Gamification of an n-back working memory task - Is it worth the effort? An EEG and eye-tracking study

Scharinger, Christian; Prislan, Laura; Bernecker, Katharina; Ninaus, Manuel

Biological Psychology 179 (2023) 108545

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Gamification of an n-back working memory task – Is it worth the effort? An EEG and eye-tracking study

Christian Scharinger [a,*,1], Laura Prislan [b], Katharina Bernecker [c], Manuel Ninaus [d,e]

[a] Leibniz-Institut für Wissensmedien Tübingen, Germany
[b] Department of Teaching and Learning with Intelligent Systems (LLiS), University of Stuttgart, Germany
[c] Department of Psychology – Psychology of Motivation, Volition, and Emotion, University of Zurich, Switzerland
[d] Institute of Psychology, University of Graz, Austria
[e] LEAD Graduate School and Research Network, University of Tübingen, Germany

ARTICLE INFO

Keywords:
N-back task
Gamification
EEG frequency band power
Pupil dilation
Eye-tracking

ABSTRACT

Gamification of cognitive tasks might positively affect emotional-motivational factors (emotional design perspective) or negatively affect cognitive factors like working memory load (minimalistic design perspective). The current study examined the effects of gamification in a spatial n-back working memory task on task performance, task load (i.e., working memory load and effort), and subjective task experience. Task load was assessed by the physiological process measures pupil dilation and EEG theta (4–6 Hz) and alpha (8–13 Hz) frequency band power. Gamification was achieved by elements of emotional design (i.e., the visual screen design using, e.g., color, cartoon figures as n-back stimuli, and a narrative embedding of the task). While EEG and eye-tracking were recorded, participants conducted gamified and non-gamified 1-back and 2-back load levels. The gamification resulted in positive effects on subjective task experience and affect. Despite these effects, gamification did not affect task performance and task load. However, exploratory analyses revealed increased EEG theta power at right-parietal electrodes for gamified task versions compared to non-gamified ones. Potentially, this effect might indicate participants' increased effort or concentration in the gamified n-back task. In line with an emotional design perspective, gamification positively altered subjective task experience and affect without hampering task performance and therefore justify the extra effort of implementing game elements.

1. Introduction

From our everyday experience, it might be pretty indisputable that we engage in many cognitive tasks mainly because they promise some gains in the future. For instance, we study to get a university degree or work to get promoted or get a paycheck at the end of the month. Research suggests that immediate rewards (e.g., fun, positive feedback) are more relevant to people's task persistence than how important they regard the task to be in the long run (Brandstätter & Bernecker, 2022; Woolley & Fishbach, 2015, 2016).

An obvious idea to improve people's task persistence – and perhaps performance – might be to add immediate rewards. For basic cognitive tasks like working memory tasks this might be achieved by gamification, that is, the integration of game elements (Sanchez & Lee, 2022). Typical game elements used to gamify a task are, amongst others, the visual design of the task, embedding the task in a motivating narrative, and providing transparent feedback on task progress and success (Hamari et al., 2014; Landers et al., 2017; Plass et al., 2015; Robinson & Bellotti, 2013; Sailer et al., 2017).

Gamification is thought to result in positive emotional-motivational effects, hence being beneficial for task persistence and performance. This is in line with an *emotional design perspective* on task design (Plass et al., 2020; Plass & Kaplan, 2016; Um et al., 2012). However, research on gamification consistently only reported positive effects on subjective task experience, whereas effects on actual task performance are mixed (Bernecker & Ninaus, 2021; Ninaus et al., 2020; Parong & Mayer, 2019; Vermeir et al., 2020). Consequently, more detailed analyses of the different elements of gamification and their cognitive and emotional-motivational outcomes have been requested (Sailer et al., 2017). One potential disadvantage when gamifying existing task

* Correspondence to: Leibniz-Institut für Wissensmedien Tübingen, Schleichstr. 6, 72076 Tübingen, Germany.
E-mail address: c.scharinger@iwm-tuebingen.de (C. Scharinger).
[1] https://orcid.org/0000-0003-0287-8725

https://doi.org/10.1016/j.biopsycho.2023.108545
Received 16 November 2022; Received in revised form 21 March 2023; Accepted 22 March 2023
Available online 23 March 2023

elements or adding additional game elements to cognitive tasks might be an increased extraneous load on working memory. This is because working memory capacity is limited (Baddeley, 2012; Cowan, 2010). According to the cognitive load theory (CLT; Moreno & Park, 2010; Sweller et al., 1998), the task-unrelated, extraneous load should be as minimal as possible to avoid detrimental effects on task performance because of a potential overload situation of working memory. We will call this assumption the *minimalistic design perspective* on task design. Consequently, gamification could be detrimental to task performance as the integrated game elements might increase the extraneous load. In case gamification directly affects working memory, harmful cognitive effects should be most directly observable in working memory tasks and the performance therein.

In the current study, we were thus specifically interested in the effects of gamification on a working memory task. In addition to task performance, we were interested in the effects of gamification on task load, attention, and subjective task experience, allowing for a comprehensive investigation. Importantly, by utilizing physiological measures, we examined the potential effects of gamification on task load and attention on a process level. In particular, we compared gamified and non-gamified versions of a spatial n-back working memory task with two load levels (i.e., 1-back, 2-back). The theta (4–6 Hz) and alpha (8–13 Hz) frequency band power of the electroencephalogram (EEG) and pupil dilation served as process measures. Eye-tracking was used to assess participants' attention (i.e., total fixation durations) on certain game elements (i.e., a progress bar and score information on the main task screen).

1.1. Theoretical background

1.1.1. Working memory and instructional design

Working memory is considered at the heart of cognition, relevant for and defining performance in higher-order cognitive functions like information processing and learning (Diamond, 2013; St Clair-Thompson & Gathercole, 2006). In working memory, current information is temporarily held active in the focus of attention, processed, and integrated with previous information (i.e., prior knowledge) from long-term memory (Baddeley, 2012; Cowan, 2010). Working memory is closely linked to attentional executive control processes, such as shifting the attentional focus, or inhibiting irrelevant, distracting information (Baddeley, 1996; Cowan et al., 2012; Miyake & Friedman, 2012).

Concerning the visual design of cognitive task materials informed by current models and theories of instructional psychology, two orthogonal design perspectives (and derived design recommendations) can be differentiated: a *cognitive, minimalistic design perspective* and a motivational, *emotional design perspective*. The minimalistic design perspective builds on models and theories of multimedia learning like the cognitive theory of multimedia learning (CTML; Mayer, 2009) and the CLT (Moreno & Park, 2010; Sweller et al., 1998) and corresponding multimedia design principles (e.g., Mayer & Fiorella, 2014). According to the minimalistic design perspective, irrelevant design elements should be avoided to minimize extraneous cognitive load, avoid overloading working memory, and spare cognitive resources for the primary task. From that perspective, rather purist task versions would result in higher performance outcomes than gamified task versions. This is because game elements might raise extraneous cognitive load on working memory and might function as seductive details (e.g., Rey, 2012; Sundararajan & Adesope, 2020), that is, task-irrelevant but attention-capturing (i.e., distracting) elements.

In contrast to the minimalistic design perspective, which mainly focuses on the cognitive aspects of task performance, the emotional design perspective aims at incorporating emotional-motivational elements in task design (e.g., using color, pleasant forms, or anthropomorphisms). The emotional design perspective underlines the importance of such aspects as potentially beneficial for task performance (Brom et al., 2018; Heidig et al., 2015; Plass & Kaplan, 2016; Um et al., 2012). Conceptually, gamification and emotional design are closely related (Plass et al., 2020). From the emotional design perspective, a gamified task version might result in better performance. This is because gamified elements might positively affect users' mood and motivation during the task, increasing users' task engagement (Bernecker & Ninaus, 2021; Vermeir et al., 2020). Consequently, a gamified task version might reduce the subjectively perceived effort.

Indeed, previous research showed slower declines in positive affect and lowered subjective effort for a gamified compared to a non-gamified n-back task (Bernecker & Ninaus, 2021). Further, gamification did not affect task performance (i.e., correct responses) but task engagement (i.e., no responses; Bernecker & Ninaus, 2021). One explanation for the null finding on task performance might be the opposing effects of gamification on cognition (e.g., higher working memory load) and motivation (i.e., more positive affect). Gamification might cause a greater working memory load, which is compensated for by its positive effects on motivation. However, this previous work focused on self-reported affect and effort only and was thus not able to detect possible negative consequences of gamification on working memory load. The present research addresses this limitation by adding physiological measures of task load to the research design.

1.1.2. N-back working memory task and physiological process measures

The n-back working memory task is an ideal candidate for the current research question aiming at assessing the effects of gamification on task performance and working memory load by physiological process measures. First, the n-back working memory task has been widely used in neuroscientific research on working memory (e.g., Bledowski et al., 2010; Chen et al., 2008; Gevins & Smith, 2000; Palomäki et al., 2012; Scharinger et al., 2015). Thus, the task is proven suited to validly assess physiological process measures. However, in neuroscientific research, the n-back task is typically presented in rather purist, non-gamified designs. Second, the n-back task paradigm has some potential for being relatively easily converted into a gamified task version. Several studies have already described gamified n-back task versions (e.g., Bernecker & Ninaus, 2021; Lu et al., 2023; Vermeir et al., 2020).

Common to all n-back task variants is that participants get a timed sequence of trials with n-back stimuli (e.g., digits, letters, spatial locations on the screen). As defined by N, participants must keep a certain number of stimuli permanently active in working memory. For each stimulus presented, they must compare a particular stimulus parameter (e.g., position, color, value) with the stimulus N steps back in the sequence and decide which response key to press as the correct reaction. Thus, N paradigmatically defines the task's working memory load and task difficulty.

The EEG theta (4–6 Hz) and alpha (8–13 Hz) frequency band power at fronto-central electrodes (typically electrode Fz) and parietal electrodes (typically electrode Pz), respectively, have both been shown as sensitive measures indexing the working memory load of different n-back load levels (e.g., Gevins et al., 1997; Palomäki et al., 2012; Scharinger et al., 2015). Naturally, for increased n-back load levels, the EEG theta power rises, whereas the EEG alpha power decreases (Gevins et al., 1997; Palomäki et al., 2012; Scharinger et al., 2015). These changes in EEG frequency band power are commonly expressed in relation to a particular baseline (i.e., as percentage of change) and labeled event-related synchronization or desynchronization (i.e., ERD/ERS%-values; Pfurtscheller & Lopes da Silva, 1999; Scharinger, Schüler, et al., 2020). Important to note, increasing the N in the n-back task not only increases the working memory load correspondingly but also task difficulty and potentially the effort participants invest in the task (Brehm & Self, 1989; Richter et al., 2016). Therefore, changes in EEG theta and alpha ERD/ERS%-values might index not only working memory load but task load in a broader sense (i.e., consisting of working memory load and invested effort). We, therefore, will use the more general term task load henceforth.

In addition to EEG alpha and theta frequency band power, pupil dilation has been reported as a valid measure of n-back task load (e.g., Scharinger et al., 2015). Pupil dilation, however, is a rather broad but highly sensitive measure reacting to increases in working memory load, effort, or emotional arousal by increases in pupil diameter (Cabestrero et al., 2009; van der Wel & van Steenbergen, 2018). On the downside, pupil dilation is prone to luminance confounds (e.g., Van Gerven et al., 2004). That is, interpreting pupil dilation as a measure of task load has to be done with some caution. Nevertheless, being easily acquired alongside eye-tracking data, pupil dilation might be an interesting measure of the global task load and the effects of gamification therein. Therefore, in addition to the EEG frequency band power, we analyzed pupil dilation data in the current study. In sum, the physiological measures are thought to provide objective information on the effects of gamification on participants' task load while performing the n-back task.

1.1.3. The current study

In the current study, we were interested in the effects of gamification in a spatial working memory n-back task on task performance (i.e., behavioral data), task load (i.e., physiological data), and subjective task experience (i.e., self-report data). Specifically, we were interested in assessing the physiological process measures pupil dilation and EEG theta and alpha frequency band power. To the best of our knowledge, these physiological measures have not been reported before in studies on n-back working memory tasks for comparing gamified and non-gamified task versions.

The task we implemented was built on the spatial working memory n-back task developed by Bernecker and Ninaus (2021) with additional 1-back load levels. In the gamified version of the n-back task, we utilized popular game elements (e.g., Hamari et al., 2014) such as a narrative and a corresponding visual design, which aligns with the principles of the emotional design perspective (Pang et al., 2021; Plass et al., 2020). In the non-gamified task version, we forego any motivating narrative, and the overall visual design was kept simple concerning colors and forms. Importantly, in both task versions, the layout of the main task screen and the elements presented were identical (i.e., in both task versions, a score counter and a progress bar were shown, yet in a different visual design). Any potential differences observed in the physiological measures between gamified and non-gamified task versions should be specifically due to the emotional design but not due to having a different number of elements on the screen or other task mechanics (i.e., different feedback provided to the participants). This differs from Bernecker and Ninaus (2021), who compared a fully gamified condition against a condition without any game elements.

We assessed task performance via accuracy and reaction times, extending the study by Bernecker and Ninaus (Bernecker & Ninaus, 2021). Task load was assessed via EEG theta and alpha frequency band power and pupil dilation. Subjective task experience was evaluated in seven dimensions, namely *task success*, *effort*, *frustration*, *affect*, *user experience*, *flow*, and *attention*. The study was preregistered on aspredicted.org (Scharinger, Prislan, et al., 2020). Our main hypothesis was that compared to a non-gamified version, the gamified n-back task would reduce subjects' perceived effort. Further, we expected that this would be reflected in the neurophysiological measures as well because, at this time, we expected these measures to reflect effort and unfortunately disregarded at the time that they might as well reflect task load more generally rather than effort alone. Consequently, we hypothesized higher parietal EEG alpha power, a lower frontal EEG theta power, and decreased pupil dilation in the gamified n-back task compared to the non-gamified n-back task. However, given that gamification might as well lead to increased working memory load that is as well reflected in these measures, one would expect to observe increased pupil dilation and EEG theta power and a decreased EEG alpha power for gamified as compared to non-gamified task versions. Nevertheless, this was not the hypothesis we preregistered originally.

In addition, we expected that a higher n-back difficulty level (i.e., a 2-back) as compared to a lower n-back difficulty level (i.e., a 1-back) would generally result in increased effort as measured by subjective ratings, performance measures (i.e., reaction times and accuracies), and physiological measures (i.e., increased pupil dilation, decreased parietal EEG alpha power, increased frontal EEG theta power). We ran exploratory analyses: First, we did not preregister any hypothesis for task performance due to the null effect in previous work (Bernecker & Ninaus, 2021) and thus explored the effect in the present study. Additionally, we did not preregister any hypothesis on subjective experience outcomes except for effort. Finally, by analyzing the eye-tracking data, we explored the effects of the emotional vs. minimalistic design of the two elements in the task that provided additional feedback on the main task screen (i.e., the progress bar and the score information). For each of these two areas of interest (AOIs), we calculated the total fixation durations as a measure of visual attention (Alemdag & Cagiltay, 2018; Just & Carpenter, 1976, 1980; van Gog & Scheiter, 2010).

2. Method

2.1. Preregistration

This study's desired sample size, included variables, hypotheses, and planned analyses were preregistered at aspredicted.org (https://aspredicted.org/blind.php?x=QNN_GEE; Scharinger, Prislan, et al., 2020).

2.2. Participants

As preregistered, we recruited 20 healthy subjects (mean age = 23.50 years, SD = 3.10, 16 females, four males) who participated in the study for 10 € per hour. Our preregistered sample size has been mainly justified by resource constraints (i.e., by the available remuneration for participants and by time constraints for data collection because of the Covid-19 pandemic situation in Europe in 2020; for a discussion of valid justifications to determine sample size see Lakens, 2022). Note that physiological studies on n-back back working memory tasks typically report sample sizes between seven (Gevins et al., 1997) and 36 participants (e.g., Lu et al., 2023; Pesonen et al., 2007). All participants were German native speakers, right-handed as indicated by the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal visual acuity, and reported having neither known visual or neurological disorders (e.g., color blindness or epilepsy) nor having recently consumed relevant medications with impact on cognitive performance (e.g., tranquilizers). The study has been approved by the local ethics committee and was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. All participants gave their written informed consent before their inclusion in the study.

2.3. Materials

2.3.1. N-back tasks

Two versions of a spatial n-back task (i.e., a gamified and a non-gamified version) were used, each with two difficulty levels (i.e., a 1-back and a 2-back load level). The n-back task versions and difficulty levels were presented in a complete within-subject task design. While the n-back difficulty levels were presented in a fixed sequence within the task versions (i.e., 1-back load level followed by 2-back load level), the sequence of task versions was counter-balanced across participants (i.e., half of the participants started with the gamified n-back versions, half of the participants with the non-gamified versions). The main differences between the gamified and the non-gamified versions were the visual design of the screen (i.e., the use of color and drawings) that corresponded to the used narrative of the task (i.e., the task instructions).

In the gamified version, the task was introduced by a written story explaining to the participants that zombies invaded their local city and that only they could fight against the zombies and potentially rescue the city by successfully performing the following task (i.e., the n-back task). The main task screen showed a graveyard with four graves and tombstones (see Fig. 1). A zombie (i.e., a cartoon figure) could appear at each of the four locations. To each position, a specific key on a conventional computer keyboard was assigned (i.e., the key 'W', 'D', 'P', 'L'). The pseudo-random occurrence of a zombie at one of the four positions for 900 ms followed by an inter-stimulus interval of about 1400 ms showing the graveyard only defined a trial. The n-back task conditions were presented in blocks, each consisting of 180 trials. Note that within a block, each of the four positions was used equally often for the occurrence of a zombie. In the 1-back task condition, participants were instructed to press the key that matched the position of the occurrence of the zombie in the previous trial (i.e., the trial 1-step back in the sequence) each time when a zombie occurred (see Fig. 1, upper part for a schematic depiction). Participants were told that in doing so, they would be able to kill that specific zombie. In the 2-back condition, participants were instructed to press the key that matched the zombie's position in the trial two steps back in the sequence to kill the corresponding zombie. Note that no reactions were possible at the beginning of a block for the first trial (1-back condition) and the first two trials (2-back condition), respectively. A numeric score was shown on the screen's upper-right, counting correct key presses. On the upper left of the screen, a status bar in the form of a brain that filled up was shown to provide participants with information on the remaining length of the n-back block.

The non-game version introduced the task as a classical working memory task. The primary screen layout was identical to the gamified task version. Yet, a simple gray background replaced the graveyard, gray rectangles replaced the graves and tombstones, and rectangles of lighter gray replaced the zombies pseudo-randomly popping up at one of the four positions. The fully functional score counter and the status bar were shown at the same positions as in the gamified task version, yet in a rather puristic design (i.e., geometrical forms in gray color, see Fig. 1, right-hand side). The number of trials, the timing, the difficulty levels (i.e., a 1-back and a 2-back block), and the required reactions were identical to the gamified task version.

2.3.2. Questionnaires

A modified German version of the PANAS (Janke, & Glöckner-Rist, 2014) was used to assess participants' affect. In this version, the PANAS subscale (i) *activity* was defined by the items (German adjectives) "active", "interested", "awake", the subscale (ii) *joy* by the items "excited", "joyfully aroused", the subscale (iii) *fatigue* by the items "tired", "debilitated", "weary", "exhausted", the subscale (iv) *fear* by the items "anxious," "distressed," "jumbled", and the subscale *upset* by the items "irritated," "annoyed". Participants were asked to judge their current feeling with respect to each item on slider rating scales ranging from 0 ("not at all") to 100 ("extremely"). The PANAS was administered after each n-back block.

Three modified items of the NTLX (Hart & Staveland, 1988) were used to assess participants' subjective ratings of perceived task *success* ("How successful do you think you were in the task you just completed?"), *effort* ("How hard did you have to work to achieve this performance?"), and *frustration* ("How frustrated did you feel during the task?"), each item using slider rating scales ranging from 0 ("not at all") to 100 ("extremely"). The NTLX was administered after each n-back block.

A German version of the UEQ (Laugwitz et al., 2008; Schrepp et al., 2017) was used to assess participants' user experience for the gamified and the non-gamified n-back versions (i.e., without differentiating between 1-back and 2-back blocks). All six UEQ scales were used (i.e., *attractiveness*, *perspicuity*, *efficiency*, *dependability*, *stimulation*, and *novelty*). Each scale consisted of 4 items, except for the scale *attractiveness*, which consisted of 6 items. Each item consisted of an orthogonal word pair at the left and right end of a slider rating scale (e.g., "efficient – non-efficient", subscale *efficiency*). By positioning the slider of the rating scale accordingly, participants had to indicate whether they would agree with the negative or positive terms concerning the previously conducted n-back task. For statistical analyses, the output of the slider rating scales was converted to the typical UEQ range from − 3 at the negative end to + 3 at the positive end.

Using the German short-version of the flow questionnaire (FKS; Rheinberg et al., 2003), participants' flow experience during the gamified and non-gamified n-back versions (without differentiating between 1-back and 2-back blocks) was assessed on three dimensions. Three items assessed the dimension *fear of failure* (FoF) (e.g., "I was not allowed to make any mistakes."), the dimension *task fluency* by six items (e.g., "My thoughts ran smoothly and fluidly."), and the dimension *task absorption* by four items (e.g., "I was completely absorbed in the task."). Using slider rating scales labeled "does not apply" on the left end, "partly applies" in the middle, and "very much applies" on the right end, participants were asked to rate their task experience for each of the flow items.

Finally, participants' attention on the task (without differentiating between 1-back and 2-back blocks) was assessed using a mind-wandering scale (1 item, "My mind has often wandered.") with the same slider rating scale as used in the flow questionnaire.

2.4. Procedure

At the beginning of the study, the EEG system was prepared (i.e., setting up the EEG cap, applying electrode gel, and impedance measurement), and some essential demographic variables were assessed (i.e., age, sex, handedness). As an initial task, participants performed a short eye-movement task to record typical patterns of the electro-occulogram (EOG, horizontal and vertical eye movements, blinks) within the EEG, followed by a relaxation task of about 4 min. The

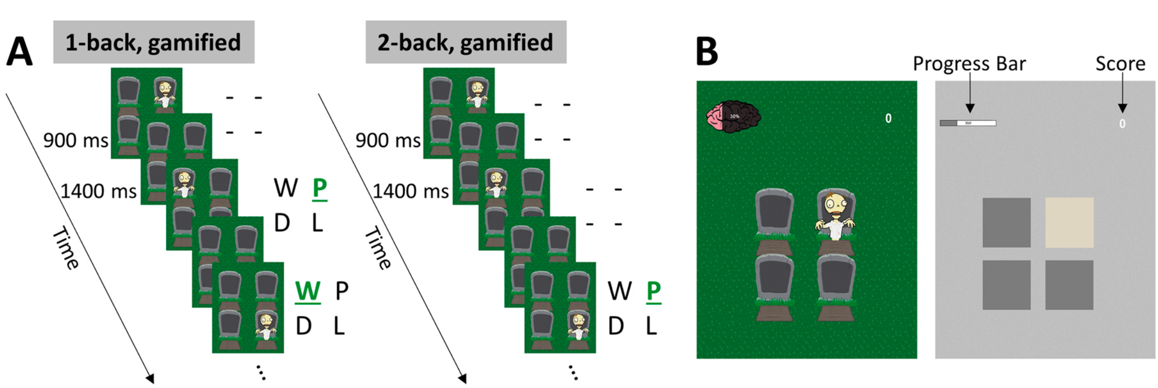

Fig. 1. (A) Schematic sequence of n-back trials of a 1-back and a 2-back block with corresponding response keys (correct key = green, bold, underlined). (B) Schematic depiction of the gamified and non-gamified screen layout. Note that the width of the screen depicted here is reduced for reasons of space limits.

relaxation task consisted of presenting aesthetically pleasant landscape pictures (8 pictures, presented in random order for 15 s each, each picture presented twice) combined with relaxing music. Participants were instructed to watch the pictures and to relax. At the end of the relaxation task, participants' current affective state was assessed using the PANAS.

Participants then performed two blocks of the n-back task, a 1-back and a 2-back block in a fixed sequence, both presented in the gamified or non-gamified version. The sequence of the gamified and the non-gamified n-back version was counter-balanced across participants to avoid sequential effects. At the beginning of each n-back block, participants performed a short training session on the task. During that, direct feedback (correct or incorrect) was given when participants pressed a key. If participants failed to reach an accuracy of at least 50% correct responses for the training task block, the n-back training was repeated. Doing so ensured that participants correctly understood the task before the actual task block started. After successful eye-tracking calibration, the actual n-back task block started. At the beginning of each block, the task was instructed once more. Before the first trial, the basic scenery (i.e., the graveyard or the gray background) was shown for 8 s. This pre-block baseline was used to baseline-correct the pupil dilation data. Participants had to fill in the NTLX and PANAS after each n-back. Additionally, after the 2-back blocks, the UEQ, flow, and mind-wandering scales were administered (i.e., allowing only to differentiate between gamified and non-gamified n-back versions). In total, the n-back task consisted of 4 blocks (i.e., 1-back and 2-back, each in a gamified and a non-gamified version). Including breaks, the study summed up to about 2–2 ½ hours (depending on the length of the breaks and the time needed for the EEG preparation).

2.5. Apparatus

Note that the apparatus, as well as the data pre-processing pipeline described below, has been used and validated in previous studies of the first author (Scharinger et al., 2017; Scharinger et al., 2015; Scharinger, Schüler, et al., 2020; Scharinger et al., 2015). Participants sat in a comfortable chair in front of a 24-inch Dell monitor (Dell U 2412 M, 1920 ×1200 px screen resolution, eye-to-monitor distance approx. 70 cm), in an evenly lit, quiet room. Stimuli were presented using PsychoPy2 (V. 1.86; Peirce et al., 2019). At the start of each trial, triggers were sent out to the EEG and the eye-tracking system for later synchronization of the data streams.

A 250 Hz SMI remote eye-tracking system (SMI RED 250m, SensoMotoric Instruments, Teltow, Germany) recorded participants' gaze behavior (binocular) at a sampling rate of 250 Hz (SMI iViewRed 4.2.77). A chin rest was used to avoid head movements during data recording and to ensure the eyes remained at a fixed distance from the eye-tracking device. The eye-tracker was calibrated before the start of each task using the built-in calibration routines (SMI, 9-point calibration, and validation; the calibration was repeated in case of gaze deviations $> 1.0^{\circ}$ visual angle).

EEG data were recorded (PyCorder 1.0.9) at a 1000 Hz sampling rate (ActiCHamp amplifier, Brainproducts, Ltd., Gilching, Germany) using 32 active electrodes (actiCap slim electrodes, Brainproducts, Ltd., Gilching, Germany). Impedances were kept below 10 kΩ. 30 electrodes were placed on the scalp (i.e., Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP3, CP4, CP6, P7, P3, Pz, P4, P8, POz, O1, O2) positioned according to the international 10/20 system (Jasper, 1958). Two electrodes were placed on the mastoids, with the right mastoid as the reference during recording. The ground electrode was positioned at AFz.

2.6. Data pre-processing and analyses

All data pre-processing steps were conducted using customized Matlab scripts (Matlab 2018b, MathWorks, Inc., Natick, MA, USA) and the toolbox EEGLAB (v. 14.1.2; Delorme & Makeig, 2004)) with the EYE-EEG plugin (v.0.85; Dimigen et al., 2011). Statistical analyses were conducted within Matlab (EEGLAB function 'statcond') and using R (v. 4.0.5, R Core Team, 2020), with the packages 'ez', (Lawrence, 2016), 'emmeans', (Lenth, 2021), and 'schoRsch', (Pfister & Janczyk, 2016).

First, the single eye-tracking and EEG data recordings were synchronized (EYE-EEG). The triggers at the onset of each trial in both the EEG data and the eye-tracking data served as synchronization events. Overall, the average jitter between the triggers in the EEG and the corresponding ones in the eye-tracking data was within ± 6 data sample (i.e., ± 6 ms). The eye-tracking data were integrated into the EEG data as additional channels (i.e., channels containing each eye's pupil dilation data and the raw gaze positions), with the original sampling rate of the eye-tracking data being up-sampled to match the sampling rate of the EEG data. All single recordings of a participant were then combined in one EEG data file. Blinks in the pupil dilation data were interpolated using an algorithm by Siegle and colleagues (Siegle et al., 2003).

The continuous EEG data were filtered (low-pass 48 Hz, high-pass 0.25 Hz, linear finite impulse response filters). EOG artifacts (eye movements, blinks) were corrected by using independent component analysis (ICA) decompositions (Infomax ICA, EEGLAB function 'runica'). Independent components (ICs) visually identified as EOG-ICs were rejected (Delorme et al., 2007; Jung et al., 2000). The EEG data were then re-referenced to average reference. The two electrodes positioned at the mastoids were excluded, resulting in 30 symmetrically distributed EEG channels over the scalp in the final data set used for statistical analyses.

Finally, the continuous EEG data were divided into epochs of 2.5 s length, including a 250 ms pre-stimulus period, time-locked to stimulus-onset of the n-back stimuli. Using these epochs, an automatic artifact removal was performed: Epochs that exceeded ± 100 μV were excluded from further analyses (Duncan et al., 2009; Pesonen et al., 2007). On average, this procedure removed 1.37% (SD = 2.78) of all data epochs. No further artifact removal or correction was performed on the EEG data. Using fast-Fourier transforms (FFT; 500 ms sliding window, Hanning tapered), the frequency band power values (absolute values) in the frequency range between 2 and 30 Hz (0.125 Hz frequency spacing) were calculated for the data epochs cut to 2 s length, averaged for the different task conditions. The event-related desynchronization and synchronization (ERD/ERS%; Pfurtscheller & Lopes da Silva, 1999) were then calculated. This was done separately for each electrode and frequency band. As baselines served the frequency band power of all four task conditions averaged together (i.e., a whole epoch, global condition baseline; Cohen, 2014). Consequently, the ERD/ERS%-values indicate a relative increase or decrease in frequency band power of a specific task condition in relation to all task conditions. Based on literature (e.g., Gevins & Smith, 2000), frontal-midline electrode Fz and parietal-midline electrode Pz were chosen for statistical analyses of the EEG theta and alpha frequency band ERD/ERS%, respectively.

For the eye-tracking data, as a measure of visual attention, total fixation durations on two areas of interest (AOIs) were calculated by summing up for each AOI separately all single fixation durations on that specific AOI. The two AOIs were rectangular, positioned around the score display and the progress bar. Notably, the dimensions of the AOIs were identical for all task conditions.

Note that the first two trials of each block were excluded from all analyses. Reaction times, pupil dilation, and the EEG theta and alpha ERD/ERS% were calculated for correctly solved trials only. Accuracy was calculated as the percentage of trials with correct key-presses in a block in relation to all trials of the block (excluding the first two trials). Two-factorial repeated-measures ANOVAs with the factors load (1-back, 2-back) and gamification (gamified, non-gamified) were conducted for each dependent variable. For post hoc pairwise comparisons (*t*-tests, two-tailed) of significant ANOVA effects, *p*-values were Bonferroni-Holm corrected for multiple comparisons. The significance level was set at $\alpha = 0.05$ for all analyses, and partial eta-square (η_p^2) is reported as a measure of effect size for the ANOVAs. In addition, for exploratory data analyses and visualization, to inspect the localization of the theta and alpha ERD/ERS%-changes on the scalp, topoplots showing the ERD/ERS %-values at all 30 electrode sites were calculated.

Note that some of the data pre-processing techniques detailed here differed slightly from those we preregistered. Instead of calculating EEG frequency band power using Morlet wavelet convolution, we used the fast-Fourier transform (FFT). FFT is also a standard methodology for calculating EEG frequency band power. We used FFT for reasons of time efficiency (i.e., the FFT could be calculated on a standard laptop computer requiring less time) and to avoid edge artifacts (i.e., Morlet wavelet convolution typically requires longer time windows for analysis as the epochs of about 2 s we had in the current study). Instead of a pre-stimulus baseline, we used a combined whole epoch baseline for calculating the ERD/ERS%-values.

3. Results

3.1. Behavioral task performance data

For accuracy, the 2-factorial repeated-measures ANOVA revealed a significant main effect of *load*, $F(1,\ 19) = 34.88$, $p < .001$, $\eta_p^2 = .65$. Accuracies were generally decreased for the 2-back as compared to the 1-back load levels (Table I). Neither the factor *gamification*, $F(1,\ 19) = 2.39$, $p = .138$, $\eta_p^2 = .11$, nor the interaction between *gamification* and *load* were significant, $F(1,\ 19) = 2.61$, $p = .123$, $\eta_p^2 = .12$.

Although numerically, the reaction times were slightly lower in the game as compared to the non-game conditions, as well as in the 1-back as compared to the 2-back load levels (Table 1), none of these effects were statistically significant, *load*, $F < 1$, $p > .661$, *gamification*, $F(1,\ 19) = 1.78$, $p = .198$, $\eta_p^2 = .09$, and no significant interaction between *gamification* and *load*, $F(1,\ 19) = 2.82$, $p = .109$, $\eta_p^2 = .13$. Overall, our findings replicate the null effect of gamification on task performance reported by Bernecker and Ninaus (2021).

3.2. Subjective ratings

3.2.1. NTLX

The NTLX (Table 2) showed outcomes in line with behavioral performance. Participants judged their task success as being higher in the 1-back as compared to the 2-back load levels, main effect of *load*, $F(1,\ 19) = 29.93$, $p < .001$, $\eta_p^2 = .61$. There was no effect of *gamification*, $F(1,\ 19) = 1.29$, $p = .270$, $\eta_p^2 = .06$, and no interaction between *gamification* and *load*, $F < 1$, $p > .989$. Comparably, participants judged their invested effort being higher in the 2-back as compared to the 1-back load levels, main effect *load*, $F(1,\ 19) = 10.56$, $p = .004$, $\eta_p^2 = .36$, no effect for *gamification*, $F < 1$, $p > .974$, no interaction, $F < 1$, $p > .560$. This was not in line with our main hypothesis, which stated a main effect of gamification on subjective effort ratings. Participants were more frustrated in the 2-back as compared to the 1-back load levels, as revealed by a main effect of *load*, $F(1,\ 19) = 4.55$, $p = .046$, $\eta_p^2 = .19$, with both the factor *gamification* and the interaction between *gamification* and *load* being non-significant, all $F < 1$, $p > .344$.

3.2.2. PANAS

The PANAS subscale *activity* showed a main effect of *load*, $F(1,\ 19) = 5.69$, $p = .028$, $\eta_p^2 = .23$, and *gamification*, $F(1,\ 19) = 5.33$, $p = .032$, $\eta_p^2 = .22$, with the interaction being not significant, $F < 1$, $p > .427$. The score for activity was overall higher in the game compared to the non-game conditions and higher for the 2-back as compared to the 1-back load levels (Fig. 2). The PANAS subscale *joy* (Fig. 2) showed a main effect of *gamification*, $F(1,\ 19) = 9.97$, $p = .005$, $\eta_p^2 = .34$, no effect of *load*, $F(1, 19) = 1.56$, $p = .227$, $\eta_p^2 = .08$, and no interaction between the two, $F < 1$, $p > .227$. The PANAS score on the dimension joy was increased in the gamified compared to the non-gamified task versions. The PANAS subscale *fatigue* (Fig. 2) showed no significant effect, all $F < 1$, $p > .345$. The PANAS subscale *fear* did not show any significant effect (1-back, non-game: $M = 10.45$, $SD = 8.34$, 1-back, game: $M = 10.15$, $SD = 11.31$, 2-back, non-game: $M = 11.67$, $SD = 10.01$, 2-back, game: $M = 11.25$, $SD = 10.01$) neither for *gamification*, $F < 1$, $p > .856$, nor *load*, $F(1,\ 19) = 1.16$, $p = .295$, $\eta_p^2 = .06$, with the interaction being also not significant, $F < 1$, $p > .965$. Also, there were no effects for the PANAS subscale *upset*, all $F < 1$, $p > .344$ (1-back, non-game: $M = 10.45$, $SD = 12.38$, 1-back, game: $M = 10.72$, $SD = 10.87$, 2-back, non-game: $M = 11.82$, $SD = 15.49$, 2-back, game: $M = 11.80$, $SD = 12.90$). Overall, the analyses replicate previous findings on the positive effects of gamification on positive affect (i.e., activity, joy) and a null effect of gamification on negative affect (i.e., fatigue, fear, upset; Bernecker & Ninaus, 2021).

3.2.3. UEQ

1-factorial repeated-measures ANOVAs revealed significant main effects for *gamification* on the UEQ scales *attractiveness*, $F(1,\ 19) = 9.67$, $p = .006$, $\eta_p^2 = .34$, *stimulation*, $F(1,\ 19) = 11.72$, $p = .003$, $\eta_p^2 = .38$, and *novelty*, $F(1,\ 19) = 22.19$, $p < .001$, $\eta_p^2 = .54$. The gamified n-back versions were perceived as more attractive ($M = 1.00$, $SD = 0.94$) as compared to the non-gamified versions ($M = 0.26$, $SD = 0.47$), more stimulating (gamified: $M = 0.66$, $SD = 0.88$, non-gamified: $M = -0.07$, $SD = 0.69$), and more novel (gamified: $M = 0.60$, $SD = 1.15$, non-gamified: $M = -0.78$, $SD = 1.11$). The UEQ scales *perspicuity* (gamified: $M = 1.57$, $SD = 0.72$, non-gamified: $M = 1.48$, $SD = 0.99$), $F < 1$, $p > .650$, *efficiency* (gamified: $M = 1.00$, $SD = 0.78$, non-gamified: $M = 1.27$, $SD = 0.59$), $F(1,\ 19) = 2.47$, $p = .133$, $\eta_p^2 = .11$, and *dependability* (gamified: $M = 1.02$, $SD = 0.74$, non-gamified: $M = 1.13$, $SD = 0.62$), $F < 1$, $p > .468$, did not show any significant difference.

3.2.4. Flow questionnaire

The flow questionnaire did not show any significant difference between the gamified and the non-gamified n-back versions (see Table 3) neither on the dimension *fluency*, *absorption*, or *fear of failure (FoF)* as revealed by 1-factorial repeated-measures ANOVAs; all $F < 1$, $p > .361$.

3.2.5. Participants' attention on the task

Participants' attention on the task was not significantly altered by gamification (game: $M = 35.70$, $SD = 30.42$, non-game: $M = 29.85$, $SD = 25.68$), $F(1,\ 19) = 1.33$, $p = .263$, $\eta_p^2 = .07$.

3.3. Physiological process measures

3.3.1. Pupil dilation

Pupil dilation (baseline-corrected) was significantly increased for the 2-back task conditions as compared to the 1-back task conditions (Fig. 3), as revealed by a main effect of *load*, $F(1,\ 19) = 42.79$, $p < .001$, $\eta_p^2 = .69$. There was no main effect of *gamification* and no interaction between *load* and *gamification*, all $F < 1$, $p > .808$.

3.3.2. Eye-tracking

Interestingly, the average total fixation duration on the score element (upper right side on the task screen) was larger in the 2-back (gamified: $M = 24.75$ s, $SD = 25.17$, non-gamified: $M = 20.39$ s, $SD = 24.53$) as compared to the 1-back conditions (gamified: $M = 11.87$ s, $SD = 20.68$, non-gamified: $M = 7.80$ s, $SD = 12.28$), main effect *load*, F(1,

Table 1
Behavioral Task Performance Measures.

Load	Gamification	Accuracy [%]		RT [ms]	
1-back	game	91.71	(9.35)	568	(314)
1-back	non-game	92.44	(6.13)	649	(369)
2-back	game	81.71	(14.93)	602	(333)
2-back	non-game	78.23	(13.80)	637	(381)

Note. Mean accuracy and reaction times (RT) for n-back trials. SD in brackets.

Table 2
Descriptive Statistics of the NTLX by condition.

Load	Gamification	Success		Effort		Frustration	
1-back	**game**	69.80	(25.09)	57.70	(21.76)	25.85	(25.23)
1-back	**non-game**	64.70	(20.80)	55.80	(21.60)	34.00	(21.74)
2-back	**game**	51.75	(23.21)	68.65	(16.49)	38.75	(26.08)
2-back	**non-game**	46.75	(23.39)	70.30	(11.85)	37.95	(23.62)

Note. Mean rating values, *SD* in brackets.

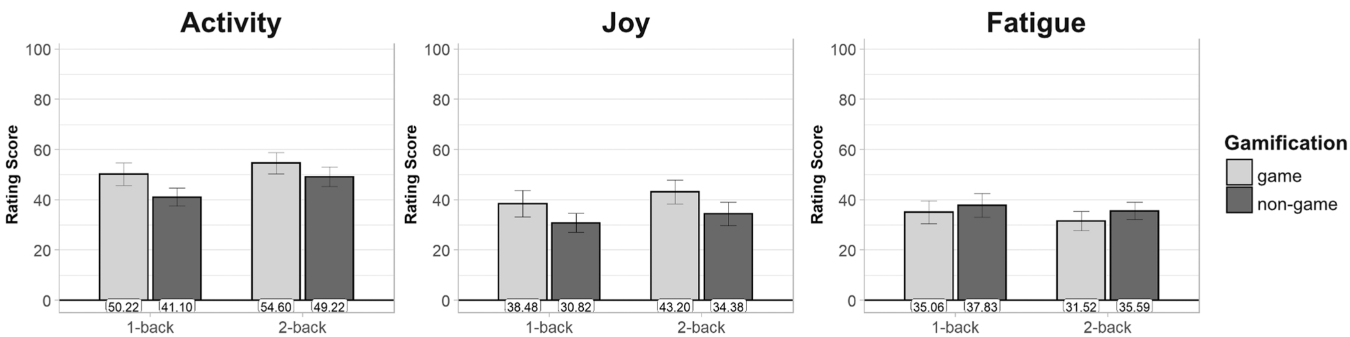

Fig. 2. Barplots, depicting the PANAS rating scores for the dimensions *activity*, *joy*, *fatigue*. Note. Error bars = ± 1SEM.

Table 3
Descriptive Statistics of the Flow Questionnaire by condition.

Gamification	Fluency		Absorption		FoF	
game	54.49	(16.34)	57.63	(20.03)	40.22	(23.95)
non-game	51.63	(16.75)	56.49	(14.11)	37.87	(22.94)

Note. Mean rating values for the flow dimensions *fluency*, *absorption*, and *fear of failure (FoF)*. *SD* in brackets.

19) = 14.21, $p = .001$, $\eta_p^2 = .43$, with neither an effect for *gamification*, $F(1, 19) = 1.19$, $p = .290$, $\eta_p^2 = .06$, nor an interaction, $F < 1$, $p > .965$. The average total fixation duration on the progress bar showed an orthogonal pattern with decreased average total fixation durations in the 2-back (gamified: $M = 2.16$ s, $SD = 2.55$, non-gamified: $M = 1.75$ s, $SD = 1.96$) as compared to the 1-back load levels (gamified: $M = 5.10$ s, $SD = 6.07$, non-gamified: $M = 5.27$ s, $SD = 7.11$), main effect *load*, $F(1, 19) = 9.26$, $p = .007$, $\eta_p^2 = .33$, no effect of *gamification* and no interaction, all $F < 1$, $p > .643$.

3.3.3. *EEG frequency band power*

The EEG theta ERD/ERS% at frontal electrode Fz (Fig. 3) was not significantly affected by the experimental manipulations, albeit numerically, the EEG theta power was relatively increased for the 2-back as compared to the 1-back load levels, as revealed by a trend for *load*, $F(1, 19) = 4.05$, $p = .059$, $\eta_p^2 = .18$. There was no effect of *gamification*, $F < 1$, $p > .636$, and no interaction, $F(1, 19) = 1.64$, $p = .216$, $\eta_p^2 = .08$ for theta ERD/ERS%-values on frontal electrode Fz. Interestingly, however, the exploratory topoplots (Fig. 4, left-hand side) indicated differences in EEG theta power between gamified and non-gamified n-back tasks predominantly at right-parietal electrodes. The EEG theta power at right-parietal electrodes was increased in the gamified task conditions of both, the 1-back and 2back load levels as compared to the non-gamified task conditions.

The alpha frequency band power at parietal electrode Pz (Fig. 3) showed an apparent effect of the n-back load manipulations, main effect *load*, $F(1, 19) = 19.57$, $p < .001$, $\eta_p^2 = .51$, with decreased EEG alpha ERD/ERS%-values in the 2-back as compared to the 1-back load levels. There was no effect for *gamification*, $F < 1$, $p > .701$, and no interaction, $F(1, 19) = 2.04$, $p = .169$, $\eta_p^2 = .10$. Exploratory topoplots (Fig. 4, right-hand side) indicated the decrease in alpha power for 2-back as compared to 1-back load levels to be widely distributed across the scalp. Gamification did not result in an alpha frequency band power effect at any electrode.

4. Discussion

In the current study, we were interested in the effects of gamification on task performance, task load as assessed by physiological measures, and subjective task experience. From an emotional design perspective, we expected participants' perceived effort to be reduced in the game

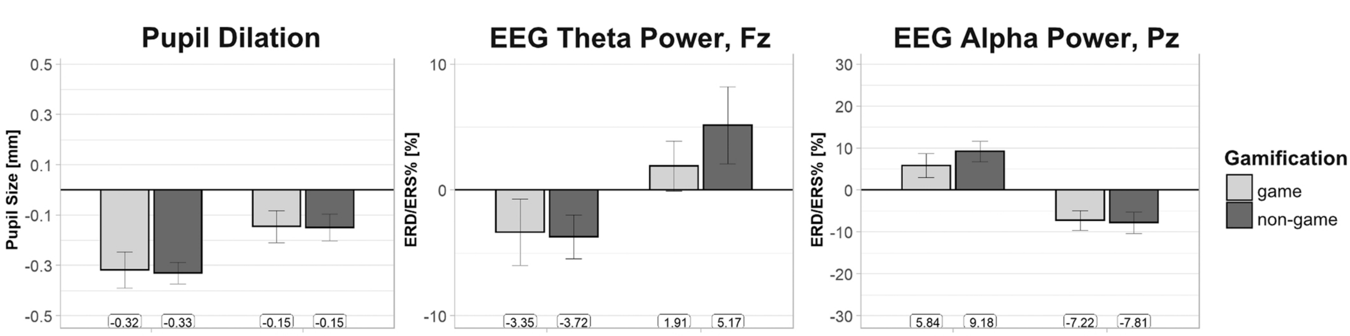

Fig. 3. Barplots, depicting the mean values of the physiological measures: pupil dilation, EEG theta ERD/ERS%-values at frontal electrode Fz, and EEG alpha ERD/ERD%-values at parietal electrode Pz. Note. Error bars = ± 1SEM.

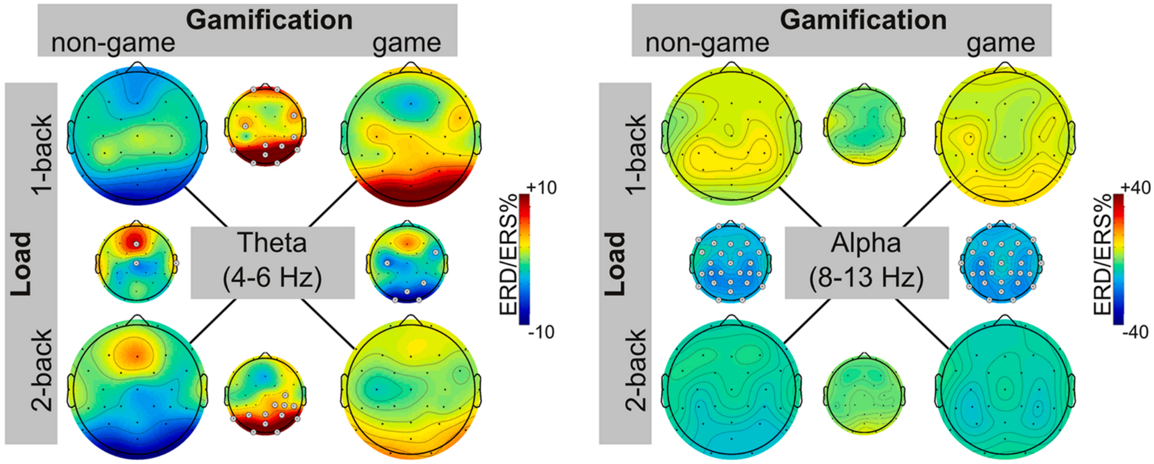

Fig. 4. Topoplots depicting the EEG theta and alpha frequency band power (ERD/ERS%-values) of each electrode for the n-back tasks of both, the gamified and non-gamified n-back task version as well as the 1-back and the 2-back load levels. Smaller topoplots adjacent to the larger ones indicate the difference between pairs of conditions (i.e., pairwise comparisons). The difference plots were created by subtracting for each electrode the ERD/ERS%-values of the non-game condition from the game condition and the 1-back condition from the 2-back condition, respectively. Electrodes marked by gray disks surrounded by black circles indicate electrode locations showing significant differences between task conditions (t-tests, two-sided, $p < .05$, permutation-based statistics, using false-discovery rate to correct for multiple comparisons).

compared to the non-game task variants. This is because gamification might positively alter emotional-motivational aspects of task experience. Yet, from a minimalistic design perspective, we expected that gamification might result in increased extraneous working memory load (i.e., task load), as reflected in increased pupil dilation, increased EEG theta frequency band power at frontal electrode Fz, and a decreased EEG alpha frequency band power at parietal electrode Pz when comparing gamified and non-gamified n-back task versions. Concerning task performance, our expectations were ambiguous. From the minimalistic design perspective, the potentially increased task load under gamification might result in decreased task performance, especially in the high (2-back) working memory load conditions when working memory resources are already depleted. Yet, from an emotional design perspective, because of the potentially positive effects of gamification on participants' mood and motivation, task performance might be increased in gamified as compared to non-gamified n-back task versions.

The main results were the following. First, as indicated by the subjective measures of task experience, the gamified n-back version was perceived as more attractive, novel, and stimulating than the non-gamified task versions (UEQ). The PANAS also indicated positive effects on the dimensions *activity* and *joy* for the gamified as compared to the non-gamified n-back versions. In sum, gamification improved subjective task experience and resulted in beneficial effects on participants' affective experience.

Second, the physiological process measures did not differ between gamified and non-gamified task versions concerning the EEG theta frequency band power - at least not at the expected electrode locations. Based on this outcome, we might conclude that the gamification of the n-back task in the current study did not result in increased working memory load as the minimalistic design perspective would have suggested. Notably, gamification did not significantly affect the subjective ratings of effort (NTLX), which was not in line with our main hypothesis and previous findings (Bernecker & Ninaus, 2021).

Third, the n-back load levels resulted in the typical effects known from the literature (e.g., Palomäki et al., 2012; Scharinger et al., 2015). In the high (2-back) as compared to the low (1-back) n-back load levels, pupil diameter was increased, EEG theta power at electrode Fz (at least numerically) increased, EEG alpha power at electrode Pz decreased, behavioral performance decreased (i.e., decreased accuracies, at least numerically increased reaction times), and subjective ratings of effort (NTLX) increased. Interestingly, these effects were modulated by gamification for none of these measures. In the following, we will discuss these primary outcomes and some additionally observed results and insights in more detail.

4.1. Effects of gamification on task experience

The UEQ indicated increased user experience for gamified n-back versions on the dimensions *attractivity*, *novelty*, and *stimulation* but not on the dimensions *perspicuity*, *dependability*, and *efficiency*. Observing no-effects on the latter three dimensions, which are considered to reflect pragmatic qualities (Laugwitz et al., 2008; Schrepp et al., 2017), is not surprising as the n-back task mechanics (i.e., the task per se), as well as the number of task elements (i.e., the score and the progress feedback), were not altered between the task versions. Furthermore, the dimensions *perspicuity* and *dependability* might not be very informative for an n-back task. This is because the n-back task typically is system paced and does not require (or allow) much interaction with the system. Yet, the effects on the dimensions *attractivity*, *novelty*, and *joy* indicated that participants' user experience was positively affected by gamification.

The observed effects on the PANAS for the dimensions *joy* and *activity* underline the positive impact of gamification on subjective experience. Replicating previous findings (Bernecker & Ninaus, 2021), gamification had no effect on the negatively connotated PANAS dimensions (i.e., *fatigue*, *fear*, and *upset*). Consequently, rather than attenuating negative affect, gamification seems to affect positive affect dimensions.

Gamification in the current study did not result in an effect concerning mind-wandering or flow experience. This outcome is potentially due to the system-paced speed of the n-back task. Irrespective of the n-back levels or gamification, participants had to be continuously attentive to react correctly to the stimuli and successfully finish the task. A recent study by Lu and colleagues (Lu et al., 2023) indicated that in a gamified n-back task with three load levels (i.e, 1-back, 2-back, and 3-back), subjective flow experience fitted to two physiological measures of the phasic locus coeruleus norepinephrine (LC-NE) system in the form of an inverted U-shape with (subjective) task difficulty. This indicates that subjects experienced optimal flow in challenging but not too hard task conditions. Unfortunately, the study by Lu and colleagues did not directly compare gamified and non-gamified n-back task versions but only used a gamified n-back task version. Therefore, it remains an open question whether their observed effects for flow were due to the additional 3-back load level and their analysis related to subjective task difficulty rather than to gamification per se. Nevertheless, previous studies did not always identify differences in flow experience when investigating gamification in working memory tasks (e.g., Ninaus et al., 2015). As flow is a rather broad concept with several antecedents (Kiili et al., 2012), the current manipulation might not have been strong enough to change participants' perceived flow experience.

Important to note, despite the apparent effects of gamification on emotional-motivational aspects of subjective task experience, in contrast to our main hypothesis, participants' perceived effort was not reduced by gamification. At least descriptively, the results showed that participants reported less effort for the 2-back in the game vs. non-game condition, but the effect was minimal. Compared to previous studies, the present study manipulated gamification within participants, which might have reduced the impact of the manipulation due to carry-over effects. Further, the non-gamified version of the task had two game elements (i.e., score and progress bar) that were omitted from previous non-game conditions (Bernecker & Ninaus, 2021). In this regard, one could argue that previous effects of gamification on effort might have been mainly due to the motivational function of the elements in terms of feedback on progress and performance rather than the emotional design of the game (e.g., task embedded in a story, colors). However, this needs to be tested in future research, which systematically varies these elements of gamification within one study.

4.2. Effects of gamification on task performance and task load

Despite the apparent effects of gamification on subjective task experience and positive affect, there were no effects of gamification on behavioral performance – neither in the 1-back nor in the 2-back levels. This is in line with the physiological process measures, indicating no effect of gamification concerning task load. Thus, the manipulation of gamification we used in the current study did not result in increased task load (i.e., working memory load or effort). Potentially, this is because we manipulated gamification by visual design (i.e., emotional design) and narrative embedding only. Task elements were modified between gamified and non-gamified task versions, but the number and functions of the task elements presented were not altered. Nevertheless, the visual complexity might still be rated higher in the gamified version than the non-gamified version of the n-back task because of the visual representation of the narrative (see Fig. 1).

Interestingly, as indicated by the exploratory topoplots (Fig. 4, left-hand side), the EEG theta frequency band power at right parietal electrodes was increased in the gamified n-back task versions compared to the non-gamified task versions. We can only speculate on this somewhat unexpected outcome concerning the localization on the scalp. Increases in EEG theta frequency band power are described for increased working memory load or task difficulty, albeit typically at frontal-midline electrodes (e.g., Antonenko et al., 2010; Gevins et al., 1997; Klimesch, 1999; Pesonen et al., 2007; Scharinger et al., 2015). Neuronal activity in the EEG theta range has been mainly associated with (working) memory and executive functioning (Haciahmet et al., 2021; Hanslmayr et al., 2009; Sauseng et al., 2005, 2009, 2010). Yet some studies also reported relations between the EEG theta frequency band and effort and concentration (DeLosAngeles et al., 2016; Smit et al., 2004; Yang et al., 2017). For example, DeLosAngeles and colleagues (DeLosAngeles et al., 2016) reported increased EEG theta power at parietal electrodes for states of concentrative meditation. Specifically, parietal theta power seems to be affected in visuospatial attention tasks (Yang et al., 2017), that is, to reflect aspects of visuo-spatial attention (Harris et al., 2017; Van der Lubbe et al., 2023) and bottom-up processing (Bastos et al., 2015). Thus, although somewhat speculative at this point, the observed differences in theta power at parietal electrodes between gamified and non-gamified n-back task versions in the current study might be due to the spatial n-back task design and might indicate participants' increased visuo-spatial attention or effort in the gamified as compared to the non-gamified task version. However, we need to note that participants did not judge their invested effort as being higher in the gamified condition on a subjective level. More research is necessary to substantiate these somewhat speculative interpretations.

Furthermore, although being numerically increased for the 2-back load levels compared to the 1-back load levels, reaction times did not differ significantly between these two load levels. This is a rather unexpected outcome (cf. 1.1.2). However, one plausible explanation might be the specific nature of the current spatial n-back task requiring participants to react to stimuli via four different target keys. This contrasts with the binary decision typically required in most n-back tasks (Owen et al., 2005; Redick & Lindsey, 2013). Having to react with one of four different fingers might have substantially increased the variance in the reaction time data masking the actual load level effect.

Finally, another exploratory observation is noteworthy. The eye-tracking data did not show significant differences in the total fixation durations on the feedback elements of the task when comparing gamified and non-gamified task versions. Yet, when comparing 1-back and 2-back load levels, the total fixation durations on these elements varied significantly. In the 1-back load levels, the progress bar was looked at longer as compared to the 2-back load levels, whereas for the score feedback, the opposite pattern of results could be observed. It seems that in the relatively straightforward 1-back load levels, participants did not require much information on their task performance (i.e., the score information). Instead, participants seemed to be more interested in the task's progress (and upcoming end) when the task was simple and perhaps even dull. In contrast, in the somewhat tricky, 2-back load levels, the score information gained importance, potentially aligning one's task performance.

4.3. Limitations and future research

Some limitations of the current study and suggestions for future research have to be addressed briefly. First, as detailed in the method section, sample size of the current study was determined by resource constraints. Future studies might use an a priori power analysis to determine adequate sample sizes. Nevertheless, we want to underline that the sample size of the current study was well within the typical sample size range of physiological (i.e., EEG) studies on the n-back working memory task (cf. method section).

Second, gamification might show (positive) effects on behavioral performance on the long run (i.e., when a task is repeated over days and weeks) rather than on the short run (i.e., when a task is conducted only once a day like in the current study). Positive effects of gamification on motivation might especially show up when any novelty effect of the n-back task has worn out, that is, when subjects have become highly familiar with the task and thus might have become increasingly bored in repeatedly performing the task. Therefore, over the course of several n-back task sessions, participants in an adequately gamified n-back task might stay more motivated and, hence, might show less deteriorated task performance. Future research might address these hypotheses.

5. Conclusion

In conclusion, the current study showed apparent positive effects of gamification on subjective task experience and affect. Physiological process measures did not clearly show increased task load for gamification, albeit theta frequency band power might indicate subtle differences in visuospatial attention or effort for gamification in the current task. Behavioral performance and subjective effort ratings were unaffected by gamification in the present study. In sum, the outcomes primarily favor an emotional design perspective; that is, a gamified n-back working memory task increases task experience without hampering task performance. Therefore, the extra effort of designing and implementing game elements for an n-back working memory task seems to be well justified.

Ethical approval

All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Funding

This study was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), SCHA 2137/1-1. The funding source was not involved in the study design, in the collection, analysis and interpretation of data, in the writing of the report, nor in the decision to submit the article for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

NA.

References

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education, 125*, 413–428. https://doi.org/10/gdrbwp.
Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review, 22*(4), 425–438. https://doi.org/10/bjdvnb.
Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A, 49*(1), 5–28. https://doi.org/10/ch8mb6.
Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*(1), 1–29. https://doi.org/10/cw5hjb.
Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H., & Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron, 85*(2), 390–401. https://doi.org/10.1016/j.neuron.2014.12.018
Bernecker, K., & Ninaus, M. (2021). No pain, no gain? Investigating motivational mechanisms of game elements in cognitive tasks. *Computers in Human Behavior, 114*, Article 106542. https://doi.org/10/gjqqc3.
Bledowski, C., Kaiser, J., & Rahm, B. (2010). Basic operations in working memory: Contributions from functional imaging studies. *Behavioural Brain Research, 214*(2), 172–179. https://doi.org/10/bnccs8.
Brandstätter, V., & Bernecker, K. (2022). Persistence and disengagement in personal goal pursuit. *Annual Review of Psychology, 73*(1), 271–299. https://doi.org/10/gm39k5.
Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology, 40*, 109–131. https://doi.org/10/b3mzs2.
Brom, C., Stárková, T., & D'Mello, S. K. (2018). How effective is emotional design? A meta-analysis on facial anthropomorphisms and pleasant colors during multimedia learning. *Educational Research Review, 25*, 100–119. https://doi.org/10/gjk2vx.
Cabestrero, R., Crespo, A., & Quirós, P. (2009). Pupillary dilation as an index of task demands. *Perceptual and Motor Skills, 109*(3), 664–678. https://doi.org/10/c7kcm9.
Chen, Y.-N., Mitra, S., & Schlaghecken, F. (2008). Sub-processes of working memory in the n-back task: An investigation using ERPs. *Clinical Neurophysiology, 119*(7), 1546–1559. https://doi.org/10/ds4vjj.
Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. MIT Press. https://doi.org/10.7551/mitpress/9609.001.0001
Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51–57. https://doi.org/10/bftsc5.
Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review, 119*(3), 480–499. https://doi.org/10/f34kfz.
Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10/bqr2f2.
Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage, 34*(4), 1443–1449. PubMed. https://doi.org/10/dvrjq7.
DeLosAngeles, D., Williams, G., Burston, J., Fitzgibbon, S. P., Lewis, T. W., Grummett, T. S., Clark, C. R., Pope, K. J., & Willoughby, J. O. (2016). Electroencephalographic correlates of states of concentrative meditation. *International Journal of Psychophysiology, 110*, 27–39. https://doi.org/10/f9jcqq.
Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135–168. https://doi.org/10/b2m2.
Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General, 140*(4), 552–572. https://doi.org/10/ck6s6b.
Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology, 120*(11), 1883–1908. https://doi.org/10/d3gz5h.
Van Gerven, P. W. M., Paas, F., Van Merrienboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*(2), 167–174. https://doi.org/10/cp66gp.
Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex, 10*(9), 829–839. https://doi.org/10/dkd4bf.
Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex, 7*(4), 374–385. https://doi.org/10/dbjw7j.
van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95–99. https://doi.org/10/dt6k3j.
Haciahmet, C. C., Frings, C., & Pastötter, B. (2021). Target amplification and distractor inhibition: Theta oscillatory dynamics of selective attention in a flanker task. *Cognitive, Affective & Behavioral Neuroscience, 21*(2), 355–371. PubMed. https://doi.org/10/gqmrkz.
Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does gamification work? A literature review of empirical studies on gamification*. 2014 47th Hawaii International Conference on System Sciences. https://doi.org/10/gfj8gf.
Hanslmayr, S., Spitzer, B., & Bauml, K.-H. (2009). Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cerebral Cortex, 19*(7), 1631–1640. https://doi.org/10.1093/cercor/bhn197
Harris, A. M., Dux, P. E., Jones, C. N., & Mattingley, J. B. (2017). Distinct roles of theta and alpha oscillations in the involuntary capture of goal-directed attention. *NeuroImage, 152*, 171–183. https://doi.org/10.1016/j.neuroimage.2017.03.008
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183.
Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behavior, 44*, 81–95. https://doi.org/10/f6zwnv.
Janke, S., & Glöckner-Rist, A. (2014). Deutsche Version der Positive and Negative Affect Schedule (PANAS). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/zis146.
Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol, 10*, 370–375.
Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*, 163–178. https://doi.org/10/bddntq.
Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*(4), 441–480. https://doi.org/10/fhm73x.
Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354. https://doi.org/10/dmtdrk.
Kiili, K., de Freitas, S., Arnab, S., & Lainema, T. (2012). The design principles for flow experience in educational games. *Procedia Computer Science, 15*, 78–91. https://doi.org/10/ggckds.
Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29*(2–3), 169–195. https://doi.org/10/bdhf72.
Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology, 8*(1), 33267. https://doi.org/10.1525/collabra.33267
Landers, R. N., Armstrong, M. B., & Collmus, A. B. (2017). How to use game elements to enhance learning: Applications of the theory of gamified learning. *Serious Games and Edutainment Applications*, 457–483. https://doi.org/10/gqmrk3.
Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. *Lecture Notes in Computer Science*, 63–76. https://doi.org/10/fgsrh6.
Lawrence, M.A. (2016). *Package 'ez'* (4.4–0). ⟨https://CRAN.R-project.org/package=ez⟩.
Lenth, R.V. (2021). *Package 'emmeans'* (1.7.0). ⟨https://CRAN.R-project.org/package=emmeans⟩.
Lu, H., van der Linden, D., & Bakker, A. B. (2023). Changes in pupil dilation and P300 amplitude indicate the possible involvement of the locus coeruleus-norepinephrine (LC-NE) system in psychological flow. *Scientific Reports, 13*(1), 1908. https://doi.org/10.1038/s41598-023-28781-z

Van der Lubbe, R. H. J., Panek, B., Jahangier, I., & Asanowicz, D. (2023). Lateralized connectivity in the alpha band between parietal and occipital sources when spatial attention is externally and internally directed. *Frontiers in Cognition, 2*, Article 1145854. https://doi.org/10.3389/fcogn.2023.1145854

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Mayer, R. E., & Fiorella, L. (2014). Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed..,, pp. 279–315). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.015.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. https://doi.org/10/fzdfvs.

Moreno, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 9–28). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.003.

Ninaus, M., Kiili, K., Wood, G., Moeller, K., & Kober, S. E. (2020). To add or not to add game elements? Exploring the effects of different cognitive task designs using eye tracking. *IEEE Transactions on Learning Technologies, 13*(4), 847–860. https://doi.org/10/ghjv5h.

Ninaus, M., Pereira, G., Stefitz, R., Prada, R., Paiva, A., Neuper, C., & Wood, G. (2015). Game elements improve performance in a working memory training task. *International Journal of Serious Games, 2*(1). https://doi.org/10/ggcj5m.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*(1), 46–59. https://doi.org/10/cxh645.

Palomäki, J., Kivikangas, M., Alafuzoff, A., Hakala, T., & Krause, C. M. (2012). Brain oscillatory 4–35Hz EEG responses during an n-back task with complex visual stimuli. *Neuroscience Letters, 516*(1), 141–145. https://doi.org/10/f3w4mm.

Pang, W.M., Wei Liew, T., & Leow, M.C. (2021). Emotional design for educational multimedia: A mini-review. *2021 14th International Conference on Human System Interaction (HSI)* , 1–8. ⟨https://doi.org/10.1109/HSI52170.2021.9538667⟩.

Parong, J., & Mayer, R. E. (2019). Cognitive consequences of playing brain-training games in immersive virtual reality. *Applied Cognitive Psychology, 34*(1), 29–38. https://doi.org/10/gqmrk5.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. PubMed. https://doi.org/10/gft89w.

Pesonen, M., Hämäläinen, H., & Krause, C. M. (2007). Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load. *Brain Research, 1138*, 171–177. https://doi.org/10/b3×46d.

Pfister, R., & Janczyk, M. (2016). schoRsch: An R package for analyzing and reporting factorial experiments. *The Quantitative Methods for Psychology, 12*(2), 147–151. https://doi.org/10/gfvmw6.

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology, 110*(11), 1842–1857. https://doi.org/10/ffwmrh.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258–283. https://doi.org/10/gfgxhw.

Plass, J. L., Homer, B. D., MacNamara, A., Ober, T., Rose, M. C., Pawar, S., Hovey, C. M., & Olsen, A. (2020). Emotional design for digital games for learning: The effect of expression, color, shape, and dimensionality on the affective quality of game characters. *Learning and Instruction, 70*, Article 101194. https://doi.org/10/ggjgtf.

Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. *Emotions, Technology, Design, and Learning* (pp. 131–161). Elsevier,. https://doi.org/10.1016/b978-0-12-801856-9.00007-4

R Core Team. (2020). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. ⟨https://www.R-project.org/⟩.

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review, 20*(6), 1102–1113. https://doi.org/10/f5krsq.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7*(3), 216–237. https://doi.org/10/gf4ztn.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Flow short scale. In *PsycTESTS Dataset*. American Psychological Association (APA). ⟨https://doi.org/10.1037/t47787–000⟩.

Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). Three Decades of Research on Motivational Intensity Theory. In *Advances in Motivation Science* (Vol. 3, pp. 149–186). Elsevier,. https://doi.org/10.1016/bs.adms.2016.02.001

Robinson, D., R., & Bellotti, V. (2013). A preliminary taxonomy of gamification elements for varying anticipated commitment. *Proc. ACM CHI 2013 Workshop on Designing Gamification: Creating Gameful and Playful Experiences.*

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380. https://doi.org/10/f9tr2z.

Sanchez, D.R., & Lee, C.A. (2022). Understanding the challenges of game-based training: Recommendations for moving research forward in game-based learning. In O. Bernardes, V. Amorim, & A.C. Moreira (Eds.), *Handbook of Research on the Influence and Effectiveness of Gamification in Education* (pp. 541–578). IGI Global. ⟨https://doi.org/10.4018/978–1-6684–4287-6.ch027⟩.

Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience & Biobehavioral Reviews, 34*(7), 1015–1022. https://doi.org/10/bwp34d.

Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., Glennon, M., Gerloff, C., Birbaumer, N., & Hummel, F. C. (2009). Brain oscillatory substrates of visual short-term memory capacity. *Current Biology, 19*(21), 1846–1852. https://doi.org/10/fb975q.

Sauseng, P., Klimesch, W., Schabus, M., & Doppelmayr, M. (2005). Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International Journal of Psychophysiology, 57*(2), 97–103. https://doi.org/10/cfdt6j.

Scharinger, C., Kammerer, Y., & Gerjets, P. (2015). Pupil dilation and EEG alpha frequency band power reveal load on executive functions for link-selection processes during text reading. *PLOS ONE, 10*(6), Article e0130608. https://doi.org/10/f7qzdj.

Scharinger, C., Prislan, L., Bernecker, K., & Ninaus, M. (2020). *"N-back gamified" (AsPredicted #45524)*. ⟨https://aspredicted.org/blind.php?x=QNN_GEE⟩.

Scharinger, C., Schüler, A., & Gerjets, P. (2020). Using eye-tracking and EEG to study the mental processing demands during learning of text-picture combinations. *International Journal of Psychophysiology, 158*, 201–214. https://doi.org/10/ghs95q.

Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2015). When flanker meets the n-back: What EEG and pupil dilation data reveal about the interplay between the two central-executive working memory functions inhibition and updating. *Psychophysiology, 52*(10), 1293–1304. https://doi.org/10/f7qn24.

Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2017). Comparison of the working memory load in n-back and working memory span tasks by means of eeg frequency band power and P300 amplitude. *Frontiers in Human Neuroscience, 11*. https://doi.org/10/gd3vqp.

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Construction of a benchmark for the user experience questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence, 4*(4), 40. https://doi.org/10/ggccnt.

Siegle, G. J., Steinhauer, S. R., Stenger, V. A., Konecky, R., & Carter, C. S. (2003). Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. *NeuroImage, 20*(1), 114–124. https://doi.org/10/fjbvpd.

Smit, A. S., Eling, P. A. T. M., & Coenen, A. M. L. (2004). Mental effort affects vigilance enduringly: After-effects in EEG and behavior. *International Journal of Psychophysiology, 53*(3), 239–243. https://doi.org/10/ccb5wd.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59*(4), 745–759. https://doi.org/10/dxm65b.

Sundararajan, N., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review, 32*(3), 707–734. https://doi.org/10/gjqf6t.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10/fxd3d5.

Um, E. Rachel, Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology, 104*(2), 485–498. https://doi.org/10/b9rrpj.

Vermeir, J. F., White, M. J., Johnson, D., Crombez, G., & Van Ryckeghem, D. M. L. (2020). The effects of gamification on computerized cognitive training: Systematic review and meta-analysis. *JMIR Serious Games, 8*(3). e18644–e18644. PubMed. https://doi.org/10/ghjsnm.

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25*(6), 2005–2015. PubMed. https://doi.org/10/gfkpvp.

Woolley, K., & Fishbach, A. (2015). The experience matters more than you think: People value intrinsic incentives more inside than outside an activity. *Journal of Personality and Social Psychology, 109*(6), 968–982. https://doi.org/10/f74kqn.

Woolley, K., & Fishbach, A. (2016). For the fun of it: Harnessing immediate rewards to increase persistence in long-term goals. *Journal of Consumer Research, 42*(6), 952–966. https://doi.org/10/btbc.

Yang, F.-C., Jacobson, T. K., & Burwell, R. D. (2017). Single neuron activity and theta modulation in the posterior parietal cortex in a visuospatial attention task. *Hippocampus, 27*(3), 263–273. PubMed. https://doi.org/10/gmnsmf.