

Gamifying a BCSS: Rehearsal and reflection in reinforcing a health message response

Piiastiina Tikka*, Miia Laitinen, Iikka Manninen, Harri Oinas-Kukkonen

Oulu Advanced Research on Service and Information Systems, University of Oulu, Finland

* piiastiina.tikka@oulu.fi, P.O.Box 3000, FI-90014 Oulu, Finland

Gamifying a BCSS: Rehearsal and reflection in reinforcing a health message response

Completed Research Paper

A gamified approach to promoting reflection and to engaging users in rehearsing decision-making in a dietary context was studied. The game was based on the principles of the Implicit Attitude Test, but was not a complete IAT in itself like the original test at projectimplicit.net: by categorizing food items under positive or negative associations the players gained points according their categorization speed but the score was merely indicative of attitudinal alignment with the target response. Showing the scores was a vehicle for reflection, and repeated playing constituted response rehearsal. Research questions: 1) does a gamified process of drawing attention to implicit attitudes evoke self-reflection, 2) does gamification of response rehearsal contribute positively to behaviour change? We expected the exposure to one's own choices in the game to heighten the awareness of personal food choices. Experiment participants (N=58) played the game over a five-day period. Constructs of Rehearsal, Reflection, Perceived Persuasiveness, and Perceived Health Behavior Change were analyzed using PLS-SEM. The findings point to rehearsal having a role in how reflection and perceived persuasiveness are connected. Reflection was involved in the self-reported behaviour change, and perceived persuasiveness can promote behaviour change. Open-ended questions showed awareness of choice having a behavioural effect.

Persuasive technology, Behavior change, Gamification, Self-reflection, Perceived persuasiveness, PSD, PLS-SEM

1 Introduction

The role of reflection in behaviour change is well established (Ploderer, Reitberger, Oinas-Kukkonen, and van Gemert-Pijnen, 2014; Li, Dey, and Forlizzi, 2011; Rogers, 2001) and typically the approach is to allow system users to monitor their own behaviour – a common device for such self-monitoring being activity bracelets or smart watches. A plethora of sensors carried by modern information technology products ensures that there is no shortage of means for monitoring what people do physically, but how can technology support turning that monitoring into actual reflection

and, in turn, potential for behaviour change? What mechanisms are required in order to help turn reflection into action?

In health behaviour, monitoring behaviour provides a good view of whether a person is doing what he or she is meant to do: a food diary (provided it is accurate) will reveal if the diet is what it should be, or an activity record from a bracelet will show if there has been enough activity and exercise in a day. It is also known that the attrition rates with systems for behaviour change is rather high (Kelders, Kok, Ossebaard, and van Gemert-Pijnen, 2012). To even start using such a system a person first needs to become aware of a problem. Further to acknowledging a problem as the first step, learning and developing ways of addressing the problem is quite another. Activity monitoring or food diaries are good methods for tracking behaviour and a source for feedback, which a person can use as the basis for reflection. They can be, however, overwhelming in the active effort required.

In the present study we wanted to explore the potential of gamifying a method of feedback provision on decision-making and even implicit attitudes as means of supporting a change in existing thinking. Gaming has been associated with rehearsal (Liborius, 2013; Ghazisaeidi, Safdari, Goodini, Mirzaiee, and Farzi, 2017) and combining a rehearsal with an immediate feedback through the engagement in a game offers an appealing environment for learning and reinforcing target behaviours. We developed a mobile game 'Implicit - The Food Game' based on the principles of the Implicit Association Test (IAT) (Greenwald, McGhee and Schwartz, 1998) where the prevailing implicit associations and attitudes are presumed to be automatic and reactions to relevant stimuli are faster than to stimuli opposite to the automatic attitudes. The game, then, was built to offer the player a choice of food items and positive/negative association words (Manninen and Tikka, 20018), illustrated further in section Materials: the game. We hypothesized that by playing a game where the player has to quickly categorize food items into positive and negative categories would a) expose the player to a possible attitude bias in their thinking as they see the scores of their games and seeing the immediate correct/incorrect feedback as they play the game, and b) through repetition the player would be rehearsing responses to types of food as if learning by rote. Bargh, Chen and Burrows (1996) studied automatic responses in social cognition and concluded that in order for one to change their automatic responses, awareness of the automatic effect is necessary in order for any motivation to change it to be engaged. For example, when the player is consistently slower in placing vegetables into the positive category than placing bacon in the negative category, the message to the player would be that they need to be aware of their automatic choices and preferences, and by playing the game more the player can rehearse and learn the target response. The reflection that takes place can teach the player to observe his or her own thinking as regards target behaviour. Health behaviour in the present study is vegetable and fruit consumption.

In the present study we then ask 1) does a gamified process of drawing attention to implicit attitudes evoke self-reflection, and 2) does gamification of response rehearsal contribute positively to behaviour change? By means of an explorative quasi-experiment the present study examines the relationships between how much a game is played, how much players engage in reflection, how persuasive they eventually deem the game to be as regards the behavioural domain, and how much they perceived the game to affect their own behaviour afterwards. Perceiving a system to be persuasive has been found to impact not only the intention to use the system further, but also the system user's perception of the system's impact on him or her (Lehto, Drozd and Oinas-Kukkonen,

Formatted: English (United States)

2012). Accepting a persuasive effort from a system is one part of processing messages intended to persuade, and depends somewhat on an attitude that allows such an effort (Crano and Prislin, 2006). Without acceptance of the medium, reception of the message is not likely. Behaviour change itself is studied in this experiment through self-reported observation on change, accompanied with open-ended comments from the participants. The study explores the relationships of the constructs and how they potentially lead to the acceptance of the system and to, ultimately, changes in behaviour.

2 Background

The theoretical cornerstones for the present study are found in the Behaviour Change Support System (BCSS) framework and the Persuasive Systems Design (PSD) model (Oinas-Kukkonen, 2013; Oinas-Kukkonen and Harjuma, 2009). Building on this base we used gamification principles (Hamari and Koivisto, 2014) and the Implicit Association Test (IAT) (Greenwald, McGhee and Schwartz, 1998) to create a game mechanism, feedback provisioning (for reflection) and engagement elements in the form of a highly gamified BCSS.

2.1 Persuasive Systems Design and Behavior Change Support Systems

A system that from the onset aims at behavioural and psychological outcomes, but does this openly and without coercion or deception is, by definition, a Behaviour Change Support System (BCSS) (Oinas-Kukkonen, 2013). The aims and goals to be defined when developing a behaviour change support system involve the type and expected outcome of the system: should the system form, reinforce or change a user's compliance, behaviour or attitude? Further to that, the Persuasive Systems Design (PSD) model is a tool for analysing and implementing those goals through system features (Oinas-Kukkonen and Harjuma, 2009).

The development of the game used in the present study and the selection of its persuasive system features was guided by a PSD analysis of the goals, intentions and persuasion context. The main goals of the system were identified as changing behaviour and/or attitude, or reinforcing behaviour and/or attitude where change was not necessary. Gamification aspects of the system used Dialogue Support features *praise* and *rewards*. More at the core of the persuasiveness were Primary Task Support features of *rehearsal*, *self-monitoring* and *reduction*. System credibility support features included reliable sources for health information, transparency in sources, and use of authorities. While the game includes this selection of features, our focus in terms of research is on rehearsal and self-monitoring. Features such as rewards and reduction can be seen as natural components that will be present in such a system as a result of the selected technology context (mobile device) and delivery strategy (gamified BCSS).

Of the core persuasive features the role of rehearsal should be highlighted: playing a game can involve numerous repetitions of the set tasks, which in the present case means numerous repetitions of quick decision-making as regards food items. Such repetition means that the target response is rehearsed at a quick pace in volumes that a person would not readily encounter in real life. A player of such a game also rehearses the target response.

Many persuasive systems are built to encourage reflection (cf. especially self-monitoring in the PSD model) that is expected to lead to change in behaviours while others engage in a more prescriptive approach (Polderer et al., 2014). Open-ended reflection can, perhaps, be seen as a less obtrusive means to an end than a purposefully prescriptive one. Naturally systems can and do use other features and mechanisms as well, but reflection is one approach present in many systems. The PSD model (Oinas-Kukkonen and Harjumaa, 2009) postulates that one of the key elements of a persuasive system is that it is unobtrusive in the sense that user does not require interaction with a system when performing the behaviour: it does not get in the way of a system user's primary task. Reflection, when a by-product of an activity, can be seen as a subtle approach to paying attention to behaviours and to changing them. The feedback from the game allows the player to observe his or her own thinking in a game context. The present study focuses on the effects of rehearsal (repetition of the game) and how exposure to the message (food selections) might on one hand induce reflection and on another support decision-making, and as such have a role in possible behaviour change. In the present study the persuasion features were selected based on context analysis in the development phase (Manninen and Tikka, 2018)

In the area of personal informatics, the data that is used as the basis for reflection tends to include behaviour data and psychological data (Li, Day, Forlizzi, 2011), such as frequency of exercise and heart rate. The purpose of collecting such data is an increase in self-knowledge (Li, Dey, Forlizzi, 2011; Halttu and Oinas-Kukkonen, 2017). Open-ended facilitation of self-reflection allows for interpretative sense-making (Ploderer et al. 2014) and as such, it need not require immediate compliance in a specific target behaviour activity. In the game used in the present study the system users merely play the game, and they are not specifically asked to change their behaviour. The outcomes – the score based feedback from the game – allows the user to observe his or her own thinking in a game context.

Another perspective on self-reflection through feedback is the Cognitive Dissonance theory (Festinger 1957) which, as also taken into account in the PSD model, can be harnessed into behaviour change when a system user is presented with actual and factual feedback about his or her own behaviour. The options, according to Cognitive Dissonance Theory, are to accept the feedback and perhaps act upon the new knowledge, or to reject it in order to reduce the cognitive dissonance (ibid.). That a person would decide to reduce the dissonance by behaviour change is by no means guaranteed. In the game developed for this study the premise is that a) users know from the start that the game is involved with health behaviour and so the message is by no means covert and b) playing the game is a rehearsal of the target behavior (making food choices), where just playing the game already manifests the behaviour advocated by the game. In the latter the conflict between condition and behaviour can be adjusted internally by explaining behavior (making good food choices) with a change in thinking (Festinger and Carlsmith, 1959).

2.2 Gamification

At its simplest, gamification can be defined as using characteristics of gaming (self-purposefulness and hedonistic use) with an ultimately utilitarian goal such as behaviour change or learning (Hamari and Koivisto, 2013). Gamification repurposes the intrinsic motivation that goes with game play as a tool for utilitarian use by using typical game elements of points, badges, leader boards, goals, narratives, feedback and achievements in the system design (Hamari and Koivisto, 2013;

Formatted: English (United States)

Huotari and Hamari, 2012). At the highly gamified end the utilitarian benefits come almost as a side product of a pleasurable activity. On the one hand, gamification of behaviour change could be seen as a form of “sugaring the pill”, but on the other hand the appeal of the gamification approach may be more closely related to the pursuit of unobtrusiveness. Using gamification as an element of a BCSS is not unheard of, and it can be done successfully (Alahäivälä, Oinas-Kukkonen, 2016; Kawachi, 2017; Schoech, Boyas, Blac, and Elias-Lambert, 2013; King, Greaves, Exeter, and Darzi, 2013; Kappen and Orji, 2017). Further to that, cognitive tests have also been used before as the basis of a gamified system Lumsden, Edwards, Lawrence, Coyle and Munafo, 2016).

2.3 *Implicit Association Test*

The IAT (Greenwald, McGhee and Schwartz, 1998) was developed to test implicit attitudes by revealing automatic associations. In the test subjects’ response times to test objects are used to determine automatic associations. While in the present study the developed game does not follow the IAT concept fully so that it could be used to determine a player’s implicit associations as such, the concept of reaction times as an indicator of automatic associations is used as the basic mechanism for scoring and game progression. In research IAT has been used also in healthy eating and diet related studies (Maison, Greenwald, Bruin, 2001; Roefs and Jansen, 2002), which provides an encouraging signal that to base a healthy-eating domain gamified BCSS on the IAT concept is not unjustified. In the present paper the focus is on the effect of rehearsal, and the IAT scores over the test period are not included. This selection was necessary in order to limit the scope of the present paper.

3 Method

3.1 *Model, hypotheses, and measurement instruments*

As an explorative experimental design, the research model (Fig 1) illustrates the constructs of Reflection (REFL), Rehearsal (REH), Perceived Persuasiveness (PEPE), and Perceived Health Behaviour Change (PHBC) and their relationships, as hypothesized. In addition, a further hypothesis (H8) is presented regarding the effect of the study on reported fruit and vegetable consumption. Measurement instruments used in the study are described in Table 5.

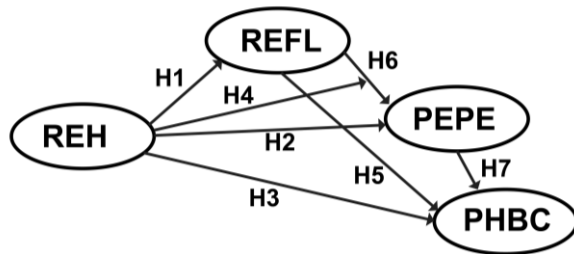


Figure 1. Research model and hypotheses.

Rehearsal (REH) is the number of game sets played by each participant. The PSD model (Oinas-Kukkonen and Harjumaa, 2009) presents rehearsal as a means in a system for a user to rehearse a behaviour. Participants were asked to play a minimum of three sets per day on five consecutive days, and they were told that they were free to play as much as they wished. The more an individual plays the game the more he or she is exposed to the underlying health message and the feedback (score) on his or her (implicit) associations. We hypothesize that the exposure to the feedback through repetition of games will lead to reflection, higher perceived persuasiveness of the system, as well as have a positive effect on post-experiment health behavior. We also hypothesize that REH has a moderating effect on the relationship between REFL and PEPE.

H1	<i>Higher number of game sets played means a higher level of exposure to the health message and feedback, leading to a positive effect on Reflection.</i>
H2	<i>Higher number of game sets played offers a higher level of repetition of the target response and it directs a player to associate healthier foods with positive words (or vice versa). Such exposure will have a positive effect of Rehearsal on Perceived Persuasiveness of the system.</i>
H3	<i>Higher number of game sets will directly enforce the concept of healthier food choices in a repeated way as the player is expected to categorize foods and words. The volume of game play will thus have a positive effect on the Perceived Health Behaviour Change.</i>
H4	<i>The higher exposure to the health message and reinforced target response (correct choices) will have a moderating effect on the relationship between Reflection and Perceived Persuasiveness, as the awareness of player's own choices and their game consequences increase.</i>

Table 1 Hypotheses regarding REH construct.

Reflection (REFL) construct indicates engagement in reflection (Grant, Franklin and Langford, 2002), how individuals inspect and evaluate their own thoughts, feelings and behaviours (Roberts and Stark, 2008). While H1 supposes that playing a game would trigger users to think about their thinking more, it is likely this tendency is already present in some degree in the participants. However, regardless of its origin, reflection is a component in measuring an individual's readiness for purposeful behavior change (Grant, Franklin and Langford, 2002).

H5	<i>Reflection on one's behavior and choices in the game is directly connected to evaluating the qualities of various foods, and high engagement (more games played) in the game will lead to higher reflection and thus have a positive effect on Perceived Health Behaviour Change.</i>
H6	<i>Reflection on one's own behavior and choices in the game is directly connected to evaluating one's own responses to an openly influencing system and will have a positive effect on Perceived Persuasiveness.</i>

Table 2 Hypotheses regarding REFL construct.

Perceived Persuasiveness (PEPE) is a construct combining system user's perception of the system itself and its message into an attitude item that consists of cognitions and affect (Crano and Prislin, 2006; Lehto, Oinas-Kukkonen, and Drozd, 2012). A favourable attitude and a person's assessment of the system and its effect on his or her own behavior or thinking is a factor in a system's ability to support behavior change (Lehto, Oinas-Kukkonen, and Drozd, 2012; Petty, and Cacioppo, 1986).

H7	<i>Perceived Persuasiveness of a system has a positive effect on the readiness of an individual to immediately evaluate the direct impact of the system on his or her behaviour, and supports system users in engaging in target behaviour.</i>
----	---

Table 3 Hypotheses regarding Perceived Persuasiveness construct.

Perceived Health Behavior Change (PHBC) is a construct made of open-ended evaluations by participants on the actual changes they may have noticed after the experiment. The construct is close to PEPE in that it is based on an individual's perception of behavior change impact as regards the system. However, PHBC assesses the direct result of system use, in the present case increased use of fruit and vegetables or contemplation of existing dietary habits with a view to engage in behavior change. Where PEPE as an instrument focuses on evaluating how the user relates to the system and on the respondent's view of the role of the system in any change and personal relevance, the PHBC instrument asks the participants in an open-ended question to describe the exact and actual change (or lack of it). In addition to the PHBC as a measure of behavioural outcome, we collect a one-day food diary at the start and finish of the study period.

<i>H8</i>	<i>The gameplay and the increased reflection triggered by gameplay will result in an increase in actual fruit and vegetable intake after the study period.</i>
-----------	--

Table 4 Hypothesis regarding Perceived Health Behavior Change construct

Measurement	Instruments
Reflection (REFL)	The experiment used a previously validated and published self-assessment scale (Grant, et al., 2002), which was further studied by (Roberts and Stark, 2008). The original scale included three components: need for reflection, engagement in reflection, and insight. Relevant questions from the scale on engagement in reflection were included in the post-test questionnaire.
Perceived persuasiveness (PEPE)	The perceived persuasiveness scale used in this experiment has been developed specifically for assessing BCSSs (Lehto et al. 2012) and as such is a validated scale. It is a self-assessment scale for assessing the impact a persuasive system.
Perceived Health Behavior Change (PHBC)	The instrument is based on an open-ended question presented at the end of the experiment asking participants what (if any) changes they have noticed in their food choices. The scale is then formed by two researchers individually categorizing all the statements into five categories based on the strength and type of effect from no impact to raised awareness, contemplation of change and actual behavioural impact. The ground-up approach of the scale allows users to freely describe their experience rather than having to evaluate their responses against a ready-set frame. The scale has been used earlier in (Tikka and Oinas-Kukkonen, 2017).
Rehearsal (REH)	Collected from the game.
Food intake (1-day food diary)	The present study used a self-administered fruit and vegetable portion estimation form (Cox, Anderson, Reynolds, McKellar, Lean, and Mela, 1998) The questionnaire instruction included guidance for estimating a portion of vegetables and fruit. After guidance, portion-based estimations have been found to be realistic and reliable (Cox et al. 1998). The guidance given in this study is not directly

comparable to (Cox et al., 1998), but it provides a basis for comparable variables for pre- and post-test assessments.

Table 5. Measurement instruments in the study

3.2 Sample selection and study procedure

Sample selection. The sample, N=58, is one of opportunity, largely made up of university students. The researchers distributed fliers at the university campus as well as in shopping centres. A simple advert was also spread through social media (namely Facebook and Twitter). As a thank you, every participant received a cinema ticket. To be applicable, the participants had to be over 18 years old. Cohen's (1992) rules of thumb for sample size are applicable with PLS-SEM analysis (Hair, Hult, Ringle and Sarstedt, 2014) and in order to detect a minimum R^2 level of 0.50 at a 5% significance level, assuming statistical power of 80%, the presented model would require at a minimum 38 observations, which the present study exceeds. To detect an R^2 level of 0.25 at a 5 % significance level, a 59 observations would be required.

Procedure. The study comprised three sections: starting questionnaires (background information and one-day food diary), 5-day period for game play, and final questionnaires (on Reflection, Perceived Persuasiveness, Perceived Health Behaviour Change, and another one-day food diary). Immediately after signing up the participants were given information about the experiment setup (what it involved for them), explicitly asking for their informed consent, and also informing the participants that they were naturally free to withdraw from the experiment at any time and have their data removed. Identity protection procedures were also explained. A unique user ID, obtained from the mobile game, was used in all the forms in order to maintain anonymity.

The participants were instructed to play the game on five consecutive days, a minimum three sets per day, and then finally send their gaming data to the researchers by using a send button in the application. 2-3 days after sending the data, participants received the post-test questionnaire. The flexibility in sending the instruction for the final questionnaire aimed at allowing the users at least one full day after completing the game period before filling in the food diary but also at avoiding the reporting landing on weekend days. By encouraging working days for reporting the intention was to collect the food diaries under as similar conditions as possible.

The interface of the game was in English, which was not the native language for most of the anticipated participants. However, given the prevalence of English in all forms of electronic gaming these days, the expected age range of volunteering participants, and the typical English as a foreign language competence in the test location, language was not deemed to be a major issue in the study setup.

3.3 Materials: the game

The game (example screens in Figure 2) in the present study was developed from the start as a highly gamified BCSS with persuasive features selected based on a Persuasive Systems Design (PSD) analysis (Oinas-Kukkonen and Harjumaa, 2009). As a means of producing the selected PSD

Formatted: English (United States)

features the game adapted the Implicit Association Test (IAT) (Greenwald et al., 1998) as a mechanism for a) producing feedback for the users and b) as the basis for scoring logic (Manninen and Tikka, 2018). The game is not intended to be an IAT in itself, but rather uses the mechanism as a means of providing instant feedback that is entirely based on the player's own reactions and responses. In other words, the score is not based on any game-originating random factors, surprises or obstacles that are not under a player's control.

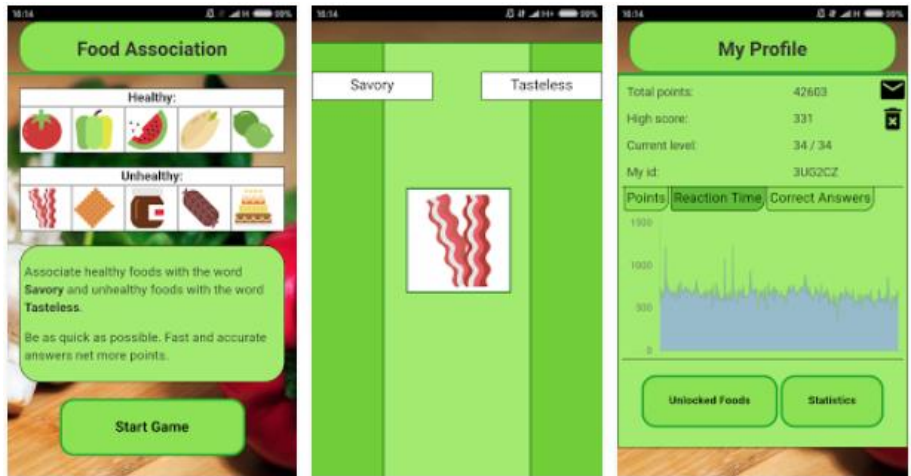


Figure 2. 'Implicitly - The Food Game' start screen (left), categorization task item (centre) where left and right margins are the touch target, and score screen (right).

The game contains two categorization tasks. In the first task, the user has to associate healthy foods with a positive word and unhealthy foods with negative words. In the second task, positive words are associated with a healthy food and negative words with an unhealthy food. The pictures/words appear at centre screen and by tapping on the sides of the screen the player aims at categorizing the items as quickly as possible. Reaction times from the point of presenting the food/word are measured. After each set of ten items a score screen is shown. Scoring was based on reaction thresholds described in (Greenwald et al. (1998) and Nosek and Banaji (2001). Gamification elements included points, levels, and content unlocking (Hamari and Koivisto, 2013).

4 Data analysis and results

4.1 Sample characteristics

Fifty-eight participants took part in the study. Of the total, 33 were female and 25 male. Average age for the whole sample was 24 years (median 23), average for females being 24 (18-35 min/max,

Formatted: English (United States)

median 22) and for males 23 (18-33 min/max, median 23). Eight participants reported that they do not usually play any computerized games, 19 used a computer as their main gaming platform, 20 used a gaming console, and 11 reported a mobile device as their primary gaming platform. This informs us that the participants were, at least, somewhat familiar with typical gaming activities and mechanisms from the player perspective. The sample size satisfies the general rule of thumb of ten times the largest number of paths directed at a particular construct (Hair, et al., 2014), and also at minimum R2 level of 0.50 at a 5% significance level (assumed statistical power 80%) the sample size exceeds the minimum requirement of 38.

4.2 Measurement model

Using PLS-SEM analysis the relationships between latent variables were examined. The analysis, carried out using SmartPLS software, demonstrates the explained variance (R^2 values) in the latent variables, and indicates the strength (β -values) and their statistical significance of the relationships in the model (Hair et al. 2014; Hair, Ringle, and Sarstedt, 2011). See Figure 3 for the R^2 -values and β -values. There were no missing values in the dataset. The measurement model was assessed, following the guidelines presented by Hair et al. (2014), for internal consistency, the reliability of the indicators, and the convergent validity and the discriminant validity. Composite reliability and Cronbach's alpha were both used in the assessment of internal consistency of the model. Indicator reliability was checked with average variance extracted (AVE). Indicator loadings and internal consistency were above 0.708, which is considered acceptable, though for explorative studies values over 0.6 are also acceptable (Hair et al. 2014).

	CA	CR	AVE	1	2	3	4	5
1. REH	1.000	1.000	1.000	<i>1.000</i>				
2. Moderator (REH on REFL>PEPE)	1.000	1.000	1.000	-0.272	<i>1.000</i>			
3. PEPE	0.692	0.868	0.692	0.179	0.325	<i>0.832</i>		
4. PHBC	1.000	1.000	1.000	-0.009	0.080	0.835	<i>1.000</i>	
5. REFL	0.668	0.923	0.668	-0.161	0.191	0.281	0.338	<i>0.817</i>

Convergent reliability is indicated with average variance extracted (AVE) and Fornell-Larcker analysis. Square root of AVE and inner-construct correlations are shown in italics.

Table 6. Internal consistency and indicator reliability assessment.

4.3 Structural model

The research model (Figure 1) presented the hypotheses regarding the relationships in the model.

The results of the PLS analysis are illustrated in Figure 3.

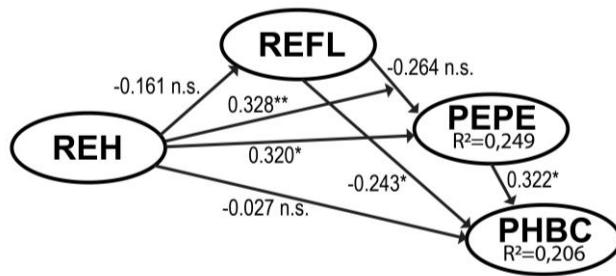


Figure 3. Structural model. ***p<.01,**p<.02,*p<.05, n.s. for non-significant paths.

Formatted: English (United States)

Field Code Changed

Formatted: English (United States)

Formatted: English (United States)

The path coefficients for the research model were obtained using parametric bootstrapping with 5000 subsamples (parallel processing, no sign changes). The constructs in the model were reflective. In the complete model we see that 21% of the variance in perceived health behaviour change (PHBC) was explained by engagement in reflection (REFL) and perceived persuasiveness (PEPE) (supporting H5 and H7), and 25% of the variance in PEPE was explained by amount of rehearsal (REH) (supporting H2 and H4); REH was a significant moderator in the relationship between REFL and PEPE. Hypotheses H1, H3 and H6 were not supported.

4.4 Total effects and effect sizes, and predictive validity of the model

Practical relevance of the model was determined by assessing the total effects and effect sizes (Cohen f^2), assessing an exogenous construct's contribution to the R^2 value of an endogenous latent variable (Table 7). Assessment guide values were 0.02 (small), 0.15 (medium), and 0.35 (large). In the present model there were five small, one non-relevant and one medium f^2 , indicating that on the whole there is some practical relevance in the model. The non-relevant f^2 was REH to PHBC, and the medium f^2 was REH as a moderator for REFL and PEPE.

	PEPE	PHBC	REFL
REH	0.320 (0.125)	n.s.	n.s.
Moderator (REH on REFL>PEPE)	0.328 (0.157)		
PEPE		0.322 (0.113)	
PHBC			

REFL	n.s.	0.243 (0.065)	
------	------	------------------	--

Table 7 Total effects and effect sizes (non-significant not reported).

A blindfolding procedure was used to observe the predictive validity of the model (Table 8). The Stone-Geisser cross-validated redundancy value (Q^2) above 0 is considered to indicate predictive validity of endogenous constructs. All endogenous constructs (PEPE, PHBC and REFL) demonstrate $Q^2 > 0$, and thus indicate the path model's predictive relevance to each of the constructs (Hair et al. 2014).

Construct	Q^2
PEPE	0.108
PHBC	0.122
REFL	0.011

Table 8 . Predictive validity of the model: Q^2 values for endogenous constructs.

4.5 Fruit and vegetable consumption before and after the study period

A dependent means t-test was used to compare the effect of timing of self-reported fruit and vegetable consumption before the game play period and immediately after it (Table 9). One significant outlier was identified and removed, after which the data sets were tested to be normally distributed (Shapiro-Wilk test $p > .05$ for both pre-test and post-test data). On average, at α -level of .05, participants reported consuming more vegetables after participating in the experiment ($M=5.00$, $SE=.350$) than before the experiment ($M=3.78$, $SE=3.16$, $t(56)=-4.091$, $p<.001$, Cohen's $d=0.54$) The result provides indirect support for the use of PEPE and PHBC constructs in assessing the effect of system use on behavior through showing that the self-reported behavior observations are in line with the food diaries. Cohen's d (0.54) indicates a medium effect size. The result supports H8.

	N	Mean	Std.Dev.
Pre-study vegetable consumption	57	3.78	2.385
Post-study vegetable consumption	57	5.00	2.639

Table 9 . Descriptive statistics for dependent means t-test.

4.6 Correct and incorrect categorizations and reaction times

We checked simple correlations between correct and incorrect categorizations and reaction times to healthy and unhealthy foods. There was a moderate positive correlation (Pearson's $r=0.6$) of categorizations of healthy foods with positive terms or positive words with healthier foods and reaction times to good foods (higher response time), that is, increase in correct responses was accompanied with an increase in response time where good foods were concerned. There was also a moderate negative correlation (Pearson's $r=-0.6$) between incorrect categorization and reaction time to healthier foods; in other words increase in incorrect responses is accompanied by a reduction in reaction time – faster responses producing more incorrect responses. Considering how implicit measures of attitudes assume that implicit attitudes are revealed through reaction times, these moderate correlations are more than a nod in the direction of the game revealing a slight bias in the present sample towards the less-healthy foods. To categorize healthier foods correctly, the participants took more time, and when reaction times were faster, there were more incorrect categorizations. Given the relatively low levels of incorrect categorizations, 7,5% of all responses were incorrect, it should be fair to assess that the health message and the basis of expected categorizations to healthy and less-healthy foods was clear to the participants. From the correlations, then, we interpret that with a little bit of thinking (that is, higher reaction time), it was possible for the participants to provide the correct response most of the time.

4.7 Open-ended responses: did the system have an effect on me?

The Perceived Health Behavior Change (PHBC) construct in the study was built on the basis of open-ended responses to the question "did playing the game have any effect on you?" The responses were categorized according to change strength (or lack of change) into five categories where the one indicated no effect at all, and five indicated a clear, identified change in health behavior. The categorization was done by two researches independently and the statements were handled separately from any other data. The majority of the responses fell in the bottom and the middle of the scale: while some participants were very adamant the game had not affected them in any way, some claimed there was no effect except they were more mindful of what they were eating or buying in the shop as a consequence.

When considering the following examples of the open-ended responses, the results of the t-test on actual fruit and vegetable consumption before and after the test makes for an interesting backdrop: on average, fruit and vegetable consumption went up from fewer than four to five portions on the polled days. The statements have been translated from their original language by a professional. The sample statements come predominantly from female participants – they seemed to have been more willing to elaborate on their experience than the male participants.

No effect (Rating 1):

Responses in this category typically simply stated "No", "Not in the least", "Not really", or "Not that I noticed", with not much if any elaboration. Some participants wanted to explain how there was no effect because they were already engaging in the promoted behavior:

"Taking part in this experiment did not change my dietary behavior, because I already pay attention to eating vegetables." (Female participant played the game 116 times).

"There was no effect. I eat healthy food every day already." (Female participant played the game 256 times).

Another participant did not wish to be persuaded:

"No [the game had no effect], because I would not let it affect my opinions." (Female participant played the game 14 times).

Two participants found the game too forward and manipulative, which made them object to it – one of them did not find the game to have any effect:

"Not really. The game was in a way doing a bit of a brainwash, which made me want to object to its motivation." (Female participant played the game 63 times).

Minor effect (Rating 2):

The second participant to describe the game to have brainwash-tendencies did not object so strongly:

"Not much change, this almost felt like brainwash. Good game. If nothing else it got me to remember certain adjectives about different foods and then remember that some food was good or bad according to the game." (Female participant played the game 220 times).

Otherwise in this category the game was experienced as perhaps increasing awareness of foods, or participants admitted to a small effect.

"I don't think there was much change, but it did result in some active thinking with a view to healthier eating." (Female participant played the game 55 times).

"There was only little effect, if at all. I've never had a sweet tooth and I've always since I was little eaten a lot of vegetables. So in that respect I may not be an ideal subject for this study." (Male participant played the game 78 times).

Moderate effect (Rating 3):

This category includes statements where participants express increased thinking of healthier eating, more so than in the Rating 2 category.

"This got me thinking about fruit more, and in the supermarket I was thinking I really should buy something healthy. Mind, that may also have been because I've been ill and I know I should eat more vitamins. [The game] may have reinforced my existing thoughts. I do think about food choices quite actively as it is. Probably

Formatted: English (United States)

no huge impact on my choices as such." (Female participant played the game 77 times).

Some responses also revealed that identifying change is not always easy, as seen by a participant who does not admit to change while stating that he did do something differently as a consequence:

"Taking part in the study did not make me change my diet. It did perhaps make me buy a wider variety of vegetables, but it didn't affect the volumes." (Male participant played the game 39 times).

"Didn't notice [a change] but this did get me to thinking about my own dietary habits and could well have some effect on me later." (Male participant played the game 29 times).

Medium effect (Rating 4):

When a participant was able to point to a direct behavioral change that they connected with taking part in the study, the statements were rated at 4.

"Yes [there was an effect], I did not eat my sweets because I thought of this game." (Female participant played the game 73 times).

"This encouraged me to buy more greens at the store." (Female participant played the game 191 times).

"I have consumed more fruit." (Male participant played the game 34 times).

Strong effect (Rating 5):

No statements were categorized at 5 in this study. The statements in this category would need to express a clearly verbalized development of insight or intention to change current dietary habits or to increase fruit and vegetable consumption with a conscious effort.

5 Discussion

Successful design of Behavior Change Support Systems involves selecting the right persuasive tools for the purpose (Oinas-Kukkonen and Harjuma, 2009; Fogg, 2003). The forms of BCSSs are various from simple information sharing websites to interactive mobile applications to, as in the present paper, gamified systems. Whatever the format, the basic persuasive design principles can be applied. In the game presented in this paper the selected persuasive design principles (Oinas-Kukkonen and Harjuma, 2009) were built on top of a cognitive test concept and they were implemented as gaming features, following game development principles (Hamari and Koivisto, 2013).

In our study the game provided the users with an opportunity to rehearse a target behavior and also to reflect upon the rehearsal results with the help of immediate feedback. The main interest in the paper was in the effect of rehearsal on reflection leading on to perceived persuasiveness and health behavior change. Research into reflection suggests that unless the

reflection leads to insight, behavior change is less likely to take place (Fogg, 2003; Lyke, 2009; Halttu and Oinas-Kukkonen, 2017), and dysfunctional attitudes can hinder the path from insight to personal well-being (Stein and Grant, 2014). A review into use of digital games in cancer management found potential (and also challenges) in both treatment training but more notably for example when active participation or behavioural rehearsal for physical and psychosocial activity are needed (Ghazisaeidi, Safdari, Goodini, Mirzaiee, and Farzi, 2017).

The moderating role of rehearsal on the relationship between reflection and perceived persuasiveness would appear to be an element persuasive systems should consider utilizing when intending to facilitate behaviour change through reflection. Triggering reflection, as can be seen from the present study, is not entirely straight-forward. Internal state awareness, the engagement and need for reflection (Grant, et al. 2002) is a characteristic that needn't only be triggered and thus it may be difficult to distinguish in small experiments how much of such engagement occurs naturally and how much is triggered through interactions with people or with technology. In the present experiment a significant relationship between amount of rehearsal and engagement in reflection was not established, yet reflection affected the perceived health behavior change construct in the model. Was the key factor here the conspicuous lack of goal-setting, which would give self-awareness a clear function of observing one's own actions in order to establish self-regulation (Grant et al., 2002)? In order to cater for a path from self-monitoring (feedback on actions in the system in the present experiment) to reflection, whether on-action or in-action (Ploderer et al. 2014), it seems necessary to understand what other parameters **besides** dose are necessary in order for feedback to turn into self-reflection? The amount of gameplay in the present experiment varied from barely over 10 sets to some 260 sets, yet there was no link between exposure to the feedback or the health message and engagement in reflection.

While rehearsal did not affect self-reflection directly, some reflection on the established problem domain seemed to take place. The open-ended responses from participants, when simply asked if the system affected them at all, repeatedly mentioned that the game nudged the players to be more aware of their eating, to notice differences between foods, and even select foods differently. In some comments, even when participants insisted the game had had no effect, they then went on to admit that it has made them think about foods or their diet at least in some degree. This, admittedly, can be the result of any intervention or study where a specific problem domain is suddenly brought to people's attention, but the present experiment goes to show that a fairly light-hearted, fully gamified and simple system can achieve this new attention as well without diluting the core health message too much. There is also the population starting level to consider as regards informed food choices: many of those participants who stated the game had no effect on their behaviour also said they felt they already knew how to make healthy choices and to eat vegetables and fruit. Such responses, then, beg the question if the impact of gameplay might be different amongst a population not quite so well versed in healthy eating as the sample in the present study?

What we can learn from the results is that rehearsal can amplify the effect of reflection as regards a user's perception of the persuasiveness of a system. We can also see that rehearsal alone may not be enough to result in positive behavioural outcomes: volume of game play did not affect the immediate health behaviour directly in the way it did Perceived Persuasiveness. In turn, Perceived Persuasiveness had an impact on the immediate health behavior, as did Reflection. In other words, for a change to take place, both reflection and perceived persuasiveness are necessary and they can be made more effective through rehearsal. The perceived outcomes were supported in

Formatted: English (United States)

the study with the actual fruit and vegetable consumption estimates, as seen in the t-test on the consumption of vegetable and fruit before and after the test period. Implicit responses tend to manifest under time pressure whereas without the time pressure a person has the opportunity to make considered choices (Friese, Wänke and Plessner, 2002). The results in the present study as regards the perceived health behaviour change (PHBC) very much reflect on the situation where time pressure is not present: the respondents report on a time after playing the game as they have carried on with everyday life. In this study the score of correct/incorrect choices was the time-pressured variable considered a vehicle for reflection, and the outcome was the self-reported effect of the period of playing the game (perceived health behaviour change).

The presented study naturally does not provide insights into long-term effects. The gameplay period was also relatively short (5 days) and game content was limited (only 34 levels). Further insights into the direct role of various persuasive features (such as rewards and praise) would have made for an even richer study set-up, but owing to the necessity to have participants commit to play the game for five days it was necessary to be parsimonious as regards the extent of the questionnaire batteries and demands on participant time. As existing research shows, for example (Roefs and Janesen, 2002), we can see that the relationship between implicit attitudes and behavior can be complicated. Therefore, future research directions leading on from the present study should involve the actual implicit association directions and strengths and their development after reflection. Other further research directions could be found in operationalising implicit attitude awareness and behaviour change through information systems.

We cannot emphasize enough the preliminary and exploratory nature of the presented study. Using a game setting outside strictly controlled laboratory setting allowed, at this stage, for a setup where we were able to observe initial reactions and potential for further theorizing on the discussed topics. Any further research in an experimental setting will require controlling not only the key variables (feedback and rehearsal) more closely, but also involving a control group for more robust observations on actual behavioural effects. In the present study the effects of feedback have not been fully separated from the effects of rehearsal, given that the game is available only in one format that always provided the feedback to the players. As it is not possible to fully distinguish between the effects of feedback and rehearsal, we can only conclude that the observable outcomes of the present study may be attributed to any or both of these two behaviour change mechanisms. As it is, then, the present paper offers insights into the relationships between feasible factors in the process of reflection and gamified rehearsal based behaviour change supporting further exploration of the topics for both academics and practitioners.

As regards the theoretical foundations of this study, the interconnectedness of rehearsal and feedback are present there as well. The study assumes that repeated exposure to feedback (scores and correct/incorrect response feedback) that comes from repeatedly playing the game (rehearsal) promotes reflection. The results in the present study were inconclusive as to this relationship: why would exposure to one's performance not evoke reflection? If a system relies on simple feedback on user actions to generate such reflection that should support behaviour change, how well-founded are such expectations? With any future research on the topic, it will be necessary to address the balance between automatizing responses through rehearsal as opposed to evoking reflection through repeated feedback on that rehearsal. Such consideration should be of interest and great relevance to the field of behaviour change systems. Further research needs arising from the learnings of this paper include the exploration of feasible sub-groups as regards significance of

Formatted: English (United States)

reflection upon perceived persuasiveness (to which at averages rehearsal acted as a significant moderator at present) and understanding the role of goal-setting even in a context where no behavior target was explicitly set.

References

- Alahäivälä, T., and Oinas-Kukkonen, H. 2016. "Understanding persuasion contexts in health gamification: a systematic analysis of gamified health behavior change support systems literature." *International journal of medical informatics*, 96, 62-70
- Bargh, J.A., Chen, M., and Burrows, L. 1996. Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action. *Journal of Personality and Social Psychology*, Vol. 71, No. 2, 230 – 244.
- Cox, D.N., Anderson, A.S., Reynolds, J., McKellar, S., Lean, M.E.J., Mela, D.J. 1998. "Take five, a nutrition education intervention to increase fruit and vegetable intakes: impact on consumer choice and nutrient intakes." *Br J Nutr* 80(2):123-131
- Crano, W. D., and Prislin, R. 2006. "Attitudes and Persuasion." *Annu. Rev. Psychol* 57, pp. 345-74
- Festinger, L.A. and Carlsmith, J.M. 1959. "Cognitive consequences of forced compliance." *The Journal of Abnormal and Social Psychology* 58(2), 203.
- Fogg B.J. 2003. *Persuasive technology: using computers to change what we think and do*. San Francisco: Morgan Kaufmann.
- Fornell C., Larcker D. 1981. "Evaluating structural equation models with unobservable variables and measurement error." *J Mark Res* 1(18):39-50.
- Friese M., Wänke, M., and Plessner, H. 2006. Implicit Consumer Preferences and Their Influences on Product Choice. *Psychology and Marketing*, Vol. 23(9): 727-740.
- Gefen D., Ringdon E.E., and Straub D.W. 2011." Editor's comment: an update and extension to SEM guidelines for administrative and social science research." *MIS Q* 35(2):3-14.
- Ghazisaeidi, M., Safdari, R., Goodini, A., Mirzaiee, M., & Farzi, J. 2017. "Digital games as an effective approach for cancer management: Opportunities and challenges." *Journal of Education and Health Promotion* 6, 30.
- Grant, A.M., Fraklin, J. and Langford, P. 2002. "The Self-reflection and Insight Scale: A new measure of private self-consciousness." *Social Behavior and Personality: an international journal*, Vol. 30, November 8, pp. 821-835(15).
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. 1998. "Measuring individual differences in implicit cognition: the implicit association test." *Journal of personality and social psychology* 74(6), 1464.

Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M. 2014. *A primer on partial least squares structural equation modeling (PLS-SEM)*. Los Angeles: SAGE Publications..

Formatted: English (United States)

Hair J.F., Ringle C.M., Sarstedt M. 2011. "PLS-SEM: indeed a silver bullet." *The Journal of Marketing Theory and Practice* 19(2):139-152

Halttu, K. and Oinas-Kukkonen, H. 2017. "Persuading to Reflect: Role of Reflection and Insight in Persuasive Systems Design for Physical Health." *Human-Computer Interaction*, Vol. 00, pp. 1- 32 .

Formatted: English (United States)

Hamari, J., and Koivisto, J. 2013. "Social motivations to use gamification: An empirical study of gamifying exercise." *Proceedings of the ECIS 2013 European Conference on Information Systems*.

Huotari, K. and Hamari, J. 2012. "Defining Gamification – A Service Marketing Perspective." In proceedings of *MindTrek 2012*, October 3-5 (2012).

Kappen, D.L. and Orji, R. 2017. "Gamified and Persuasive Systems as Behavior Change Agents for Health and Wellness." In *XRDS: Crossroads, The ACM Magazine for Students* 24, 1: 52-55. ACM

Kawachi, I. 2017. "It's All in the Game – The Uses of Gamification to Motivate Behavior Change." *JAMA Internal Medicine* 177 (11).

Kelders, S. M., Kok, R. N., Ossebaard, H. C. and Van Gemert-Pijnen, J. E. 2012. "Persuasive System Design Does Matter: A Systematic Review of Adherence to Web-Based Interventions." *Journal of Medical Internet Research*, 14(6)

King, D., Greaves, F., Exeter, C., & Darzi, A. 2013. "Gamification Influencing health behaviours with games." *Journal of the Royal Society of Medicine*, 106(3), 76-78

Ploderer, B., Reitberger, W., Oinas-Kukkonen, H., and van Gemert-Pijnen, J. 2014. "Social interaction and reflection for behaviour change." *Personal and Ubiquitous Computing*.
<http://doi.org/10.1007/s00779-014-0779-y>

Lehto, T., Oinas-Kukkonen, H., and Drozd, F. 2012. "Factors affecting perceived persuasiveness of a behavior change support system." *Proc. of the International Conference on Information Systems (ICIS 2012)*.

Li, I., Dey, A. K., & Forlizzi, J. 2011. "Understanding my data, myself: Supporting selfreflection with ubicomp technologies." *Proceedings of the UbiComp 2011 International Conference on Ubiquitous Computing*. New York, NY: ACM.

Liboriussen, B. 2013. "Craft, creativity, computer games: The fusion of play and material consciousness." *Philosophy & Technology*, 26(3), 273-282 .

Formatted: English (United States)

- Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., and Munafò, M. R. 2016. "Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy." *JMIR Serious Games*, 4(2), e11.
- Lyke, J. A. 2009. "Insight, but not self-reflection, is related to subjective well-being." *Personality and Individual Differences*, 46, 66–70
- Oinas-Kukkonen, H. 2013. "A foundation for the study of behavior change support systems." *Pers Ubiquit Comput* 17(6):1223–1235.
- Maison, D., Greenwald, A.G. and Bruin, R. 2001. "The Implicit Association Test as a Measure of Implicit Consumer Attitudes." *Polish Psychological Bulletin*, 32, 1-9.
- Manninen, I. and Tikka, P. 2018. "Developing a Gamified Behavior Change Support System: Case Implicity – The Food Game." *Sixth International Workshop on Behavior Change Support Systems (BCSS 2018)*.
- Nosek, B. A., and Banaji, M. R. 2001. "The go/no-go association task." *Social cognition*, 19(6), 625-666.
- Nunnally J.C, Bernstein I. 1994. *Psychometric theory*. New York: McGraw-Hill.
- Oinas-Kukkonen, H. and Harjumaa, M. 2009. "Persuasive systems design: key issues, process model, and system features." *Communications of the Association for Information Systems*, 24.
- Oinas-Kukkonen, H. 2013. "A foundation for the study of behavior change support systems." *Pers Ubiquit Comput* 17(6):1223–1235
- Petty R.E. and Cacioppo J.T. 1986. "The elaboration likelihood model of persuasion." *Advances in Experimental Social Psychology*, Vol 19.
- Roefs, A., and Jansen, A. 2002. "Implicit and explicit attitudes toward high-fat foods in obesity." *Journal of Abnormal Psychology*, 111(3), 517-521.
- Roberts, C., and Stark, P. 2008. "Readiness for self-directed change in professional behaviours: Factorial validation of the self-reflection and insight scale." *Medical Education*, 42, 1054–1063 .
- Rogers, R. R.2001. "Reflection in higher education: A concept analysis." *Innovative Higher Education*, 26, 37–57
- Schoech, D., Boyas, J.F., Black, B.M., and Elias-Lambert, N.2013. "Gamification for Behavior Change: Lesons from Developing a Social, Multiuser, Web-Tablet Based Prevention Game for Youths." *Journal of Technology in Human Services*, 31:2, pp. 197-217.

Formatted: English (United States)

Formatted: English (United States)

Stein, D., and Grant, A. M. 2014. "Disentangling the relationships among self-reflection, insight, and subjective well-being: The role of dysfunctional attitudes and core self-evaluations." *The Journal of Psychology*, 148, 505–522.

Tikka, P. and Oinas-Kukkonen, H. 2017. "Contributing or receiving – the role of social interaction styles in persuasion over a social networking platform." *Pers Ubiquit Comput* 21:705–721.