value	Cost per Subject
FR-now	14	EUR 450.07	N.A.	4	EUR 65.00	EUR 4.64
VR-now	26	EUR 397.91	213	9	EUR 120.00	EUR 4.62
VR-delay	21	EUR 300.00	123	3	EUR 255.00	EUR 12.14
Total	61	EUR 1147.98	336	16	EUR 440.00	EUR 7.21

3.3. Evaluation of User Performance

3.3.1. Engagement Analyses

For subsequent statistical analyses we have distinguished between subjects that won a monetary incentive, and subjects that did not win a voucher. Figure 6 displays both the number of days participants visited the web app, as well as the number of activities participants performed. From the third set of statistical analyses based on Figure 6a, it

was found that the number of days participants checked the app was significantly higher for participants that won a voucher in the second to sixth week. Particularly, our contrast analysis showed that participants that won a voucher were 0.924 days more online at $p = 0.0017$ in the second week; 1.989 days more online at $p < 0.0001$ in week three; 1.862 days more online at $p < 0.0001$ in week four; 2.376 days more online at $p < 0.0001$ in week five; 1.818 days more online at $p < 0.0001$ in week six, than subjects that did not win a voucher. The difference between both groups in the first week was not significant at $p = 0.1571$. Furthermore, the number of days participants that won a voucher checked the app did not change significantly over time, compared to the first week. However, the number of days participants that did not win a voucher checked the app changed significantly in the second to the sixth week, compared to the first week (i.e., -0.105 days at $p = 0.0111$ in the second week; -0.600 days at $p < 0.0001$ in week three; -0.630 days at $p < 0.0001$ in week four; -0.808 days at $p < 0.0001$ in week five; and -1.562 days at $p < 0.0001$ in week six).

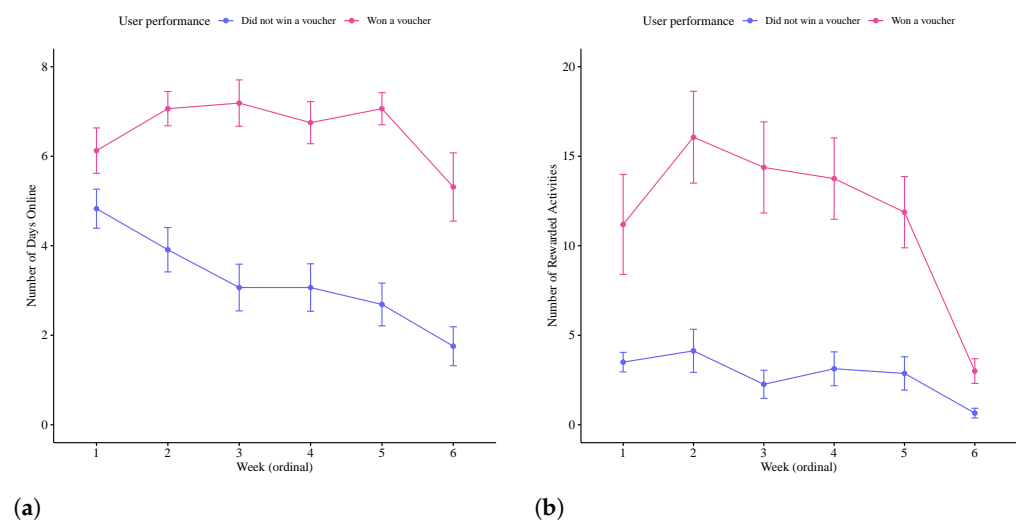


Figure 6. Mean plots per participants that either did, or did not win a voucher of: (a) the number of days participants visited the web app (all participants, $N = 61$), and (b) the number of activities participants registered (active participants, $N = 39$).

Subsequently, from our statistical analyses based on Figure 6b, it was found that the number of activities participants performed was significantly higher for participants that won a voucher in all weeks. Particularly, our contrast analysis showed that participants that won a voucher performed 1.408 more activities at $p = 0.0046$ in the first week; 5.306 more activities at $p < 0.0001$ in the second week; 6.465 more activities at $p < 0.0001$ in week three; 4.980 more activities at $p < 0.0001$ in week four; 4.538 more activities at $p < 0.0001$ in week five; 1.161 more activities at $p < 0.0083$ in week six, than subjects that did not win a voucher. Furthermore, the number of activities participants that won a voucher performed increased significantly from the first to the second week (i.e., $+0.687$ activities at $p = 0.0244$). From the third to the fifth week, the number of activities participants that won a voucher performed did not change significantly, compared to the first week. However, in the sixth week, the number of activities participants that won a voucher performed dropped significantly (i.e., -2.224 activities at $p < 0.0001$), compared to the first week. Finally, the number of activities participants that did not win a voucher performed dropped significantly in the third, fifth, and sixth week, compared to the first week (i.e., -0.582 activities at $p = 0.0172$ in week three; -0.455 activities at $p = 0.0349$ in week five; and -1.910 activities at $p < 0.0001$ in week six).

3.3.2. Analyses of Individual Differences

In the fourth set of statistical analyses, we have explored whether the group of participants that won a voucher and the group of participants that did not win a voucher had different personal characteristics. From our statistical analyses of personality traits based on Figure 7a, we did not find any significant differences between these two groups in terms of Big Five personality traits. However, from our statistical analysis of participant's self-reported capability to defer gratification, based on Figure 7b, it was found that participants that won a voucher were significantly more capable to defer gratification (i.e., +0.312, at $p = 0.0426$). Furthermore, we found that participants that won a voucher reported significantly lower levels of intrinsic motivation with the program (i.e., -0.757 , at $p = 0.0095$).

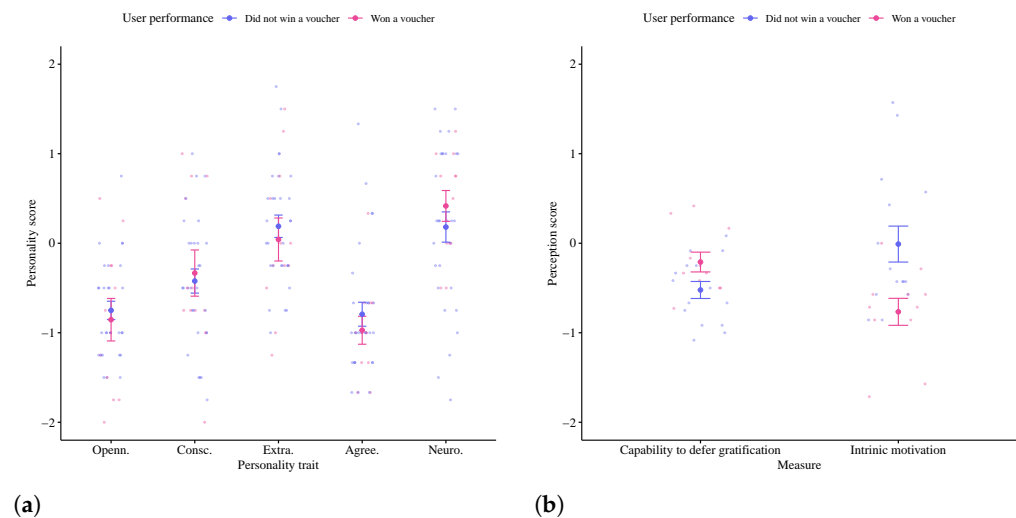


Figure 7. Mean plots per participants that either did, or did not win a voucher of: (a) self-reported personality scores (pre-tested participants, $N = 41$), and perceived deferment of gratification, (b) perceived intrinsic motivation (post-tested participants, $N = 26$).

4. Discussion

4.1. Evaluation Outcomes

From this study, it is found that (variable) monetary incentives can foster health behavior, which is in line with findings from earlier studies [7,8,10,11,13]. However, as opposed to findings from Berardi and colleagues [13], no single reinforcement schedule was found to have a more profound impact on app engagement or the adoption of health behaviors in our study: In our study, every reinforcement schedule we evaluated had a similar impact on user engagement levels. Hence, although variable ratio schedules have been found in earlier studies to be more effective [12], we could not confirm our hypothesis **H1**, as participants that were assigned to a variable ratio reinforcement schedule were not necessarily more engaged, in our study. Moreover, we could also not (fully) confirm our hypothesis **H2**. Reinforcement that occurs immediately after a targeted behavior has previously been found to be a more effective reinforcer [9,13], but we could not replicate that finding in this study. Nevertheless, we did find that the immediate, variable ratio schedule can foster engagement levels at relatively low costs (i.e., at EUR 4.62 per participant, the lowest cost per participant in our study). Hence, we would suggest practitioners to experiment with this schedule.

Although all reinforcement schedules have fostered engagement levels to a similar extent, it seems that this outcome is produced by only a small subset of our sample: the group of subjects that won a voucher. It makes sense to assume that people that won a voucher, were particularly triggered by the monetary incentives. We saw huge differences in engagement levels between subjects that did win a voucher, and subjects that did not.

Particularly, subjects that won a voucher were consistently engaged with the app, both passively and actively, throughout the entire program (6 weeks). Only in the last week, the number of activities this group of subjects performed (i.e., active engagement) dropped significantly. This drop was likely explained by the upcoming Christmas break, that started immediately after our program had ended. Subjects that did not win a voucher (and were likely not triggered by the monetary incentives) were significantly less engaged with the program. First, the average level of active engagement in this group was consistently lower than the average active engagement levels that were reported in subjects that did win a voucher. Second, although passive engagement levels were similar for all subjects in the first week, passive engagement dropped significantly (and rapidly) in subjects that did not win a voucher in the weeks that followed. Note that it makes sense that the average number of days participants visited our app in the first week was similar among all participants: It shows that all participants were equally curious to join the program at baseline, and only after learning the specific intervention strategies were employed in the program (e.g., monetary incentives), engagement levels in subjects that were not triggered by these strategies dropped.

Finally, subjects that won a voucher, and subjects that did not win a voucher were not different in terms of Big Five personality traits. Therefore, we could not confirm our hypothesis H3, suggesting that subjects with extraverted personalities are especially triggered by financial rewards, in general. However, the participants that won a voucher were distinguishable in terms of other personal characteristics. Users that won a voucher were significantly more capable to defer gratification (self-reported) and significantly less intrinsically motivated (self-reported) to engage with the program. This makes sense, as individuals with a larger ability to defer gratification are more likely to be achievers [23], and therefore strive to obtain a financial reward. Furthermore, it makes sense that individuals that were triggered by the monetary incentive were less intrinsically motivated. Instead, these participants were more extrinsically motivated to participate in our program (and hence triggered by the financial incentive). Although these results are obvious in hindsight, these insights may be used to categorize participants at the start of a health promotion program, to decide whether an individual could be triggered with monetary incentives, or not. Admittedly, deriving intrinsic motivation before the program has started relies on judgements about your future self, which may be biased. However, deferment of gratification can easily be measured beforehand, without judgements about your future self, using standardized survey instruments [23]. This measure may be used to select monetary incentives as an intervention strategy, for a specific subject at the start of a health promotion program, and effectively tailor end user experiences, although further research is required.

4.2. Study Weaknesses

This study was subject to several weaknesses. First, this study evaluated the impact of our intervention program on a particular target group (i.e., university staff members and students) within a specific context (i.e., a university environment). Based on the current study, findings cannot be generalized yet, as results were derived from a relatively small sample (i.e., $N = 61$, from a population of 15,000). Similarly, due to low post-test response rates, some analyses, and the last set of statistical analyses in particular, were based on even smaller sample sizes. Although measures were taken to attract as many participants as possible to complete the post-test survey (e.g., via e-mail invitations and reminders), taking the post-test—and participation in the program in general—was voluntary. Hence, the post-test response rate and program participation rate could not be controlled, although these rates did influence study power.

4.3. Future Work

Future studies should explore whether a measure of deferment of gratification can be used to effectively identify users that are generally triggered by financial rewards. Particu-

larly, in an intervention trial, participants should take the deferment of gratification survey as a pre-test. It can be expected that, subjects with a higher ability to defer gratification are more engaged with the study than subjects with a lower ability to defer gratification, when the study employs financial rewards to engage end users. Additionally, future work should focus on improving our understanding of the impact of personality traits in a health promotion context. Although personality trait data did not explain the impact of the financial rewards that we evaluated in this study, personality trait data may still be relevant to explain overall engagement levels with a mHealth app, or overall intrinsic motivation with a health promotion program.

5. Conclusions

Financial rewards have traditionally been employed in mHealth apps to promote end user engagement levels, and foster health behaviors. However, the optimal reinforcement schedule for distributing monetary incentives have remained unclear. In this study, we have aimed at identifying a reinforcement schedule with a high impact on end user engagement levels, and low overall costs for the study organizer. We have evaluated the impact of different reinforcement schedules on engagement levels with our mHealth app in a six-week, three-arm randomized intervention trial. In line with findings from earlier studies, we found that monetary incentives can effectively stimulate mHealth app engagement levels. In our study, every reinforcement schedule we evaluated had a similar impact on end user engagement levels. Still, we suggest practitioners to experiment with variable ratio reinforcement schedules with immediate payouts, as this schedule was reported to have the lowest cost per participant in our study. Finally, we emphasize that financial incentives may not be effective to stimulate engagement levels in everyone. Particularly, we found that individuals that were triggered by the financial incentives generally had a greater ability to defer gratification.

Author Contributions: Conceptualization, R.N., P.V.G. and A.K. (Alireza Khanshan); methodology, R.N., P.V.G. and A.K. (Alireza Khanshan); software, R.N., P.V.G. and A.K. (Alireza Khanshan); validation, R.N. and P.V.G.; formal analysis, R.N.; investigation, R.N. and A.K. (Alireza Khanshan); resources, R.N. and A.K. (Alireza Khanshan); data curation, R.N.; writing—original draft preparation, R.N.; writing—review and editing, R.N., P.V.G., A.K. (Alireza Khanshan), P.L.B., A.K. (Astrid Kemperman), P.v.d.B. and M.S.; visualization, R.N.; supervision, P.V.G., P.L.B., A.K. (Astrid Kemperman), P.v.d.B. and M.S.; project administration, R.N., P.V.G. and A.K. (Astrid Kemperman); funding acquisition, P.V.G., P.L.B., A.K. (Astrid Kemperman), P.v.d.B. and M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the research program ‘Gamification for Overweight Prevention and Active Lifestyle’ (443001101), which is partly financed by The Netherlands Organization for Health Research and Development (ZonMw).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Eindhoven University of Technology (protocol code ERB2021IEIS7, approved 9 March 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available in Figshare at <https://doi.org/10.6084/m9.figshare.16692271> (accessed on 25 November 2021), reference number [17].

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

mHealth Mobile Health

Appendix A. Overview of Available Activities

Table A1 displays the list of 18 activities that participants were suggested. Participant themselves selected a total of 4 activities to focus on, which could be changed at any

moment in time throughout the program. Activities were categorized as either impacting: (1) physical activity (PA), (2) dietary intake (DI), or meaning & happiness (MH). For some activities, the intensity of the activity could be tailored by the participants themselves. For example, participants could set the minimum number of steps for the activity “Make a total of $\{\minSteps\}$ steps per day”. The default value of $\{\minSteps\}$ was 6000 steps, and the minimum value of this variable was 5000 steps.

Table A1. List of activities that were suggested to participants.

Cat.	Description of Activity	Proof	Frequency	Points
PA	Make a total of $\{\minSteps\}$ steps per day. ¹	Automatic smartwatch	via Max. 7 × per week	+5
PA	Make a total of $\{\minSteps\}$ steps per day walking. ²	Automatic smartwatch	via Max. 7 × per week	+5
PA	Make a total of $\{\minSteps\}$ steps per day running. ²	Automatic smartwatch	via Max. 7 × per week	+5
PA	Join the digital Monday Morning Workout Session.	Automatic	via on-line attendance Max. 1 × per week	+25
PA	Do $\{\minRepetitions\}$ lunges (i.e., $\{\minRepetitions\} \div 2$ on each side). ³	Video	Max. 7 × per week	+5
PA	Do $\{\minRepetitions\}$ sit-ups. ⁴	Video	Max. 7 × per week	+5
PA	Do $\{\minRepetitions\}$ squats (while holding as many rolls of toilet paper as possible). ⁵	Video	Max. 7 × per week	+5
PA	Keepy-uppies: Juggle a soccer ball at least $\{\minRepetitions\}$ times. ⁵	Video	Max. 7 × per week	+5
PA	Go for a walk with your dog.	Selfie	Max. 2 × per week	+7
DI	Eat a piece of fruit.	Selfie	Max. 7 × per week	+3
DI	Make a healthy sandwich.	Selfie	Max. 7 × per week	+3
DI	Make a healthy salad.	Selfie	Max. 7 × per week	+3
DI	Cook a healthy dish.	Selfie	Max. 7 × per week	+3
DI	Picture your shopping cart with healthy products only.	Selfie	Max. 2 × per week	+7
MH	Mimic this yoga pose for a few minutes: 🧘.	Selfie	Max. 7 × per week	+3
MH	Write down three things you are grateful for.	Text input	Max. 7 × per week	+4
MH	Take a selfie at the park, or in the forest.	Selfie	Max. 7 × per week	+3
MH	Share your most beautiful picture outdoors (i.e., picture has to be taken recently).	Selfie	Max. 2 × per week	+7
MH	Skype with friends of yours while you work on a resolution of yours.	Selfie or screenshot	Max. 2 × per week	+7

¹ $\{\minSteps\}$ defaulted to 6000 steps, with a minimum of 5000 steps. ² $\{\minSteps\}$ defaulted to 5000 steps, with a minimum of 4000 steps. ³ $\{\minRepetitions\}$ defaulted to 30 repetitions, with a minimum of 20 repetitions. ⁴ $\{\minRepetitions\}$ defaulted to 20 repetitions, with a minimum of 10 repetitions. ⁵ $\{\minRepetitions\}$ defaulted to 15 repetitions, with a minimum of 10 repetitions.

Appendix B. Software to Simulate Study Costs

```
import random
import math
import copy
import json
from statistics import mean

# Duration of the program in waves (e.g., weeks)
nrOfWaves = 6

# Expected active engagement of participants (i.e., engagement * points per week = estimated points)
engagement = { 'min': .5, 'max': .9 }

treatments = [
  {
    'name': 'Treatment FR-now',
    'sample': 30,
    'maxPointsPerWave': 100,
    'lotteries': [
      {
        'name': 'vouchers',
        'cost': 80,
        'rewards': [
          { 'name': 'Voucher', 'odd': 1, 'monetaryValue': 5, 'maxPerPerson': 5 },
        ],
      },
    ],
  },
  {
    'name': 'Treatment VR-now',
    'sample': 30,
```

```

        'maxPointsPerWave': 100,
        'lotteries':[
            {
                'name': 'vouchers',
                'cost': 10,
                'rewards':[
                    { 'name': 'Voucher', 'odd': .1, 'monetaryValue': 5, 'maxPerPerson': 5,},
                ],
            },
        ],
    ],
]

# Necessary configuration : Transform treatment object
setBoundaries()

# Simulate once
simtreatments = simulate()
list(map(lambda t: {'treatment': t['name'], 'expense': t['simulatedMonetaryValue']}, simtreatments))

# Simulate 10,000 times
expenses = motecarlo(10000)

def motecarlo(n):
    expenses = []
    for i in range(n):
        simtreatments = simulate()
        e = list(map(lambda t: t['simulatedMonetaryValue'], simtreatments))
        expenses.append(e)

    avgs = list(map(mean, zip(*expenses)))
    mins = list(map(min, zip(*expenses)))
    maxs = list(map(max, zip(*expenses)))

    result = []
    for t, treatment in enumerate(treatments):
        result.append( {'name': treatment['name'], 'mean': avgs[t], 'min': mins[t], 'max': maxs[t],})

    print(json.dumps(result, indent=2))
    return expenses

def setBoundaries():
    for treatment in treatments:
        for lottery in treatment['lotteries']:
            lottery = boundaries(lottery)

def boundaries(lottery):
    left = 0
    for reward in lottery['rewards']:
        reward['left'] = left
        reward['right'] = left + reward['odd']
        left += reward['odd']
    return lottery

def play(lottery):
    prize = None

    ticket = random.random()
    for reward in lottery['rewards']:
        if ticket >= reward['left'] and ticket < reward['right']:
            prize = reward
            break

    return prize

def simulate():
    users = []
    for i in range(max(list(map(lambda t: t['sample'], treatments)))):
        user = {
            'activeness': random.uniform(engagement['min'], engagement['max']),
            'points': 0,
            'rewards': [],
        }
        users.append(user)

    for treatment in treatments:
        treatment['users'] = copy.deepcopy(users)[treatment['sample']]
        treatment['simulatedMonetaryValue'] = 0

        for w in range(nrOfWaves):
            for user in treatment['users']:
                user['points'] += math.floor(user['activeness'] * treatment['maxPointsPerWave'])

                canPlay = True

                while canPlay:
                    l = 0
                    # Check if user can still play the lottery
                    if user['points'] < treatment['lotteries'][l]['cost']:
                        canPlay = False
                    else:
                        user['points'] -= treatment['lotteries'][l]['cost']

                    if canPlay:
                        reward = play(treatment['lotteries'][l])
                        if reward and len(list(map(lambda r: r['name'] == reward['name'], user['rewards']))) < reward['maxPerPerson']:
                            user['rewards'].append(reward)
                            treatment['simulatedMonetaryValue'] += reward['monetaryValue']

    return treatments

```


Appendix C. Pre-Test and Post-Test Surveys

Appendix C.1. Post-Test Survey

Table A2. Details of post-test survey ¹.

#	Question or Statement	Type
Sample demographics		
1.	What is your e-mail address?	Open
2.	How are you affiliated to TU/e?	One of: Student, Ph.D. candidate, Staff member, Alumnus, or Other
3.	What is your gender?	One of: Female, Male, or Undisclosed
4.	What is your age?	One of: 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, 75 or older, or Undisclosed
Health impact		
5.	I have walked more often than usual due to the TU/e Samen Gezond lifestyle program.	5-point Likert
6.	I have engaged in more physical activity than usual due to the TU/e Samen Gezond lifestyle program.	5-point Likert
7.	I have improved my diet due to the TU/e Samen Gezond lifestyle program.	5-point Likert
8.	I have been in touch with my colleagues more often than usual due to the TU/e Samen Gezond Lifestyle Program.	5-point Likert
Intrinsic Motivation Inventory: <i>Interest/enjoyment</i> , as a proxy of intrinsic motivation ¹		
9.	I enjoyed participating in the TU/e Samen Gezond lifestyle program very much.	5-point Likert
10.	The TU/e Samen Gezond lifestyle program was fun to participate in.	5-point Likert
11.	I thought participating in the TU/e Samen Gezond lifestyle program was boring.	5-point Likert
12.	The TU/e Samen Gezond lifestyle program did not hold my attention at all.	5-point Likert
13.	I would describe the TU/e Samen Gezond lifestyle program as very interesting.	5-point Likert
14.	I thought participating in the TU/e Samen Gezond lifestyle program was quite enjoyable.	5-point Likert
15.	While I was participating in the TU/e Samen Gezond lifestyle program, I was thinking about how much I enjoyed it.	5-point Likert
Deferment of gratification survey ²		
16.	I am good in saving my money instead of spending it at once.	5-point Likert
17.	I enjoy something more when I have to wait for it and plan for it.	5-point Likert
18.	When I was a child, I saved any pocket money that I had.	5-point Likert
19.	When I am in the supermarket, I usually buy a lot of things that I had not planned to buy.	5-point Likert
20.	I am constantly without money.	5-point Likert
21.	I agree with the philosophy: "Eat, drink, and be happy, for tomorrow we may all be dead".	5-point Likert
22.	I would describe myself as often acting without thinking for my own good.	5-point Likert
23.	I often think it is important to wait and think things over before deciding.	5-point Likert
24.	I like spending my money immediately after I get it.	5-point Likert
25.	It is hard for me to avoid losing my temper when someone gets me very angry.	5-point Likert
26.	Most of the time, it is easy for me to be patient when I am kept waiting for things.	5-point Likert
27.	I am good in planning things ahead.	5-point Likert

¹ Questions 9 to 15 were displayed in random order. ² Questions 16 to 27 were displayed in random order.

Appendix C.2. Pre-Test Survey

Table A3. Details of pre-test survey.

#	Question or Statement	Type
Sample demographics		
1.	What is your e-mail address?	Open
Big Five personality trait: <i>Extraversion</i> ¹		
2.	In general, I am the life of the party.	5-point Likert
3.	In general, I do not talk a lot.	5-point Likert
4.	In general, I talk to a lot of different people at parties.	5-point Likert
5.	In general, I keep in the background.	5-point Likert
Big Five personality trait: <i>Agreeableness</i> ¹		
6.	In general, I sympathize with others' feelings.	5-point Likert
7.	In general, I am not interested in other people's problems.	5-point Likert
8.	In general, I feel others' emotions.	5-point Likert
9.	In general, I am not really interested in others.	5-point Likert
Big Five personality trait: <i>Conscientiousness</i> ¹		
10.	In general, I get chores done right away.	5-point Likert
11.	In general, I often forget to put things back in their proper place.	5-point Likert
12.	In general, I like order.	5-point Likert
13.	In general, I make a mess of things.	5-point Likert
Big Five personality trait: <i>Neuroticism</i> ¹		
14.	In general, I have frequent mood swings.	5-point Likert
15.	In general, I am relaxed most of the time.	5-point Likert
16.	In general, I get upset easily.	5-point Likert
17.	In general, I seldom feel blue.	5-point Likert
Big Five personality trait: <i>Openness to experience</i> ¹		
18.	In general, I have a vivid imagination.	5-point Likert
19.	In general, I Am not interested in abstract ideas.	5-point Likert
20.	In general, I have difficulty understanding abstract ideas.	5-point Likert
21.	In general, I do not have a good imagination.	5-point Likert

¹ Questions 2 to 21 were displayed in random order.

References

- Eckerstorfer, L.V.; Tanzer, N.K.; Vogrincic-Haselbacher, C.; Kedia, G.; Brohmer, H.; Dinslaken, I.; Corcoran, K. Key Elements of mHealth Interventions to Successfully Increase Physical Activity: Meta-Regression. *JMIR MHealth UHealth* **2018**, *6*, e10076. [[CrossRef](#)] [[PubMed](#)]
- DeSmet, A.; Rycckeghem, D.V.; Compennolle, S.; Baranowski, T.; Thompson, D.; Crombez, G.; Poels, K.; Lippevelde, W.V.; Bastiaensens, S.; Cleemput, K.V.; et al. A meta-analysis of serious digital games for healthy lifestyle promotion. *Prev. Med.* **2014**, *69*, 95–107. [[CrossRef](#)] [[PubMed](#)]
- Hamari, J.; Koivisto, J.; Sarsa, H. Does Gamification Work?—A Literature Review of Empirical Studies on Gamification. In Proceedings of the 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 6–9 January 2014; IEEE: New York, NY, USA, 2014. [[CrossRef](#)]
- Deterding, S.; Dixon, D.; Khaled, R.; Nacke, L. From game design elements to gamefulness. In Proceedings of the MindTrek '11: 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland, 28–30 September 2011; ACM Press: New York, NY, USA, 2011; doi:10.1145/2181037.2181040. [[CrossRef](#)]
- Huotari, K.; Hamari, J. A definition for gamification: Anchoring gamification in the service marketing literature. *Electron. Mark.* **2016**, *27*, 21–31. [[CrossRef](#)]
- Lewis, Z.H.; Swartz, M.C.; Lyons, E.J. What is the Point?: A Review of Reward Systems Implemented in Gamification Interventions. *Games Health J.* **2016**, *5*, 93–99. [[CrossRef](#)] [[PubMed](#)]
- Mitchell, M.S.; Goodman, J.M.; Alter, D.A.; John, L.K.; Oh, P.I.; Pakosh, M.T.; Faulkner, G.E. Financial Incentives for Exercise Adherence in Adults. *Am. J. Prev. Med.* **2013**, *45*, 658–667. [[CrossRef](#)] [[PubMed](#)]
- Wall, J.; Mhurchu, C.N.; Blakely, T.; Rodgers, A.; Wilton, J. Effectiveness of Monetary Incentives in Modifying Dietary Behavior: A Review of Randomized, Controlled Trials. *Nutr. Rev.* **2006**, *64*, 518–531. [[CrossRef](#)] [[PubMed](#)]
- Miltenberger, R.G. Reinforcement. In *Behavior Modification: Principles and Procedures*; Cengage Learning: Wadsworth, OH, USA, 2015; Chapter 4, pp. 63–86.

10. Haisley, E.; Volpp, K.G.; Pellathy, T.; Loewenstein, G. The Impact of Alternative Incentive Schemes on Completion of Health Risk Assessments. *Am. J. Health Promot.* **2012**, *26*, 184–188. [[CrossRef](#)] [[PubMed](#)]
11. van der Swaluw, K.; Lambooi, M.S.; Mathijssen, J.J.P.; Schipper, M.; Zeelenberg, M.; Berkhout, S.; Polder, J.J.; Prast, H.M. Physical activity after commitment lotteries: Examining long-term results in a cluster randomized trial. *J. Behav. Med.* **2018**, *41*, 483–493. [[CrossRef](#)] [[PubMed](#)]
12. Ferster, C.B.; Skinner, B.F. *Schedules of Reinforcement*; Appleton-Century-Crofts: Norwalk, CT, USA, 1957; doi:10.1037/10627-000. [[CrossRef](#)]
13. Berardi, V.; Hovell, M.; Hurley, J.C.; Phillips, C.B.; Belletiere, J.; Todd, M.; Adams, M.A. Variable Magnitude and Frequency Financial Reinforcement is Effective at Increasing Adults' Free-Living Physical Activity. *Perspect. Behav. Sci.* **2020**, *43*, 515–538. [[CrossRef](#)] [[PubMed](#)]
14. Depue, R.A.; Collins, P.F. Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behav. Brain Sci.* **1999**, *22*, 491–517. [[CrossRef](#)] [[PubMed](#)]
15. Huber, M.; van Vliet, M.; Giezenberg, M.; Winkens, B.; Heerkens, Y.; Dagnelie, P.C.; Knottnerus, J.A. Towards a 'patient-centred' operationalisation of the new dynamic concept of health: A mixed methods study. *BMJ Open* **2016**, *6*, e010091. [[CrossRef](#)] [[PubMed](#)]
16. Khanshan, A.; Van Gorp, P.M.E.; Nuijten, R.C.Y.; Markopoulos, P. Assessing the Influence of Physical Activity Upon the Experience Sampling Response Rate on Wrist-Worn Devices. *Int. J. Environ. Res. Public Health* **2021**, *18*, 10593. [[CrossRef](#)] [[PubMed](#)]
17. Nuijten, R.C.; Van Gorp, P.M. Supplementary Material: Health Promotion through Monetary Incentives: Evaluating the Impact of Different Reinforcement Schedules on Engagement Levels with a mHealth App. 2021. Available online: https://figshare.com/articles/dataset/Supplementary_material_Health_promotion_through_monetary_incentives_Evaluating_the_impact_of_different_reinforcement_schedules_on_engagement_levels_with_an_mHealth_app/16692271 (accessed on 19 September 2021).
18. Perski, O.; Blandford, A.; West, R.; Michie, S. Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. *Transl. Behav. Med.* **2016**, *7*, 254–267. [[CrossRef](#)] [[PubMed](#)]
19. Scherer, E.A.; Ben-Zeev, D.; Li, Z.; Kane, J.M. Analyzing mHealth Engagement: Joint Models for Intensively Collected User Engagement Data. *JMIR MHealth UHealth* **2017**, *5*, e1. [[CrossRef](#)] [[PubMed](#)]
20. Donnellan, M.B.; Oswald, F.L.; Baird, B.M.; Lucas, R.E. The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of Personality. *Psychol. Assess.* **2006**, *18*, 192–203. [[CrossRef](#)] [[PubMed](#)]
21. John, O.P.; Srivastava, S. The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of Personality: Theory And Research*; Guilford Press: New York, NY, USA, 1999; pp. 102–138.
22. Ryan, R.M.; Mims, V.; Koestner, R. Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *J. Personal. Soc. Psychol.* **1983**, *45*, 736–750. [[CrossRef](#)]
23. Ray, J.J.; Najman, J.M. The Generalizability of Deferment of Gratification. *J. Soc. Psychol.* **1986**, *126*, 117–119. [[CrossRef](#)]
24. Daugherty, J.R.; Brase, G.L. Taking time to be healthy: Predicting health behaviors with delay discounting and time perspective. *Personal. Individ. Differ.* **2010**, *48*, 202–207. [[CrossRef](#)]
25. Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. *Akaike Information Criterion Statistics*; D. Reidel Publishing Company: Heidelberg, Germany, 1986.
26. Kaltenbach, H.M. Comparing Treatment Groups with Linear Contrasts. In *Statistical Design and Analysis of Biological Experiments*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; Chapter 5, pp. 97–120.5. [[CrossRef](#)]