



| | |
|-------------------------------------|---|
| Title | Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques |
| Authors(s) | Lumsden, Jim, Skinner, Andy, Coyle, David, et al. |
| Publication date | 2017-11-22 |
| Publication information | Lumsden, Jim, Andy Skinner, David Coyle, and et al. "Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques" 19, no. 11 (November 22, 2017). |
| Publisher | JMIR |
| Item record/more information | http://hdl.handle.net/10197/9115 |
| Publisher's statement | This work was published in the Journal of Medical Internet Research, and is licensed under the Creative Commons Attribution License CC-BY 2.0. |
| Publisher's version (DOI) | 10.2196/jmir.8473 |

Downloaded 2023-10-05T14:16:07Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)



© Some rights reserved. For more information

Original Paper

Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques

Jim Lumsden^{1,2}, MSc; Andy Skinner^{1,2}, PhD.; David Coyle³, PhD; Natalia Lawrence⁴, PhD; Marcus Munafò^{1,2}, PhD

¹UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, United Kingdom

²The MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom

³School of Computer Science, University College Dublin, Dublin, Ireland

⁴School of Psychology, College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom

Corresponding Author:

Jim Lumsden, MSc

UK Centre for Tobacco and Alcohol Studies

School of Experimental Psychology

University of Bristol

12a Priory Rd

Bristol, BS8 1TU

United Kingdom

Phone: 44 117 92 88450

Email: jim.lumsden@bristol.ac.uk

Abstract

Background: The prospect of assessing cognition longitudinally and remotely is attractive to researchers, health practitioners, and pharmaceutical companies alike. However, such repeated testing regimes place a considerable burden on participants, and with cognitive tasks typically being regarded as effortful and unengaging, these studies may experience high levels of participant attrition. One potential solution is to gamify these tasks to make them more engaging: increasing participant willingness to take part and reducing attrition. However, such an approach must balance task validity with the introduction of entertaining gamelike elements.

Objective: This study aims to investigate the effects of gamelike features on participant attrition using a between-subjects, longitudinal Web-based testing study.

Methods: We used three variants of a common cognitive task, the Stop Signal Task (SST), with a single gamelike feature in each: one variant where points were rewarded for performing optimally; another where the task was given a graphical theme; and a third variant, which was a standard SST and served as a control condition. Participants completed four compulsory test sessions over 4 consecutive days before entering a 6-day voluntary testing period where they faced a daily decision to either drop out or continue taking part. Participants were paid for each session they completed.

Results: A total of 482 participants signed up to take part in the study, with 265 completing the requisite four consecutive test sessions. No evidence of an effect of gamification on attrition was observed. A log-rank test showed no evidence of a difference in dropout rates between task variants ($\chi^2_2=3.0$, $P=.22$), and a one-way analysis of variance of the mean number of sessions completed per participant in each variant also showed no evidence of a difference ($F_{2,262}=1.534$, $P=.21$, partial $\eta^2=0.012$).

Conclusions: Our findings raise doubts about the ability of gamification to reduce attrition from longitudinal cognitive testing studies.

(*J Med Internet Res* 2017;19(11):e395) doi:[10.2196/jmir.8473](https://doi.org/10.2196/jmir.8473)

KEYWORDS

behavioral research/methods; games, experimental; computers; cognition; Internet; play and playthings/psychology; boredom; task performance and analysis; executive function; inhibition (psychology)

Introduction

The prospect of assessing cognition remotely and longitudinally is attractive to researchers, health practitioners, and pharmaceutical companies alike. However, traditionally, assessments of cognitive functioning are performed in a laboratory or clinical setting, making multiple testing sessions expensive and a burden to both researchers and participants. At the time of writing, the use of Internet-based platforms for crowdsourcing participants, such as Amazon MTurk [1] and Prolific Academic [2], combined with the growing number of platforms for delivering Web-based cognitive assessments, such as Testable [3] and Gorilla [4], has newly given researchers the ability to gather data on large numbers of people within very short time spans [5-8]. These new technologies have allowed psychological experiments and interventions to be delivered via the Web easily and inexpensively [9-11].

However, one issue for Web-based studies (and particularly longitudinal studies) is that they must compete against the wealth of entertainment and distraction available on the Internet to attract and retain their participants. This is made more difficult by the fact that dropping out of a Web-based study is easier than doing so in the laboratory, that is, a participant needs to only close their browser window [12]. Many authors have reported difficulties sustaining participant numbers for the duration of their Internet-based studies [13,14], and reviews of adherence to intervention trials have documented dropout rates of around 50% [15,16], considerably higher than in laboratory studies where dropout rates are around 13% [17]. The gradual reduction in the number of participants who continue to provide study data over time is known as attrition [16,18]. High levels of attrition may cause studies to suffer from smaller than intended sample sizes, incomplete datasets, wasted participant compensation, and potentially biased results [19-21].

Attrition is often characterized as a “lack of participant engagement” [22,23], but the definition of *engagement* is unclear. One potential definition conceptualizes engagement in a twofold sense [24], both referring to participants’ subjective experience of taking part in a study (ie, their enjoyment of the procedure) and participants’ behavior when interacting with the study (ie, how often they return to the study website, or how quickly they drop out). Under this definition, attrition is a subcomponent of engagement: an objective behavioral measure that could be assumed to relate to the concept of engagement as a whole.

Gamification has been heralded as a potential mechanism for increasing participant engagement with Web-based studies and interventions [25-27]. The premise is that by adding gamelike features (points, graphics, levels, competition, etc) to an otherwise mundane task, we might be able to create a more enjoyable and compelling experience for the user [28-30]. By utilizing games’ ability to engage individuals, it may be possible to make the testing experience less burdensome, thereby reducing attrition. In previous studies, self-report questionnaires of participants’ enjoyment have found that gamelike experiments are typically rated as more enjoyable than their nongamelike counterparts [25,31-35]. There are also some examples of

gamification increasing *objective* measures of engagement, such as number of optional trials completed [34] or the number of optional testing blocks chosen [36].

Two systematic reviews looked at the effect of gamification on engagement with “online programs” (mostly e-learning) [37] and Web-based mental health interventions [38]. Drawing on the data from 15 studies comparing engagement with gamified programs with nongamified programs, Looyestyn et al [37] found medium to large effects of gamification on objective measures of engagement such as time spent using the program, number of website visits, and volume of contributions. In contrast, Brown et al [38] assessed the impact of gamification on adherence to 61 Internet-based mental health interventions and found that not only was gamification applied fairly lightly (most studies used only one gamelike element) but there was also little evidence for its efficacy. These conflicting findings could be the result of the reviews’ different scopes, the lack of studies in Brown’s review which specifically assessed the impact of gamification on adherence, or the very minimal gamification found to have been applied in the reviewed mental health interventions.

The reluctance of researchers to liberally apply gamification to precisely designed mental health interventions is understandable; any small change might impact the intervention’s efficacy. Within our own field of gamified cognitive assessment, efforts to increase participant engagement and reduce attrition must be implemented carefully to avoid introducing additional cognitive load and affecting the cognitive constructs under test, thus invalidating the task. Although some studies have reported a positive effect of game mechanics on participant performance [39-41], others have found evidence that gamelike tests *do not* improve performance, and may in fact worsen it [31-33,42-44]. For example, Katz et al [42] found that adding a point scoring system to a working memory training task negatively impacted the task’s ability to train cognition. These contrasted findings are likely due to the diverse range of cognitive tasks being used and the variety of gamification approaches applied to them, thus highlighting the need for research, which systematically manipulates gamification techniques within a single type of task [25].

We previously conducted a study exploring the impact of two simple game mechanics (points and theme) on the data collected by, and subjective participant ratings of, a response inhibition task [43]. The points variant rewarded participants with points in accordance with their performance on the task, whereas the theme variant utilized a variety of narratively themed stimuli and task backgrounds. A nongame variant was included as a control condition. This was comparable with a clinical version of the task, with some minor graphical changes to ensure suitability for online use. We found that points were rated highest of the three variants on a subjective questionnaire of enjoyment and engagement and did not negatively affect participant performance on the test. However, we found that the narratively themed task was less liked and negatively affected participant performance. We also observed ceiling effects on participant accuracy in all three task variants because of the ease of the response inhibition task we used.

In this study, we aimed to investigate whether simple gamification could reduce participant attrition from a Web-based longitudinal cognitive testing study. Building on our previous study, we used three variants of a response inhibition task, but we switched to using the Stop Signal Task (SST) to increase task difficulty and avoid ceiling effects. We used the same gamelike features (nongame, points, and theme) as in the previous study. We aimed to assess the effect of gamification on attrition using a longitudinal design whereby participants signed up to four compulsory test sessions over 4 consecutive days before entering a 6-day voluntary period where they could continue to take part once per day if they desired. Participants were told that they would receive £4 for completing all compulsory sessions and an additional 50p for each optional session they completed.

We hypothesized that the nongame variant would have the highest attrition rate, losing participants quickly once the fourth session was complete. We expected the points variant initially to maintain high numbers before falling rapidly around day 6 and 7. Finally, we expected the theme variant to lose participants steadily at first before stabilizing to a low attrition rate, eventually retaining a higher number of participants than either the nongame or points variants. For more information on why we predicted these hypotheses, see [Multimedia Appendix 1](#).

Methods

Design and Overview

We used a between-subjects repeated measures experimental design that took place online over 4 to 10 days. The independent variable was SST variant (nongame, points, and theme). The dependent variables of interest were participant attrition, scores on a questionnaire of enjoyment and engagement, two pilot objective measures of engagement, and stop signal reaction times (SSRTs). We preregistered the study on the Open Science Framework [45].

Participants and Procedure

Participants were recruited from the user base of Prolific Academic [2], which handles the process of checking inclusion criteria, displaying study information, and participant reimbursement. We required participants to be older than 18 years and to have English as a first language but had no further criteria. Once registered, participants were directed to the *Mindgames platform* where they entered their prolific ID and received a unique link, which they used to access the study thereafter. They were then randomly assigned to a single task variant for the duration of the study and completed a Web-based consent form before the testing commenced.

Participants were required to complete one 10-min session per day for the first 4 days of the study to receive £4 as compensation for their time. If participants dropped out of the study before completing four sessions and did not contact us with a reason (technical difficulties, etc), then they did not receive any compensation. This was made clear on the information sheet, which participants read before they signed up to the study, and on the study website itself. For the first four sessions, participants were sent daily reminders via the Prolific Academic messaging system. On the fourth day, participants were informed that there would be no more reminders, and that they were free to either drop out or continue to take part in the study each day thereafter for up to 6 days, with each additional session earning them 50p, for a total of between £4 and £7.

The appropriate compensation for the optional sessions was determined by way of a pilot study using the nongame variant only. We randomly allocated participants to one of the three levels of compensation, that is, 50p, £1, or £2 per optional session completed (the base compensation was still £4) and found that the average number of sessions completed per participant was 7.1, 8.4, and 9.4, respectively. Given that we anticipated the nongame variant to be the least motivating of the three variants, that we wanted to avoid ceiling effects, and that we wanted to minimize the motivational influence of the compensation, we opted for a reward of 50p per optional session.

Ethics approval was obtained from the Faculty of Science Research Ethics Committee at the University of Bristol (40361), and the study was conducted according to the revised Declaration of Helsinki [46].

Materials

The Mindgames Platform

Aside from participant recruitment, daily reminders, and reimbursement, all other elements of the study were hosted on a custom website [47]. The website was a single-page Web app written in JavaScript, with a JSON-based Firebase database [48] and PixiJS [49] as the 2D renderer. The site opened to a main menu screen from which the participant could view the number of sessions they had completed and the amount of money they had earned so far (Figures 1-3). Participants had access to a *history* screen, which allowed them to view their previous progress and monitor their results over time. Clicking the start button displayed a series of instruction screens, followed by the SST task and a short questionnaire. The session ended on the history screen, and the main menu's *start* button became inactive until midnight that night. Each session took approximately 10 min to complete. On the first day of taking part, participants also completed a short demographic questionnaire, which collected data on age, sex, ethnicity, level of education, and the number of hours spent playing video games each week.

Figure 1. Menu screen of the nongame task variant.



Figure 2. Menu screen of the points task variant.



Figure 3. Menu screen of the theme task variant.



Stop Signal Task: Nongame Variant

The SST measures response inhibition [50,51], a key feature of executive control [52]. It tests the participant's "action restraint" by presenting a series of stimuli to which the participant must respond as quickly as possible but are occasionally required to withhold a response. These *stop trials* are indicated by a visual warning that presented a brief delay after stimulus presentation. The primary outcome measure of the SST is the SSRT, which is the number of milliseconds of warning a participant needs for them to be able to successfully inhibit their planned response [52].

In this study, we decided to use the SST as opposed to the Go-NoGo task from our previous study [43]. We did this because we found many participants to be performing at ceiling in the Go-NoGo task, which limited our ability to detect differences between the task variants. The SST is more challenging than the Go-NoGo task because it dynamically adjusts the task's difficulty to match the inhibitory control of the user, therefore reducing the likelihood of a participant performing at ceiling.

We based our SST on the widely used CANTAB SST [53,54] albeit with a visual rather than auditory stop signal and some graphical upgrades to make the task more suitable for the Web. Each trial began with a fixation cross that was displayed in the middle of the screen, with two colored zones on the left and right of the fixation cross (Figure 4). After 500 ms, a colored circle appeared over the fixation cross and participants had to respond as rapidly as possible by pressing either the left or right arrow key to indicate which colored zone matched the color of the circle (Multimedia Appendix 2). In 25% of trials, white brackets appeared around the circle after it was shown: when this occurred, the subject had to withhold their response and wait until the next trial began (each trial was displayed for 900 ms). If the participant responded before the stop signal was displayed, then the trial was recorded as failed, but white brackets were not displayed. Between each trial, there was a random intertrial interval ranging from 500 to 1000 ms. The delay between the circle onset and the bracket onset is called the stop signal delay (SSD), and was varied according to a four-staircase tracking algorithm, designed to sample across the *inhibition-probability by SSD* space (see Multimedia Appendix 1) [55,56]. The task consisted of five blocks of 48 trials each, with a 10s break between each block. If the participant minimized the browser window or changed tabs, then the task was paused (because of the default way in which timers in JavaScript operate). However, if the browser window was not in focus but was still visible (eg, on a second monitor), then the task was not paused.

In the nongame variant, the participant's history was presented as a list of previous sessions, with median reaction times and estimated SSRTs (Figure 5). Hovering over a column displayed a brief explanation of the variable (eg, "The reaction time

column shows the average time in milliseconds, which it took you to respond to the circles appearing in each session.").

Stop Signal Task: Points Variant

The points variant was similar to the nongame variant but with the addition of a points mechanic and the task being framed as a game. Points are a common feature of gamified tasks [25] and are classed as "1st Step" gamification [57]. In our task, the participant's points score was displayed at the bottom of the screen throughout (Figure 6) (Multimedia Appendix 3). The scoring system was very similar to that used in our previous study [43], which in turn was based on that used by Miranda et al [33]. The scoring system also incorporates the findings of Guitart-Masip et al [58] who found that subjects were much more successful in learning active (go) choices when rewarded for them and passive choices (stop) when punished. On each successful nonstop trial, the participant earned points equal to $0.2 \times \text{bonus} \times (800 - \text{reaction time})$, and the number of points gained was displayed briefly in the intertrial interval. This bonus was a multiplier ($\times 2, \times 3, \times 4 \dots$), which increased by 1 every 3 trials but decreased by 3 when the participant failed a stop trial. The bonus was not lost on stop trials to which the participant responded before the stop signal was displayed (to all appearances, the trial was not a stop trial). On a successful inhibition to a stop signal, the bonus was not lost, but no points were awarded (as there was no reaction time on which to base the score for that trial). Scores were maintained over blocks but not over sessions. The scoring system was outlined to the participants in the instructions for the task.

The participant's history was presented as a list of median reaction times, SSRTs, and scores from each testing session (Figure 7). Additionally, the participant's highest score was saved as a high score and was displayed in the top right-hand corner throughout every testing session.

Stop Signal Task: Theme Variant

The theme variant was similar to the nongame variant but with the addition of a graphical theme and a sense of progression. The task was framed as a game and featured themed graphics and stimuli, with the yellow and blue stimuli being replaced by images of objects, though still predominantly blue or yellow (Figure 8, Multimedia Appendix 4). The task was presented on a series of different graphical backgrounds (Multimedia Appendix 5) but with some shared elements: a conveyor belt on which objects appeared and two bins to the left and right into which these objects were sorted. The stop signal was explained as an automatic *fault detector*, which scanned objects as they sat on the conveyor.

The participant's history was presented as a map (Figure 9), and previous sessions' summary data were displayed when the user hovered over the corresponding icon. Each level on the map had a unique name and thematic instruction text, with the intention of creating an overarching goal, perceptual curiosity, and fostering a sense of participant progression [59-61].

Figure 4. Screenshot of the nongame stop signal task.

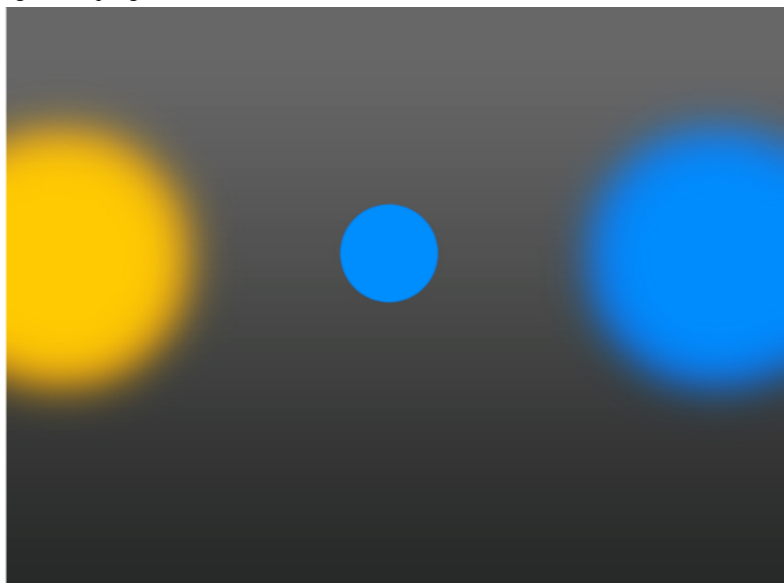


Figure 5. Screenshot of the nongame variant history screen.

Your History

| Session# | Session Date | RT | SSRT | Stop Accuracy | Go Accuracy |
|----------|--------------|-------|-------|---------------|-------------|
| 1 | 25-07-2016 | 510ms | 202ms | 50% | 96% |
| 2 | 26-07-2016 | 517ms | 226ms | 55% | 98% |
| 3 | 27-07-2016 | 447ms | 257ms | 41% | 88% |

Hover over a column heading for more info

Figure 6. Screenshot of the points variant stop signal task.

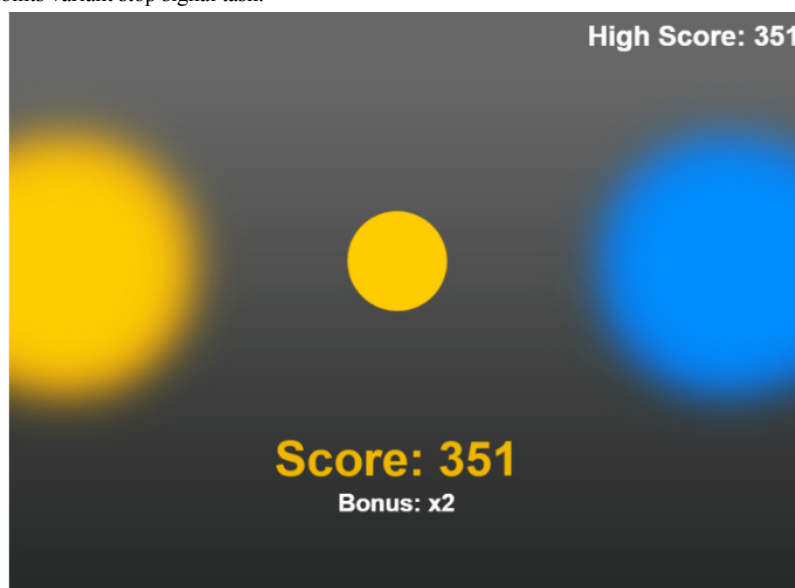


Figure 7. Screenshot of the points variant history screen.

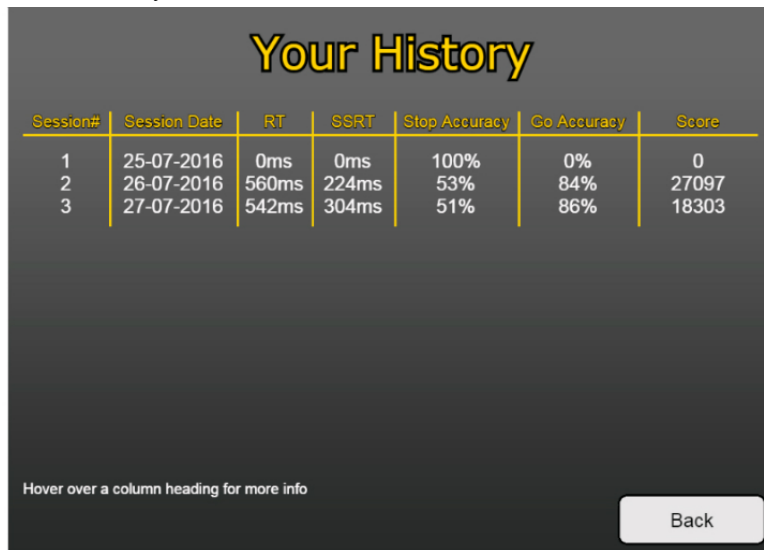


Figure 8. Screenshot of the theme variant stop signal task.



Figure 9. Screenshot of the theme variant history (map) screen.



Enjoyment and Engagement Questionnaire

The enjoyment and engagement questionnaire was designed to collect subjective ratings of the task and was delivered after every session for all three variants. Sessions 1, 4, 7, and 10 delivered the full 10-item questionnaire, whereas the remaining sessions delivered a shorter 5-item questionnaire. These items were answered using a continuous visual analog scale, presented as a horizontal line, 500 pixels long, with a label at either end and no subdivisions. Participants marked a point between these two labels using their mouse.

The following questions were based on those used in previous studies [31,33,43] and were presented in a random order: (1) How enjoyable did you find that?, (2) How frustrating did you find that?, (3) How difficult was it to concentrate for the duration of that?, (4) How well do you think you performed on that?, (5) How mentally stimulating did you find that to be?, (6) How boring did you find that?, (7) How much effort did you put in throughout that?, (8) How repetitive did you find that?, (9) How willing would you be to do that again tomorrow?, and (10) How willing would you be to recommend the study to a friend? Questions 3, 4, 6, 7, and 10 appeared only in the long version of the questionnaire.

Dependent Variable Calculation

Attrition

Attrition was measured in two ways: First, we calculated the mean number of sessions completed per participant (sessions which were started but not finished were excluded from this calculation). Second, we calculated the percentage of participants that completed at least one session, two sessions, etc.

Subjective Measures of Engagement

Subjective engagement with the task was measured by calculating a mean score from the 10-item enjoyment and engagement questionnaire. Questions 2, 3, 6, and 8 were reverse-scored in this calculation. This measure was calculated for each participant's first and fourth sessions, and we also created a *combined score* by taking the mean of the participant's scores from sessions 1 and 4.

Objective Measures of Engagement

We piloted two measures that could potentially serve as objective proxies for engagement: we counted the number of times that participants hid the browser window or moved focus to another window while completing the SST, hypothesizing that unengaged participants would be more likely to briefly visit other websites while testing. We combined the counts of both these events into a single measure: loss-of-focus events. We then created an overall measure of loss-of-focus events for each participant by calculating the mean number of loss-of-focus events from their first four sessions.

We also investigated coefficients of variation, which quantify reaction time intra-individual variability with respect to mean reaction time, as there is some evidence that changes in motivation can be reflected in reaction time variation [62,63]. Coefficients of variation were calculated by dividing the

standard deviation (SD) of nonstop trial reaction times by the mean nonstop trial reaction time. Similarly, we created an overall measure of reaction time variation for each participant by calculating the mean coefficient of variation from their first four sessions.

Stop Signal Reaction Times

We calculated SSRTs for each session separately, excluding sessions where the assumptions of the race model did not hold. The race model is a commonly used model of inhibitory control and aims to describe the relationship between stop and go processes [64]. The race model is used to derive the SSRT and so if the assumptions underlying the race model are broken, then the resultant SSRTs are not good representations of the data [50,64]. To that end, we excluded sessions where the median nonstop trial reaction time was longer than the median failed stop-trial reaction time, where SSDs were not positively correlated with their corresponding median failed stop reaction times, and where stop-trial accuracy was not negatively correlated with SSD.

For the sessions that did meet the assumptions of the race model, SSRTs were calculated by modeling an inhibition function, and using it to estimate the SSD at which the participant's probability of inhibiting a stop signal was 50% [56]; we then used this SSD to calculate the SSRT for that session [50,51]. We also created a combined measure of SSRT for each participant by taking the mean SSRT of their first four sessions.

Statistical Analysis

The data that form the basis of our results are available on request from the University of Bristol Research Data Repository [65].

Attrition

Differences in attrition curves were assessed visually using the Kaplan–Meier method to estimate survival functions, a log-rank test, and a one-way analysis of variance (ANOVA) of number of sessions completed.

Subjective Measures of Engagement

We assessed differences in subjective ratings both visually, using bar charts, and using a repeated measures ANOVA of the total score with session number as the time factor and task variant as the between-subjects factor. Where there was evidence of a difference between task variants, we used post hoc *t* tests to investigate further.

Objective Measures of Engagement

We assessed differences in coefficient of variation and website loss-of-focus events between task variants using one-way ANOVAs with data combined across the first four sessions. Where there was evidence of a difference between task variants, we used post hoc *t* tests to investigate further.

Stop Signal Reaction Times

We used boxplots and a one-way ANOVA with task variant as the between-subjects factor to investigate the effects of gamification on SSRT.

Table 1. Interpreting Bayes factors.

| Hypothesis 0: the effect size is 0 | Strength of evidence | Hypothesis 1: the absolute effect size is between 0 and X^a |
|------------------------------------|-----------------------|---|
| $.33 \leq BF^b \leq 1$ | No support either way | $1 \leq BF \leq 3$ |
| $.1 \leq BF \leq .33$ | Positive | $3 \leq BF \leq 10$ |
| $.01 \leq BF \leq .1$ | Strong | $10 \leq BF \leq 100$ |
| $BF < .01$ | Decisive | $BF > 100$ |

^aX: Cauchy prior width.

^bBF: Bayes factor.

Bayesian Analyses

The three task variants were designed with the aim of minimizing differences in primary task reaction time and nonstop trial accuracy. Therefore, given that frequentist statistics are not ideal for testing equivalences [66,67], we used Bayesian *t* test to assess the evidence for equality of means where frequentist methods failed to find a difference [68,69]. A Bayesian *t* test produces a Bayes Factor (BF), which compares the evidence for two hypotheses. If the evidence favors one hypothesis over the other, then the BF will reflect that, but if the evidence is equal for both hypotheses, then the BF will imply that the data are insensitive [69-71] (Table 1). We used the Bayesian *t* test procedure in JASP [72], with a Cauchy prior width of 0.707. Setting the Cauchy prior width to 0.707 means that in our analysis, one hypothesis is *the effect size is zero* and the other is *the effect size is between -0.707 and 0.707*. Although both hypotheses are centered on an effect size of 0, the former makes a stronger claim than the latter. As such, effect sizes that are not close to 0 are better represented by the latter hypothesis. A prior width of 0.707 was selected for our analysis because it represents the expectation of a medium-large effect, thus weighting the BF against small effects and reducing the likelihood of a false positive.

Sample Size Determination

At the time of study design, to the best of our knowledge, no other studies had investigated the impact of gamification on attrition from a cognitive testing program, and therefore, we had no previous effect size on which to base a sample size determination. Instead, we hypothesized attrition curves (see Multimedia Appendix 1) for each variant and calculated the anticipated effect size ($\phi=0.231$) resulting from a Kaplan–Meier method and log-rank test (ie, a chi-square test) on those attrition curves. To detect this difference with $\alpha=.05$ and 95% power, a sample size of 290 was required. We set this to 291 to allow for equal group sizes.

Results

Characteristics of Participants

Participants were recruited in two waves: one starting in October 2016 and another starting in January 2017. In both waves, the intended sample size was met within 3 days of the study being posted on Prolific Academic. A total of 482 participants signed up to take part in the study, with 419 (86.9%) of those completing at least one session. A total of 265 (54.9%)

participants completed four sessions over 4 consecutive days as was required by the study criteria (henceforth called *conforming participants*). We excluded 5 participants from the analysis because their reaction times or stimulus categorization accuracy scores were more than 4 interquartile ranges away from the group median. We excluded data from sessions that were started but not completed, and we removed trials from the analysis where participants responded in less than 150 ms.

The analysis below presents data from 260 participants, that is, less than our intended sample size of 291. This was because 32 participants failed to complete the required four sessions in 4 days but instead managed to complete four sessions within 5 days. During the study, we intended on including these *loosely conforming* participants in the analysis, and so stopped recruitment once our intended sample size was achieved. However, for simplicity and adherence to the protocol, we have now decided to present only strictly conforming participant's data below. Analysis of the nonconforming and loosely conforming participants' attrition is presented in Multimedia Appendix 1.

Excluding outliers, 260 conforming participants took part: 91 in the nongame variant, 86 in the points variant, and 83 in the theme. The number of hours spent playing video games was comparable between the groups, and participants typically had a high level of education (Table 2). The most common browser used to complete the experiment was Google Chrome ($n=184$, 70.8%), with others including Firefox ($n=50$, 19.2%), Netscape ($n=13$, 5%), Safari ($n=11$, 4.2%), Opera ($n=1$, 0.5%), and Internet Explorer ($n=1$, 0.5%).

Attrition

Figure 10 shows the attrition of conforming participants, whereas Table 3 shows the mean number of sessions completed per participant in each variant. A log-rank test showed no evidence of a difference between the distributions ($\chi^2_2=3.0$, $P=.22$), and a one-way ANOVA of the number of sessions completed also found no clear evidence of a difference between task variants ($F_{2,262}=1.534$, $P=.21$, partial $\eta^2=0.012$). Given the similarity between nongame and points in mean number of sessions completed, we used a Bayesian *t* test to assess their equality and found substantial evidence that they were equal ($BF=0.16$), but there was no evidence of equality between the theme and the points variant ($BF=0.49$) or the nongame variant ($BF=0.43$).

Table 2. Conforming participants' demographic information, shown separately by task variant.

| Demographic | Nongame | Points | Theme |
|-------------------------------------|-------------------|-------------------|-------------------|
| Age, mean (SD) | 36 (12) | 35 (12) | 34 (11) |
| Male, n (%) | 43 (47) | 49 (57) | 42 (51) |
| Mean video game hours per week (SD) | 6 (12) | 8 (16) | 8 (14) |
| Median level of education | Bachelor's degree | Bachelor's degree | Bachelor's degree |
| Mode ethnicity, n (%) | White, 80 (88) | White, 74 (86) | White, 75 (90) |

Figure 10. Percentage of conforming participants plotted against the number of sessions they completed, shown separately by task variant.

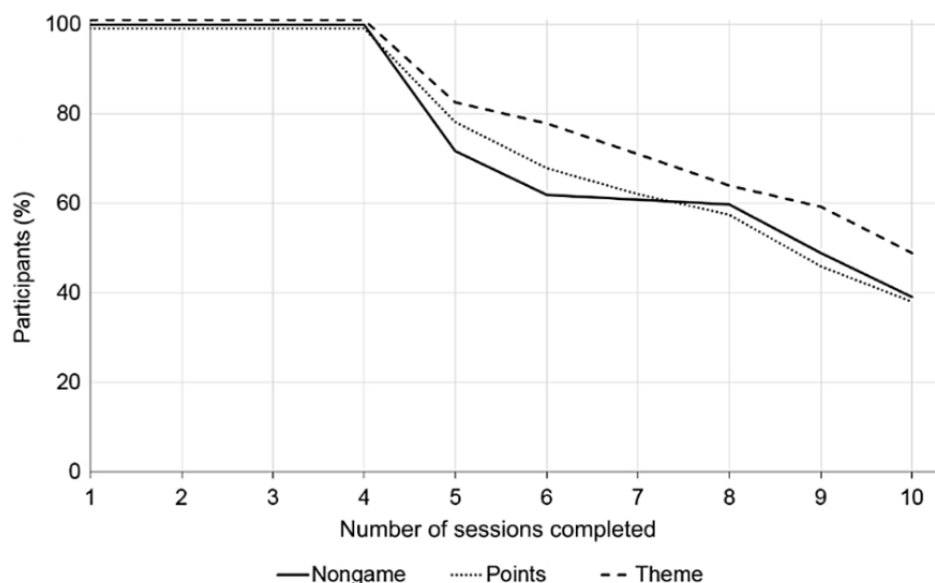


Table 3. Mean number of sessions completed per participant, shown separately by task variant. Conforming participants are those who completed their first four sessions within 4 days as required. All participants includes all who signed up, regardless of their number of sessions completed.

| Variant | All participants (95% CI) | Conforming participants (95% CI) |
|---------|---------------------------|----------------------------------|
| Nongame | 4.9 (4.4-5.5) | 7.4 (6.8-8.0) |
| Points | 5.1 (4.5-5.6) | 7.5 (7.0-8.0) |
| Theme | 5.3 (4.7-5.9) | 8.0 (7.5-8.6) |

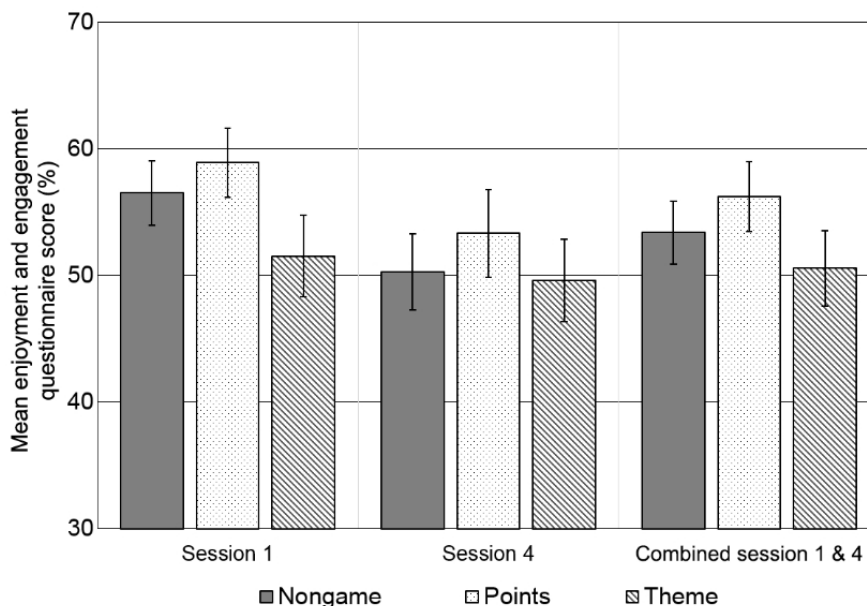
Subjective Measures of Engagement

We used a repeated measures ANOVA of mean score from the enjoyment and engagement questionnaire with session number (1 and 4) as the within-subjects factor and task variant as the between. We used only the two full-length questionnaires completed on the first and the fourth sessions and completed by all participants (for short-form questionnaire results, see [Multimedia Appendix 1](#)). We saw evidence for small effects of both task variant ($F_{2,261}=3.805, P=.02$, partial $\eta^2=0.028$) and time ($F_{1,261}=35.693, P<.001$, partial $\eta^2=0.120$), and weak evidence of an interaction ($F_{2,261}=3.014, P=.05$, partial $\eta^2=0.023$). We noted that ratings of all task variants decrease between the first (M=56, 95% CI 54-57) and fourth sessions (M=51, 95% CI 49-53), but it appears that the nongame and points variants were the main drivers of the interaction effect: dropping by 6% (95% CI 4-8) between sessions 1 and 4, whereas ratings of the theme task decreased only by 2% (95% CI -1 to

5). Post hoc *t* tests on the combined scores showed no evidence for differences between nongame and points, or for nongame and theme ($ps>.15$) but did show points and theme to be different (mean difference=5.7%, 95% CI 1.6-9.7, $t_{171}=2.749, P=.01, d=0.42$). [Figure 11](#) shows the mean scores from each task variant at two time points and a combined score taking the averages of both sessions. A breakdown of ratings by individual questions is presented in [Multimedia Appendix 1](#).

As an unplanned exploratory analysis, we were interested in whether a participant's rating on one day predicted their return to the study on the following day. We ran a logistic regression with *returned following day* as the binary dependent variable and the previous day's score on the subjective questionnaire as the predictor variable. However, we saw no evidence that subjective questionnaire scores predicted return the following day (beta=.008; standard error=0.005; Wald₁=2.166; $P=.14$; odds ratio=1.001, 95% CI 0.997-1.019).

Figure 11. Overall scores from the subjective enjoyment and engagement questionnaire. Mean responses of visual analog scale scores from questionnaires delivered on sessions 1 and 4, and the average scores from sessions 1 and 4, shown separately by task variants and time point. Error bars represent 95% CIs.



Objective Measures of Engagement

We analyzed reaction time coefficient of variation and website loss-of-focus events from the four compulsory sessions combined (Table 4). A one-way ANOVA of coefficient of variation showed strong evidence for a medium effect of task variant ($F_{2,260}=3.131, P=.045, \text{partial } \eta^2=0.024$) on participants' reaction time variability, with lower coefficients indicating that there was less variability. Post hoc *t* tests showed strong evidence of a difference between the points and theme variants (mean difference=1.5%, 95% CI 0.2-2.7; $t_{170}=2.349; P=.02; d=0.36$), but no clear evidence for other differences were found ($ps>.06$).

Loss-of-focus events were rare in all task variants, with each participant switching away from the task less than once per session on average. Regardless, we assessed differences in loss-of-focus events between the three task variants using a one-way ANOVA but found no evidence of any difference ($F_{2,260}=1.137; P=.32; \text{partial } \eta^2=0.008$).

Stop Signal Reaction Times

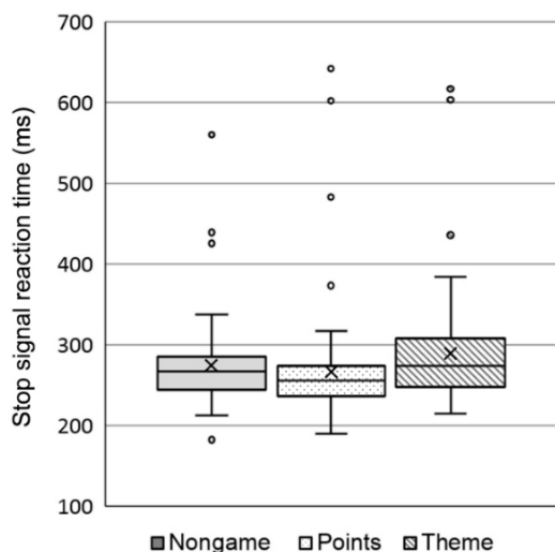
We checked the data from each session against the assumptions of the race model. Of the 1050 sessions assessed, we excluded 161 sessions: 75 from the nongame variant, 37 from points, and 49 from theme. A total of 3 participants failed to meet the assumptions of the race model in all four compulsory sessions, resulting in their exclusion from this analysis. We then analyzed each participant's mean SSRT, with boxplots shown in Figure 12.

A one-way ANOVA showed weak evidence for a small effect of task variant on SSRT ($F_{2,255}=2.954; P=.05; \text{partial } \eta^2=0.022$) with post hoc *t* tests showing a difference between the theme variant (M=289; SD=67) and points variant (M=266, SD=66; mean difference=23, 95% CI 5-42; $t_{169}=2.386; P=.05; d=0.35$). There was no evidence for other differences ($ps>.24$). Bayesian *t* tests showed no evidence of equality between the SSRTs of the nongame and theme variants (BF=0.59) but found substantial evidence for equality between the nongame (M=274, SD=55) and the points variants (BF=0.22).

For brevity, not all the analyses planned in the study protocol have been presented. For more detailed methods and analyses, please see Multimedia Appendix 1.

Table 4. Mean objective measures of participant engagement from the first four sessions, shown separately by task variant.

| Variant | Coefficient of variation (%) (95% CI) | Loss-of-focus events (95% CI) |
|---------|---------------------------------------|-------------------------------|
| Nongame | 18.7% (17.9-19.6) | 0.85 (0.50-1.19) |
| Points | 19.0% (18.1-19.8) | 0.82 (0.43-1.20) |
| Theme | 17.5% (16.7-18.4) | 1.21 (0.75-1.67) |
| Overall | 18.4% (17.9-18.9) | 0.95 (0.72-1.18) |

Figure 12. Boxplots of mean stop signal reaction time. Data combined per participant over the first four sessions and shown separately by task variants.

Discussion

Principal Findings

Contrary to our hypotheses, we saw no clear evidence of an effect of task variant on participant attrition. This was further strengthened when we included data from loosely conforming participants (see [Multimedia Appendix 1](#)), which showed strong evidence that the mean number of sessions completed was equal in all task variants. To the best of our knowledge, this is the first empirical study examining the effects of gamification on participant attrition within a cognitive testing context, and our results raise doubts about the efficacy of gamelike tasks for reducing participant dropout.

Despite there being no difference in usage between the variants, we did find an effect of task variant on the subjective ratings of the tasks, with the points variant having the highest *combined sessions* mean, followed by the nongame variant and the theme. One possible explanation for these findings relates to self-determination theory, a popular theory of motivation that centers around the concept of psychological needs and need satisfaction. Self-determination theory posits that human beings have three needs, which are competence, autonomy, and relatedness and that we find activities to be intrinsically motivating if they help us to fulfill these needs [73]. In the case of our gamelike variants, the points variant would seem to address competency needs by providing constant feedback on their performance, which reinforces the player's success [74], but we do not consider the theme variant or the nongame variant to adequately meet any of the three needs. As the points variant was the only variant to address any of these needs, this may explain why it was rated as the most enjoyable in both this study and our previous study [43].

The theme variant was rated as the worst of the three tasks, which was surprising as it maintained the highest percentage of participants until day 10. One potential explanation is that the task was framed as a game and looked like a game but offered no actual gameplay. Moreover, the map screen and changing graphical backgrounds may have hinted at player

autonomy and exploration as is typical in other games, but ultimately, the player experience was railroaded. These two factors may have undermined autonomy and violated participant expectations, resulting in a dissatisfying experience [75,76]. Despite this, it is possible that the clear end goal on the map and novelty of changing backgrounds could explain the maintenance of participants in the theme variant while still not being a very satisfying or enjoyable experience.

One additional factor to consider, in the light of self-determination theory, is that paying participants in attrition studies such as this may be counterproductive to measuring true engagement. There is evidence that providing extrinsic rewards for otherwise motivating tasks may undermine participant autonomy, therefore affecting the task's ability to meet our psychological needs [77,78]. In this study, it is not possible to determine whether intrinsic motivation to take part was affected by the incentive of 50p per additional session. This is further complicated by the potentially unrepresentative nature of a Prolific Academic sample: all of whom have voluntarily signed up to take part in science experiments over the Internet but can choose studies based on the amount of monetary compensation awarded in exchange for their data. Given these issues, one potentially informative avenue for future research in this area would be to explore the effects of these same gamification mechanisms on attrition but without providing financial incentives.

Money can be a powerful motivator; for example, Khadjesari et al [79] found that offering a £10 Amazon voucher to each participant in a longitudinal study resulted in a 9% increased response rate. In this study, it may simply be that money was the most important factor for taking part and that the similar attrition rates were driven by the identical incentives.

We also found no evidence that participant ratings of engagement and enjoyment could predict the number of optional sessions they would complete. This, combined with the disconnect between the theme variant ratings and theme variant usage, serves to highlight the split in different types of engagement that has begun to be conceptualized in the literature

[24]. In short, the word engagement has been used in the past to refer to both *engagement as subjective experience* and *engagement as usage*, and this study is further evidence that the two concepts are not as closely related as one might assume. Evidence from the video game literature has found that game enjoyment does not relate strongly to game usage, and that game usage can be driven by many other factors, including boredom, loneliness, and need for escapism [80,81]. This highlights the need for future studies of engagement, which collect both subjective and objective measures.

Our two pilot objective measures of engagement, reaction time variation (coefficients of variation) and loss-of-focus events, were difficult to interpret. We saw no evidence that losses of focus differed between the task variants, and this is likely because such events were rare (less than one loss-of-focus event per session on average). This is a positive finding, as it shows that participants are willing to fully concentrate on Web-based cognitive tasks. With respect to coefficients of variation, the pattern of results is directly in contrast with our subjective measures of engagement: the points variant had the most variable response times but the highest subjective rating, whereas the theme variant had the lowest variability and the lowest rating. This is either evidence contrary to the premise that reaction time variability is related to motivation [62,63] or signals that our subjective ratings are not good measures of motivation. Regardless, further research is necessary to understand whether these objective measures provided are related to the broader concept of engagement.

When assessing cognitive data, we found evidence that SSRTs were equivalent between the points variant and the nongame variant. Although the points variant introduced additional elements to the task, which may have increased cognitive load, it is possible that the highly salient feedback and motivational effect of points served to increase participant performance as has been found in a number of previous studies [41,82-84].

Limitations and Conclusions

First, we consider the fact that we did not achieve our intended sample size, an important limitation of this study. However, we maintain that the results of our supplementary analyses including the loosely conforming participants are quite conclusive and strengthen our finding that there was no effect of gamification on attrition. However, we accept that a balanced group analysis would be preferable. Second, we acknowledge that our sample, recruited from Prolific Academic, with high levels of education, may not be representative of the wider population. Third, we

acknowledge that the design of study used is not suitable to validate our gamelike variants as measures of response inhibition, as that would require a within-subjects design to test predictive validity [57,85]. Fourth, the gamelike features we implemented were very lightweight and certainly would not constitute a full game. Indeed, neither of our games were likely enjoyable enough that a participant would consider doing them for their own sake. Though this was necessary to try to reduce the impact of gamification on the cognitive data, it likely reduced any effect of gamification that we might have seen. Fifth, the time course of our study, which took place over days, may not be informative about attrition in studies that take place over longer periods of weeks or months. Sixth, as mentioned previously, there are issues relating to motivation and incentives, as in reality, participants completing cognitive assessments will be presented with requests to complete a study over a fixed period for a fixed fee and not with the option to continue for additional recompense. Finally, both incentives and reminders have been well established as effective methods of increasing engagement, and we used both in this study [86]. Although all three task variants had the same incentives and the same program of reminders (which stopped on day 4), it is possible that these baseline engagement strategies acted as confounders, potentially muddying the effect of gamification on attrition.

In conclusion, the theme variant had negative effects on the cognitive data and showed no clear evidence of reducing attrition. It was also rated as the least enjoyable and was the task switched away from most often. This suggests that themed gamelike tasks, at least those that use graphics alone, are nonoptimal for use in cognitive assessment studies. In contrast, and replicating our previous finding [43], subjective ratings showed the points variant to be well received. We found SSRTs from the points and nongame variants to be equal, showing that points can be an effective way of increasing participant enjoyment of a cognitive task while still collecting valid data.

Despite differences in subjective ratings between the task variants, we saw no effect of gamification on participant attrition over the 6-day optional testing period. Gamification has been promoted as a potential solution to engagement problems in both psychology and digital health care for several years, but we found no effect of gamification on *engagement as usage* in this case. The term gamification may have existed for a decade, but the formalization of gamification's implementation and effectiveness is only just beginning, and there is clearly further work to be conducted to understand how we can translate differences in subjective ratings to differences in usage.

Acknowledgments

The authors gratefully acknowledge the artistic contribution of Melissa Groves who provided graphical resources for the themed task variant. The authors are members of the United Kingdom Centre for Tobacco and Alcohol Studies, a United Kingdom Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UKCRC, is gratefully acknowledged. This work was supported by the Medical Research Council (MC_UU_12013/6 and MC_UU_12013/7) and a PhD studentship to JL funded by the Economic and Social Research Council and Cambridge Cognition Limited. The funders had no role in review design, data extraction and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

JL and MM designed the study. JL programmed the task software and collected the data. JL analyzed the data. AS, NL, and MM contributed to the development of the manuscript.

Conflicts of Interest

DC is a director of Handaxe CIC, a not-for-profit company that develops technology, including computer games, to support mental health interventions for children and adolescents.

Multimedia Appendix 1

A document containing additional details on the methods used in the study and supplementary analyses not core to the message of the paper.

[[PDF File \(Adobe PDF File\), 747KB - jmir_v19i11e395_app1.pdf](#)]

Multimedia Appendix 2

Animated GIF showing several trials of the Non-Game variant of the Stop Signal Task.

[[GIF File, 438KB - jmir_v19i11e395_app2.gif](#)]

Multimedia Appendix 3

Animated GIF showing several trials of the Points variant of the Stop Signal Task.

[[GIF File, 798KB - jmir_v19i11e395_app3.gif](#)]

Multimedia Appendix 4

Animated GIF showing several trials of the Theme variant of the Stop Signal Task.

[[GIF File, 1MB - jmir_v19i11e395_app4.gif](#)]

Multimedia Appendix 5

PDF document showing all possible task backgrounds from the theme variant.

[[PDF File \(Adobe PDF File\), 447KB - jmir_v19i11e395_app5.pdf](#)]

References

1. Mturk. Amazon MTurk URL: <https://www.mturk.com/mturk/welcome> [accessed 2017-07-11] [WebCite Cache ID 6rsMtCfuJ]
2. Prolific. Bringing people together to power the world's research URL: <https://www.prolific.ac/> [accessed 2017-07-11] [WebCite Cache ID 6rsN1syuF]
3. Testable. The web platform for creating, running and sharing behavioral experiments URL: <https://www.testable.org/> [accessed 2017-07-11] [WebCite Cache ID 6rsNDoV2R]
4. Gorilla. Behavioural science in the cloud URL: <https://gorilla.sc/> [accessed 2017-07-11] [WebCite Cache ID 6rsNEk4Uj]
5. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. *J Exp Soc Psychol* 2017 May;70:153-163. [doi: [10.1016/j.jesp.2017.01.006](https://doi.org/10.1016/j.jesp.2017.01.006)]
6. Woods AT, Velasco C, Levitan CA, Wan X, Spence C. Conducting perception research over the internet: a tutorial review. *PeerJ* 2015;3:e1058 [FREE Full text] [doi: [10.7717/peerj.1058](https://doi.org/10.7717/peerj.1058)] [Medline: [26244107](https://pubmed.ncbi.nlm.nih.gov/26244107/)]
7. Crump MJ, McDonnell JV, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One* 2013;8(3):e57410 [FREE Full text] [doi: [10.1371/journal.pone.0057410](https://doi.org/10.1371/journal.pone.0057410)] [Medline: [23516406](https://pubmed.ncbi.nlm.nih.gov/23516406/)]
8. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. *J Med Internet Res* 2006;8(2):e10 [FREE Full text] [doi: [10.2196/jmir.8.2.e10](https://doi.org/10.2196/jmir.8.2.e10)] [Medline: [16867965](https://pubmed.ncbi.nlm.nih.gov/16867965/)]
9. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspect Psychol Sci* 2011 Jan;6(1):3-5. [doi: [10.1177/1745691610393980](https://doi.org/10.1177/1745691610393980)] [Medline: [26162106](https://pubmed.ncbi.nlm.nih.gov/26162106/)]
10. Shapiro DN, Chandler J, Mueller PA. Using Mechanical Turk to study clinical populations. *Clin Psychol Sci* 2013 Jan 31;1(2):213-220. [doi: [10.1177/2167702612469015](https://doi.org/10.1177/2167702612469015)]
11. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. *Annu Rev Public Health* 2009;30:273-292. [doi: [10.1146/annurev.publhealth.031308.100235](https://doi.org/10.1146/annurev.publhealth.031308.100235)] [Medline: [19296777](https://pubmed.ncbi.nlm.nih.gov/19296777/)]
12. Birnbaum MH. Human research and data collection via the internet. *Annu Rev Psychol* 2004 Feb;55:803-832. [doi: [10.1146/annurev.psych.55.090902.141601](https://doi.org/10.1146/annurev.psych.55.090902.141601)] [Medline: [14744235](https://pubmed.ncbi.nlm.nih.gov/14744235/)]

13. Etter JF. Comparing the efficacy of two Internet-based, computer-tailored smoking cessation programs: a randomized trial. *J Med Internet Res* 2005;7(1):e2 [[FREE Full text](#)] [doi: [10.2196/jmir.7.1.e2](https://doi.org/10.2196/jmir.7.1.e2)] [Medline: [15829474](#)]
14. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. *J Med Internet Res* 2005;7(1):e7 [[FREE Full text](#)] [doi: [10.2196/jmir.7.1.e7](https://doi.org/10.2196/jmir.7.1.e7)] [Medline: [15829479](#)]
15. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in Internet-based interventions. *Patient Prefer Adherence* 2008;2:57-65 [[FREE Full text](#)] [Medline: [19920945](#)]
16. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. *J Med Internet Res* 2012;14(6):e152 [[FREE Full text](#)] [doi: [10.2196/jmir.2104](https://doi.org/10.2196/jmir.2104)] [Medline: [23151820](#)]
17. Lodwick R. Qmro.qmul.ac.: Queen Mary University of London; 2016. Crossover designs: issues in construction, use, and communication URL: <http://qmro.qmul.ac.uk/xmlui/handle/123456789/12860> [accessed 2017-09-15] [[WebCite Cache ID 6ueOotap6](#)]
18. Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomised controlled trials. *Br Med J* 2006 Apr 22;332(7547):969-971 [[FREE Full text](#)] [doi: [10.1136/bmj.332.7547.969](https://doi.org/10.1136/bmj.332.7547.969)] [Medline: [16627519](#)]
19. Zhou H, Fishbach A. The pitfall of experimenting on the web: how unattended selective attrition leads to surprising (yet false) research conclusions. *J Pers Soc Psychol* 2016 Oct;111(4):493-504. [doi: [10.1037/pspa0000056](https://doi.org/10.1037/pspa0000056)] [Medline: [27295328](#)]
20. Christensen H, Mackinnon A. The law of attrition revisited. *J Med Internet Res* 2006;8(3):e20 [[FREE Full text](#)] [doi: [10.2196/jmir.8.3.e20](https://doi.org/10.2196/jmir.8.3.e20)] [Medline: [17032636](#)]
21. Eysenbach G. The law of attrition. *J Med Internet Res* 2005;7(1):e11 [[FREE Full text](#)] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](#)]
22. Saul JE, Amato MS, Cha S, Graham AL. Engagement and attrition in Internet smoking cessation interventions: insights from a cross-sectional survey of "one-hit-wonders". *Internet Interv* 2016 Sep;5:23-29. [doi: [10.1016/j.invent.2016.07.001](https://doi.org/10.1016/j.invent.2016.07.001)]
23. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 Steps Australia. *J Med Internet Res* 2015;17(7):e176 [[FREE Full text](#)] [doi: [10.2196/jmir.4339](https://doi.org/10.2196/jmir.4339)] [Medline: [26180040](#)]
24. 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* 2017;7(2):254-267. [doi: [10.1007/s13142-016-0453-1](https://doi.org/10.1007/s13142-016-0453-1)] [Medline: [27966189](#)]
25. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. *JMIR Serious Games* 2016;4(2):e11 [[FREE Full text](#)] [doi: [10.2196/games.5888](https://doi.org/10.2196/games.5888)] [Medline: [27421244](#)]
26. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. *Am J Prev Med* 2012 Jun;42(6):630-638 [[FREE Full text](#)] [doi: [10.1016/j.amepre.2012.02.023](https://doi.org/10.1016/j.amepre.2012.02.023)] [Medline: [22608382](#)]
27. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. *JMIR Serious Games* 2013;1(1):e3 [[FREE Full text](#)] [doi: [10.2196/games.3139](https://doi.org/10.2196/games.3139)] [Medline: [25658754](#)]
28. Deterding S, Sicart M, Nacke L, O'Hara K, Dixon D. Gamification. Using game-design elements in non-gaming contexts. In: CHI '11 Extended Abstracts on Human Factors in Computing Systems. USA: ACM; 2011 Presented at: CHI; May 7-11, 2011; Vancouver, BC, Canada p. A. [doi: [10.1145/1979742.1979575](https://doi.org/10.1145/1979742.1979575)]
29. Rigby S, Ryan RM. *Glued to Games: How Video Games Draw Us In and Hold Us Spellbound (New Directions in Media)*. Santa Barbara, CA: Praeger Publishers Inc; 2010.
30. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. *J R Soc Med* 2013 Mar;106(3):76-78. [doi: [10.1177/0141076813480996](https://doi.org/10.1177/0141076813480996)] [Medline: [23481424](#)]
31. Hawkins GE, Rae B, Nesbitt KV, Brown SD. Gamelike features might not improve data. *Behav Res Methods* 2013 Jun;45(2):301-318. [doi: [10.3758/s13428-012-0264-3](https://doi.org/10.3758/s13428-012-0264-3)] [Medline: [23055169](#)]
32. McPherson J, Burns NR. Gs invaders: assessing a computer game-like test of processing speed. *Behav Res Methods* 2007 Nov;39(4):876-883. [Medline: [18183904](#)]
33. Miranda AT, Palmer EM. Intrinsic motivation and attentional capture from gamelike features in a visual search task. *Behav Res Methods* 2014 Mar;46(1):159-172. [doi: [10.3758/s13428-013-0357-7](https://doi.org/10.3758/s13428-013-0357-7)] [Medline: [23835649](#)]
34. Prins PJ, Dovis S, Ponsioen A, ten BE, van DO. Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD? *Cyberpsychol Behav Soc Netw* 2011 Mar;14(3):115-122. [doi: [10.1089/cyber.2009.0206](https://doi.org/10.1089/cyber.2009.0206)] [Medline: [20649448](#)]
35. Tong T, Chignell M. Developing a serious game for cognitive assessment: choosing settings and measuring performance. 2014 Apr 26 Presented at: Chinese CHI '14: Proceedings of the Second International Symposium of Chinese CHI; April 26-27, 2014; Toronto, Canada p. 70-79. [doi: [10.1145/2592235.2592246](https://doi.org/10.1145/2592235.2592246)]
36. Dörrenbächer S, Müller PM, Tröger J, Kray J. Dissociable effects of game elements on motivation and cognition in a task-switching training in middle childhood. *Front Psychol* 2014 Nov 13;5:1275 [[FREE Full text](#)] [doi: [10.3389/fpsyg.2014.01275](https://doi.org/10.3389/fpsyg.2014.01275)] [Medline: [25431564](#)]

37. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. *PLoS One* 2017;12(3):e0173403 [FREE Full text] [doi: [10.1371/journal.pone.0173403](https://doi.org/10.1371/journal.pone.0173403)] [Medline: [28362821](https://pubmed.ncbi.nlm.nih.gov/28362821/)]
38. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. *JMIR Ment Health* 2016;3(3):e39 [FREE Full text] [doi: [10.2196/mental.5710](https://doi.org/10.2196/mental.5710)] [Medline: [27558893](https://pubmed.ncbi.nlm.nih.gov/27558893/)]
39. Delisle J, Braun C. A context for normalizing impulsiveness at work for adults with attention deficit/hyperactivity disorder (combined type). *Arch Clin Neuropsychol* 2011 Nov;26(7):602-613. [doi: [10.1093/arclin/acr043](https://doi.org/10.1093/arclin/acr043)] [Medline: [21653627](https://pubmed.ncbi.nlm.nih.gov/21653627/)]
40. Dovis S, Van der Oord S, Wiers RW, Prins PJ. Can motivation normalize working memory and task persistence in children with attention-deficit/hyperactivity disorder? The effects of money and computer-gaming. *J Abnorm Child Psychol* 2012 Jul;40(5):669-681 [FREE Full text] [doi: [10.1007/s10802-011-9601-8](https://doi.org/10.1007/s10802-011-9601-8)] [Medline: [22187093](https://pubmed.ncbi.nlm.nih.gov/22187093/)]
41. Ninaus M, Pereira G, Stefitz R, Prada R, Paiva A, Neuper C, et al. Game elements improve performance in a working memory training task. *Int J Serious Games* 2015 Feb 10;2(1):10. [doi: [10.17083/ijsg.v2i1.60](https://doi.org/10.17083/ijsg.v2i1.60)]
42. Katz B, Jaeggi S, Buschkuhl M, Stegman A, Shah P. Differential effect of motivational features on training improvements in school-based cognitive training. *Front Hum Neurosci* 2014 Apr;8:242 [FREE Full text] [doi: [10.3389/fnhum.2014.00242](https://doi.org/10.3389/fnhum.2014.00242)] [Medline: [24795603](https://pubmed.ncbi.nlm.nih.gov/24795603/)]
43. Lumsden J, Skinner A, Woods AT, Lawrence NS, Munafò M. The effects of gamelike features and test location on cognitive test performance and participant enjoyment. *PeerJ* 2016;4:e2184 [FREE Full text] [doi: [10.7717/peerj.2184](https://doi.org/10.7717/peerj.2184)] [Medline: [27441120](https://pubmed.ncbi.nlm.nih.gov/27441120/)]
44. McPherson J, Burns NR. Assessing the validity of computer-game-like tests of processing speed and working memory. *Behav Res Methods* 2008 Nov;40(4):969-981. [doi: [10.3758/BRM.40.4.969](https://doi.org/10.3758/BRM.40.4.969)] [Medline: [19001388](https://pubmed.ncbi.nlm.nih.gov/19001388/)]
45. Lumsden J, Munafò M, Lawrence N. *Osf*. 2016 Oct 11. Effects of game mechanics on participant attrition, data quality and engagement in an online longitudinal cognitive assessment program Internet URL: <https://osf.io/ysaqe/> [accessed 2017-11-01] [WebCite Cache ID 6ueQNJIV5]
46. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *J Am Med Assoc* 2013 Nov 27;310(20):2191-2194. [doi: [10.1001/jama.2013.281053](https://doi.org/10.1001/jama.2013.281053)] [Medline: [24141714](https://pubmed.ncbi.nlm.nih.gov/24141714/)]
47. Lumsden J. *Mindgames*. firebaseapp. 2016. *Mindgames* URL: <https://mindgames.firebaseio.com/> [accessed 2017-07-11] [WebCite Cache ID 6rsNag7Fy]
48. Firebase.google. *Firestore* helps you build better mobile apps and grow your business URL: <https://firebase.google.com/> [accessed 2017-11-01] [WebCite Cache ID 6ueSN3dkz]
49. Pixi.js. The HTML5 creation engine URL: <http://www.pixijs.com/> [accessed 2017-07-11] [WebCite Cache ID 6rsNHAhV4]
50. Logan GD. On the ability to inhibit thought and action: a users' guide to the stop signal paradigm. In: *Inhibitory Processes in Attention, Memory, and Language*. San Diego: Academic Press; 1994:189-239.
51. Logan GD, Cowan WB. On the ability to inhibit thought and action: a theory of an act of control. *Psychol Rev* 1984;91(3):295-327. [doi: [10.1037/0033-295X.91.3.295](https://doi.org/10.1037/0033-295X.91.3.295)]
52. Verbruggen F, Logan GD. Response inhibition in the stop-signal paradigm. *Trends Cogn Sci* 2008 Nov;12(11):418-424 [FREE Full text] [doi: [10.1016/j.tics.2008.07.005](https://doi.org/10.1016/j.tics.2008.07.005)] [Medline: [18799345](https://pubmed.ncbi.nlm.nih.gov/18799345/)]
53. Cambridgecognition. *Cantab research suite* URL: <http://www.cambridgecognition.com/cantab/> [accessed 2017-11-01] [WebCite Cache ID 6ueSfdK6W]
54. Cambridgecognition. *Cantab stop signal task (SST)* URL: <http://www.cambridgecognition.com/cantab/cognitive-tests/executive-function/stop-signal-task-sst/> [accessed 2017-09-13] [WebCite Cache ID 6tRv9eeZj]
55. Logan GD, Schachar RJ, Tannock R. Impulsivity and inhibitory control. *Psychol Sci* 2016 May 06;8(1):60-64. [doi: [10.1111/j.1467-9280.1997.tb00545.x](https://doi.org/10.1111/j.1467-9280.1997.tb00545.x)]
56. Band GP, van der Molen MW, Logan GD. Horse-race model simulations of the stop-signal procedure. *Acta Psychol (Amst)* 2003 Feb;112(2):105-142. [doi: [10.1016/S0001-6918\(02\)00079-3](https://doi.org/10.1016/S0001-6918(02)00079-3)] [Medline: [12521663](https://pubmed.ncbi.nlm.nih.gov/12521663/)]
57. Boendermaker WJ, Prins PJ, Wiers RW. Cognitive bias modification for adolescents with substance use problems--can serious games help? *J Behav Ther Exp Psychiatry* 2015 Dec;49(Pt A):13-20 [FREE Full text] [doi: [10.1016/j.jbtep.2015.03.008](https://doi.org/10.1016/j.jbtep.2015.03.008)] [Medline: [25843611](https://pubmed.ncbi.nlm.nih.gov/25843611/)]
58. Guitart-Masip M, Huys QJ, Fuentemilla L, Dayan P, Duzel E, Dolan RJ. Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage* 2012 Aug 01;62(1):154-166 [FREE Full text] [doi: [10.1016/j.neuroimage.2012.04.024](https://doi.org/10.1016/j.neuroimage.2012.04.024)] [Medline: [22548809](https://pubmed.ncbi.nlm.nih.gov/22548809/)]
59. Malone T. What makes things fun to learn? Heuristics for designing instructional computer games. USA: ACM; 1980 Presented at: Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems; September 18, 1980; Palo Alto, California p. 162-169.
60. Malone TW. Toward a theory of intrinsically motivating instruction. *Cogn Sci* 1981 Dec;5(4):333-369. [doi: [10.1016/S0364-0213\(81\)80017-1](https://doi.org/10.1016/S0364-0213(81)80017-1)]
61. Schell J. *The Art of Game Design: A Book of Lenses*. San Francisco, California: Morgan Kaufmann; 2008.

62. Andreou P, Neale BM, Chen W, Christiansen H, Gabriels I, Heise A, et al. Reaction time performance in ADHD: improvement under fast-incentive condition and familial effects. *Psychol Med* 2007 Dec;37(12):1703-1715 [FREE Full text] [doi: [10.1017/S0033291707000815](https://doi.org/10.1017/S0033291707000815)] [Medline: [17537284](https://pubmed.ncbi.nlm.nih.gov/17537284/)]
63. Garrett DD, Macdonald SW, Craik FI. Intraindividual reaction time variability is malleable: feedback- and education-related reductions in variability with age. *Front Hum Neurosci* 2012;6:101 [FREE Full text] [doi: [10.3389/fnhum.2012.00101](https://doi.org/10.3389/fnhum.2012.00101)] [Medline: [22557954](https://pubmed.ncbi.nlm.nih.gov/22557954/)]
64. Verbruggen F, Logan GD. Models of response inhibition in the stop-signal and stop-change paradigms. *Neurosci Biobehav Rev* 2009 May;33(5):647-661 [FREE Full text] [doi: [10.1016/j.neubiorev.2008.08.014](https://doi.org/10.1016/j.neubiorev.2008.08.014)] [Medline: [18822313](https://pubmed.ncbi.nlm.nih.gov/18822313/)]
65. Lumsden J, Skinner A, Coyle D, Lawrence N, Munafo M. Dataset: No effect of gamification on attrition from a web-based longitudinal cognitive testing study. Bristol: University of Bristol; 2017 Nov 01. [doi: [10.5523/bris.2lh5oz3p93q7p2uyx4w2h70a0v](https://doi.org/10.5523/bris.2lh5oz3p93q7p2uyx4w2h70a0v)]
66. Berger JO, Sellke T. Testing a Point Null Hypothesis: The Irreconcilability of Values and Evidence. *J Am Stat Assoc* 1987 Mar 01;82(397):112-122. [doi: [10.1080/01621459.1987.10478397](https://doi.org/10.1080/01621459.1987.10478397)]
67. Blackwelder WC. "Proving the null hypothesis" in clinical trials. *Control Clin Trials* 1982 Dec;3(4):345-353. [Medline: [7160191](https://pubmed.ncbi.nlm.nih.gov/7160191/)]
68. Wetzels R, Raaijmakers JG, Jakab E, Wagenmakers EJ. How to quantify support for and against the null hypothesis: a flexible WinBUGS implementation of a default Bayesian t test. *Psychon Bull Rev* 2009 Aug;16(4):752-760. [doi: [10.3758/PBR.16.4.752](https://doi.org/10.3758/PBR.16.4.752)] [Medline: [19648463](https://pubmed.ncbi.nlm.nih.gov/19648463/)]
69. Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon Bull Rev* 2009 Apr;16(2):225-237. [doi: [10.3758/PBR.16.2.225](https://doi.org/10.3758/PBR.16.2.225)] [Medline: [19293088](https://pubmed.ncbi.nlm.nih.gov/19293088/)]
70. Raftery AE. Bayesian model selection in social research. *Sociol Methodol* 1995;25:111-163. [doi: [10.2307/271063](https://doi.org/10.2307/271063)]
71. Jeffreys H. *Theory of Probability*. Oxford: Oxford University Press; 1998.
72. JASP Team. Jasp-stats. 2017. JASP v0.8.1.2 URL: <https://jasp-stats.org/> [accessed 2017-11-01] [WebCite Cache ID [6ueTM2lhH](https://www.webcitation.org/6ueTM2lhH)]
73. Ryan RM, Deci EL, editors. *Handbook of Self-Determination Research*. Rochester: University of Rochester Press; 2002.
74. Sailer M, Hense JU, Mayr SK, Mandl H. How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. *Comput Hum Behav* 2017 Apr;69:371-380. [doi: [10.1016/j.chb.2016.12.033](https://doi.org/10.1016/j.chb.2016.12.033)]
75. Boendermaker WJ, Sanchez Maceiras S, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. *JMIR Serious Games* 2016 Dec 06;4(2):e20 [FREE Full text] [doi: [10.2196/games.6464](https://doi.org/10.2196/games.6464)] [Medline: [27923780](https://pubmed.ncbi.nlm.nih.gov/27923780/)]
76. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. *Motiv Emot* 2006;30(4):344-360. [doi: [10.1007/s11031-006-9051-8](https://doi.org/10.1007/s11031-006-9051-8)]
77. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol Bull* 1999 Nov;125(6):627-68; discussion 692. [Medline: [10589297](https://pubmed.ncbi.nlm.nih.gov/10589297/)]
78. Deterding S. Situated motivational affordances of game elements: a conceptual model. In: *CHI Gamification Workshop 2011*. 2011 May 07 Presented at: CHI 2011; 2011; Vancouver, Canada.
79. Khadjesari Z, Murray E, Kalaitzaki E, White IR, McCambridge J, Thompson SG, et al. Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials. *J Med Internet Res* 2011;13(1):e26 [FREE Full text] [doi: [10.2196/jmir.1523](https://doi.org/10.2196/jmir.1523)] [Medline: [21371988](https://pubmed.ncbi.nlm.nih.gov/21371988/)]
80. Boyle EA, Connolly TM, Hailey T, Boyle JM. Engagement in digital entertainment games: a systematic review. *Comput Hum Behav* 2012 May;28(3):771-780. [doi: [10.1016/j.chb.2011.11.020](https://doi.org/10.1016/j.chb.2011.11.020)]
81. Lee D, LaRose R. A socio-cognitive model of video game usage. *J Broadcast Electron Media* 2007;51(4):632-650. [doi: [10.1080/08838150701626511](https://doi.org/10.1080/08838150701626511)]
82. Attali Y, Arieli-Attali M. Gamification in assessment: do points affect test performance? *Comput Educ* 2015 Apr;83:57-63. [doi: [10.1016/j.compedu.2014.12.012](https://doi.org/10.1016/j.compedu.2014.12.012)]
83. Leotti LA, Wager TD. Motivational influences on response inhibition measures. *J Exp Psychol Hum Percept Perform* 2010 Apr;36(2):430-447 [FREE Full text] [doi: [10.1037/a0016802](https://doi.org/10.1037/a0016802)] [Medline: [20364928](https://pubmed.ncbi.nlm.nih.gov/20364928/)]
84. Mekler ED, Brühlmann F, Tuch AN, Opwis K. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Comput Hum Behav* 2017 Jun;71(C):525-534. [doi: [10.1016/j.chb.2015.08.048](https://doi.org/10.1016/j.chb.2015.08.048)]
85. Kato PM. What do you mean when you say your serious game has been validated? Experimental vs Test Validity. 2013 Apr 25. URL: <https://pamkato.com/2013/04/25/what-do-you-mean-when-you-say-your-serious-game-has-been-validated-experimental-vs-test-validity/> [accessed 2016-04-19] [WebCite Cache ID [6gt9POLIu](https://www.webcitation.org/6gt9POLIu)]
86. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013 Aug;46(1):81-95. [doi: [10.1007/s12160-013-9486-6](https://doi.org/10.1007/s12160-013-9486-6)] [Medline: [23512568](https://pubmed.ncbi.nlm.nih.gov/23512568/)]

Abbreviations

ANOVA: analysis of variance
BF: Bayes Factor
SD: standard deviation
SSD: stop signal delay
SSRT: stop signal reaction time
SST: Stop Signal Task

Edited by G Eysenbach; submitted 19.07.17; peer-reviewed by W Boendermaker, O Perski; comments to author 03.08.17; revised version received 25.09.17; accepted 09.10.17; published 22.11.17

Please cite as:

Lumsden J, Skinner A, Coyle D, Lawrence N, Munafo M

Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques

J Med Internet Res 2017;19(11):e395

URL: <http://www.jmir.org/2017/11/e395/>

doi: [10.2196/jmir.8473](https://doi.org/10.2196/jmir.8473)

PMID: [29167090](https://pubmed.ncbi.nlm.nih.gov/29167090/)

©Jim Lumsden, Andy Skinner, David Coyle, Natalia Lawrence, Marcus Munafo. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 22.11.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.