

Original Paper

Effective Gamification of the Stop-Signal Task: Two Controlled Laboratory Experiments

Maximilian Achim Friehs¹, PhD; Martin Dechant², MSc; Sarah Vedress², BA; Christian Frings¹, PhD; Regan Lee Mandryk², PhD

¹Department of Cognitive Psychology and Methodology, Trier University, Trier, Germany

²Human-Computer-Interaction Lab, Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada

Corresponding Author:

Maximilian Achim Friehs, PhD

Department of Cognitive Psychology and Methodology

Trier University

Universitätsring 15

Trier, 54292

Germany

Phone: 49 651201 ext 3250

Email: friehs@uni-trier.de

Abstract

Background: A lack of ability to inhibit prepotent responses, or more generally a lack of impulse control, is associated with several disorders such as attention-deficit/hyperactivity disorder and schizophrenia as well as general damage to the prefrontal cortex. A stop-signal task (SST) is a reliable and established measure of response inhibition. However, using the SST as an objective assessment in diagnostic or research-focused settings places significant stress on participants as the task itself requires concentration and cognitive effort and is not particularly engaging. This can lead to decreased motivation to follow task instructions and poor data quality, which can affect assessment efficacy and might increase drop-out rates. Gamification—the application of game-based elements in nongame settings—has shown to improve engaged attention to a cognitive task, thus increasing participant motivation and data quality.

Objective: This study aims to design a gamified SST that improves participants' engagement and validate this gamified SST against a standard SST.

Methods: We described the design of our gamified SST and reported on 2 separate studies that aim to validate the gamified SST relative to a standard SST. In study 1, a within-subject design was used to compare the performance of the SST and a stop-signal game (SSG). In study 2, we added eye tracking to the procedure to determine if overt attention was affected and aimed to replicate the findings from study 1 in a between-subjects design. Furthermore, in both studies, flow and motivational experiences were measured.

Results: In contrast, the behavioral performance was comparable between the tasks ($P < .87$; $BF_{01} = 2.87$), and the experience of flow and intrinsic motivation were rated higher in the SSG group, although this difference was not significant.

Conclusions: Overall, our findings provide evidence that the gamification of SST is possible and that the SSG is enjoyed more. Thus, when participant engagement is critical, we recommend using the SSG instead of the SST.

(*JMIR Serious Games* 2020;8(3):e17810) doi: [10.2196/17810](https://doi.org/10.2196/17810)

KEYWORDS

video games; games, experimental; proof of concept study; cognition; psychology; motivation

Introduction

Background

Gamification is the process of applying game design elements (eg, scoring systems, graphical interface, narrative) to nongame

environments (eg, cognitive tasks, work context) to increase task performance and engagement [1]. Gamification has been used in a variety of settings, such as in business [2] and education [3]. Serious games are also used in the context of health care education to support desirable behavior [1,4-7]. The use of games or game-like tasks makes it possible to enhance

voluntary engagement and decrease participant drop-out rates [8,9]; in fact, a recent study showed that the experience of playing digital games as compared with standard cognitive tasks was perceived as less stressful [10]. A high dropout rate, especially in difficult-to-obtain samples, can lead to difficulties interpreting the results, for example, due to decreased statistical power [11-14]. Increased task engagement is especially important when cognitive tasks are used as a diagnostic tool because they rely upon the participant to perform the task to the best of their ability. Data obtained from individuals who lack the motivation to perform the task will not be representative of their ability, and this can lead to misinterpretations [15-17]. Although it seems that gamification is generally useful, it can also change task performance in an undesired direction [18]. For example, adding a simple scoring or reward system creates a motivational pull that can interact and interfere with the to-be-measured variable and change behavior [19-21]. However, a reward can even capture attention when it is counterproductive to the task performance, which might make simple reward elements, for example, not always suitable for all gamification purposes [22].

Cognitive Task Gamification

There have been efforts to gamify cognitive tasks for the purposes of training and assessment [23,24]. The interpretation of cognitive task data depends on the assumption that individuals are putting forth their best effort and are fully attentive to the task, but cognitive tasks are often repetitive and boring, so unfocused effort is a common problem [15,16]. An individual's true ability will not be represented if they are not engaged and fully attentive, which can lead to inaccurate interpretations of cognitive task performance data [17]. To improve engagement with cognitive tasks, researchers have looked to games [25,26], with Aeberhard et al [27] noting that "leveraging gamification to repeatedly obtain behavioral samples paves the way for next-generation high-throughput psychometric toolset."

However, caution must be taken when introducing game elements to cognitive tasks owing to the risk of muddying the measurement of the targeted cognitive process [28]. Cognitive tasks are very sensitive to manipulation—even basic tasks (eg, Stroop task, dot-probe task) are extensively studied to understand the effects of making small changes to the task paradigm [29]. Adding game-based elements to basic cognitive tasks could affect performance and experience in unintended ways [28]. Studies on how gamification of cognitive tasks affect behavior have shown mixed results. For example, adding points (a common gamification technique) to a task has been shown to increase engagement [28,30-32] and improve performance, such as by facilitating faster reaction times [24,28]. However, the inclusion of points has also shown to increase error rates in a dot-probe task [28]. Adding thematic elements and complex graphics has been shown to lead to decreased performance: for example, in a go-no-go task, the use of cowboy characters resulted in worse performance compared with a control (green and red objects) [31], and the use of zombie characters resulted in worse performance compared with a control (circles and squares) [33], likely because the stimuli were not as simple to discriminate. However, the inclusion of thematic elements and graphical stimuli have been shown to increase enjoyment [31]

but also decrease enjoyment [28,30,33], relative to a control task.

As there is little agreement on how typical gamification approaches affect performance on, and engagement with, cognitive tasks [24,28], it is imperative that gamified cognitive tasks, intended for use in research, be validated against the basic version before use. Especially, in the context of cognitive psychology or clinical diagnostics, it is important to maintain internal validity [24,34].

Theoretical Underpinnings of Gamification

There are many theories that go beyond the mantra of "games are fun" as to why game design elements are so successful in shaping behavior. Although there is still an open debate regarding the understanding of what makes games enjoyable [35], two of the most prominent theories are the Flow Theory of Motivation [36] and the Self-Determination Theory (SDT) [37].

The Flow Theory states that there are some factors that facilitate flow experience. Specifically, the activity must have clear goals, there must be immediate and unambiguous feedback during task performance, and the perceived challenges of the activity must be balanced with the individual's own skills [38-40]. A flow experience itself differs from individual to individual but is generally characterized by a high concentration on the task at hand, a loss of self-consciousness, a loss of sense of time, and deriving personal purpose from the task performance (ie, autotelic experience) [36,38-40]. In games and player experience research, flow is a key concept and has been proven, among other factors, to be important for player motivation and retention [41-46].

SDT is based upon 3 basic needs: the need for competence (ie, experiencing mastery over challenges); the need for autonomy (ie, doing something owing to an individual's own volition); and the need for relatedness (ie, experiencing meaningful social relations) [47-49]. Importantly, games have been shown to be capable of addressing those needs and enhancing intrinsic motivation [37]. If one or ideally all 3 needs are satisfied, the motivation to engage in the task will increase [50-53]. SDT has been mirrored in the gamification classification system developed by Nicholson [54], in which he proposed 2 types of gamification: reward-based gamification and meaningful gamification. Although reward-based gamification aims to modify extrinsic motivation, meaningful gamification aims to increase intrinsic motivation. Thus, SDT can be used to explain the underlying components of intrinsic motivation, which has been shown to be an important predictor of task engagement [37,55].

In summary, flow theory and SDT are 2 promising theories that can explain an individual's motivation for and experience while performing a task. Importantly, the 2 perspectives are not mutually exclusive but rather complement each other. Thus, gamification based on these theories can inform certain design guidelines for developing gamified versions of cognitive tasks. [54,56].

The Stop-Signal Task

One such cognitive task that is valuable to assess is the ability to inhibit an already initiated action. For example, a basketball player on defense might have to suppress his or her jumping response to avoid falling for the pump-fake of the offensive player or a person might have to stop crossing the street to avoid a speeding car. This type of response inhibition process can be measured using the stop-signal task (SST), which is an established measure of response inhibition and has been used in laboratories now for over 50 years [57,58]. The ability to inhibit a response is also modulated by inter- and intrapersonal differences in humans. For example, a reduction in inhibitory control and a general increase in impulsivity can be seen in people with attention-deficit/hyperactivity disorder (ADHD) [59,60] or patients with schizophrenia [61,62]. In addition, evidence suggests that training or certain types of sports [63,64], as well as noninvasive brain stimulation, can modulate an individual's ability to stop a response [65,66]. As response inhibition has been consistently associated with certain disorders, it has been proposed that response inhibition capabilities can be used as a form of objective diagnostic indicator, especially in ADHD but also in other disorders such as obsessive-compulsive disorder (OCD) [60,67-69]. Individuals affected by mental disorders, especially in the case of ADHD, may have problems focusing on the cognitive task, which makes it particularly important to develop a task that is more engaging to properly assess their cognitive functioning. However, consideration must be taken as gamified tasks have been shown to normalize the performance of individuals with ADHD, meaning that the gamified cognitive task no longer differentiates between people with and without ADHD [70].

The SST requires the participant to withhold their response on a random subset of trials during a choice reaction time task. The delay after which the stopping cue is presented (aptly termed stop-signal delay [SSD]) is fitted to the individual so that in approximately half of all stopping trials, the response inhibition will fail. In detail, when a participant successfully stops their response, the SSD is increased, making a successful stopping less likely on the subsequent trial (vice versa for unsuccessful stop-trials). Usually, participants are tested in a controlled, distraction-free environment, and the stimuli are presented on a monochromatic screen without any irrelevant or interfering elements. Although this leads to a very precise and clean measurement of an individual's response inhibition capabilities, it is not comparable with everyday situations in which the stopping of an already initiated response is required.

In other areas dealing with inhibition of information or responses, an effort has been made to transfer fundamental research principles to applied settings. For example, it was shown that the conflict resolution process as measured by classical cognitive psychological tasks such as the Stroop task or Eriksen flanker task [71,72] is conceptually similar and abides by the same rules as deceptive actions in sports [73-75]. Interestingly, recent studies provide evidence that even the underlying neural generators of these 2 conceptually analog tasks are similar [76,77]. However, as previously mentioned, caution must be taken when adding visual complexity to cognitive tasks due to the potential effects on performance.

In the case of the SST, previous work has shown that changing the stimuli from colored circles to colored fruits (along with an accompanying narrative) resulted in greater stop-signal reaction times (SSRTs; ie, worsened performance) relative to a version gamified with points, but no narrative [30]; however, enjoyment was also reduced in the thematic version relative to the points version, suggesting that engagement may not have been facilitated through the particular theme and stimuli chosen. Research on other tasks has suggested that a poorly implemented theme that offers little gameplay might be worse for engagement than including no theme at all [28].

This Study

To better understand everyday human behavior or, in the case of this paper, specifically stopping the behavior, gamification might be helpful to aid researchers in gathering large data sets over time. In this case, a gamified version of the SST would allow researchers to enhance the ecological validity of the inhibition measurement by presenting the task in a visually complex environment, while also keeping participants motivated to perform well. This ties into the proposition that modern technology can be used to enhance mundane and experimental realism while keeping experimental control high and potentially even increase the effect of experimental manipulation [78,79]. A gamified SST can mirror a more natural setting and therefore elicit more natural responses without sacrificing experimental control. Thus, it is important to choose a task design that not only reliably taps into the targeted processes (ie, the response inhibition process) but also leads to increased participant engagement [80,81]. However, the game must be validated against a basic task to ensure the efficacy of gamification and validity of measurement. In this paper, we present the design of a gamified SST (the stop-signal game [SSG]) and evaluate it relative to the basic task in 2 experiments that consider the effect of gamification on both performance and player experience.

Methods

Overall Procedure

This study sets out to evaluate a gamified version of the SST—termed SSG—along the 2 dimensions of performance and experience. We employed 2 studies to show that performance in a standard SST and in the new SSG was comparable within (study 1) and between (study 2) participants. Thus, comparing performance data in study 1 would give insight into the comparability of both tasks without adding unexplained variance in the form of interindividual variability. Study 2 aimed to replicate the results from study 1; a robust result should still hold even for between-group comparisons. Furthermore, we measured motivation and flow using the Intrinsic Motivation Inventory (IMI) [49] and the Flow State Scale (FSS) [82] in both studies to measure participant experience. Finally, in study 2, we employed an eye-tracking protocol to explain the influence of complex graphics on gaze behavior, and ultimately participant performance. We are of the opinion that the risk of sequence and carry-over effects in eye-tracking studies is especially high as participants pay increased attention to their eye movements. Thus, eye tracking was only employed in study 2. The whole

eye-tracking procedure was comparable with earlier studies utilizing eye tracking in combination with the SST [83]. This paper aims to show that by leveraging gamification, cognitive tasks can be redesigned to produce more realistic and better data, without compromising internal validity [30,78].

Study 1: Within-Subject Design Participants

A total of 30 young, healthy adults were recruited for the study (16 female, 13 male, and 1 nonbinary). The mean age was 23.6 years (SD 4.51; range 17-35 years). The study was approved by the behavioral research ethics board of the University of Saskatchewan. All participants provided written informed consent.

Power Analysis

A power analysis was carried out using G.Power 3.1.3 [84]. For a medium-sized effect ($\eta^2=0.15$ or $f=0.42$), a medium-sized correlation between repeated measures, a power of $1-\beta=0.95$, and an α value of .05, a minimum sample of 22 participants was needed to detect a significant difference between

performance or subjective experience. Thus, failure to find a significant effect will support the null hypothesis.

SSG Design

Both the SST and SSG were implemented using the Unity3D engine. The basic SST as well as the SSG consisted of 3 blocks, each containing 100 trials, 75% of which were go-trials and 25%, stop-trials. Between separate blocks, a pause of 15 seconds was granted. The go-stimulus was presented for a maximum of 1500 msec or until reaction. The stop signal was played over headphones following a variable delay (SSD), which was initially set to 250 msec. The SSD was continuously adjusted with the staircase procedure to obtain a probability of responding to 50%. After the reaction was successfully stopped (ie, button press was inhibited), the SSD was increased by 50 msec, whereas when the participants did not stop successfully, the SSD was decreased by 50 msec. The intertrial interval was set to a random value between 500 msec and 1500 msec. In the basic SST, participants had to respond to a left- or right-pointing arrow, which was presented in the upper third portion of the screen (Figure 1).

Figure 1. The stop-signal task (left) vs stop-signal game (right) trial appearance.

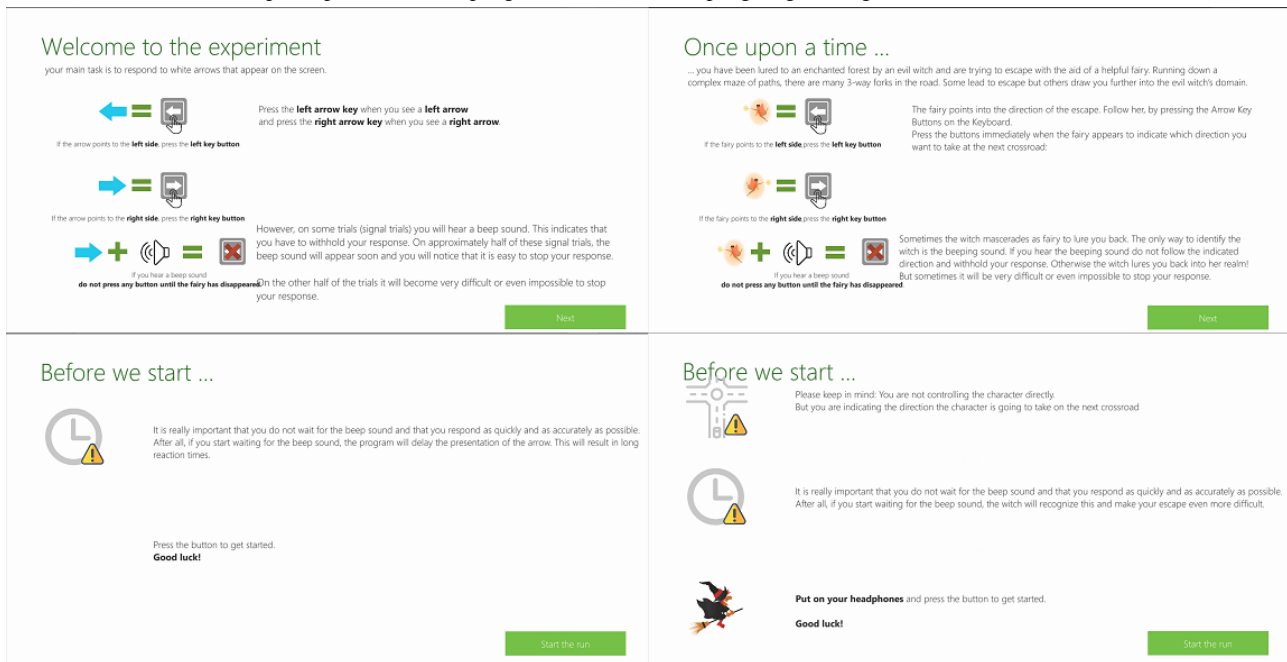


Our goals in designing the SSG were to make the game as identical to the task as possible, while also providing enjoyment. As many gamified cognitive tasks end up being experienced as disappointing in terms of enjoyment [24,28], we built our game around a popular game genre and ensured professional quality graphics. The SSG was built on the 3D infinite runner genre, in which the player sees a third-person view of their avatar running down a path (similar to the popular mobile game Temple Run by Imani Studios). The game premise was integrated with the task instructions, as shown in Figure 2: “Once upon a time you have been lured to an enchanted forest by an evil witch are trying to escape with the aid of a helpful fairy.” In contrast to the arrows used in the SST, the SSG presented arrows in the form of a magical fairy who was pointing to the left or right and would guide them out of the

forest (Figure 1). However, players were told that the evil witch sometimes masqueraded as the fairy, and the only way to know was through a beeping sound (ie, the auditory stop signal); in this case, they were to withhold their response or be lured deeper into the forest. After a choice was made, the avatar turned in the direction selected by the player, regardless of whether or not it was correct (technically, the camera rotated the world and the avatar continued straight). If they failed to respond or correctly withheld their response, the avatar continued on straight. Each choice occurred at a crossroad so that all options were possible, regardless of player response. The terrain was procedurally generated and shaded so that the forest was very dark (matching the dark background of the SST). As shown in Figure 1, we used “low poly” game art, which refers to meshes in 3D computer graphics that contain a small number of

polygons, to give a professional appearance in real time apps (ie, games), while optimizing performance.

Figure 2. Instructions shown to participants for the stop-signal task (left) and stop-signal game (right).



The task was implemented using the Game Engine by Unity3D (version 2019.01) in a single version with a toggle button to switch between the SST and SSG (to keep the implementation of the underlying task identical). The differences between the SST and the SSG were the inclusion of a narrative theme and premise and the presence of the graphical elements, which included the background, the player avatar, and the stimulus (arrow or pointing fairy). The pointing fairy was designed to make the direction easily discriminable to avoid effects from overhead movement interfering with processing the intended movement direction, as replacing basic stimuli with more visually complex ones has been suggested to influence cognitive task performance [28]. In terms of gamification elements employed, we did not use points, scores, or a win or loss condition in the SSG but employed narrative elements including a backstory, a theme, and characters along with immersive elements of a 3D world and theme-appropriate graphics.

Stimuli and Apparatus

Participants were seated in front of a 27-inch color monitor with a viewing distance of approximately 80 cm. The study took place in an ordinarily lit room, 1 participant at a time. Participants were tasked to respond to the signals on screen by using 2 marked keys on the keyboard and withhold their response when an auditory stop signal (900 Hz) was presented over headphones. Participants were instructed to react as fast and accurately as possible. They were tasked to complete both SSTs: the basic SST and the SSG, as previously described. Both SST and SSG took approximately 15 min to complete.

Questionnaire Measures

A total of 2 established questionnaires were used to assess participant experience. The IMI measures motivation on 4 different subscales: interest-enjoyment, effort-importance, perceived competence, and tension-pressure. Each item was

rated through an agreement with a statement on a 7-point scale (higher=greater agreement). The FSS assesses the subjective flow experience and factors influencing it using 9 subscales: challenge-skill balance, action-awareness merging, clear goals, unambiguous feedback, concentration on the task at hand, paradox of control, loss of self-consciousness, transformation of time, and autotelic experience. The FSS items were measured through an agreement with statements on a 5-point scale (higher=greater agreement).

Procedure

Participants were tasked to complete both the basic SST and the SSG, each followed immediately by the FSS and the IMI. The order of presentation of the SST and SSG was counterbalanced across participants and included as a factor in the analysis. After completion of both tasks and questionnaire sets, participants completed a demographic questionnaire.

Design

The study was based on a 2 (task: SST, SSG) x 2 (task-order: SST-SSG vs SSG-SST) mixed measures design with task as a within-subjects factor and task-order as a between-subjects factor.

Data Exclusion

In the data reduction phase, participants were excluded if they were uncooperative or produced faulty data. Initially, it was checked that all participants had normal or corrected-to-normal vision and hearing. For participant exclusion based on the SST and SSG performance, we followed the recommendations in the literature [85,86] that the SSRT can be reliably estimated for each participant in both sessions. Specifically, $p(\text{response}|\text{signal})$ had to be .4-.6, the horse-race model had to be satisfied, and the participant should not display strategic behavior (eg, waiting for the stop signal to appear). Furthermore,

outliers based on the Tukey outlier criterion [87] within the data were identified and removed if necessary. After these procedures, 6 participants were excluded resulting in a sample of 24 with valid behavioral data. Furthermore, 1 additional participant had to be removed from the questionnaire analysis owing to a data collection error.

Dependent Measures

The main dependent variable for performance was the SSRT, that is, the estimate of time needed to respond to the stop signal and to cancel the movement, which measures the covert inhibition process. The estimation of the SSRT was based on the integration method with the replacement of omissions [85,88]. We also measured the SSD, the overall reaction time (RT) for both signal and no-signal trials, the probability of correct inhibition ($p(\text{response}|\text{signal})$), and the omission and commission errors, as standard measures within the stop-signal paradigm [85]. The main dependent variables for experience were measured by using the IMI and FSS, as previously described.

Hypotheses

We hypothesized that there would be no difference in performance measures between the SSG and SST. Questionnaire data were analyzed to test our hypothesis that the SSG would elicit a more positive subjective experience as compared with the basic SST in terms of motivation and flow.

Study 2: Between-Subject Design

Sample

A total of 39 healthy subjects (20 female and 19 male) aged between 18 and 36 years (mean age 24.26, SD 4.99 years) were recruited for the study. All participants had normal or corrected-to-normal hearing and vision. The study was approved by the behavioral research ethics board of the University of Saskatchewan. All participants provided written informed consent.

Eye Tracking

We used a 60 Hz Tobii 4C eye tracker to measure the user's gaze focus. Areas of interest (AOIs) were mapped inside the app for subsequent analysis. The most important AOI was the instruction location (ie, stop-and-go signal location). For the SSG, additional AOIs were defined, including the path, the avatar, and the background.

Stimuli and Apparatus

These were identical as in study 1, apart from the eye-tracking device, which was mounted below the monitor.

Questionnaires

These were identical as in study 1.

Procedure

Upon entering the laboratory, participants were randomly divided into 2 groups: (1) basic SST and (2) gamified SSG. The eye tracker was calibrated for each participant; after calibration,

participants started with the assigned task. After task completion, participants completed the questionnaires.

Design

The experiment was based on a two-group design. Each group was tasked to only complete either the basic SST or the gamified SSG.

Data Analysis

The experiment was based on a two-group (task: SSG vs SST) design. All other details are identical to study 1.

Data Exclusion

The procedure was identical to study 1. A total of 9 participants had to be excluded during the data reduction process resulting in a final sample of 30, evenly split between the 2 groups.

Results

Overview

A summary of the inference statistics in table form can be found in [Multimedia Appendix 1](#). The tables (A1; stop-signal measures), A2 (IMI), and A3 (Flow) show means and SDs of measures in study 1, whereas tables A4 (stop-signal measures), A5 (IMI), and A6 (Flow) show the means and SDs for study 2.

Study 1: Within-Subject Design

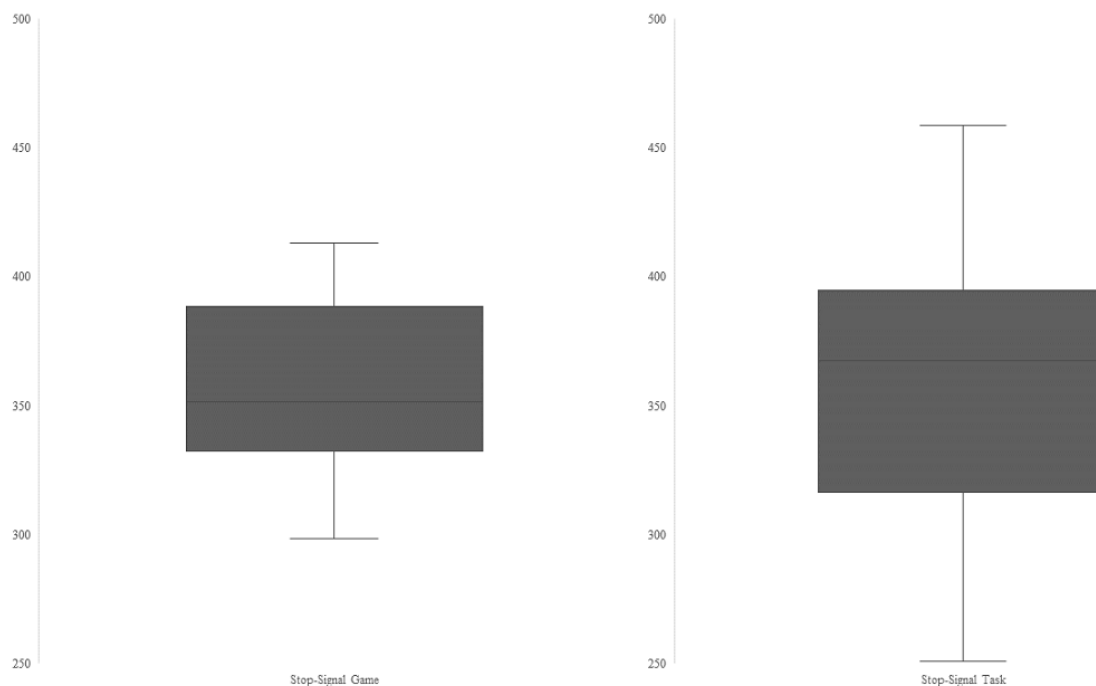
Performance Results

Control Analysis

It is recommended to validate the obtained stop-signal data by showing a significant difference between the average signal RT and the average no-signal RT, with higher RTs for no-signal trials. To this end, a 2 (task-order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) x 2 (trial-type: signal vs no-signal) multivariate analysis of variance (MANOVA) was calculated. Only the main effect trial type ($F_{1,22}=300.38$; $P<.001$; $\eta^2=0.92$) was statistically significant, which shows that signal RT and no-signal RT were different in the expected direction. The main effect task ($F_{1,22}=2.55$; $P=.13$) and the main effect order ($F_{1,22}=1.28$; $P=.27$) were nonsignificant. Furthermore, all the two-way interactions and the three-way interactions did not yield a statistically significant result (all $F<1$).

SSRT

SSRT is an indirect estimate for the duration of the cognitive inhibition process, in which lower values represent higher inhibition speeds and efficiency. SSRTs were analyzed using a 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA. The main effect task type ($F_{1,22}=0.03$; $P=.86$) and the main effect order ($F_{1,22}=0.55$; $P=.47$) as well as the interaction ($F_{1,22}=0.02$; $P=.88$) were not significant. To illustrate the comparable SSRT values for both task types, [Figure 3](#) shows the SSRT distribution depending on the task type. Thus, the speed of the inhibition process was not altered by any experimental manipulation, providing support for the equivalence of the 2 task types concerning their measurement properties ([Table 1](#)).

Figure 3. Stop-signal reaction time (in milliseconds) distribution depending on task type for study 1.**Table 1.** Mean reaction time in milliseconds dependent on task type for study 1. The measurements are collapsed across the order of task performance.

Task	Signal reaction time	No-signal reaction time	Stop-signal delay	Stop-signal reaction time
Stop-signal task, mean (SD)	714.69 (164.83)	817.90 (181.60)	445.97 (185.87)	360.03 (53.86)
Stop-signal game, mean (SD)	750.67 (166.65)	851.55 (181.17)	476.02 (179.39)	358.33 (33.28)

SSD

A 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA was computed. The main effect task type ($F_{1,22}=1.28$; $P=.27$) and the main effect order ($F_{1,22}=1.51$; $P=.23$) as well as the interaction ($F_{1,22}=0.02$; $P=.89$) were not significant. Thus, there was no performance difference on the SSD depending on order or the type of SST (Table 1).

Signal RT

The incorrect signal RTs were analyzed using a 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA. Neither the main effect task ($F_{1,22}=2.68$; $P=.12$) nor the main effect order ($F_{1,22}=1.59$; $P=.22$) or the interaction ($F_{1,22}=0.26$; $P=.31$) were significant. This indicates that signal RT was not dependent on order or task type (Table 1).

No-Signal RT

Correct no-signal RTs were analyzed using a 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA. The main effect task ($F_{1,22}=2.18$; $P=.15$), the main effect order ($F_{1,22}=1.02$; $P=.33$), and the interaction ($F_{1,22}=0.004$; $P=.95$) were all not significant. This result illustrates that overall correct RTs were not dependent on order or task type (Table 1).

Correct Inhibition

The probability of correctly inhibiting a response ($p(\text{response}|\text{signal})$) was analyzed using a 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVA. The main effect task type ($F_{1,22}=0.01$; $P=.93$) and the main effect order ($F_{1,22}=1.10$; $P=.31$) as well as the interaction ($F_{1,22}=0.01$; $P=.93$) were not significant. Thus, there was no performance difference in correct inhibition depending on order or the type of SST employed (Table 2).

Table 2. Mean error rates and accuracy in their relative proportion to the total trial count dependent on task type for study 1. The measurements are collapsed across the order of task performance.

Task	$p(\text{response} \text{signal})$	Omission error	Choice error
Stop-signal task, mean (SD)	0.47 (0.03)	0.04 (0.074)	0.0043 (0.007)
Stop-signal game, mean (SD)	0.47 (0.03)	0.016 (0.023)	0.0024 (0.0046)

Error Analysis

A total of 2 types of errors can be made during go-trials: omission errors (ie, missing a response) and commission errors (ie, choosing the wrong directional reaction). Both were analyzed using 2 (order: SST-SSG vs SSG-SST) x 2 (task: SSG vs SST) repeated measures MANOVAs. For omission errors, the main effects task ($F_{1,22}=3.40$; $P=.08$), the main effect order ($F_{1,22}=0.10$; $P=.76$) and the two-way interaction ($F_{1,22}=0.61$; $P=.44$) were not significant. Similarly, the main effect task ($F_{1,22}=2.85$; $P=.11$) and the main effect order ($F_{1,22}=0.69$; $P=.42$) and the interaction ($F_{1,22}=0.22$; $P=.64$) were not significant with regard to the commission errors. Taken together, order and task did not influence error rates (Table 2).

Experience Results

IMI

As a first step, reliability scores for all 4 IMI subscales were calculated, using Cronbach alpha. The 4 subscales interest-enjoyment, perceived competence, effort-importance, and tension-pressure showed reliability scores of $\alpha_{ie}=.90$, $\alpha_{pc}=.91$, $\alpha_{ei}=.82$, and $\alpha_{tp}=.70$, which were deemed satisfactory. As a second step, all scores were analyzed using separate 2 (task: SSG vs SST) x 2 (order: SST-SSG vs SSG-SST) MANOVAs. For the subscale interest-enjoyment, a significant main effect of task was observed ($F_{1,21}=16.35$; $P=.001$; $\eta^2=0.44$), whereas the main effect order ($F_{1,21}=0.03$; $P=.88$) and the

two-way interaction ($F_{1,21}=0.03$; $P=.86$) were not significant. In detail, participants rated interest-enjoyment on average 0.8 points higher (SD 0.914) for the SSG compared with the SST. The type of task did not affect ratings on all other subscales. For perceived competence, the main effect task ($F_{1,21}=0.69$; $P=.41$), the main effect order ($F_{1,21}=0.56$; $P=.46$), and the two-way interaction ($F_{1,21}=0.81$; $P=.38$) were not significant. For the subscale effort-importance, the main effect task ($F_{1,21}=0.02$; $P=.90$) and the main effect order ($F_{1,21}=0.004$; $P=.95$) as well as the interaction ($F_{1,21}=0.02$; $P=.90$) were not statistically significant. Finally, the subscale tension-pressure was not modulated by task ($F_{1,21}=0.71$; $P=.71$), order ($F_{1,21}=0.85$; $P=.37$), or the interaction between the 2 variables ($F_{1,21}=1.10$; $P=.31$). In summary, participants rated the game higher in interest-enjoyment compared with the basic version; the order in which tasks were completed did not affect the results. Owing to a lack of an overall scale score, the between-task difference values (Δ SSG-SST) were submitted to a multivariate analysis to determine whether or not overall IMI ratings differed among the tasks. The analysis revealed that when considering all subscales simultaneously, the SSG scored significantly higher compared with the SST ($F_{4,18}=6.35$; $P=.002$; $\eta^2=0.59$). Taken together, the analysis shows that the SSG scored significantly higher on the subscale interest-enjoyment (Cohen $d=0.601$) and was overall rated higher on the IMI (Cohen $d=1.109$). For scale means, refer to Table 3.

Table 3. Mean scale values for each Intrinsic Motivation Inventory subscale depending on task variant and the study (study 1).

Subscale	Stop-signal task, mean (SD)	Stop-signal game, mean (SD)
Interest-enjoyment	3.25 (1.45)	4.05 (1.20)
Perceived competence	3.83 (1.39)	3.99 (1.41)
Effort-importance	4.68 (1.39)	4.71 (1.32)
Tension-pressure	3.29 (1.17)	3.49 (1.11)

Flow

As a first step, reliability scores for all 9 flow subscales and the complete scale were calculated using the Cronbach alpha. The 9 subscales, challenge-skill balance ($\alpha_{csb}=.85$), action-awareness merging ($\alpha_{awm}=.78$), clear goals ($\alpha_{cg}=.86$), unambiguous feedback ($\alpha_{uf}=.81$), concentration on the task at hand ($\alpha_c=.66$), paradox of control ($\alpha_{pc}=.90$), loss of self-consciousness ($\alpha_{isc}=.82$), transformation of time ($\alpha_{tt}=.82$), autotelic experience ($\alpha_{ae}=.87$), and the overall scale ($\alpha_{overall}=.91$), showed satisfactory reliability scores. As a second step, all scores were analyzed using separate 2 (task: SSG vs SST) x 2 (order: SST-SSG vs SSG-SST) MANOVAs. For the subscales challenge-skill balance, action-awareness merging, clear goals, paradox of control, loss of self-consciousness, and transformation of time, no significant effects emerged. In detail, with regard to the action-awareness merging subscale, the main effect task ($F_{1,21}=3.38$; $P=.08$), the main effect order ($F_{1,21}=0.85$; $P=.36$), and the interaction ($F_{1,21}=0.30$; $P=.59$) were not significant. The main effects task ($F_{1,21}=0.48$; $P=.50$), order ($F_{1,21}=0.04$; $P=.85$), and their interaction ($F_{1,21}=0.001$; $P=.98$) were not

significant with regard to the action-awareness merging subscale. The analysis of the subscale clear goals neither revealed a significant main effect task ($F_{1,21}=2.24$; $P=.15$) nor a main effect order ($F_{1,21}=0.06$; $P=.81$) and no interaction ($F_{1,21}=0.16$; $P=.69$). Neither the task type ($F_{1,21}=3.83$; $P=.06$) nor the order ($F_{1,21}=0.32$; $P=.58$) or the interaction between task x order ($F_{1,21}=2.06$; $P=.17$) were significant for the subscale paradox of control. Loss of self-consciousness was not modulated by the task ($F_{1,21}=1.72$; $P=.20$) or by the order ($F_{1,21}=0.10$; $P=.76$), and there was no interaction between the 2 variables ($F_{1,21}=2.73$; $P=.11$). The ratings for transformation of time were neither influenced by the task ($F_{1,21}=0.14$; $P=.71$) nor by the order ($F_{1,21}=0.002$; $P=.96$) or their interaction ($F_{1,21}=0.02$; $P=.91$). All effects with regard to the subscale unambiguous feedback were significant. In detail, the main effect task ($F_{1,21}=5.76$; $P=.04$; $\eta^2=0.22$) and the interaction of task x order ($F_{1,21}=5.76$; $P=.04$; $\eta^2=0.22$) displayed equally large effects whereas the effect for the main effect order ($F_{1,21}=3.77$; $P=.06$; $\eta^2=0.15$) was slightly smaller. Taken together, these results show that unambiguous feedback was

rated higher in the game version than the basic task and this effect was enhanced when participants first worked on the basic version and then played the game. The interaction of task x order for the concentration on the task at hand score was significant ($F_{1,21}=6.81$; $P=.02$; $\eta^2=0.25$) whereas the main effect task ($F_{1,21}=1.06$; $P=.31$) and order ($F_{1,21}=0.11$; $P=.74$) were not, showing that concentration decreased in the second session regardless of which task version was done first or second. For an autotelic experience, the main effect task ($F_{1,21}=6.79$; $P=.02$; $\eta^2=0.24$) was significant whereas the main effect order ($F_{1,21}=0.40$; $P=.53$) and the interaction ($F_{1,21}=0.0002$; $P=.99$) were not, showing that participants were more internally driven playing the game version over the basic version. Most importantly, the overall Flow scale score was significantly

higher for the SSG compared with the SST as indicated by the main effect task ($F_{1,21}=5.92$; $P=.02$; $\eta^2=0.22$) and there was no main effect order ($F_{1,21}=0.54$; $P=.47$) or an interaction between task x order ($F_{1,21}=0.14$; $P=.71$). To summarize, concentration on the task at hand decreased in the second session, which can be attributed to fatigue. In addition, the experience of unambiguous feedback increased when participants did the basic task first and then played the game. This result illustrates the participants' feelings that the performance feedback was better and more responsive in the game version compared with the basic task. Furthermore, the results show that the overall experience of flow and the autotelic experience in particular were rated higher in the gamified version of the task. For mean values of the scale, refer to [Table 4](#).

Table 4. Mean scale values for each Flow subscale depending on the task variant for study 1.

Subscale	Stop-signal task, mean (SD)	Stop-signal game, mean (SD)
Challenge-skill balance	3.10 (.90)	3.28 (.81)
Action-awareness merging	3.38 (.75)	3.26 (.84)
Clear goals	3.89 (.71)	4.03 (.62)
Unambiguous feedback	3.37 (.73)	3.52 (.83)
Concentration on the task	3.04 (.80)	3.26 (.75)
Paradox of control	3.12 (.96)	3.32 (.88)
Loss of self-consciousness	3.16 (1.03)	3.35 (1.00)
Transformation of time	3.05 (1.01)	3.00 (.86)
Autotelic experience	2.43 (.84)	2.78 (.79)
Overall	3.17 (.54)	3.31 (.49)

Bayesian Analysis

We employed the Bayesian analysis to put our results to an additional test and support any eventual interpretation of our data. Task version difference scores were calculated for each dependent variable (ie, all performance measures and scale values) to reflect the difference between SST and SSG. The difference scores for all performance variables (eg, SSRT, interest-enjoyment, challenge-skill balance) were submitted to Bayesian paired sample *t* tests using JASP. For performance measures, two-tailed tests were used, and for questionnaires, one-tailed tests were used. We used a Cauchy prior distribution with $r=0.707$. This prior was chosen because it reflects the range of most psychological effects [89] but given our hypothesis of a nonsignificant difference between the 2 task variations, it is a somewhat conservative prior. For the behavioral performance measures, an unspecific alternate hypothesis was specified ($H1$:

$SST \neq SSG$), whereas for the questionnaire data, a hypothesis-conform alternate hypothesis was chosen ($H1$: $SSG > SST$). The Bayesian analysis showed that there were no performance differences between SST and SSG. In detail, results showed weak-to-moderate support for the null hypothesis, and $H0$ was up to 4.63 times as likely as the alternative hypothesis depending on the behavioral performance measure in question [90-92]. With regard to the IMI, the analysis revealed decisive evidence for the subscale interest-enjoyment ($BF_{10}=168.11$), whereas for all other IMI subscales, the null hypothesis was more likely (BF_{01} ranging from 2.4 to 4.19). Flow analysis showed mostly indecisive BFs, but there was moderate support for $H0$ with regard to the 2 subscales, action-awareness merging ($BF_{01}=7.21$) and transformation of time ($BF_{01}=5.91$). In contrast, the analysis revealed moderate support for $H1$ with regard to the autotelic experience ($BF_{10}=7.61$) and the overall flow experience ($BF_{10}=4.91$; [Table 5](#)).

Table 5. The Bayes factor table for study 1 shown by BF01 and BF10.

Dependent measures	Study 1	
	BF ₀₁ ^a	BF ₁₀
Behavioral Performance Measures^b		
No-signal reaction time	1.72	0.58
Signal reaction time	1.48	0.68
Stop-signal reaction time	4.60	0.22
Stop-signal delay	2.52	0.40
Omission errors	1.16	0.86
Choice errors	1.37	0.73
p(response signal)	4.63	0.22
Intrinsic Motivation Inventory		
Interest-enjoyment	0.006	168.11
Perceived competence	2.40	0.42
Effort-importance	4.19	0.24
Tension-pressure	2.44	0.41
Flow State Scale		
Challenge-skill balance	0.60	1.68
Action-awareness merging	7.21	0.14
Clear goals	0.82	1.22
Unambiguous feedback	0.51	1.96
Concentration	1.34	0.75
Paradox of control	0.66	1.52
Loss of self-consciousness	1.65	0.61
Transformation of time	5.91	0.17
Autotelic experience	0.13	7.61
Overall	0.20	4.92

^aBF: Bayes factor.

^bFor the behavioral performance measures H0: SSG=SST and H1: SSG≠SST. For questionnaire measures H1: SSG>SST.

Study 2: Between-Subject Design

Performance Measures

Control Analysis

To establish that signal RT and no-signal RT significantly differ from each other, a 2 (task: SSG vs SST) x 2 (trial-type: signal vs no-signal) MANOVA was calculated. Only the main effect trial type ($F_{1,28}=156.14$; $P<.001$; $\eta^2=0.85$) was statistically significant, which showed that signal RT and no-signal RT trials were different in the expected direction. The main effect of task

($F_{1,28}=0.14$; $P=.71$) as well as the two-way interaction ($F_{1,28}=1.37$; $P=.25$) were not significant.

SSD

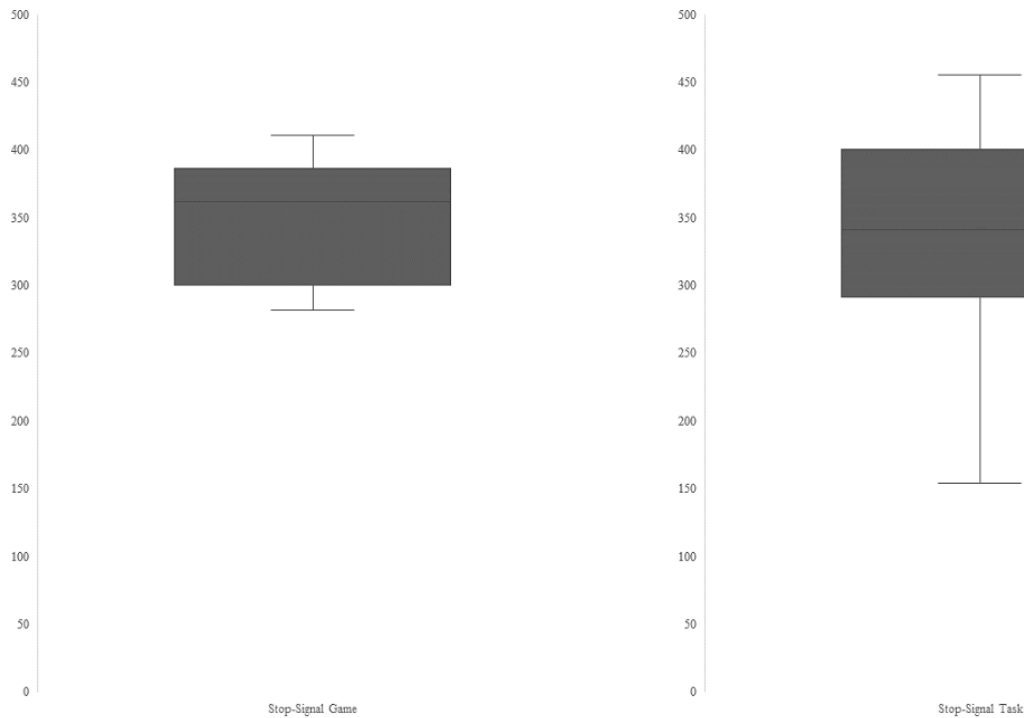
There was no significant difference between the game and the basic version with regard to SSD ($F_{1,28}=0.00006$; $P=.99$).

SSRT

A one-way between-subjects analysis of variance (ANOVA) was conducted to compare the effect of task type (SSG vs SST) on SSRT. The main effect task ($F_{1,28}=0.03$; $P=.87$) was not significant. Thus, the speed of the inhibition process did not depend on the task (Table 6; Figure 4).

Table 6. Mean reaction time in milliseconds dependent on the task type for study 2.

Task	Signal reaction time	No-signal reaction time	Stop-signal delay	Stop-signal reaction time
Stop-signal task, mean (SD)	648.12 (137.99)	759.90 (171.10)	406.03 (207.00)	344.33 (75.73)
Stop-signal game, mean (SD)	681.49 (183.49)	774.10 (204.26)	405.43 (208.52)	348.08 (46.11)

Figure 4. Stop-signal reaction time (in milliseconds) distribution depending on task type for study 2.**Signal RT**

A one-way between-subjects ANOVA showed no difference between the game and basic version of the SST ($F_{1,28}=0.32$; $P=.58$; Table 6).

No-Signal RT

Correct no-signal RTs did not differ between the task types ($F_{1,28}=0.04$; $P=.84$; Table 6).

Correct Inhibition

The probability of correctly inhibiting a response ($p(\text{response}|\text{signal})$) did not differ between the basic and game version ($F_{1,28}=0.79$; $P=.38$; Table 7).

Table 7. Mean error rates and accuracy in their relative proportion to the total trial count dependent on the study and task type for study 2.

Task	$p(\text{response} \text{signal})$	Omission error	Choice error
Stop-signal task, mean (SD)	.49 (.04)	0.019 (0.034)	0.0021 (0.0038)
Stop-signal game, mean (SD)	.47 (.04)	0.020 (0.035)	0.0033 (0.0043)

Error Analysis

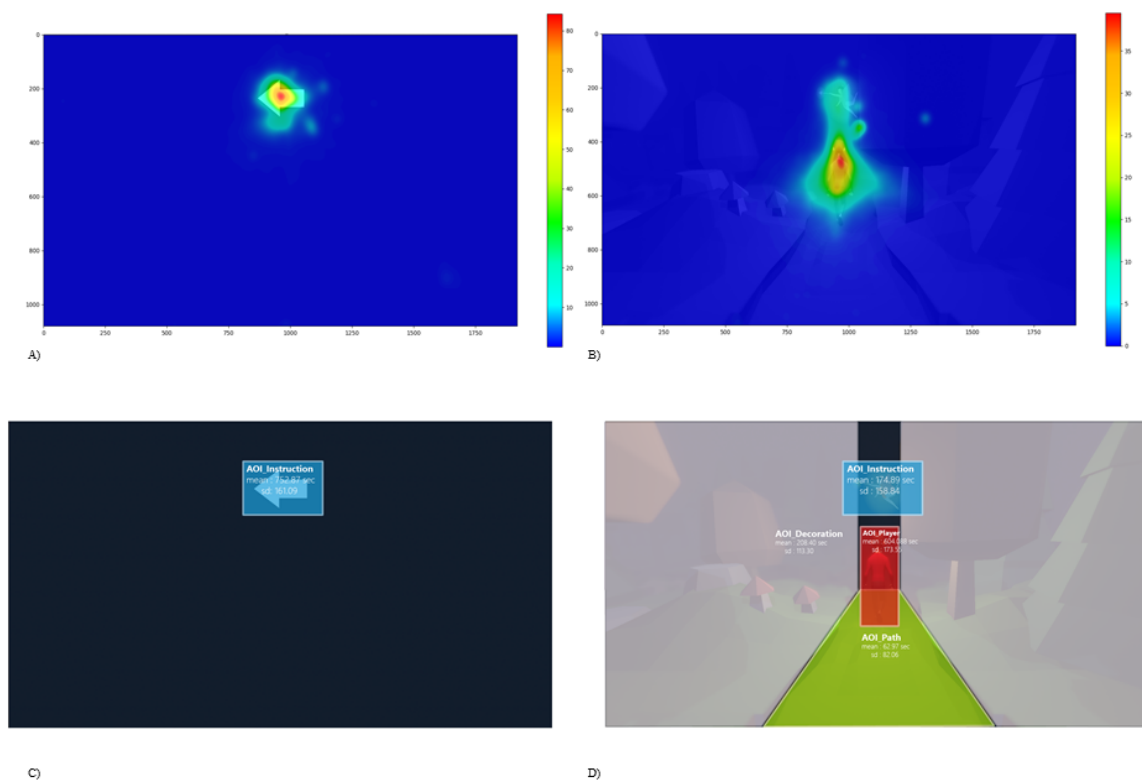
Neither the omission errors ($F_{1,28}=0.005$; $P=.94$) nor the commission errors ($F_{1,28}=0.66$; $P=.42$) differed between the 2 task versions. Taken together, order and task type did not influence error rates (Table 7).

Eye Tracking

We recorded the estimated gaze fixation per user, which is the average fixation of the 2 eyes of the user. This screen coordinate

was mapped on the previously introduced AOIs. Per user, we calculated the average focused time per AOI for both conditions over the complete experiment. In the gamified condition, users mostly focused on the avatar (mean 604.09 seconds, SD 173.55), less on the environment (mean 208.41 seconds, SD 113.31), and least on the instruction location (mean 174.89 seconds, SD 158.84), whereas they mostly looked at the instruction location in the basic version of the task (mean 752.88, SD 161.09). For an illustration of the results, see Figure 5.

Figure 5. Visualization of the eye-tracking results. Parts (A) and (B) display the gaze focus as a heat map. Parts (C) and (D) display the mean time that the participants spent focused on parts of the display.



Experience Measures

IMI

The reliability for all subscales was calculated. The 4 subscales interest-enjoyment, perceived competence, effort-importance, and tension-pressure showed reliability scores of $\alpha_{ie}=.87$, $\alpha_{pc}=.89$, $\alpha_{ei}=.77$, and $\alpha_{tp}=.81$, respectively. All subscale scores

were submitted to a one-way between-subject ANOVA comparing the basic and the gamified task version. There were no significant differences for interest-enjoyment ($F_{1,28}=0.001$; $P=.98$), perceived competence ($F_{1,28}=0.28$; $P=.60$), effort-importance ($F_{1,28}=1.28$; $P=.27$) and tension-pressure ($F_{1,28}=0.57$; $P=.46$; Table 8).

Table 8. Mean scale values for each Intrinsic Motivation Inventory subscale depending on task variant and the study.

Subscale	Stop-signal task, mean (SD)	Stop-signal game, mean (SD)
Interest-enjoyment	3.09 (1.21)	3.11 (1.29)
Perceived competence	3.6 (1.29)	3.84 (1.19)
Effort-importance	4.3 (1.39)	4.78 (.89)
Tension-pressure	3.43 (1.44)	3.80 (1.22)

Flow

The reliabilities for all subscales and the overall reliability for the Flow scale was calculated. In detail, the 9 subscales, challenge-skill balance ($\alpha_{csb}=.66$), action-awareness merging ($\alpha_{awm}=.66$), clear goals ($\alpha_{cg}=.71$), unambiguous feedback ($\alpha_{uf}=.74$), concentration on the task at hand ($\alpha_c=.82$), paradox of control ($\alpha_{pc}=.86$), loss of self-consciousness ($\alpha_{isc}=.71$), transformation of time ($\alpha_{tt}=.79$), autotelic experience ($\alpha_{ae}=.88$), and the overall scale ($\alpha_{overall}=.86$), showed satisfactory reliability scores. Overall, there were no statistical differences between

the game and the basic version. In detail, the subscales for challenge-skill balance ($F_{1,28}=0.05$; $P=.83$), action-awareness merging ($F_{1,28}=0.07$; $P=.80$), clear goals ($F_{1,28}=3.20$; $P=.08$), unambiguous feedback ($F_{1,28}=0.72$; $P=.40$), the concentration on the task at hand ($F_{1,28}=3.27$; $P=.08$), paradox of control ($F_{1,28}=1.36$; $P=.25$), loss of self-consciousness ($F_{1,28}=1.56$; $P=.22$), transformation of time ($F_{1,28}=0.65$; $P=.43$), and autotelic experience ($F_{1,28}=0.01$; $P=.91$) as well as all subscales combined ($F_{1,28}=0.83$; $P=.37$) did not differ between the basic and game version (Table 9).

Table 9. Mean scale values for each Flow subscale depending on task variant and the study (study 2).

Subscale	Stop-signal task, mean (SD)	Stop-signal game, mean (SD)
Challenge-skill balance	3.08 (.71)	3.13 (.52)
Action-awareness merging	3.20 (.71)	3.27 (.71)
Clear goals	3.93 (.50)	3.60 (.52)
Unambiguous feedback	3.47 (.77)	3.27 (.79)
Concentration on the task	3.17 (.78)	2.67 (.73)
Paradox of control	2.95 (.83)	3.28 (.73)
Loss of self-consciousness	3.53 (.92)	3.13 (.83)
Transformation of time	3.42 (.83)	3.18 (.75)
Autotelic experience	2.28 (.96)	2.32 (.59)
Overall	3.23 (.44)	3.09 (.34)

Bayesian Analysis

Similar to study 1, we tested the 2 stopping task types against each other using a Bayesian independent sample *t* test with the same parameters as in study 1. For the behavioral performance measures, an unspecific alternate hypothesis was specified ($H_1: SST \neq SSG$), whereas for the questionnaire data, a hypothesis-conform alternate hypothesis was chosen ($H_1: SSG > SST$). We obtained moderate evidence for the null

hypothesis with regard to the behavioral performance measures, confirming that there is no performance difference between the two. The analysis of IMI scores revealed no conclusive BFs and only a tendency toward the null hypothesis. Results of the FSS analysis showed moderate evidence for the null hypothesis in several subscales (ie, clear goals, unambiguous feedback, concentration, loss of self-consciousness, and transformation of time) as well as the overall scale score (Table 10).

Table 10. The Bayes factor (BF) table showing BF01 and BF10.

Dependent measures	Study 2	
	BF ₀₁ ^a	BF ₁₀
Behavioral Performance Measures^b		
No-signal reaction time	2.86	.35
Signal reaction time	2.57	.39
Stop-signal reaction time	2.87	.35
Stop-signal delay	2.90	.34
Omission errors	2.90	.35
Choice errors	2.26	.44
p(response signal)	2.15	.47
Intrinsic Motivation Inventory		
Interest-enjoyment	2.85	.35
Perceived competence	1.94	.52
Effort-importance	1.08	.93
Tension-pressure	1.58	.63
Flow State Scale		
Challenge-skill balance	2.48	.40
Action-awareness merging	2.41	.42
Clear goals	6.83	.15
Unambiguous feedback	4.73	.21
Concentration	6.78	.15
Paradox of control	1.04	.97
Loss of self-consciousness	5.63	.18
Transformation of time	4.63	.22
Autotelic experience	2.68	.37
Overall	4.87	.21

^aBF: Bayes factor.

^bFor the behavioral performance measures H0: SSG=SST and H1: SSG≠SST. For questionnaire measures H1: SSG>SST.

Discussion

Principal Findings

Overall results show that our newly developed SSG can be used to measure response inhibition as well as SST while being more enjoyable. Specifically, in 2 studies employing a within-subject (study 1) and between-subject (study 2) design, we showed that there were no significant differences between the 2 tasks across all behavioral performance measures. Furthermore, we obtained strong evidence that the SSG was more enjoyable and led to higher experiences of flow but only when participants were able to compare the 2 tasks with each other.

In detail, the results of study 1 showed that performance did not differ between the SSG and the basic SST and that the order of tasks did not influence performance. Concerning the experience of flow and intrinsic motivation, the SSG was superior to the standard SST paradigm, with the largest effect being shown by the interest-enjoyment subscale in the IMI, in

which 44% of the variance was explained by the game versus task manipulation. Importantly, effect sizes suggest the existence of a large difference between SST and SSG with regard to interest-enjoyment ($\eta^2=0.44$; Cohen $d=0.601$) and overall intrinsic motivation ($\eta^2=0.59$; Cohen $d=1.109$). Furthermore, the SSG scored higher on the flow subscales for an autotelic experience and unambiguous feedback, and the overall flow score was significantly higher for the SSG, with the game elements explaining 22% to 24% of the variance in experienced flow. These frequentist results were confirmed by the Bayesian analysis. First, there was evidence against performance differences between the 2 tasks. Second, we obtained decisive evidence for a higher interest-enjoyment rating in the SSG compared with the SST, whereas all other IMI subscales were not affected by the type of stopping task. Third, there was evidence for a higher level of autotelic and overall flow experience in the SSG compared with the SST. Overall, our findings suggest that the SSG can be used as a reliable

measurement of the response inhibition process, while being experienced as more enjoyable for participants.

In a second study, we aimed to extend and replicate our findings. As there is evidence that stopping is influenced by perceptual distractors [83], we implemented an eye-tracking procedure to assess the gaze differences between the SST and the SSG. The eye-tracking implementation mirrors the exploratory analysis by Verbruggen et al [83]. Their exploratory analysis showed that the frequency of eye movements was increased in the condition where the stop signal was presented peripherally. If we had found a significant performance difference between the 2 task versions, we could have used the eye-tracking data to explain this result. On the contrary, our results show that despite a more visually complex environment, which modified gaze and eye movements, the SSG leads to a comparable performance with the SST. We opted for a between-subjects design, which has the additional benefit of eliminating any sequence effects on the eye-tracking data; especially, when people are aware that their behavior is tracked across different task versions, they might behave differently. With that being said, we expected the differences in subjective experience (ie, differences in questionnaire scores) to be smaller owing to the lack of a direct comparison in a between-subjects design.

The results of study 2 partially replicated the results of study 1. We did not find performance differences between SST and SSG in any performance measure. Contrary to study 1, an analysis of questionnaire scores showed that there was no difference between SST and SSG in either the IMI or Flow scale. A Bayes analysis confirmed these findings and revealed small-to-moderate evidence for the null hypothesis (H_0 : SST=SST) with regard to performance measures as well as the IMI and Flow scales. The lack of differences in questionnaire scores was somewhat expected. This result is likely due to the fact that participants in study 2 had no chance to compare the 2 task versions coupled with a regression toward the mean and a tendency of participants to avoid the more extreme scale ratings [93,94]. In addition, we think that the lack of a significant difference between tasks in study 2 is positive. It reflects that only when an implicit comparison between SST and SSG can be made are the 2 versions perceived differently, but overall, the influence of the gamification on motivation is not exceedingly large, as task performance was still comparable. Game elements that overly influence task performance can in turn make it difficult to gather an individual's exact baseline performance. The average fixation time on the previously presented AOIs showed that the gaze focus differed between the 2 tasks. However, this crucially did not seem to affect performance. Interestingly, this hints at the possibility that foveal focus and attention is not required to effectively process a simple stop signal.

Comparison With Previous Work

To the authors' knowledge, there only has been one other study that tried to gamify the SST [30]. The aforementioned study compared 3 SST variants—standard, theme, and scoring—in participants who were recruited and tested on the web. They found no effect of task variant on attrition, and although the variant with the scoring system had higher ratings, the theme

variant scored lower compared with the standard SST paradigm. Importantly however there are several differences between the study by Lumsden et al [30] and this paper. We employed the SST and SSG in a controlled lab environment and not on the web, which has the clear advantage of control over the environment, the experimental set-up, and participant compliance. Furthermore, Lumsden et al [30] focused on the rounds played by participants after 4 required initial sessions but found no effect of task-variant playtime. Although the amount played might be a good measure of motivation, the reward for playing was low—monetarily and intrinsically.

In detail, the gamification used in the study by Lumsden et al [30] consisted of a scoring system without any graphical changes to the task or a thematic variation of the SST; in this case, the player had to sort fruit into different buckets. The theme version of the SST did not implement a scoring system, which is similar to the SSG in this study. Although this was not tested and is pure speculation, the authors of this study think it is reasonable to assume that the haunted forest cover story provides a higher sense of urgency and might be more engaging than sorting fruit by color. Furthermore, Lumsden et al [30] did award participants with only 50 cents for every session after the fourth session, which may not have been enough to keep players motivated. With that being said, the authors mirrored our results by showing that there were no performance differences between the task versions. We decided against the more volatile measure of play sessions and aimed to directly capture performance and motivation. Nevertheless, we think that the study provided important initial evidence and that it might be interesting in the future to validate our SSG on the web.

Limitations

This study has 3 important limitations. First, overall reaction times and inhibition speeds were elevated in both the SST and SSG compared with the ordinarily observed values [65,66,83,85,86,95]. Furthermore, there is evidence that SSRT is unaffected by the demands of the go task [95], but this is still up for debate [83,86]. However, RTs as reported in this study are not completely unusual, and as both tasks produced comparable performance measures, this elevation might be traced back to the samples. Second, we only found a reliable difference in flow and motivation between SSG and SST in study 1 (ie, within-subject design). As task-order did not affect the evaluation of SST or SSG, we speculate that the increased motivation and flow experience in the within-subject study was because both tasks could be compared side-by-side. This illustrates that those kinds of subjective questionnaire measures are somewhat context-dependent. Third, one could also take the neuroscientific approach to compare SST and SSG. In detail, if our claim is that the SSG is a methodologically valid and more enjoyable substitute of the SST that can measure response inhibition accurately, then similar neural correlates should be obtainable. Specifically, similar to the SST, we would expect the right prefrontal cortex to play a crucial role in stopping performance during the SSG and, in contrast to the SST, other areas more responsible for visual information processing should show increased activity during the SSG [65,66,96-100]. In addition, recent evidence suggests that performance differences in video games translate to differential brain activity [101].

Thus, future neuropsychological studies might have to take the individual baseline performance and brain activity into account.

Outlook

There are several directions in which future research could be taken. As already mentioned in the introductory section of this paper, the ability to inhibit an already initiated action is linked to mental health conditions such as ADHD, OCD, schizophrenia, and posttraumatic stress disorder [59,69,102,103]. As video games are accessible, motivating, and can be custom built to capture behavior, such as the SSG in this paper, it has been proposed that digital games or game-like tasks can be useful for the assessment and treatment of mental health issues [104,105]. Although the use of cognitive psychological testing in a clinical setting is well established and those experimental approaches produce reliable between-group (ie, clinical vs nonclinical sample) differences, they are not necessarily reliable on an individual level over time [106]. Nevertheless, we propose that the SSG should be experimented with its use in a more

applied setting. This can be especially important in cases where obtaining a valid response inhibition measurement is difficult. For example, in some clinical subsamples, the ability to focus on the task at hand is limited, and the SSG is more easily accessible and motivating for participants although validly measuring the stopping ability. To this end, a first step could be to validate the present results on a larger scale in a remote web-based assessment. A web-based assessment of behavioral as well as psychophysiological measures via game-like tasks has been done before and shown to be promising for the future [107,108].

Conclusions

Taken together, our results suggest that the newly developed SSG is an effective tool to measure the response inhibition process. The SSG compared with the regular SST has 2 clear advantages. First, the SSG leads to higher enjoyment and flow, and second, it assesses an individual's stopping capabilities in a more realistic, ecologically valid setting.

Acknowledgments

The authors thank the Natural Sciences and Engineering Research Council of Canada and Saskatchewan-Waterloo Games User Research program for funding. They also thank Jeremy Storrington, Parker Neufeld, Chase Crawford, and Katelyn Wiley, along with the rest of the members at the Interaction Lab at the University of Saskatchewan, and the participants.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of the performance analysis results.

[\[DOCX File , 26 KB-Multimedia Appendix 1\]](#)

References

1. Deterding S, Dixon D, Khaled R, Nacke L. Privacy by Design: Informed Consent and Internet of Things for Smart Health from Game Design Elements to Gamefulness: Defining 'Gamification'. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek'11; October 6-8, 2011; Tampere, Finland. [doi: [10.1145/2181037.2181040](https://doi.org/10.1145/2181037.2181040)]
2. Robson K, Plangger K, Kietzmann JH, McCarthy I, Pitt L. Is it all a game? Understanding the principles of gamification. *Bus Horiz* 2015 Jul;58(4):411-420. [doi: [10.1016/j.bushor.2015.03.006](https://doi.org/10.1016/j.bushor.2015.03.006)]
3. Lee J, Hammer J. Gamification in education: what, how, why bother? *Acad Exch Q* 2011;15(2):146.
4. Orji R, Vassileva J, Mandryk RL. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Model User-Adap Inter* 2014 Jul 14;24(5):453-498. [doi: [10.1007/s11257-014-9149-8](https://doi.org/10.1007/s11257-014-9149-8)]
5. Grund CK. How games and game elements facilitate learning and motivation: a literature review. *Lect Notes Informatics* 2015:- [FREE Full text]
6. Birk MV, Mandryk RL, Atkins C. The Motivational Push of Games: The Interplay of Intrinsic Motivation and External Rewards in Games for Training. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. 2016 Presented at: CHI PLAY'16; October 21-23, 2016; Austin, Texas, USA. [doi: [10.1145/2967934.2968091](https://doi.org/10.1145/2967934.2968091)]
7. David R, Michael L, Chen S. -. In: *Serious Games: Games That Educate, Train, and Inform*. New York, USA: Muska & Lipman/Premier-Trade; 2006.
8. 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 Nov 14;14(6):e152 [FREE Full text] [doi: [10.2196/jmir.2104](https://doi.org/10.2196/jmir.2104)] [Medline: [23151820](https://pubmed.ncbi.nlm.nih.gov/23151820/)]
9. Vaibhav A, Gupta P. Gamification of MOOCs for Increasing User Engagement. In: *International Conference on MOOC, Innovation and Technology in Education*. 2014 Presented at: MITE'14; December 19-20, 2014; Patiala, India. [doi: [10.1109/mite.2014.7020290](https://doi.org/10.1109/mite.2014.7020290)]

10. Khalili-Mahani N, Assadi A, Li K, Mirgholami M, Rivard M, Benali H, et al. Reflective and reflexive stress responses of older adults to three gaming experiences in relation to their cognitive abilities: mixed methods crossover study. *JMIR Ment Health* 2020 Mar 26;7(3):e12388 [FREE Full text] [doi: [10.2196/12388](https://doi.org/10.2196/12388)] [Medline: [32213474](https://pubmed.ncbi.nlm.nih.gov/32213474/)]
11. 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](https://pubmed.ncbi.nlm.nih.gov/27295328/)]
12. Bados A, Balaguer G, Saldaña C. The efficacy of cognitive-behavioral therapy and the problem of drop-out. *J Clin Psychol* 2007 Jun;63(6):585-592. [doi: [10.1002/jclp.20368](https://doi.org/10.1002/jclp.20368)] [Medline: [17457848](https://pubmed.ncbi.nlm.nih.gov/17457848/)]
13. Flick SN. Managing attrition in clinical research. *Clin Psychol Rev* 1988 Jan;8(5):499-515. [doi: [10.1016/0272-7358\(88\)90076-1](https://doi.org/10.1016/0272-7358(88)90076-1)]
14. Geraghty AW, Wood AM, Hyland ME. Attrition from self-directed interventions: investigating the relationship between psychological predictors, intervention content and dropout from a body dissatisfaction intervention. *Soc Sci Med* 2010 Jul;71(1):30-37. [doi: [10.1016/j.socscimed.2010.03.007](https://doi.org/10.1016/j.socscimed.2010.03.007)] [Medline: [20400220](https://pubmed.ncbi.nlm.nih.gov/20400220/)]
15. Kirkwood MW, Kirk JW, Blaha RZ, Wilson P. Noncredible effort during pediatric neuropsychological exam: a case series and literature review. *Child Neuropsychol* 2010;16(6):604-618. [doi: [10.1080/09297049.2010.495059](https://doi.org/10.1080/09297049.2010.495059)] [Medline: [20628928](https://pubmed.ncbi.nlm.nih.gov/20628928/)]
16. DeRight J, Jorgensen RS. I just want my research credit: frequency of suboptimal effort in a non-clinical healthy undergraduate sample. *Clin Neuropsychol* 2015;29(1):101-117. [doi: [10.1080/13854046.2014.989267](https://doi.org/10.1080/13854046.2014.989267)] [Medline: [25494327](https://pubmed.ncbi.nlm.nih.gov/25494327/)]
17. Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR, Conference Participants. American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *Clin Neuropsychol* 2009 Sep;23(7):1093-1129. [doi: [10.1080/13854040903155063](https://doi.org/10.1080/13854040903155063)] [Medline: [19735055](https://pubmed.ncbi.nlm.nih.gov/19735055/)]
18. 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](https://pubmed.ncbi.nlm.nih.gov/23055169/)]
19. Bogacz R, Hu PT, Holmes PJ, Cohen JD. Do humans produce the speed-accuracy trade-off that maximizes reward rate? *Q J Exp Psychol (Hove)* 2010 May;63(5):863-891 [FREE Full text] [doi: [10.1080/17470210903091643](https://doi.org/10.1080/17470210903091643)] [Medline: [19746300](https://pubmed.ncbi.nlm.nih.gov/19746300/)]
20. Hinest NS, Anderson IM. The effects of real versus hypothetical reward on delay and probability discounting. *Q J Exp Psychol (Hove)* 2010 Jun;63(6):1072-1084. [doi: [10.1080/17470210903276350](https://doi.org/10.1080/17470210903276350)] [Medline: [19890767](https://pubmed.ncbi.nlm.nih.gov/19890767/)]
21. Hickey C, Kaiser D, Peelen MV. Reward guides attention to object categories in real-world scenes. *J Exp Psychol Gen* 2015 Apr;144(2):264-273. [doi: [10.1037/a0038627](https://doi.org/10.1037/a0038627)] [Medline: [25559653](https://pubmed.ncbi.nlm.nih.gov/25559653/)]
22. le Pelley ME, Pearson D, Griffiths O, Beesley T. When goals conflict with values: counterproductive attentional and oculomotor capture by reward-related stimuli. *J Exp Psychol Gen* 2015 Feb;144(1):158-171. [doi: [10.1037/xge0000037](https://doi.org/10.1037/xge0000037)] [Medline: [25420117](https://pubmed.ncbi.nlm.nih.gov/25420117/)]
23. Mandryk RL, Birk MV. Toward game-based digital mental health interventions: player habits and preferences. *J Med Internet Res* 2017 Apr 20;19(4):e128 [FREE Full text] [doi: [10.2196/jmir.6906](https://doi.org/10.2196/jmir.6906)] [Medline: [28428161](https://pubmed.ncbi.nlm.nih.gov/28428161/)]
24. 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 Jul 15;4(2):e11 [FREE Full text] [doi: [10.2196/games.5888](https://doi.org/10.2196/games.5888)] [Medline: [27421244](https://pubmed.ncbi.nlm.nih.gov/27421244/)]
25. Tong T, Chignell M, Lam P, Tierney MC, Lee J. Designing serious games for cognitive assessment of the elderly. *Proc Int Symp Hum Factors Ergon Health Care* 2014 Jul 22;3(1):28-35. [doi: [10.1177/2327857914031004](https://doi.org/10.1177/2327857914031004)]
26. Valladares-Rodríguez S, Pérez-Rodríguez R, Anido-Rifón L, Fernández-Iglesias M. Trends on the application of serious games to neuropsychological evaluation: a scoping review. *J Biomed Inform* 2016 Dec;64:296-319 [FREE Full text] [doi: [10.1016/j.jbi.2016.10.019](https://doi.org/10.1016/j.jbi.2016.10.019)] [Medline: [27815228](https://pubmed.ncbi.nlm.nih.gov/27815228/)]
27. Aeberhard A, Gschwind L, Kossowsky J, Luksys G, Papassotiropoulos A, de Quervain D, et al. Introducing COSMOS: a web platform for multimodal game-based psychological assessment geared towards open science practice. *J Technol Behav Sci* 2018 Sep 28;4(3):234-244. [doi: [10.1007/s41347-018-0071-5](https://doi.org/10.1007/s41347-018-0071-5)]
28. Wiley K, Vedress S, Mandryk L. How Points and Theme Affect Performance and Experience in a Gamified Cognitive Task. In: *Conference on Human Factors in Computing Systems*. 2020 Presented at: CHI'20; April 25-30, 2020; Oahu, Hawai i, USA. [doi: [10.1145/3313831.3376697](https://doi.org/10.1145/3313831.3376697)]
29. Price RB, Kuckertz JM, Siegle GJ, Ladouceur CD, Silk JS, Ryan ND, et al. Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychol Assess* 2015 Jun;27(2):365-376 [FREE Full text] [doi: [10.1037/pas0000036](https://doi.org/10.1037/pas0000036)] [Medline: [25419646](https://pubmed.ncbi.nlm.nih.gov/25419646/)]
30. Lumsden J, Skinner A, Coyle D, Lawrence N, Munafò M. Attrition from web-based cognitive testing: a repeated measures comparison of gamification techniques. *J Med Internet Res* 2017 Nov 22;19(11):e395 [FREE Full text] [doi: [10.2196/jmir.8473](https://doi.org/10.2196/jmir.8473)] [Medline: [29167090](https://pubmed.ncbi.nlm.nih.gov/29167090/)]
31. 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/)]
32. 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](https://pubmed.ncbi.nlm.nih.gov/23835649/)]
33. Birk MV, Mandryk RL, Bowey J, Buttler B. The Effects of Adding Premise and Backstory to Psychological Tasks. In: *Conference on Human Factors in Computing Systems*. 2015 Presented at: CHI'15; April 18-23, 2015; Seoul, Korea.

34. 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;71:525-534. [doi: [10.1016/j.chb.2015.08.048](https://doi.org/10.1016/j.chb.2015.08.048)]
35. Boyle EA, Connolly TM, Hainey 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)]
36. Csikszentmihalyi M, Rathunde K. The measurement of flow in everyday life: toward a theory of emergent motivation. *Nebr Symp Motiv* 1992;40:57-97. [Medline: [1340523](https://pubmed.ncbi.nlm.nih.gov/1340523/)]
37. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. *Motiv Emot* 2006 Nov 29;30(4):344-360. [doi: [10.1007/s11031-006-9051-8](https://doi.org/10.1007/s11031-006-9051-8)]
38. Elliot AJ, Dweck CS. Competence and motivation: competence as the core of achievement motivation. In: *Handbook of Competence and Motivation*. New York, USA: Guilford Publications; 2005.
39. Moneta GB. On the measurement and conceptualization of flow. In: *Advances in Flow Research*. New York, USA: Springer; 2012.
40. Csikszentmihalyi M. The flow experience and its significance for human psychology. In: *Optimal Experience: Psychological Studies of Flow in Consciousness*. Cambridge, UK: Cambridge University Press; 1988.
41. Sweeter P, Wyeth P. GameFlow: a model for evaluating player enjoyment in games. *Comput Entertain* 2005 Jul 1;3(3):3. [doi: [10.1145/1077246.1077253](https://doi.org/10.1145/1077246.1077253)]
42. Brown E, Cairns A. A Grounded Investigation of Game Immersion. In: *CHI '04 Extended Abstracts on Human Factors in Computing Systems*. 2004 Presented at: CHI EA'04; April 24-29, 2004; Vienna, Austria. [doi: [10.1145/985921.986048](https://doi.org/10.1145/985921.986048)]
43. Bowey JT, Friehs MA, Mandryk RL. Red or Blue Pill: Fostering Identification and Transportation Through Dialogue Choices in RPGs. In: *Proceedings of the 14th International Conference on the Foundations of Digital Games*. 2019 Presented at: FDG'19; August 26-30, 2019; San Luis Obispo, California. [doi: [10.1145/3337722.3337734](https://doi.org/10.1145/3337722.3337734)]
44. Chen J. Flow in games (and everything else). *Commun ACM* 2007 Apr 1;50(4):31. [doi: [10.1145/1232743.1232769](https://doi.org/10.1145/1232743.1232769)]
45. Jin SA. 'Toward integrative models of flow': effects of performance, skill, challenge, playfulness, and presence on flow in video games. *J Broadcast Electron Media* 2012;56(2):169-186 [[FREE Full text](#)]
46. Alexandrovsky D, Friehs A, Birk MV, Yates RK, Mandryk RL. Game Dynamics that Support Snacking, not Feasting. In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 2019 Presented at: CHI PLAY'19; October 22-25, 2019; Barcelona, Spain. [doi: [10.1145/3311350.3347151](https://doi.org/10.1145/3311350.3347151)]
47. Deci EL, Ryan RM. The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior. *Psychol Inq* 2000 Oct;11(4):227-268. [doi: [10.1207/s15327965pli1104_01](https://doi.org/10.1207/s15327965pli1104_01)]
48. Ryan R, Deci E. Self-determination theory: an organismic dialectical perspective. In: *The Handbook of Self-Determination Research*. New York, USA: Springer; 2002.
49. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol* 2000 Jan;55(1):68-78. [doi: [10.1037//0003-066x.55.1.68](https://doi.org/10.1037//0003-066x.55.1.68)] [Medline: [11392867](https://pubmed.ncbi.nlm.nih.gov/11392867/)]
50. Peng W, Lin J, Pfeiffer KA, Winn B. Need satisfaction supportive game features as motivational determinants: an experimental study of a self-determination theory guided exergame. *Media Psychol* 2012 May 18;15(2):175-196. [doi: [10.1080/15213269.2012.673850](https://doi.org/10.1080/15213269.2012.673850)]
51. Birk MV, Atkins C, Bowey JT, Mandryk RL. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016 Presented at: CHI'16; May 7-12, 2016; San Jose, USA. [doi: [10.1145/2858036.2858062](https://doi.org/10.1145/2858036.2858062)]
52. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. *J Educ Psychol* 2013;105(2):249-265. [doi: [10.1037/a0031311](https://doi.org/10.1037/a0031311)]
53. Birk MV, Friehs MA, Mandryk RL. Age-Based Preferences and Player Experience: A Crowdsourced Cross-sectional Study. In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 2017 Presented at: CHI PLAY'17; October 15-18, 2017; Amsterdam, The Netherlands. [doi: [10.1145/3116595.3116608](https://doi.org/10.1145/3116595.3116608)]
54. Nicholson S. A recipe for meaningful gamification. In: *Gamification in Education and Business*. New York, USA: Springer; 2015.
55. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp Educ Psychol* 2000 Jan;25(1):54-67. [doi: [10.1006/ceps.1999.1020](https://doi.org/10.1006/ceps.1999.1020)] [Medline: [10620381](https://pubmed.ncbi.nlm.nih.gov/10620381/)]
56. Seaborn K, Fels DI. Gamification in theory and action: a survey. *Int J Hum Comput Stud* 2015 Feb;74:14-31. [doi: [10.1016/j.ijhcs.2014.09.006](https://doi.org/10.1016/j.ijhcs.2014.09.006)]
57. Logan GD, Cowan WB, Davis KA. On the ability to inhibit simple and choice reaction time responses: a model and a method. *J Exp Psychol Hum Percept Perform* 1984 Apr;10(2):276-291. [doi: [10.1037//0096-1523.10.2.276](https://doi.org/10.1037//0096-1523.10.2.276)] [Medline: [6232345](https://pubmed.ncbi.nlm.nih.gov/6232345/)]
58. Lappin JS, Eriksen CW. Use of a delayed signal to stop a visual reaction-time response. *J. Exp. Psychol* 1966;72(6):805-811 [[FREE Full text](#)]
59. Jennings JR, van der Molen MW, Pelham W, Debski KB, Hoza B. Inhibition in boys with attention deficit hyperactivity disorder as indexed by heart rate change. *Dev Psychol* 1997 Mar;33(2):308-318. [doi: [10.1037/0012-1649.33.2.308](https://doi.org/10.1037/0012-1649.33.2.308)] [Medline: [9147839](https://pubmed.ncbi.nlm.nih.gov/9147839/)]

60. Schachar R, Logan GD. Impulsivity and inhibitory control in normal development and childhood psychopathology. *Dev Psychol* 1990 Sep;26(5):710-720. [doi: [10.1037/0012-1649.26.5.710](https://doi.org/10.1037/0012-1649.26.5.710)]
61. Hoptman MJ, Ardekani BA, Butler PD, Nierenberg J, Javitt DC, Lim KO. DTI and impulsivity in schizophrenia: a first voxelwise correlational analysis. *Neuroreport* 2004 Nov 15;15(16):2467-2470 [FREE Full text] [doi: [10.1097/00001756-200411150-00007](https://doi.org/10.1097/00001756-200411150-00007)] [Medline: [15538176](https://pubmed.ncbi.nlm.nih.gov/15538176/)]
62. Kiehl KA, Smith AM, Hare RD, Liddle PF. An event-related potential investigation of response inhibition in schizophrenia and psychopathy. *Biol Psychiatry* 2000 Aug 1;48(3):210-221. [doi: [10.1016/s0006-3223\(00\)00834-9](https://doi.org/10.1016/s0006-3223(00)00834-9)] [Medline: [10924664](https://pubmed.ncbi.nlm.nih.gov/10924664/)]
63. Wang C, Chang C, Liang Y, Shih C, Chiu W, Tseng P, et al. Open vs. closed skill sports and the modulation of inhibitory control. *PLoS One* 2013;8(2):e55773 [FREE Full text] [doi: [10.1371/journal.pone.0055773](https://doi.org/10.1371/journal.pone.0055773)] [Medline: [23418458](https://pubmed.ncbi.nlm.nih.gov/23418458/)]
64. Tsai C. The effectiveness of exercise intervention on inhibitory control in children with developmental coordination disorder: using a visuospatial attention paradigm as a model. *Res Dev Disabil* 2009;30(6):1268-1280. [doi: [10.1016/j.ridd.2009.05.001](https://doi.org/10.1016/j.ridd.2009.05.001)] [Medline: [19497707](https://pubmed.ncbi.nlm.nih.gov/19497707/)]
65. Friehs MA, Frings C. Pimping inhibition: Anodal tDCS enhances stop-signal reaction time. *J Exp Psychol Hum Percept Perform* 2018 Dec;44(12):1933-1945. [doi: [10.1037/xhp0000579](https://doi.org/10.1037/xhp0000579)] [Medline: [30299126](https://pubmed.ncbi.nlm.nih.gov/30299126/)]
66. Friehs MA, Frings C. Cathodal tDCS increases stop-signal reaction time. *Cogn Affect Behav Neurosci* 2019 Oct;19(5):1129-1142. [doi: [10.3758/s13415-019-00740-0](https://doi.org/10.3758/s13415-019-00740-0)] [Medline: [31313248](https://pubmed.ncbi.nlm.nih.gov/31313248/)]
67. Soreni N, Crosbie J, Ickowicz A, Schachar R. Stop signal and Conners' continuous performance tasks: test--retest reliability of two inhibition measures in ADHD children. *J Atten Disord* 2009 Sep;13(2):137-143. [doi: [10.1177/1087054708326110](https://doi.org/10.1177/1087054708326110)] [Medline: [19429883](https://pubmed.ncbi.nlm.nih.gov/19429883/)]
68. Lijffijt M, Kenemans JL, Verbaten MN, van Engeland H. A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: deficient inhibitory motor control? *J Abnorm Psychol* 2005 May;114(2):216-222. [doi: [10.1037/0021-843X.114.2.216](https://doi.org/10.1037/0021-843X.114.2.216)] [Medline: [15869352](https://pubmed.ncbi.nlm.nih.gov/15869352/)]
69. Lipszyc J, Schachar R. Inhibitory control and psychopathology: a meta-analysis of studies using the stop signal task. *J Int Neuropsychol Soc* 2010 Nov;16(6):1064-1076. [doi: [10.1017/S1355617710000895](https://doi.org/10.1017/S1355617710000895)] [Medline: [20719043](https://pubmed.ncbi.nlm.nih.gov/20719043/)]
70. Delisle J, Braun CM. 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/)]
71. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol* 1935;18(6):643-662. [doi: [10.1037/h0054651](https://doi.org/10.1037/h0054651)]
72. Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept Psychophys* 1974 Jan;16(1):143-149. [doi: [10.3758/bf03203267](https://doi.org/10.3758/bf03203267)]
73. Güldenpenning I, Steinke A, Koester D, Schack T. Athletes and novices are differently capable to recognize feint and non-feint actions. *Exp Brain Res* 2013 Oct;230(3):333-343. [doi: [10.1007/s00221-013-3658-2](https://doi.org/10.1007/s00221-013-3658-2)] [Medline: [23955103](https://pubmed.ncbi.nlm.nih.gov/23955103/)]
74. Kunde W, Skirde S, Weigelt M. Trust my face: cognitive factors of head fakes in sports. *J Exp Psychol Appl* 2011 Jun;17(2):110-127. [doi: [10.1037/a0023756](https://doi.org/10.1037/a0023756)] [Medline: [21604910](https://pubmed.ncbi.nlm.nih.gov/21604910/)]
75. Güldenpenning I, Schütz C, Weigelt M, Kunde W. Is the head-fake effect in basketball robust against practice? Analyses of trial-by-trial adaptations, frequency distributions, and mixture effects to evaluate effects of practice. *Psychol Res* 2020 Apr;84(3):823-833. [doi: [10.1007/s00426-018-1078-4](https://doi.org/10.1007/s00426-018-1078-4)] [Medline: [30128660](https://pubmed.ncbi.nlm.nih.gov/30128660/)]
76. Frings C, Brinkmann T, Friehs MA, van Lipzig T. Single session tDCS over the left DLPFC disrupts interference processing. *Brain Cogn* 2018 Feb;120:1-7. [doi: [10.1016/j.bandc.2017.11.005](https://doi.org/10.1016/j.bandc.2017.11.005)] [Medline: [29202318](https://pubmed.ncbi.nlm.nih.gov/29202318/)]
77. Friehs MA, Güldenpenning I, Frings C, Weigelt M. Electrify your game! Anodal tDCS increases the resistance to head fakes in basketball. *J Cogn Enhanc* 2019 Apr 24;4(1):62-70. [doi: [10.1007/s41465-019-00133-8](https://doi.org/10.1007/s41465-019-00133-8)]
78. Blascovich J, Loomis J, Beall AC, Swinith KR, Hoyt CL, Bailenson JN. Authors' response: immersive virtual environment technology: just another methodological tool for social psychology? *Psychol Inq* 2002 Apr;13(2):146-149. [doi: [10.1207/s15327965pli1302_03](https://doi.org/10.1207/s15327965pli1302_03)]
79. McDermott R. Experimental methodology in political science. *Polit Anal* 2017 Jan 4;10(4):325-342. [doi: [10.1093/pan/10.4.325](https://doi.org/10.1093/pan/10.4.325)]
80. Highhouse S. Designing experiments that generalize. *Organ Res Methods* 2007 Jul 23;12(3):554-566. [doi: [10.1177/1094428107300396](https://doi.org/10.1177/1094428107300396)]
81. Dobbins GH, Lane IM, Steiner DD. A note on the role of laboratory methodologies in applied behavioural research: don't throw out the baby with the bath water. *J Organ Behav* 1988;9(3):281-286 [FREE Full text]
82. Jackson SA, Marsh W. Development and validation of a scale to measure optimal experience: the flow state scale. *J Sport Exerc Psychol* 1996;18(1):17-35 [FREE Full text] [doi: [10.1123/jsep.18.1.17](https://doi.org/10.1123/jsep.18.1.17)]
83. Verbruggen F, Stevens T, Chambers CD. Proactive and reactive stopping when distracted: an attentional account. *J Exp Psychol Hum Percept Perform* 2014 Aug;40(4):1295-1300 [FREE Full text] [doi: [10.1037/a0036542](https://doi.org/10.1037/a0036542)] [Medline: [24842070](https://pubmed.ncbi.nlm.nih.gov/24842070/)]
84. Faul F, Erdfelder E, Lang AG, Buchner A. G*power 3.1 manual. *Behav Res Methods* 2007;39(2):175-191 [FREE Full text]
85. Verbruggen F, Aron AR, Band GP, Beste C, Bissett PG, Brockett AT, et al. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *Elife* 2019 Apr 29;8:- [FREE Full text] [doi: [10.7554/eLife.46323](https://doi.org/10.7554/eLife.46323)] [Medline: [31033438](https://pubmed.ncbi.nlm.nih.gov/31033438/)]

86. Verbruggen F, Logan GD. Evidence for capacity sharing when stopping. *Cognition* 2015 Sep;142:81-95 [[FREE Full text](#)] [doi: [10.1016/j.cognition.2015.05.014](https://doi.org/10.1016/j.cognition.2015.05.014)] [Medline: [26036922](https://pubmed.ncbi.nlm.nih.gov/26036922/)]
87. David FN, Tukey JW. Exploratory data analysis. *Biometrics* 1977 Dec;33(4):768. [doi: [10.2307/2529486](https://doi.org/10.2307/2529486)]
88. Verbruggen F, Chambers CD, Logan GD. Fictitious inhibitory differences: how skewness and slowing distort the estimation of stopping latencies. *Psychol Sci* 2013 Mar 1;24(3):352-362 [[FREE Full text](#)] [doi: [10.1177/0956797612457390](https://doi.org/10.1177/0956797612457390)] [Medline: [23399493](https://pubmed.ncbi.nlm.nih.gov/23399493/)]
89. Rouder JN, Morey RD, Verhagen J, Province JM, Wagenmakers E. Is There a Free Lunch in Inference? *Top Cogn Sci* 2016 Jul;8(3):520-547 [[FREE Full text](#)] [doi: [10.1111/tops.12214](https://doi.org/10.1111/tops.12214)] [Medline: [27489199](https://pubmed.ncbi.nlm.nih.gov/27489199/)]
90. Wagenmakers E, Marsman M, Jamil T, Ly A, Verhagen J, Love J, et al. Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. *Psychon Bull Rev* 2018 Feb;25(1):35-57 [[FREE Full text](#)] [doi: [10.3758/s13423-017-1343-3](https://doi.org/10.3758/s13423-017-1343-3)] [Medline: [28779455](https://pubmed.ncbi.nlm.nih.gov/28779455/)]
91. Wagenmakers E, Wetzels R, Borsboom D, van der Maas HL. Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *J Pers Soc Psychol* 2011 Mar;100(3):426-432. [doi: [10.1037/a0022790](https://doi.org/10.1037/a0022790)] [Medline: [21280965](https://pubmed.ncbi.nlm.nih.gov/21280965/)]
92. Wagenmakers E, Love J, Marsman M, Jamil T, Ly A, Verhagen J, et al. Bayesian inference for psychology. Part II: example applications with JASP. *Psychon Bull Rev* 2018 Feb;25(1):58-76 [[FREE Full text](#)] [doi: [10.3758/s13423-017-1323-7](https://doi.org/10.3758/s13423-017-1323-7)] [Medline: [28685272](https://pubmed.ncbi.nlm.nih.gov/28685272/)]
93. Stigler SM. Regression towards the mean, historically considered. *Stat Methods Med Res* 1997 Jun;6(2):103-114. [doi: [10.1177/096228029700600202](https://doi.org/10.1177/096228029700600202)] [Medline: [9261910](https://pubmed.ncbi.nlm.nih.gov/9261910/)]
94. Bland JM, Altman DG. Regression towards the mean. *Br Med J* 1994 Jun 4;308(6942):1499 [[FREE Full text](#)] [doi: [10.1136/bmj.308.6942.1499](https://doi.org/10.1136/bmj.308.6942.1499)] [Medline: [8019287](https://pubmed.ncbi.nlm.nih.gov/8019287/)]
95. Logan GD. The point of no return: a fundamental limit on the ability to control thought and action. *Q J Exp Psychol (Hove)* 2015;68(5):833-857 [[FREE Full text](#)] [doi: [10.1080/17470218.2015.1008020](https://doi.org/10.1080/17470218.2015.1008020)] [Medline: [25633089](https://pubmed.ncbi.nlm.nih.gov/25633089/)]
96. Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex. *Trends Cogn Sci* 2004 Apr;8(4):170-177. [doi: [10.1016/j.tics.2004.02.010](https://doi.org/10.1016/j.tics.2004.02.010)] [Medline: [15050513](https://pubmed.ncbi.nlm.nih.gov/15050513/)]
97. Aron AR, Behrens TE, Smith S, Frank MJ, Poldrack RA. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J Neurosci* 2007 Apr 4;27(14):3743-3752 [[FREE Full text](#)] [doi: [10.1523/JNEUROSCI.0519-07.2007](https://doi.org/10.1523/JNEUROSCI.0519-07.2007)] [Medline: [17409238](https://pubmed.ncbi.nlm.nih.gov/17409238/)]
98. Swann NC, Cai W, Conner CR, Pieters TA, Claffey MP, George JS, et al. Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. *Neuroimage* 2012 Feb 1;59(3):2860-2870 [[FREE Full text](#)] [doi: [10.1016/j.neuroimage.2011.09.049](https://doi.org/10.1016/j.neuroimage.2011.09.049)] [Medline: [21979383](https://pubmed.ncbi.nlm.nih.gov/21979383/)]
99. Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex: one decade on. *Trends Cogn Sci* 2014 Apr;18(4):177-185. [doi: [10.1016/j.tics.2013.12.003](https://doi.org/10.1016/j.tics.2013.12.003)] [Medline: [24440116](https://pubmed.ncbi.nlm.nih.gov/24440116/)]
100. Garavan H, Ross TJ, Murphy K, Roche RA, Stein EA. Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *Neuroimage* 2002 Dec;17(4):1820-1829. [doi: [10.1006/nimg.2002.1326](https://doi.org/10.1006/nimg.2002.1326)] [Medline: [12498755](https://pubmed.ncbi.nlm.nih.gov/12498755/)]
101. Wang M, Dong G, Wang L, Zheng H, Potenza MN. Brain responses during strategic online gaming of varying proficiencies: implications for better gaming. *Brain Behav* 2018 Aug;8(8):e01076 [[FREE Full text](#)] [doi: [10.1002/brb3.1076](https://doi.org/10.1002/brb3.1076)] [Medline: [30020566](https://pubmed.ncbi.nlm.nih.gov/30020566/)]
102. Catarino A, Küpper CS, Werner-Seidler A, Dalgleish T, Anderson MC. Failing to forget: inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. *Psychol Sci* 2015 May;26(5):604-616 [[FREE Full text](#)] [doi: [10.1177/0956797615569889](https://doi.org/10.1177/0956797615569889)] [Medline: [25847536](https://pubmed.ncbi.nlm.nih.gov/25847536/)]
103. Falconer E, Bryant R, Felmingham KL, Kemp AH, Gordon E, Peduto A, et al. The neural networks of inhibitory control in posttraumatic stress disorder. *J Psychiatry Neurosci* 2008 Sep;33(5):413-422 [[FREE Full text](#)] [Medline: [18787658](https://pubmed.ncbi.nlm.nih.gov/18787658/)]
104. Birk MV, Mandryk RL. The Benefits of Digital Games for the Assessment and Treatment of Mental Health. In: *Computing and Mental Health Workshop*. 2016 Presented at: CHI'16; May 8, 2016; San Jose, CA.
105. Birk MV, Wadley G, Abeele VV, Mandryk R, Torous J. Video games for mental health. *Interactions* 2019 Jun 26;26(4):32-36. [doi: [10.1145/3328483](https://doi.org/10.1145/3328483)]
106. Hedge C, Powell G, Sumner P. The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behav Res Methods* 2018 Jun;50(3):1166-1186 [[FREE Full text](#)] [doi: [10.3758/s13428-017-0935-1](https://doi.org/10.3758/s13428-017-0935-1)] [Medline: [28726177](https://pubmed.ncbi.nlm.nih.gov/28726177/)]
107. Hooshyar D, Ahmad RB, Yousefi M, Fathi M, Horng S, Lim H. Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. *Comput Educ* 2016 Mar;94:18-36. [doi: [10.1016/j.compedu.2015.10.013](https://doi.org/10.1016/j.compedu.2015.10.013)]
108. Bevilacqua F, Engström H, Backlund P. Game-calibrated and user-tailored remote detection of stress and boredom in games. *Sensors (Basel)* 2019 Jun 28;19(13):- [[FREE Full text](#)] [doi: [10.3390/s19132877](https://doi.org/10.3390/s19132877)] [Medline: [31261716](https://pubmed.ncbi.nlm.nih.gov/31261716/)]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ANOVA: analysis of variance
AOI: area of interest
BF: Bayes factor
FSS: Flow State Scale
IMI: Intrinsic Motivation Inventory
MANOVA: multivariate analysis of variance
OCD: obsessive-compulsive disorder
RT: reaction time
SDT: Self-Determination Theory
SSD: stop-signal delay
SSG: stop-signal game
SSRT: stop-signal reaction time
SST: stop-signal task

Edited by G Eysenbach; submitted 14.01.20; peer-reviewed by A De Marchi, N Khalili-Mahani, H Söbke; comments to author 12.06.20; revised version received 16.06.20; accepted 25.06.20; published 08.09.20

Please cite as:

Friehs MA, Dechant M, Vedress S, Frings C, Mandryk RL

Effective Gamification of the Stop-Signal Task: Two Controlled Laboratory Experiments

JMIR Serious Games 2020;8(3):e17810

URL: <https://games.jmir.org/2020/3/e17810>

doi: [10.2196/17810](https://doi.org/10.2196/17810)

PMID:

©Maximilian Achim Friehs, Martin Dechant, Sarah Vedress, Christian Frings, Regan Lee Mandryk. Originally published in JMIR Serious Games (<http://games.jmir.org>), 08.09.2020. 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 JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on <http://games.jmir.org>, as well as this copyright and license information must be included.