Computers in Human Behavior 149 (2023) 107948

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Game elements enhance engagement and mitigate attrition in online learning tasks

Stefan E. Huber [a,1], Rodolpho Cortez [a,b,1], Kristian Kiili [c], Antero Lindstedt [c], Manuel Ninaus [a,d,*]

[a] Department of Psychology, University of Graz, Graz, Austria
[b] Federal University of Rio Grande do Norte, Natal, Brazil
[c] Faculty of Education and Culture, Tampere University, Tampere, Finland
[d] LEAD Graduate School and Research Network, University of Tübingen, Tübingen, Germany

ARTICLE INFO

Handling Editor: Jen-Her Wu

Keywords:
Educational games
Game-based learning
Games
Distance education and online learning
Media in education
Human-computer interface
Gamification

ABSTRACT

A growing body of literature suggests that adding game elements to learning tasks indirectly influences the learning process by increasing engagement with the tasks. The present study aims to advance learning engagement research by examining an often neglected subcomponent of behavioral engagement, attrition. Implementing two equivalent versions of a learning task, differing solely in the presence of game elements, allowed unequivocal attribution of any effect on the presence of game elements. Conducting the study in an online learning environment allowed further a highly unconstrained examination of the effects of game elements on attrition. We found that game elements affected both participant attrition and engagement of participants who completed the learning task. Participants with low self-efficacy were particularly prone to drop out in the non-game condition. Game elements also affected both learning efficacy and efficiency. We further found task attractivity to partially mediate the effect of game elements on learning outcomes. The results suggest that by facilitating engagement via task attractivity game elements can compensate to some extent for the increased cognitive demand that the game elements induce. We finally discuss the importance of considering the interrelations between learner characteristics, game elements, and engagement for interpreting results on learning performance measures.

1. Introduction

The interest in how learning can be affected by the inclusion of game elements in digital learning tasks has grown over the last decades (Wouters & van Oostendorp, 2017). The main arguments for the implementation of game elements or even full-fledged games as an educational tool are based on potentially fostering learning outcomes (Plass et al., 2020; Wouters & van Oostendorp, 2013) via their capability to increase engagement and motivation, and to enhance learner (or user) experience (Zainuddin et al., 2020). To this aim, game elements such as visual and auditory aesthetics, rewards, narrative or game fantasy are frequently employed in the design of learning tasks (Bedwell et al., 2012). Albeit their use in making learning situations more appealing, game elements might also hinder learning, by overloading limited cognitive capacities (Baddeley, 1992; Chandler & Sweller, 1991; Mayer, 2014), resulting in the seductive detail effect (Bender et al., 2021; Rey, 2012). Hence, a careful balance between game elements and educational content is required to keep learning tasks attractive, effective, and engaging at the same time (Plass et al., 2020).

1.1. *Cognition, affect, motivation, and engagement in learning with multimedia*

Research focusing on cognitive aspects of learning with multimedia, like cognitive load theory (CLT; Sweller, 2011) or the cognitive theory of multimedia learning (Mayer, 2014), emphasize keeping learning material plain to avoid extraneous cognitive processing. In line with this view, additional design elements, not inherently required for task accomplishment, were shown to be capable of impeding learning by acting as seductive details (Bender et al., 2021; Rey, 2012). The latter attract and hold the learners' attention, thus distracting them from the task and its efficient accomplishment.

* Corresponding author. Universitätsplatz 2, 8010, Graz, Austria.
E-mail address: manuel.ninaus@uni-graz.at (M. Ninaus).
[1] Shared first authorship.

https://doi.org/10.1016/j.chb.2023.107948
Received 31 May 2023; Received in revised form 3 August 2023; Accepted 3 September 2023
Available online 9 September 2023

However, not all additional task elements lead to the seductive detail effect. For instance, emotional design elements were shown to facilitate learning (Wong & Adesope, 2021). Further, design elements affecting motivation can also indirectly affect learning outcomes (Huang & Mayer, 2016). For instance, the inclusion of game elements in learning tasks is typically justified from an emotional design perspective by referring to their potential to increase the learners' engagement with the task (Ge & Ifenthaler, 2017; Plass et al., 2020). Disengagement from a task, on the other hand, has been reported to be less likely if game elements are present (Bernecker & Ninaus, 2021).

Hence, cognitive, affective, and motivational aspects seem highly interrelated in multimedia learning. The Integrated Cognitive Affective Model of Learning with Multimedia (ICALM; Plass & Kaplan, 2016) provides a theoretical framework for capturing this intricate interplay between design elements, cognitive, affective and motivational factors, and learning outcomes in digital learning environments. Regarding cognition, the ICALM is similar to its conceptual predecessors (CLT, see e.g., Sweller, 2011; the cognitive theory of multimedia learning, see e.g., Mayer, 2014). The two central components for cognitive processing along verbal and non-verbal channels are working memory and long-term memory. Working memory is highly limited regarding the amount of information and the duration for which information can be stored. However, processing information via working memory is necessary for transferring it to long-term memory with presumably unlimited capacity. Long-term memory can, in reverse, affect working memory by allowing efficient processing using, for instance, chunking mechanisms, providing a conceptual framework for the effects of prior knowledge.

The ICALM, however, provides an additional channel influencing memory: affective states (Plass & Kalyuga, 2019). A crucial point in the ICALM is that cognitive processes are regarded as emerging from the interplay of prior knowledge, abilities, beliefs, affect, and motivation (Cloude et al., 2022). That also implies that cognitive processes are inseparably intertwined with affective processes, experienced as moods or emotions, when learners process multimedia content (Plass & Kaplan, 2016). Consequently, the model emphasizes that also affective processes demand cognitive resources, which should be considered in instructional design. According to the ICALM, affect that involves appraisal is subjectively experienced as interest or motivation influencing the construction of mental representations as well as influencing what elements of the multimedia material are processed. Interest and motivation are closely related to engagement (Axelson & Flick, 2010; Christenson et al., 2012), which has been defined as "the active and focused investment of effort in a game environment" (Schwartz & Plass, 2020), and is of main interest in the current study which investigates the effects of game elements on learners in an online learning environment.

1.2. Engagement and motivation in game-based learning

In a systematic review specifically focusing on the effectiveness of game features in digital tasks for enhancing engagement with online programs, Looyestyn et al. (2017) reported a medium to large effect on behavioral engagement metrics (e.g., time spent with a learning resource). More recently, Sailer and Homner (2020) reported a small effect of the use of game design elements on motivational learning outcomes (e.g., engagement, self-efficacy, preferences, intrinsic motivation) but noted a high degree of heterogeneity in effect sizes. Overall, the authors of both studies noted several limitations of the generalizability of their conclusions due to the broad diversity of characteristics of the analyzed studies, like the type of the game, the domain, the experimental design, or the employed instructional technique. In addition, tracing back effects on learning to specific game elements is often difficult, because of the multitude of factors that could, in principle, affect the learning process and its outcomes (Boyle et al., 2016; Sailer & Homner, 2020). Sailer and Homner (2020), for instance, noted that participants' characteristics such as differences in the individual perception of game design elements, personality traits, or player types could not be included in their analysis, because they are not examined in the included studies of their meta-analysis in the first place, but could account for some of the variance in the obtained effect sizes.

Regarding learning and educational environments, Fredricks et al. (2004) discuss the importance of the concept of engagement in schools due to the high rates of boredom, disaffection, and dropout of students and its consequences for learning. The adoption of games as an instructional approach has repeatedly been justified by their potential to increase learners' engagement and, consequently, to enhance learning outcomes (Cheng & Su, 2012; Ge & Ifenthaler, 2017; Plass et al., 2020). Measuring engagement is complicated by the fact that it is intricately related to many concepts like motivation, emotions, affect, attention, and concentration. To resolve the resulting ambiguity and overlap of concepts, a multidimensional perspective of engagement decomposed into behavioral, cognitive, and affective engagement is widely accepted (Fredricks et al., 2004; Ge & Ifenthaler, 2017; Plass et al., 2015, 2020).

Behavioral engagement encompasses the physical actions of learners such as gestures, movements while using a mouse, pressing the keyboard, or even full-body movements in motion-sensing input devices executed during an interaction with a gamified learning task (Plass et al., 2015). Cognitive engagement encompasses the deliberate and active investment of mental efforts that allows learners to select (address processing of relevant material), organize (arrange the content into a coherent mental structure), and integrate (associate the incoming content to prior knowledge) learning material. Cognitive engagement is closely related to the concept of need for cognition (Ke et al., 2016), which refers to an individual's general inclination towards and pleasure in challenging cognitive activities (Cacioppo et al., 1996). Affective engagement is described as learners' emotional responses to game elements aiming to improve their situational experience and maintain the willingness to invest high cognitive efforts over extended periods. Accordingly, engagement is thus often thought of in terms of action or, more specifically, the behavioral, cognitive, and affective manifestation of motivation (Fredricks & McColskey, 2012; Skinner et al., 2008). In the framework of the ICALM (Plass & Kaplan, 2016), engagement can hence be considered as the behavioral, cognitive, and affective components that mediate effects of interest and motivation on verbal and visual information processing during learning in digital environments.

Studies have demonstrated that different game elements are effective in promoting different types of engagement and enhancing learning (Abdul Jabbar & Felicia, 2015). For instance, game elements such as scaffolding, and control or choices are known as game features able to enhance behavioral engagement (Abdul Jabbar & Felicia, 2015; Lavoué et al., 2021; Schwartz & Plass, 2020). Narrative, visual and auditory aesthetics, virtual environments, and avatars are often discussed as features enhancing emotional engagement (Schwartz & Plass, 2020; Lavoué et al., 2021; Ninaus et al., 2019). Incentive systems like rewards or scores are known to be capable of enhancing both behavioral and emotional engagement (Schwartz & Plass, 2020).

Any form of engagement with a learning task is an essential prerequisite for learning to happen. Engagement with a task, however, requires participants not only to maintain their attention on the task, but also to regulate aversive feelings (like boredom) associated with it, and generally resist the temptation of engaging in something more pleasurable instead (Kurzban et al., 2013; Miller et al., 2012). In short: Engagement with a task can be typically conceptualized as conflicting with opportunities associated with alternative activities. The *Opportunity Cost Model of Subjective Effort and Task Performance* (Kurzban et al., 2013) has related this motivational conflict to both performance and subjective experience of effort. Interestingly, the integration of game elements or attributes into conventional (cognitive) tasks, has been shown to reduce task disengagement (Bernecker & Ninaus, 2021). Game elements – among other things – seem to improve user experience (e.g., attractiveness of the task; Javora et al., 2019; Ninaus et al., 2020), which can increase participants' (intrinsic) motivation to stay on the task at hand (Javora et al., 2019).

1.3. Attrition as an important subcomponent of engagement

The discontinuation of a task is often referred to as attrition. More specifically, attrition is the gradual loss of participants over the course of a study or intervention and can be considered as a subcomponent of task (dis-)engagement (Lumsden et al., 2017). Therefore, measuring attrition rate (e.g., when do participants cease to interact with the task) serves as an objective measure of this component of behavioral engagement, which can be a result of lacking intrinsic motivation (e.g., Suárez et al., 2019). In the laboratory, attrition rates are typically low due to demand characteristics, politeness expectations, obedience to authority, and conformity norms (Hoerger, 2010; Skitka & Sargis, 2006). In contrast, dropout rates of the order of 50% have been reported in online intervention studies (Kelders et al., 2012; Wangberg et al., 2008) and especially early dropout proves challenging for online studies (Eysenbach, 2005). For its relations to engagement, accounting for attrition systematically seems especially important for the investigation of effects of game elements on learning. High attrition indicates that some participants are especially prone to dropout. Under that condition, it is unlikely the participants completing a study still form a representative sample of the general population. This obviously affects the generalizability of results (Hoerger, 2010), and suggests a systematic interaction between specific person characteristics and study aspects. Scrutinizing the interrelations between game elements and engagement, which is exactly the focus of the present study, hence must aim for minimizing such bias when considering attrition.

Lumsden et al. (2017) provided an attempt to study attrition over the course of a web-based cognitive task, comparing a standard task variant with two gamified versions of it. Although the study resulted in comparable dropout rates between standard and gamified conditions, generalizability of its results appears limited regarding the investigation of attrition. In particular, the design of their experiment involved four compulsory test sessions over four consecutive days, realized by coupling participant compensation with their participation in this study period, before participants would enter another six-day period during which they could drop out without further restraints. However, being free to disengage from a task is an essential feature of an online learning environment. It is exactly the freedom from typical constraints in laboratory studies which may allow participation in research inherently more voluntary in online environments (Hoerger & Currell, 2012). Hence, as arriving at a less biased view on the interrelations between game elements and engagement is an objective of this work, it seems advisable to refrain as much as possible from putting any constraints on participants' engagement with or disengagement from the task. Thus, dropout must be easily realizable for participants at all stages of the task, requiring obviously to opt for a study design utilizing an online learning environment.

1.4. Present study and hypotheses

To meet the challenges elaborated above, the present study aims not only at providing an experimental paradigm allowing an unequivocal attribution of findings to the presence of game design elements, but also at providing an extensive account on behavioral engagement. To this aim, it provides an analysis of attrition over the course of the online learning study integrated with the analyses of cognitive and affective learning outcomes and learners' personality dispositions. In doing so, it aims at shedding an empirical light on the processes by which especially motivational aspects can influence learning in digital environments as proposed by the ICALM (Plass & Kaplan, 2016).

We conducted a value-added study to investigate how game elements affect motivation via their influence on quantifiable behavioral manifestations of engagement. In other words: *How* do game elements influence learner behavior in an online learning task? In particular, we compared behavioral engagement and learning outcomes of participants randomly assigned to either a non-game-based or a game-based digital task in an online environment. We addressed several methodological requirements to rigorously examine the effects of game elements. First, we designed a novel task in which prior knowledge about the task cannot play any role, because it is known that prior knowledge can influence the processing and effects of game elements (Huang et al., 2021; Wang et al., 2016; Yeo et al., 2022). Therefore, we implemented a simple association learning task in which participants cannot know any of the associations before having taken part in the study. Second, we implemented controlled task conditions that differed only in the presence of game elements. Third, we implemented the study in an online environment in order to log user behavior and investigate attrition in a highly unconstrained environment. With this particular experimental setup, we investigated the following hypotheses:

- Game elements should affect behavioral engagement with a learning task, including attrition. Accordingly, we expected that a learning task with game elements present (game task) differs from a learning task with game elements absent (non-game task) regarding attrition over the course of the task (Hypothesis H1a). However, moment-to-moment behavioral engagement was expected to change dynamically also for those participants who completed the task (i.e., participants who did not quit). In particular, we hypothesized that our measure for moment-to-moment behavioral engagement (keypress responses, see Section 2.3.1) differs between game and non-game tasks for participants completing the task (Hypothesis H1b).
- We expected the differences in behavioral engagement to be also reflected in differences in cognitive learning outcomes, with respect to both learning efficacy (Hypothesis H2a), i.e., how much learn participants in the different task versions, and learning efficiency (Hypothesis H2b), i.e., how fast learn participants in the different task versions.
- Besides the presence of game elements, the participants' personality and current (positive or negative) affect might be further factors influencing particpants' propensity to drop out of an online study. We thus hypothesized that participants disengaging at some point from the task differ from the ones completing it in affect or some personality disposition, depending eventually also on the task condition, i.e., the presence of game elements (Hypothesis H3). Concerning learning-relevant personality dispositions, we considered, besides the general inclination towards cognitive activities (need for cognition), also achievement motives (hope for success and fear of failure), and general self-efficacy.
- Differences in how (behaviorally) engaging the game task is in comparison to the non-game task should be observable in learners' subjective experience of the motivation associated with those tasks. As such, we expected that self-report measures associated with motivation differ between learners in the game and non-game conditions (Hypothesis H4).
- We finally hypothesized that the effects on cognitive learning outcomes (efficacy and efficiency) are (partially) mediated by the considered motivational outcomes (Hypothesis H5).

2. Material and methods

2.1. Sampling and data acquisition

Participants were invited to participate in the study from mid of February 2021 to the beginning of April 2021 via an e-mail broadcast providing a link to enter the pipeline of the complete study protocol described in Section 2.2. Participants were compensated for participating in the study by the option to enter a raffle with five vouchers á 10 EUR for an online distributor awarded at the end of data acquisition. Response rates of online studies can be low, especially for studies exceeding 10 min in duration (Sammut et al., 2021). The estimated total duration of our study was 30–45 min. Some small compensation in the form of a lottery incentive has repeatedly been shown to be capable of improving response rates (Sammut et al., 2021; Laguilles, Williams, & Saunders, 2011). At the same time, compensation is also likely to influence engagement and attrition to some extent. To counteract this, we chose a low winning probability as well as low cash prizes to reduce this influence. Further, we consider it still a rather mild constraint regarding engagement and attrition compared to typical laboratory settings (Skitka & Sargis, 2006; Hoerger, 2010).

Some data had to be excluded from all data analyses beyond attrition analysis for the following reasons. Data of two participants were excluded from the post-task survey data and the associated data from one participant was excluded from the pre-task survey and the learning task data, because of erroneous assignment of the same user-identifier to different participants after completion of the learning task. Data of eight participants were excluded due to unreasonably high numbers of correct responses (>6) in the first level of the game, in which participants could only guess correct responses, see Section 2.2. Five participants did not engage with the task for one or more levels but resumed engagement afterwards. These data sets were also excluded from analyses beyond attrition analysis. Pre-task questionnaire data of 23 participants were excluded due to incorrect responses to instructed response items (such as: “Please choose ‘not at all (1)’”), which constitutes a good measure to identify careless responding (Meade & Craig, 2012). Post-task questionnaire data of nine participants were excluded for the same reason.

The local university’s ethics committee approved the online study. All participants proceeding beyond the landing page of the link given in the e-mail invitation to the study, i.e., the online informed consent form, provided informed consent.

2.2. Study design

The study protocol distinguished four distinct stages of the online study, see Fig. 1(a). Stage 1 consisted of accessing the URL provided in the e-mail invitation with its landing page representing the consent form for the study. After providing consent, participants were automatically forwarded to stage 2, denoted as pre-task survey for the remainder of this work, and consisting of a set of self-report questionnaires described in Section 2.2.2. After completing the pre-task survey, participants were again automatically forwarded to the actual learning task in stage 3 of the study. The learning task was implemented in two different versions, i.e., presence or absence of game elements (see Fig. 1(b)). The learning task and its two versions are described in Section 2.2.1. Note that the assignment of participants to either the non-game or the game branch in Fig. 1(a) was automatically, randomly, and covertly executed directly after accessing the URL in stage 1. After finishing the learning task, participants were automatically forwarded to the final, fourth stage of the study consisting of another set of self-report questionnaires, denoted as post-task survey for the remainder of this work and described in Section 2.2.2.

The pre- and the post-task surveys were implemented with Limesurvey and the two versions of the learning task were based on the NumberTrace game engine (e.g., Koskinen et al., 2023; for a short video demonstration see https://www.youtube.com/watch?v=T7s7xSlLrac) which was developed for fraction instruction using JavaScript. To take part in the study, participants required access to a computer with an internet browser, a display, a keyboard, and a mouse, which was communicated in the e-mail broadcast described in Section 2.1. Apart from this, we had no control over the conditions in which participants took part in the study, as is the case with any online study.

2.2.1. Game and non-game versions of the learning task

For the present study, we devised a novel memorization task. While memorization cannot be identified with learning, memory is of paramount importance for learning according to established cognitive (Mayer, 2014; Sweller, 2011) or cognitive-affective (Plass & Kaplan,

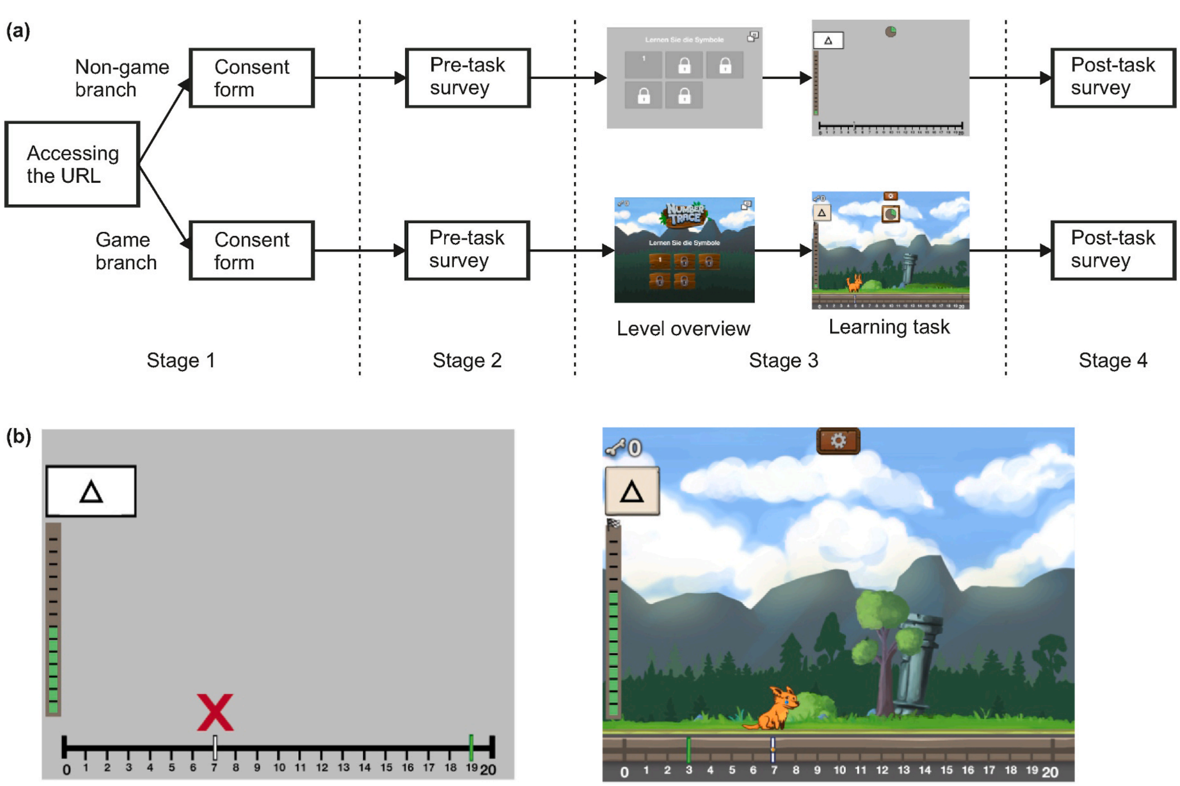

Fig. 1. (a) Study design and (b) illustrations of non-game (left) and game (right) task versions (stage 3). Note that the non-game (top row) and game (bottom row) branches in panel (a) consist of exactly the same components except for the version of the learning task in stage 3. Assignment of participants to either of the two branches was automatically, randomly, and covertly executed directly after accessing the URL in stage 1.

2016) theories as outlined in Section 1.1. In the task, the goal of the participants was to memorize associations between a set of symbols and numbers over five consecutive levels. In each level, 14 different symbols were presented to each participant. For each symbol, participants had to indicate a number, distributed spatially on a visual number line ranging from 0 to 20, see Fig. 1. Each symbol corresponded exactly to one position along the number line specified by its respective number. These associations between symbols and numbers were fixed over the entire task, but initially unknown to the participants. Participants had to learn as many as possible of these associations over the course of five consecutive levels.

Participants controlled a cursor (represented by a white vertical bar, see Fig. 1) to indicate the correct number. The left and right arrow keys of the keyboard were used to control the cursor. The spacebar was used to confirm the choice. For each symbol, participants had 20 s to select a number by moving the cursor over it and confirming their choice by pressing the spacebar. A dynamic pie chart visualized the time left to respond. Not making a choice within the given time would count as an incorrect response. After each response or after expiration of the maximum response time of 20 s, participants would receive corrective feedback illustrated in Fig. 1(b). A green vertical bar would indicate the correct position/number associated with the currently presented symbol and some visual aesthetic, which differed between game and non-game task versions (see below), would indicate if the choice had been correct. Note that the correct position/number was always shown, regardless of the response (correct or incorrect) by the participants.

Due to the lack of any prior knowledge, participants could only guess the correct position of any shown symbol at level 1. Beginning with level 2, they could use the corrective feedback from previous levels to recall correct positions. Over time, i.e., over the levels of the task, the number of correct responses should thus increase. Note that symbols were presented in random order at each level.

The differences between the non-game and game task versions were limited to the presence of game elements. The game version included a narrative, corresponding visual aesthetics, and a virtual incentive system.

The provided narrative was part of the written instruction displayed at the very end of stage 2. The instruction used in the two conditions only differed in terms of the provided narrative. In case of the game condition, the participant was supposed to help the dog “Willi” to find positions of bones in a forest. In the non-game condition, the aim was simply to find the positions of different symbols on a line. Except for differences in the use of single words the instruction for both conditions was almost exactly the same (e.g., non-game condition: “Press the arrow keys (left & right) to move (vertical white bar) on the line. Press the space bar to confirm the position on the line” vs. game condition: “Press the arrow keys (left & right) to move Willi on the forest floor. If you press the space bar, Willi digs at the position on the forest floor where you are at the moment”).

In the game version of the task, the cursor’s movement was accompanied by a walking animation of the dog. Placement of the cursor (i.e., pressing the spacebar) would initiate a digging animation for the dog. Correct positioning resulted in the dog wagging its tail, and the bone count (left upper corner in the snapshots of the game task versions in Fig. 1) would increase by one (incentive system). Given the position would be incorrect, the dog would cry, see Fig. 1(b). In the non-game version, a green check mark and a red X-symbol, see Fig. 1(b), would indicate correct and incorrect responses, respectively. The non-game version would also lack all visual aesthetics concerning the scenery illustrating the dog’s search for bones in the game version. Instead, an empty, grey background was presented in the non-game task version.

2.2.2. Pre- and post-task surveys

The pre-task survey contained questionnaires focusing on affect and personality dispositions related to motivational factors of relevance in learning contexts. In particular, the pre-task survey comprised questionnaires regarding the psychological constructs *self-efficacy, need for cognition, hope for success, fear of failure,* and *positive and negative affect.* The corresponding self-report measures were used to (i) investigate pre-task condition equivalence and to (ii) explore eventual relations between attrition, affect, and personality dispositions (Hypothesis H3). All questionnaires (within both pre- and post-task surveys) were administered in German.

Self-efficacy was measured using three items, such as “I am able to solve most problems on my own”, provided by the general self-efficacy short scale (ASKU; Beierlein et al., 2013). Self-efficacy is the degree to which individuals believe in their own abilities and competencies to achieve their goals. Items were rated on a five-point scale ranging from “doesn’t apply at all” to “applies completely”. For *self-efficacy* we obtained Cronbach’s alpha = 0.80.

Need for cognition was measured using the German, five-item short scale NFC–K-2 developed by Beiβert et al. (2019). Need for cognition refers to the tendency of a person to enjoy and engage in effortful cognitive activity (Cacioppo & Petty, 1982). Items, such as “I tend to set goals for myself that can only be achieved with considerable mental effort”, are rated on a five-point scale indicating a participant’s agreement with each item. For *need for cognition* we obtained Cronbach’s alpha = 0.66.

The achievement motives *hope for success* and *fear of failure* were assessed using the German short version of the Achievement Motive Scale (AMS) developed by Engeser (2005). Both hope for success and fear of failure are measured using five items each. Hope for success is assessed via items like “I am attracted by situations in which I can test my abilities”, whereas fear of failure is assessed by items like “Things that are a bit difficult worry me”. For *hope for success* and *fear of failure* we obtained Cronbach’s alpha of 0.78 and 0.85, respectively.

Positive and negative affect was assessed before and after the learning task using the German version of the positive and negative affect schedule (PANAS; Breyer & Bluemke, 2016). PANAS provides 20 adjectives describing feelings and emotions, e.g., “excited” or “distressed”. Participants indicate on a five-point scale the intensity with which they experience these emotions. 10 adjectives are associated with each positive and negative affect. Before the task, participants were asked for their current emotional state. After the task, participants were asked for their emotional state during the learning task. For *positive affect* we obtained Cronbach’s alpha of 0.88 and 0.91 for pre-task and post-task questionnaires, respectively. For *negative affect* we obtained Cronbach’s alpha of 0.90 and 0.82 for pre-task and post-task questionnaires, respectively.

The post-task survey comprised questionnaires focusing on task evaluation by the participants including self-reports on how motivating and how mentally demanding the task was perceived by the participants. In particular, the post-task survey comprised questionnaires regarding the psychological constructs *situational motivation*, *user experience*, and *task load*. Besides, *positive and negative affect* was also assessed as mentioned above.

Situational motivation was assessed using the German version of the Situational Motivation Scale (SIMS; Guay et al., 2000). SIMS comprises four subscales and assesses both situational *intrinsic* and *extrinsic* motivation. Within SIMS, participants are asked to respond to the question “Why are you currently engaged in this activity?” by indicating their agreement to 16 statements representing potential answers to this question on a seven-point scale ranging from “does not correspond at all” to “corresponds exactly”. Typical items for the four subscales *intrinsic motivation*, *identified regulation*, *external motivation*, and *amotivation* read “Because this activity is fun”, “Because I am doing this for my own good”, “Because I am supposed to do it”, and “I don’t know; I don’t see what this activity brings me”, respectively. The Cronbach’s alpha values obtained for *intrinsic motivation*, *identified regulation*, *external regulation*, and *amotivation*, read 0.87, 0.68, 0.80, and 0.75. Whereas intrinsic motivation is described by the subscale with the same name, the other three subscales correspond to extrinsic motivation.

We further used the subscales *attractivity* and *stimulation* of the User Experience Questionnaire (UEQ; Laugwitz et al., 2008). The subscale attractivity aims to assess how enjoyable, good, pleasing, pleasant, attractive, and friendly a product (or task) is perceived. The subscale stimulation aims to assess how valuable, exciting, interesting, and motivating a product (or task) is perceived. Regarding their face validity, the two subscales seem closely related to qualities associated with (intrinsic) motivation. According to self-determination theory (Ryan & Deci, 2020), enjoyment or pleasantness of a task or activity are characteristics of intrinsic motivation. For each of the given adjectives, participants are given a seven-point, bipolar rating scale with the endpoints given by the respective adjective and its opposite (e.g., attractive and unattractive). Participants then indicate their experience of the task for each pair of adjectives on that scale. Cronbach's alpha for the two subscales attractiveness and stimulation yielded 0.92 and 0.89, respectively.

Self-evaluations of the *mental load* during the learning task, the achieved *performance* and the required *effort* for the latter were assessed using the three respective items of the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). The three single items were each rated on an 11-point scale ranging from "very low (0)" to "very high (10)", from "very good (10)" to "very poor (0)", and from "very low (0)" to "very high (10)" for mental load, performance, and effort, respectively.

2.3. Outcome measures

According to the ICALM (Plass & Kaplan, 2016), affective processes affect cognitive processes and vice versa. This implies that learning outcomes go beyond cognitive outcomes, which aligns with the integrative model of emotional foundations of game-based learning (Loderer et al., 2020). In the present work, we also assess motivational outcomes besides cognitive learning outcomes. Besides cognitive learning outcomes, we also assess outcome measures in the form of objective, behavioral measures of moment-to-moment behavioral engagement and attrition, as well as personality dispositions.

2.3.1. Moment-to-moment behavioral engagement and attrition

As a measure of moment-to-moment *behavioral engagement* and attrition, we utilized the keypress responses required from participants to indicate their choice of position/number for each of the presented symbols during the learning task. Participants responding to an item by confirming their response by pressing the spacebar were counted as engaging with the task. Participants not responding via keypress were assigned to one of the two following categories. Participants not responding in the case of individual items or individual groups of items but resuming respondence for subsequent items or in subsequent levels, were categorized as *temporarily* disengaging from the task. That is, temporary disengagement (by not responding via pressing the spacebar) was used to monitor moment-to-moment behavioral engagement.

In contrast, attrition was assessed as follows. Participants being unresponsive for an entire task level and all subsequent task levels were categorized as *completely* disengaging from the task (i.e., attrition) at the first of the levels for which they did not respond at all. For example, a participant responding to (some) items at the first task level, but not responding to any items at the second or any subsequent levels, would be categorized as completely disengaging from the task at level 2. Accordingly, the number of participants having completely disengaged would increase by 1 at level 2 due to this participant (i.e., attrition would increase by 1). Timing data, recorded automatically upon completing an entire level of the task, was used to discern participants entering the first level, yet not completing it, from participants completing the first level, yet never engaging with the task.

2.3.2. Personality dispositions

In terms of personality dispositions, we assessed self-efficacy, need for cognition, hope for success, and fear of failure of participants using the questionnaires described in Section 2.2.2.

2.3.3. Motivational and affective outcomes

Motivational outcomes were assessed using the SIMS. Due to their theoretically apparent close relation to intrinsic motivation, we also consider the assessed subscales attractivity and stimulation of the UEQ described in Section 2.2.2 in the framework of our considered motivational outcomes.

In the ICALM (Plass & Kaplan, 2016), motivational outcomes are related to design elements via (unattributed and attributed) affect. Although not the focus of the present work, the assessment of positive and negative affect using PANAS allowed us also (to some extent) to assess affective outcomes. In particular, we explored the change of positive and negative affect from pre-to post-task survey depending on task condition.

2.3.4. Cognitive outcomes

As a measure of learning *efficacy*, we used the number of correct responses given by a participant at the last level in which they were engaging with the task. For participants completing the task, this number coincides with the number of correct responses at level 5. For a participant who, for instance, completely disengaged at level 3, this number would give the number of correct responses at level 2. Correct responses at level 1 were by design random results and not indicative in any sense of learning efficacy. They were thus not used for assessing the latter. That is, participants had to engage for at least two levels with the task for a possible assessment of learning efficacy.

As a measure of learning *efficiency*, i.e., the steepness of the increase of correctly reproduced associations for increasing levels, we modeled individual learning using an exponential learning curve (Leibowitz, Baum, Enden, & Karniel, 2013):

$$N_{\mathrm{corr},i}(L) = N_{\max}\{1 - \exp[-C_i(L-1)]\}. \tag{1}$$

In Eq. (1), $N_{\mathrm{corr},i}(L)$ denotes the number of correctly reproduced associations of the i-th participant at level L, $N_{\max} = 14$ denotes the maximum number of associations that can be correctly reproduced, and the coefficient C_i (determined by a non-linear least squares fit to the empirical data for each participant) denotes the rate constant indicating the learning efficiency of the i-th participant. The higher the rate constant, the higher learning efficiency. For instance, a rate constant of 1.0 corresponds to about 9 correctly reproduced associations at level 2, and about 12 correctly reproduced associations already at level 3. A smaller rate constant of, e.g., 0.35 corresponds to about 4 correctly reproduced associations at level 2, and about 7 correctly reproduced associations at level 3. Hence, the rate constant captures how fast learning proceeds towards the maximally possible number of 14 correctly reproduced associations in the given task. Eq. (1) reflects that learning cannot proceed beyond that number. Due to the lack of any utilizable prior knowledge, no associations can be known at level 1, which is also reflected by Eq. (1), i.e., $N_{\mathrm{corr},i}(1) = 0$. Note that more complex processes, such as forgetting or errors due to inattention or carelessness, cannot be described by the simple one-parameter model given by Eq. (1). However, we prompted for such a simple model for describing learning efficiency because for five game levels only four data points are available for each individual fit. To minimize the risk of spurious results due to overfitting, we also restricted the analysis of learning efficiency to participants completing the entire task.

Besides these behavioral measures of cognitive learning outcomes, we further assessed subjective self-report measures of mental load, achieved performance, and required effort with the respective items of the NASA-TLX described in Section 2.2.2.

2.4. Statistical analyses and software

Count data (i.e., counts of keypress responses or counts of completely disengaging participants at a specific level) were analyzed regarding statistical significance using χ^2-tests. Due to the abundance of heavily skewed data and outliers in the case of metric outcome variables, we used robust statistical methods provided by the WRS2 package (Mair and Wilcox, 2020) for comparing game and non-game conditions. In particular, we computed robust correlation coefficients in the form of percentage bend correlations ρ_{pb}, we employed the bootstrap version of Yuen's test based on trimmed means for comparing two independent groups, and we employed robust analyses of variance based on trimmed means for two-way comparisons and mixed designs. Trimming was kept at the default of 0.2 as for this value statistical power is nearly the same as is achieved by the mean when sampled from a normal distribution, while standard errors are substantially smaller for a 20% trimmed mean in the case of outliers (Mair & Wilcox, 2020). Significance levels were set at 0.05. Effect sizes are reported referring to the explanatory measure of effect size $\hat{\xi}$, whereas $\hat{\xi} = 0.1$, 0.3, and 0.5 correspond to small, medium, and large effect sizes, respectively. For comparison of dependent groups, we report instead the effect size δ_t suggested by Algina et al. (2005) as a robust version of Cohen's *d*, since the WRS2 package does not provide 95% confidence intervals for $\hat{\xi}$ in this case. Finally, our computed mediator models are based on robust regression using *M*-estimators as implemented in the MASS package (Venables & Ripley, 2002). Indirect effects of simple one-mediator models were tested using Zu and Yuan's (2010) robust approach (2010) implemented also in the WRS2 package (Mair & Wilcox, 2020). For reporting results, we followed the guidelines provided by Field and Wilcox (2017). If available, we also report 95% confidence intervals for any computed quantity, directly following the respective quantity in squared brackets.

All statistical analyses were conducted using R (R Core Team, 2022) and RStudio (RStudio Team, 2022) using the packages tidyverse (Wickham et al., 2019), ggpubr (Kassambara, 2023a), rstatix (Kassambara, 2023b), pastecs (Grosjean & Ibanez, 2018), car (Fox & Weisberg, 2019), psych (Revelle, 2022), WRS2 (Mair & Wilcox, 2020) and MASS (Venables & Ripley, 2002).

2.5. Data analysis overview

Due to the variety of outcome measures considered and the variety of different analyses employed in the present work, we provide an overview of our hypothesis and associated tests in Table 1. The table lists each hypothesis with the associated outcome measures, the assessment tools used for operationalization of measures, and the statistical tests used for testing the hypothesis for statistical significance. Given the overall number of outcome measures and hence comparisons, the choice of an overall significance level of 0.05 may seem rather liberal. However, given the exploratory nature especially of hypothesis H3 (associated with most of the conducted tests), we also do not want to miss out on eventually meaningful associations. That means, not only false positives but also false negatives are costly for future research (McDonald, 2014). However, obviously, caution must be exercised with the interpretation of results under liberal testing conditions. Therefore, we will take any eventual findings regarding those hypotheses as indications for eventual relations calling for dedicated replications or stringent testing experiments by future research.

3. Results

3.1. Attrition and moment-to-moment behavioral engagement (hypotheses H1a and H1b)

Fig. 2 gives an overview of attrition over the course of the entire online study. From 1688 persons who accessed the study URL provided via e-mail, 685 participants, equally distributed among the game and non-game branches (χ^2 (1) = 0.07, $p = 0.789$), proceeded to the actual study via signing the online consent form. During the pre-task survey, 59 and 70 participants dropped out in the non-game and game branch,

Table 1
Hypotheses, outcome measures, operationalization procedures, statistical analyses, and qualitative description of the corresponding results obtained in this work.

Hypothesis	Outcome measure	Operationalization	Statistical analysis	Result
H1a: Attrition differs between game and non-game condition.	Attrition	Counts of participants completely disengaging from the task via assessment of keypress responses	χ^2-test at each task level for 2x2-contingency table (responsive/ unresponsive participants in game/non-game condition)	Confirmed: Game elements decrease the likeliness to completely disengage from an online learning task (i.e., less attrition with game elements).
H1b: Moment-to-moment behavioral engagement differs between game and non-game condition.	Moment-to-moment behavioral engagement	Counts of missing keypress responses for participants not yet completely disengaged from the task	χ^2-test at each task level for 2x2-contingency table (keypress responses/missing keypresses in game/non-game condition)	Confirmed: Moment-to-moment behavioral engagement was higher when game elements were present.
H2a: Learning efficacy differs between game and non-game condition.	Learning efficacy	Number of correct responses at last task level in which participant was still engaging with the task	Yuen's test based on trimmed means for comparing independent groups	Not confirmed: Learning efficacy was not significantly different between conditions.
H2b: Learning efficiency differs between game and non-game condition.	Learning efficiency	Coefficient of non-linear least squares fit of Eq. (1) to individual series of numbers of correct responses from level 2–5 (behavioral measure)	Yuen's test based on trimmed means for comparing independent groups	Confirmed: Slightly higher learning efficiency was shown when game elements were absent.
H3: Affect or personality dispositions differ between participants completing/ disengaging from the task depending on task condition.	Positive and negative affect, self-efficacy, need for cognition, hope for success, fear of failure	Self-report questionnaires (PANAS; ASKU; NFC–K-2; AMS)	Robust, two-way analysis of variance (independent variables: task condition, completing/ disengaging from the task)	Partly confirmed: Participants with low self-efficacy were more likely to disengage/to drop out of the study when game elements were absent.
H4: Motivational outcomes differ between game and non-game conditions.	Intrinsic motivation, identified regulation, external regulation, amotivation, attractivity, stimulation	Self-report questionnaires (SIMS; UEQ)	Yuen's test based on trimmed means for comparing independent groups	Partly confirmed: Task attractivity was rated higher when game elements were present. External regulation was perceived higher when game elements were absent.
H5: Effects of condition (game or non-game) on cognitive outcomes are mediated by motivational outcomes.	Learning efficacy and efficiency, motivational outcomes (see H4)	See H2a, H2b, and H4	Robust mediation models (independent variable: task condition, mediator(s): motivational outcomes, dependent variable: learning efficacy or efficiency)	Confirmed: Task attractivity partially mediated the effect of game elements on learning efficacy and efficiency.

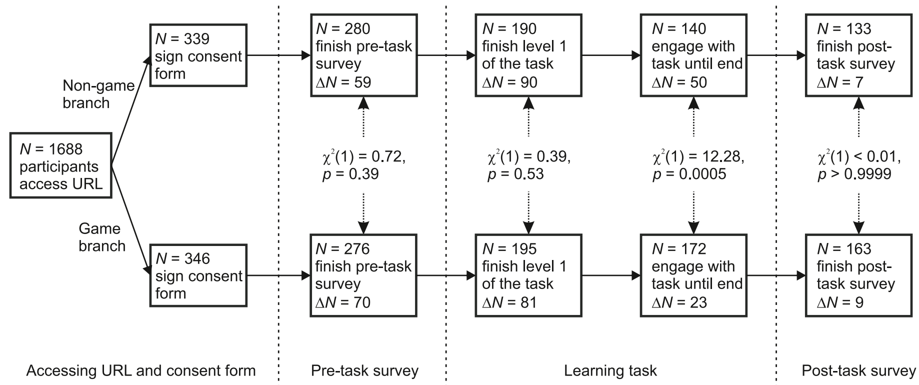

Fig. 2. Attrition over the course of the study. Note that the non-game and game branches differ only in the implementation of the learning task as illustrated in Fig. 1(a).

respectively, χ^2 (1) = 0.72, p = 0.396. Another 90 and 81 participants dropped out in the non-game and game branches, respectively, without interacting with the task (i.e., no key presses) before completing the first level of the game, χ^2 (1) = 0.39, p = 0.534. From 190 participants, who at least finished the first level of the learning task in the non-game branch, 50 dropped out over the course of the learning task. From 195 participants, who at least finished the first level of the learning task in the game branch, 23 dropped out over the course of the learning task, yielding a disproportionate dropout rate for the two branches at this stage of the study, χ^2 (1) = 12.28, p = 0.0005. Dropout rates during the post-task survey were again highly similar, χ^2 (1) < 0.01, p > 0.999.

Fig. 3 depicts the dynamics of task (dis-)engagement over the course of the learning task. Task engagement (i.e., percent of items being responded to via keypress) was substantially different between the non-game and game conditions at the first level of the learning task, χ^2 (1) = 173.31, p < 0.0001, see Fig. 3(a). Further, both the changes in task engagement from level 1 to level 2 and from level 2 to level 3 differed between the conditions, χ^2 (1) = 32.93, p < 0.0001, and χ^2 (1) = 28.47, p < 0.0001, respectively. From level 3 onwards, changes in task engagement were comparable between the conditions (p > 0.05). A substantial portion of overall task disengagement could be attributed to participant attrition, see Fig. 3(b), yielding highly differing dropout proportions between the two conditions at the first level, χ^2 (1) = 11.47, p = 0.0007, and comparable dropout rates from level 2 onwards (p > 0.05). Fig. 3(c) depicts the level-wise task engagement, renormalized with respect to the number of participants yet engaging with the task at the respective level (i.e., considering dropouts). The two conditions differed in renormalized task engagement at level 2, χ^2 (1) = 13.26, p = 0.0003.

3.2. Cognitive outcomes (hypotheses H2a and H2b)

The game and non-game conditions did not differ significantly in self-reported mental load (M_{diff} = 0.30 [−0.30, 0.90], Y_t = 0.98, p = 0.332, $\hat{\xi}$ = 0.09 [0.00, 0.27]), achieved performance (M_{diff} = −0.03 [−0.78, 0.71], Y_t = −0.09, p = 0.929, $\hat{\xi}$ = 0.02 [0.00, 0.21]), and required effort (M_{diff} = 0.25 [−0.45, 0.95], Y_t = 0.71, p = 0.474, $\hat{\xi}$ = 0.08 [0.00, 0.26]). Note that for these analyses only data of those participants (n = 285) were included, who completed both the learning task and the post-task survey completely (i.e., no missing data and not belonging to the excluded cases listed in Section 2.1).

For our behavioral measures of cognitive outcomes (i.e., learning efficacy and learning efficiency), we could use the data of all participants at least completing the learning task (excluding the data sets listed in Section 2.1), which account to n = 298. Our used measure of learning efficacy, i.e., the trimmed mean of numbers of correct responses (also shortly denoted as scores here), was also similar for the non-game and game conditions at the end of the task (score of the last level), M_{diff} = 0.28 [−1.10, 0.55], Y_t = −0.64, p = 0.501, $\hat{\xi}$ = 0.05 [0.00, 0.21]. However, we found that scores develop differently over the course of the task for the two conditions, see Fig. 4(a). In particular, we found a steeper increase of scores over consecutive task levels for the non-game than for the game condition, reflected by the interaction between task level and condition, $Q_{level,\ condition}$(3, 163.18) = 3.10, p = 0.028. Besides, we noted a significant main effect of task level, Q_{level}(3, 163.18) = 492.96, p < 0.0001, yet no significant main effect of condition, $Q_{condition}$(1191.65) = 2.58, p = 0.111.

To evaluate if this central tendency of dependence of scores on task level is also reflected in a difference between conditions regarding in-

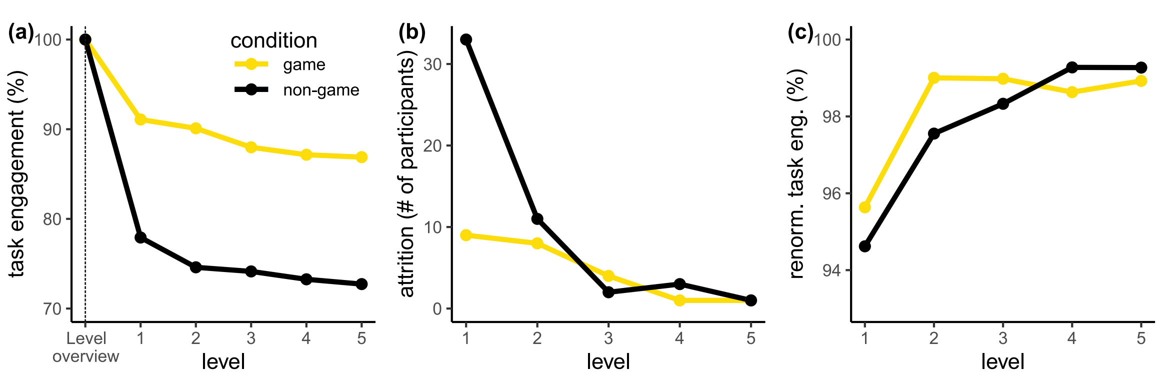

Fig. 3. Development of task engagement over the course of the learning task for game (yellow) and non-game conditions (black). Panel (a) depicts overall task engagement. Panel (b) shows dropout rates after each task level. Panel (c) depicts task engagement accounting for only those participants still engaged with the task at the respective level.

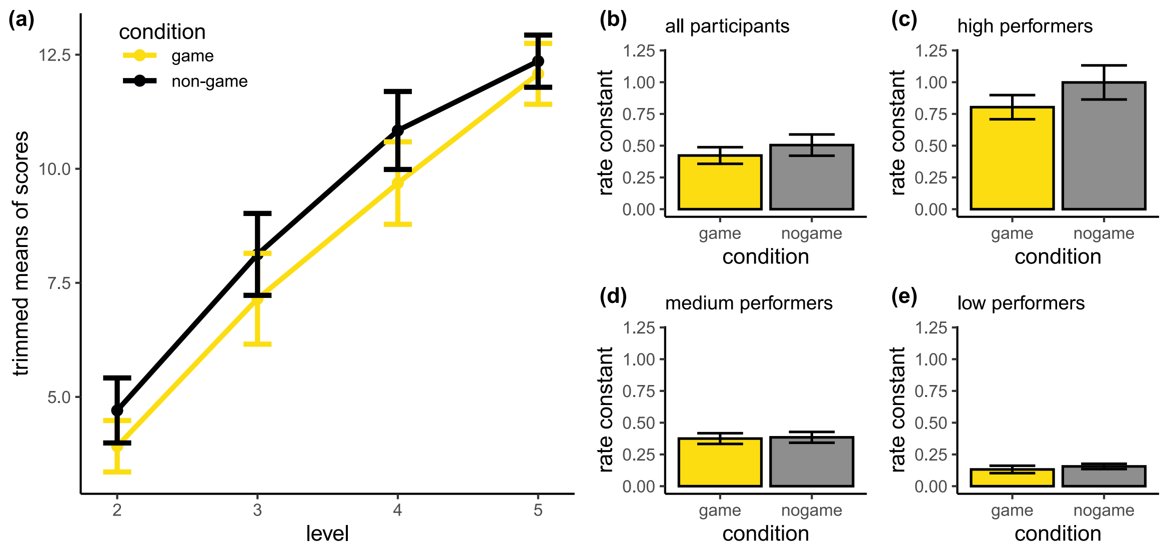

Fig. 4. (a) Trimmed means of scores, i.e., numbers of correct responses, over the course of the task for game (yellow) and non-game (black) conditions. Note that level 1 is omitted as a correct response at level 1 is not representative of learning but simply a chance result. (b)–(e) Differences between the two conditions concerning learning efficiency (described by its rate constant) for (b) all participants, (c) participants with a score of 14 at level 5, (d) participants with a score of 10–13 at level 5, and (e) participants with a score lower than 10 at level 5. Error bars always represent bootstrapped 95%-confidence intervals.

dividual learning efficiency (i.e., the steepness of the increase of correct response for increasing levels), we compared the two conditions with respect to trimmed means of individual rate constants obtained for the exponential learning curve given by Eq. (1). We found comparable rate constants for non-game and game conditions, $M_{diff} = 0.08$ [−0.03, 0.19], $Y_t = 1.49$, $p = 0.154$, $\hat{\xi} = 0.14$ [0.00, 0.28], see Fig. 4(b). However, decomposing the entire sample of participants into highly performing participants ($n_{game} = 57$; $n_{non\text{-}game} = 58$) regarding their learning outcome, i.e., yielding 14 out of 14 correct responses at level 5, participants of medium performance ($n_{game} = 64$; $n_{non\text{-}game} = 46$), i.e., yielding 10–13 correct responses at level 5, and participants of low performance ($n_{game} = 42$; $n_{non\text{-}game} = 31$), i.e., yielding 9 or less correct responses at level 5, indicated a potential dependence of learning efficiency on overall task performance. In particular, learning efficiencies were similar between conditions for participants of medium and low performance, $M_{diff} = 0.01$ [−0.05, 0.07], $Y_t = 0.33$, $p = 0.760$, $\hat{\xi} = 0.07$ [0.00, 0.30], and $M_{diff} = 0.02$ [−0.01, 0.06], $Y_t = 1.39$, $p = 0.182$, $\hat{\xi} = 0.23$ [0.00, 0.55]. Learning efficiencies for participants of high performance differed more substantially between conditions, i.e., efficiency was higher for high performers in the non-game condition than in the game condition, $M_{diff} = 0.19$ [<0.01, 0.38], $Y_t = 2.03$, $p = 0.047$, $\hat{\xi} = 0.27$ [0.04, 0.52].

However, the results presented hitherto for both learning efficacy and efficiency are limited with respect to the main objective of this work as they are based on participants completing the learning task. They hence do not account for attrition. Attrition, however, obviously matters for a comparison of task conditions with respect to learning efficacy. This is illustrated in Fig. 5. Taking only into account participants completing the task ($n = 298$), the proportion of correct responses at level 5, was slightly higher for the non-game (81.3%) than for the game (78.5%) condition, χ^2 (1) = 4.65, $p = 0.03$. Accounting, however, for the 73 dropouts by considering their scores from the last level at which they were yet engaging with the task (except for level 1), yielded a larger proportion of correct responses for the game (74.7%) than for the non-game (59.5%) condition, χ^2 (1) = 69.83, $p < 0.0001$.

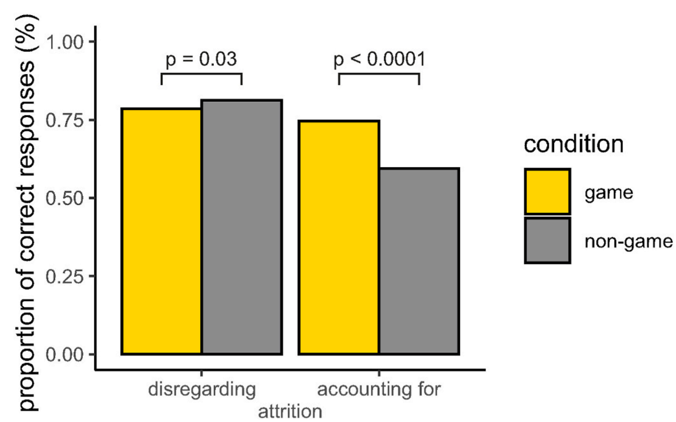

Fig. 5. Comparison between the two conditions concerning the proportion of correct responses at the last level at which participants were yet engaging with the task, either disregarding attrition (left) or accounting for attrition (right).

3.3. *Affect and personality dispositions in dependence of task engagement (hypothesis H3)*

Participants in both non-game and game branches yielded comparable trimmed means for all scales (i.e., self-efficacy, need for cognition, achievement motives, positive and negative affect) included in the pre-task survey ($p > 0.05$) irrespective of including (i) all eligible participants having completed the pre-task survey ($n = 533$), (ii) only participants engaging with the learning task for at least the first level ($n = 376$), or (iii) only participants completing the learning task ($n = 298$). However, upon decomposing subsample (ii) into participants staying engaged throughout the task and participants disengaging completely from the task at some point during the task, we identified a significant interaction between task engagement and condition affecting self-efficacy, $Q_{engagement,\ condition} = 5.02$, $p = 0.031$, illustrated in Fig. 6. While participants disengaging at some point completely from the task were more likely to score high on self-efficacy in the game condition, the opposite is the case in the non-game condition, i.e., participants disengaging from the task are more likely to score low on self-efficacy. Post-hoc computation of robust correlation coefficients between self-efficacy

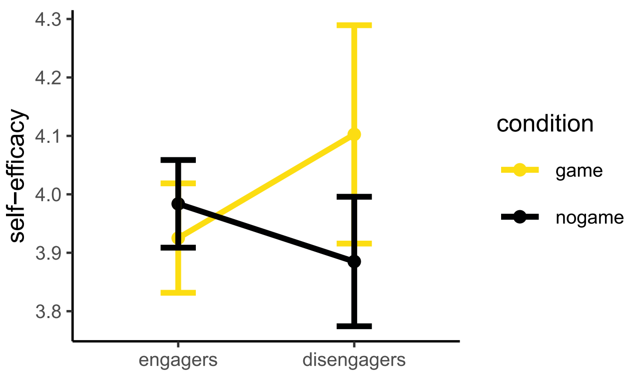

Fig. 6. Trimmed means of self-efficacy for game (yellow) and non-game (black) conditions, and participants staying engaged throughout the task (denoted as engagers) and participants disengaging completely from the task at some point during the task (denoted as disengagers). Both engagers and disengagers completed at least the first level of the learning task. Error bars represent bootstrapped 95%-confidence intervals.

and learning rate constants (i.e., efficiency) yielded larger coefficients in the non-game condition, $\rho_{pb} = 0.24$ [0.06, 0.42], $p = 0.005$, than in the game condition, $\rho_{pb} = 0.06$ [−0.09, 0.20], $p = 0.473$. A similar result was obtained for learning efficacy (i.e., scores at level 5), $\rho_{pb} = 0.20$ [0.01, 0.36], $p = 0.023$, and $\rho_{pb} = 0.06$ [−0.09, 0.19], $p = 0.472$ for non-game and game conditions, respectively.

3.4. Affective and motivational outcomes (hypothesis H4)

Positive affect in pre- and post-task surveys were compared depending on task condition for all participants completing both the learning task and the post-task survey ($n = 298$). We found a significant main effect of time, $Q_{time}(1, 153.89) = 5.39$, $p = 0.022$, yet no significant effect of condition, $Q_{condition}(1, 153.80) = 2.29$, $p = 0.132$, and also no significant interaction, $Q_{condition,\ time}(1, 153.89) = 0.06$, $p = 0.811$. In both conditions, trimmed means of positive affect were larger in the post-task than in the pre-task survey, $M_{diff} = -0.10$ [−0.25, 0.04], $Y_t(78) = -1.42$, $p = 0.161$, $\delta_t = -0.20$ [−0.40, −0.03] for the non-game condition, and $M_{diff} = -0.13$ [−0.27, 0.01], $Y_t(92) = -1.89$, $p = 0.062$, $\delta_t = -0.21$ [−0.42, −0.03] for the game condition. Concerning negative affect, we found also a significant main effect of time, $Q_{time}(1, 154.91) = 24.12$, $p < 0.0001$, yet again no significant effect of condition, $Q_{condition}(1, 148.99) = 1.74$, $p = 0.190$, and also no significant interaction, $Q_{condition,\ time}(1, 154.91) = 0.60$, $p = 0.440$. In both conditions, trimmed means of negative affect were smaller in the post-task than in the pre-task survey, $M_{diff} = 0.13$ [0.05, 0.21], $Y_t(78) = 3.12$, $p = 0.003$, $\delta_t = 0.26$ [0.12, 0.48] for the non-game condition, and $M_{diff} = 0.18$ [0.09, 0.26], $Y_t(92) = 3.96$, $p = 0.0002$, $\delta_t = 0.44$ [0.28, 0.60] for the game condition.

Concerning the scales included exclusively in the post-task survey, we obtained significant differences between conditions in user experience for task attractivity ratings and the situational motivation subscale external regulation. Task attractivity was rated significantly higher in the game than in the non-game condition, $M_{diff} = 0.46$ [0.16, 0.76], $Y_t = 2.95$, $p = 0.005$, $\hat{\xi} = 0.27$ [0.07, 0.42]. External regulation was significantly larger in the non-game than in the game condition, $M_{diff} = 0.39$ [0.01, 0.77], $Y_t = 1.96$, $p = 0.047$, $\hat{\xi} = 0.16$ [0.00, 0.34]. The trimmed means for the four subscales of situational motivation and the two subscales attractivity and stimulation of the UEQ, as well as comparisons of them for both conditions, are provided in Table 2.

3.5. Mediation analyses (hypothesis H5)

For participants completing both the learning task and the post-task survey ($n = 298$), we further investigated if the effects of task condition on learning efficacy and efficiency were partially mediated by our considered motivational outcome variables.

Before the mediation analyses, we explored associations between our considered motivational variables, see Table 3. The computed 95% confidence intervals were above or below zero for all pairs of variables, indicating that all motivational outcome variables are significantly associated with each other. Amotivation and external regulation were fairly positively associated with each other, $\rho_{pb} = 0.33$, but negatively with all other variables. Identified regulation, intrinsic motivation, attractivity, and stimulation were all positively associated with each other. Attractivity and stimulation were both strongly associated with intrinsic motivation, $\rho_{pb} = 0.69$ and $\rho_{pb} = 0.66$, respectively. Attractivity and stimulation were also strongly associated with each other, $\rho_{pb} = 0.82$.

Table 4 provides the results for a mediation model considering attractivity as a mediator for the effect of game elements on learning efficacy (top) and efficiency (bottom). For both cognitive outcome variables we obtained a significant partial mediation via attractivity. The slightly, yet non-significant, negative association between task condition and performance measures (total effect) became more negative upon taking into account the mediation by attractivity in both cases (direct effect). This indicates that game elements had a direct, negative effect on cognitive outcomes in our learning task, i.e., cognitive outcomes were impeded by the game elements. However, at the same time, game elements enhanced the attractivity of the task (pathway from condition to attractivity in Table 4) in agreement with our result of higher attractivity of the game task than the non-game task obtained in Section 3.3 (Table 2). Attractivity was in turn positively associated with cognitive outcomes (pathways from attractivity to efficacy/efficiency in Table 4). In consequence, cognitive outcomes were increased via the indirect pathway from game elements over attractivity to cognitive outcomes (indirect pathways in Table 4). Overall, part of the decrease in cognitive outcomes due to the direct effect of game elements was balanced by the increase in cognitive outcomes via the indirect pathway. In both cases of considered cognitive outcomes, this led to an overall insignificant total effect of game elements on cognitive outcomes (total pathways in Table 4).

Table 2
Comparison of situational motivation subscales amotivation (AM), external regulation (ER), identified regulation (IR), intrinsic motivation (IM), task attractivity, and task stimulation for both task conditions.

Subscale	Non-game	Game	Difference	$\hat{\xi}$	p
AM	2.84 [2.59, 3.09]	2.67 [2.44, 2.89]	0.17 [-0.13, 0.47]	0.10 [0, 0.27]	0.276
ER	2.57 [2.27, 2.88]	2.19 [1.95, 2.43]	0.39 [0.01, 0.77]	0.16 [0, 0.34]	0.047
IR	3.71 [3.54, 3.88]	3.72 [3.49, 3.94]	−0.01 [-0.29, 0.27]	0.02 [0, 0.21]	0.957
IM	4.10 [3.89, 4.31]	4.35 [4.13, 4.57]	−0.25 [-0.59, 0.09]	0.12 [0, 0.29]	0.154
Attractivity	4.68 [4.45, 4.91]	5.14 [4.90, 5.38]	−0.46 [-0.76, −0.16]	0.27 [0.07, 0.42]	0.005
Stimulation	4.54 [4.30, 4.79]	4.72 [4.49, 4.95]	−0.18 [-0.49, 0.14]	0.09 [0, 0.26]	0.218

Table 3
Pair-wise, robust correlation coefficients ρ_{pb} (below diagonal) and their 95%-confidence intervals (above diagonal) for amotivation (AM), external regulation (ER), identified regulation (IR), intrinsic motivation (IM), attractivity, and stimulation.

	1. AM	2. ER	3. IR	4. IM	5. Attractivity	6. Stimulation
1. AM	1	[0.22, 0.43]	[-0.37, −0.13]	[-0.49, −0.29]	[-0.58, −0.39]	[-0.59, −0.40]
2. ER	0.33	1	[-0.27, −0.03]	[-0.39, −0.16]	[-0.36, −0.13]	[-0.30, −0.06]
3. IR	−0.25	−0.16	1	[0.55, 0.72]	[0.34, 0.54]	[0.37, 0.57]
4. IM	−0.39	−0.28	0.64	1	[0.62, 0.75]	[0.58, 0.73]
5. Attr.	−0.49	−0.25	0.45	0.69	1	[0.78, 0.86]
6. Stim.	−0.50	−0.18	0.48	0.66	0.82	1

Table 4
Standardized coefficients (β), their standard errors (SE), *t*-values, 95%-confidence intervals, and *p*-values, for the five pathways associated with the two considered, simple mediation models with attractivity as mediator for the effect of condition on learning efficacy and efficiency.

Type	Effect	β	SE	t	95%-CI	p
Component	Condition → Attractivity	0.36	0.12	2.97	[0.12, 0.61]	0.003
	Attractivity → Efficacy	0.31	0.05	6.29	[0.22, 0.41]	<0.001
Indirect	Cond. → Attr. → Efficacy	0.11			[0.03, 0.19]	0.007
Direct	Condition → Efficacy	−0.17	0.10	−1.65	[-0.36, 0.03]	0.100
Total	Condition → Efficacy	−0.05	0.11	−0.50	[-0.27, 0.16]	0.617
Component	Condition → Attractivity	0.36	0.12	2.97	[0.12, 0.61]	0.003
	Attractivity → Efficiency	0.05	0.01	4.50	[0.03, 0.06]	<0.001
Indirect	Cond. → Attr. → Efficiency	0.02			[<0.01, 0.03]	0.003
Direct	Condition → Efficiency	−0.05	0.02	−2.65	[-0.09, −0.01]	0.009
Total	Condition → Efficiency	−0.04	0.02	−1.83	[-0.08, <0.01]	0.069

Table 5
Standardized coefficients (β), their standard errors (SE), *t*-values, 95%-confidence intervals, and *p*-values, for the five pathways associated with the two considered, simple mediation models with stimulation as mediator for the effect of condition on learning efficacy and efficiency.

Type	Effect	β	SE	t	95%-CI	p
Component	Condition → Stimulation	0.13	0.13	1.03	[-0.12, 0.38]	0.305
	Stimulation → Efficacy	0.26	0.05	5.13	[0.16, 0.36]	<0.001
Indirect	Cond. → Stim. → Efficacy	0.03			[-0.03, 0.09]	0.405
Direct	Condition → Efficacy	−0.10	0.10	−0.96	[-0.30, 0.10]	0.340
Total	Condition → Efficacy	−0.05	0.11	−0.50	[-0.27, 0.16]	0.617
Component	Condition → Stimulation	0.13	0.13	1.03	[-0.12, 0.38]	0.305
	Stimulation → Efficiency	0.03	0.01	2.85	[0.01, 0.05]	0.005
Indirect	Cond. → Stim. → Efficiency	<0.01			[>−0.01, 0.01]	0.414
Direct	Condition → Efficiency	−0.04	0.02	−1.99	[-0.08, > −0.01]	0.047
Total	Condition → Efficiency	−0.04	0.02	−1.83	[-0.08, <0.01]	0.069

For stimulation, we did not obtain a significant difference between the non-game and game conditions, see Table 2 in Section 3.3, despite its strong association with attractivity (Table 3). This was also reflected in the corresponding mediation model with stimulation as a mediator between task condition and cognitive outcomes reported in Table 5. Task condition, i.e., game elements present compared to game elements absent, was not associated with a significant change in stimulation (pathway from condition to stimulation in Table 5). Stimulation per se was positively associated with cognitive outcomes, i.e., the higher stimulation, the higher the outcomes (pathways from stimulation to efficacy/efficiency in Table 5). Overall, the indirect pathway from task condition over stimulation to cognitive outcomes did not affect cognitive outcomes significantly (indirect pathways in Table 5). Hence, the effects of game elements on cognitive outcomes remained relatively

Table 6
Standardized coefficients (β), their standard errors (SE), *t*-values, 95%-confidence intervals, and *p*-values, for the five pathways associated with the two considered, simple mediation models with intrinsic motivation (IM) as mediator for the effect of condition on learning efficacy and efficiency.

Type	Effect	β	SE	t	95%-CI	p
Component	Condition → IM	0.20	0.12	1.57	[-0.05, 0.44]	0.119
	IM → Efficacy	0.25	0.05	5.02	[0.15, 0.35]	<0.001
Indirect	Cond. → IM → Efficacy	0.06			[-0.01, 0.13]	0.079
Direct	Condition → Efficacy	−0.11	0.10	−1.14	[-0.31, 0.08]	0.256
Total	Condition → Efficacy	−0.05	0.11	−0.50	[-0.27, 0.16]	0.617
Component	Condition → IM	0.20	0.12	1.57	[-0.05, 0.44]	0.119
	IM → Efficiency	0.04	0.01	3.44	[0.02, 0.06]	0.025
Indirect	Cond. → IM → Efficiency	0.01			[<0.01, 0.02]	0.049
Direct	Condition → Efficiency	−0.05	0.02	−2.25	[-0.09, −0.01]	<0.001
Total	Condition → Efficiency	−0.04	0.02	−1.83	[-0.08, <0.01]	0.069

Table 7
Standardized coefficients (β), their standard errors (SE), t-values, 95%-confidence intervals, and p-values, for the five pathways associated with the two considered, simple mediation models with identified regulation (IR) as mediator for the effect of condition on learning efficacy and efficiency.

Type	Effect	β	SE	t	95%-CI	p
Component	Condition → IR	0.03	0.12	0.27	[-0.21, 0.27]	0.787
	IR → Efficacy	0.17	0.05	3.38	[0.07, 0.27]	<0.001
Indirect	Cond. → IR → Efficacy	<0.01			[-0.03, 0.06]	0.659
Direct	Condition → Efficacy	−0.06	0.10	−0.60	[-0.26, 0.14]	0.552
Total	Condition → Efficacy	−0.05	0.11	−0.50	[-0.27, 0.16]	0.617
Component	Condition → IR	0.03	0.12	0.27	[-0.21, 0.27]	0.787
	IR → Efficiency	0.02	0.01	2.31	[<0.01, 0.04]	0.022
Indirect	Cond. → IR → Efficiency	<0.01			[>−0.01, 0.01]	0.625
Direct	Condition → Efficiency	−0.04	0.02	−1.90	[-0.08, <0.01]	0.059
Total	Condition → Efficiency	−0.04	0.02	−1.83	[-0.08, <0.01]	0.069

Table 8
Standardized coefficients (β), their standard errors (SE), t-values, 95%-confidence intervals, and p-values, for the five pathways associated with the two considered, simple mediation models with external regulation (ER) as mediator for the effect of condition on learning efficacy and efficiency.

Type	Effect	β	SE	t	95%-CI	p
Component	Condition → ER	−0.20	0.12	−1.67	[-0.43, 0.04]	0.096
	ER → Efficacy	−0.15	0.05	−2.86	[-0.25, −0.05]	0.005
Indirect	Cond. → ER → Efficacy	0.03			[-0.009, 0.08]	0.159
Direct	Condition → Efficacy	−0.08	0.10	−0.70	[-0.28, 0.13]	0.480
Total	Condition → Efficacy	−0.05	0.11	−0.50	[-0.27, 0.16]	0.617
Component	Condition → ER	−0.20	0.12	−1.67	[-0.43, 0.04]	0.096
	ER → Efficiency	−0.02	0.01	−2.14	[-0.04, −0.002]	0.042
Indirect	Cond. → ER → Efficiency	0.004			[-0.003, 0.01]	0.248
Direct	Condition → Efficiency	−0.04	0.02	−2.05	[-0.08, −0.002]	0.033
Total	Condition → Efficiency	−0.03	0.02	−1.83	[-0.08, <0.01]	0.069

unaffected by taking into account the indirect pathways (compare direct and total pathways in Table 5 for both efficacy and efficiency).

The results obtained for the mediation models for the four subscales of situational motivation as mediators between task condition and cognitive outcomes can be easily understood on the basis of the two mediation models described so far. The models are reported in Tables 6–9 for intrinsic motivation, identified regulation, external regulation, and amotivation, respectively.

The mediation model for intrinsic motivation resembles the features of the one for attractivity. Both considered cognitive outcomes were positively associated with intrinsic motivation (pathways from intrinsic motivation to efficacy/efficiency in Table 6). Although yielding an insignificant effect of game elements on intrinsic motivation (pathway from task condition to intrinsic motivation in Table 6), the corresponding confidence interval points rather to a relatively small effect than complete absence of an effect (compared to the results for attractivity provided in Table 5). Taken together, game elements were associated with significantly higher learning efficiency via the indirect pathway over intrinsic motivation. For learning efficacy, the obtained confidence interval supports at least a tendency into the same direction. For both efficacy and efficiency, the effect of game elements on cognitive outcomes was somewhat reduced when taking into account the indirect pathway. This again means that the indirect pathway over intrinsic motivation counterbalanced a portion of the impeding effect of game elements on cognitive outcomes.

The mediation model for identified regulation resembles the features of the one for stimulation rather than the one for attractivity. While identified regulation was positively associated with increased learning efficacy and efficiency (pathways from identified regulation to efficacy/efficiency in Table 7), it appeared largely unaffected by game elements (pathway from task condition to identified regulation in Table 7). In consequence, the effects of game elements on cognitive outcomes were largely unaffected by the indirect pathway over identified regulation.

The mediation models for external regulation and amotivation resemble rather the features of the mediation model for attractivity but with an opposite sign in line with the negative association of external regulation and amotivation with all other motivational outcomes (Table 3). Hence, unsurprisingly, also the associations between external

Table 9
Standardized coefficients (β), their standard errors (SE), t-values, 95%-confidence intervals, and p-values, for the five pathways associated with the two considered, simple mediation models with amotivation (AM) as mediator for the effect of condition on learning efficacy and efficiency.

Type	Effect	β	SE	t	95%-CI	p
Component	Condition → AM	−0.15	0.12	−1.26	[-0.38, 0.08]	0.210
	AM → Efficacy	−0.19	0.05	−3.63	[-0.29, −0.08]	<0.001
Indirect	Cond. → AM → Efficacy	0.03			[-0.01, 0.08]	0.179
Direct	Condition → Efficacy	−0.08	0.10	−0.80	[-0.28, 0.11]	0.426
Total	Condition → Efficacy	−0.05	0.11	−0.50	[-0.27, 0.16]	0.617
Component	Condition → AM	−0.15	0.12	−1.26	[-0.38, 0.08]	0.210
	AM → Efficiency	−0.03	0.01	−2.60	[-0.05, −0.007]	0.010
Indirect	Cond. → AM → Efficiency	0.004			[-0.002, 0.01]	0.230
Direct	Condition → Efficiency	−0.04	0.02	−2.00	[-0.08, −0.001]	0.046
Total	Condition → Efficiency	−0.03	0.02	−1.83	[-0.08, <0.01]	0.069

Table 10
Standardized coefficients (β), their standard errors (SE), t- and z-values, 95%-confidence intervals, and p-values, for all components and indirect, direct, and total effects of the serial two-mediator model including attractivity and intrinsic motivation as mediators on the effect of task condition on learning efficacy. Note that z-values instead of t-values are only computed for the indirect effects.

Type	Effect	β	SE	t/z	95%-CI	p
Component	Condition → Attractivity	0.36	0.12	2.97	[0.12, 0.61]	0.003
	Condition → IM	−0.04	0.09	−0.51	[-0.22, 0.13]	0.613
	Attractivity → IM	0.68	0.04	15.45	[0.60, 0.77]	<0.001
	Attractivity → Efficacy	0.25	0.07	3.63	[0.11, 0.38]	<0.001
	IM → Efficacy	0.10	0.07	1.53	[-0.03, 0.24]	0.128
Indirect	Condition → Attractivity → Efficacy	0.09	0.04[a]	2.25[a]	[0.01, 0.17][a]	0.025[a]
	Condition → IM → Efficacy	−0.005	0.01[a]	−0.32[a]	[-0.03, 0.02][a]	0.750[a]
	Condition → Attractivity → IM → Efficacy	0.03	0.02[a]	1.35[a]	[-0.01, 0.06][a]	0.180[a]
Direct	Condition → Efficacy	−0.16	0.10	−1.62	[-0.36, 0.04]	0.107
Total	Condition → Efficacy	−0.05	0.11	−0.50	[-0.27, 0.16]	0.617

[a] computed based on the second order standard error and standard normal distribution, see Hayes (2022, p. 98 & 185).

regulation and cognitive outcomes were negative (pathways from external regulation to efficacy/efficiency in Table 8) and likewise for amotivation (pathways from amotivation to efficacy/efficiency in Table 9). Effects of game elements on external regulation (pathway from condition to external regulation in Table 8) and amotivation (pathway from condition to amotivation in Table 9) pointed towards relatively small negative effects, i.e., game elements are more likely to reduce external regulation and amotivation. Hence, the indirect pathways from game elements over external regulation or amotivation to cognitive outcomes again somewhat reduce the direct impeding effects of game elements on cognitive outcomes, although confidence intervals point towards small effects in both cases and considerably smaller ones in the case of amotivation.

We finally investigated two serial, two-mediator models with both attractivity and intrinsic motivation (in this order) as mediators between task conditions and either learning efficacy or learning efficiency. The results for these two models are provided in Tables 10 and 11, respectively. We found that, on top of the partial mediation of the effect of condition on learning performance by attractivity, intrinsic motivation could not explain a further, significant portion of variance. In other words: The information supplied by the participants on the intrinsic motivation of the task did not contain (much) more valuable information on the relation between task condition and learning performance than was already contained in the participants' ratings of task attractivity.

3.6. Demographics of participants completing the online study

From 1688 participants, who activated the provided link, and thus arrived at the informed consent form, 285 participants (178 female, 103 male, 4 diverse) with an age ranging from 18 to 74 years ($M = 29.20$, $SD = 12.02$, $Mdn = 25.00$, $MAD = 5.93$ all in units of years) completed both the learning task and the post-task questionnaire of the online study completely (i.e., no missing data). 177 of those participants were students, 108 were not. 155 of them were in the game condition, 130 were in the non-game condition.

These subsamples of participants in the game and non-game condition were equivalent regarding gender composition, $p = 0.741$ (using Fisher's exact test due to the small number of diverse gender). In particular, 95, 57, and 3 participants identified as female, male, and diverse in the game condition, respectively, and 83, 46, and 1 of those participants identified as female, male, and diverse in the non-game condition, respectively. They were also equivalent regarding age, $M_{diff} = 0.53$ years [−1.23, 2.30], $Y_t = 0.56$, $p = 0.557$. The subsamples were also equivalent regarding the frequency of students, χ^2 (1) = 1.83, $p = 0.176$. In particular, 98 and 57 were students and non-students in the game condition, respectively, and 71 and 59 were students and non-students in the non-game condition, respectively.

4. Discussion

4.1. Game elements increase behavioral engagement

Our results indicate that game elements positively affect behavioral task engagement in both proposed respects (Hypotheses H1a and H1b). In particular, game elements decrease the likeliness to completely disengage at some point from a digital learning task (Hypothesis H1a; Fig. 2), better known as attrition. This is in line with other studies investigating the effects of game elements on user attrition, for instance, in Massive Open Online Courses (MOOCs; e.g., Medina-Labrador et al., 2022) but also in mobile mental health interventions (e.g., Litvin et al., 2020). Importantly, in the current study, only a limited set of game elements was utilized, i.e., a narrative with corresponding visual aesthetics and a virtual incentive system, which already improved

Table 11
Standardized coefficients (β), their standard errors (SE), t- and z-values, 95%-confidence intervals, and p-values, for all components and indirect, direct, and total effects of the serial two-mediator model including attractivity and intrinsic motivation as mediators on the effect of task condition on learning efficiency. Note that z-values instead of t-values are only computed for the indirect effects.

Type	Effect	β	SE	t/z	95%-CI	p
Component	Condition → Attractivity	0.36	0.12	2.97	[0.12, 0.61]	0.003
	Condition → IM	−0.04	0.09	−0.51	[-0.22, 0.13]	0.613
	Attractivity → IM	0.68	0.04	15.45	[0.60, 0.77]	<0.001
	Attractivity → Efficiency	0.04	0.01	2.86	[0.01, 0.06]	0.005
	IM → Efficiency	0.01	0.01	0.77	[-0.02, 0.04]	0.443
Indirect	Condition → Attractivity → Efficiency	0.01	0.007[a]	2.00[a]	[0.0003, 0.03][a]	0.045[a]
	Condition → IM → Efficiency	−0.0004	0.002[a]	−0.25[a]	[-0.004, 0.003][a]	0.801[a]
	Condition → Attractivity → IM → Efficiency	0.003	0.003[a]	0.74[a]	[-0.004, 0.009][a]	0.457[a]
Direct	Condition → Efficiency	−0.05	0.02	−2.63	[-0.09, −0.01]	0.009
Total	Condition → Efficiency	−0.04	0.02	−1.83	[-0.08, 0.002]	0.069

[a] computed based on the second order standard error and standard normal distribution, see Hayes (2022, p. 98 & 185).

behavioral engagement of learners.

Furthermore, our results indicate that complete disengagement occurs particularly early upon interaction with the task [Fig. 3(b)], in line with an earlier discussion of attrition in online studies (Eysenbach, 2005). Hence, it is rather unsurprising that Lumsden et al. (2017) did not find any effects of game elements on attrition, as they considered only participants in their attrition analysis who already finished four mandatory sessions of cognitive training and only looked at attrition beyond this first mandatory phase. In contrast, our study – even though considerably shorter than the study by Lumsden et al. (2017) – enabled us to accurately track user engagement across multiple phases, facilitating a clear interpretation of the observed patterns. That is, once learners interacted with the learning task either in the game or non-game version of the task, behavioral engagement differed.

In addition, we also find that participants staying engaged with and completing the task, differ in moment-to-moment behavioral engagement over the course of the task for game and non-game versions [Hypothesis H1b; Fig. 3(c)]. From an emotional design and game-based learning perspective (Greipl, et al., 2021; Ninaus et al., 2019; Ninaus, Kiili, Wood, Moeller, & Kober, 2020; Plass et al., 2015), this result may hardly come as a surprise, as game elements are particularly known to foster all forms of task engagement, including its behavioral, cognitive, and affective components. Thus, our current results, in agreement with the ICALM model (Plass & Kaplan, 2016), support the motivational capabilities of game elements, or the behavioral results thereof, as indicated by higher learner engagement.

Self-reports on both positive and negative affect were, however, comparable between game and non-game conditions before, and more importantly, also after the learning task. While positive affect slightly increased, negative affect decreased considerably over the course of the task in both conditions. According to the ICALM model (Plass & Kaplan, 2016) motivational effects of game elements are mediated by affective processes. However, our two aggregated, point measures of affect (i.e., pre- and post-task) may be too coarse-grained to uncover more subtle affective dynamics over the course of the task and its effect on motivational state. Future studies on digital game-based learning may utilize more fine-grained, temporally denser measures to uncover affective dynamics (e.g., Ninaus et al., 2019; Greipl, Bernecker, & Ninaus, 2021; Cloude et al., 2022) and their relation to motivational variables. Scrutinizing the exact mechanisms by which objective task features influence affective dynamics and in consequence thereof, motivation and interest in the task at hand, goes, however, certainly beyond the present work, but may represent an expedient direction of future research.

Another reason for comparable affective states in both conditions, at least at an aggregated level, is that participants who would give rise to notable differences did actually not complete the task. This complication for interpretation due to attrition dynamics will be discussed in more depth below. Concerning its eventual effect on affective dynamics, it suffices to say that for an empirical resolution of this issue also temporally denser process measures seem necessary.

Finally, utilizing an online learning environment allowed acquiring objective metrics of behavioral engagement in a rather naturalistic learning setting; at least compared to laboratory studies, in which attrition can hardly be investigated (Hoerger, 2010). Hence, we conclude that the present results make a strong case for including game elements in digital learning tasks to keep users engaged on a behavioral level. In addition, they emphasize the importance of including behavioral indicators to study the effects of game elements.

The possibility to accurately monitor learners' engagement is important to better understand, and in the best case, support and facilitate the learning process in general. Attrition, or complete disengagement from a learning task is the last step in a dynamic process of varying behavioral engagement. Possibilities to monitor the dynamic evolution of behavioral engagement and understanding the dynamic changes culminating in attrition could aid to further develop just-in-time support measures for the learning process within adaptive learning environments (Ninaus & Sailer, 2022). Adaptive learning environments enable personalized learning (Bernacki et al., 2021), and personalized learning within adaptive learning environments can effectively enhance learning outcomes (Aleven et al., 2016; Bernacki et al., 2021). Within adaptive learning environments, macro- and micro-level adaptivity needs to be distinguished (Plass & Pawar, 2020). The term macro-level describes general categories of the overall learning context. Adaptivity, in this context, could, for example, refer to the suggestion of a specific course sequence or feedback given after completing a particular unit. In contrast, adaptivity at the micro-level modifies a learning task based on an individual's current needs based on real-time measurements of the learning process (Plass & Pawar, 2020). Such adaptivity at the micro-level can encompass feedback approaches (Hattie & Timperley, 2007) or adaptive scaffolding (Radkowitsch et al., 2021). In any case, micro-level adaptivity requires reliable behavioral measures of the learning process, including behavioral engagement and disengagement measures, to provide students with optimal support at the optimal time in a specific situation in personalized learning.

4.2. Indirect effects of game elements on learning outcomes

Participants in the non-game condition showed slightly higher efficacy and efficiency of learning (Fig. 4). Thus, game elements might have challenged the limited capacity of cognitive resources (Chandler & Sweller, 1991; Mayer, 2014). This result, however, requires further clarification in at least three respects.

First, from the perspective of a cognitive theory of multimedia learning (Mayer, 2014) or cognitive load theory (Chandler & Sweller, 1991) a better learning performance of participants in the non-game condition could be expected. It is intriguing, however, that we find such an effect only for the highest performing participants [Fig. 4(b-e)].

Second, even for the highest performing participants, the (statistically significant) difference in learning efficiency between the non-game (trimmed mean rate constant of 1.0) and game condition (trimmed mean rate constant of 0.8) pragmatically boils down to a slightly more relaxed learning process in the game than in the non-game condition. In concrete numbers, these two rate constants, account for about 9 and 8 correctly reproduced associations at the second game level in the non-game and in the game condition, respectively. Learners in the non-game condition would simply be ahead by one remembered association from level 2 onwards.

Third, and this point can hardly be emphasized enough, the entire comparison of efficacy and efficiency relies, for reasons of data analysis requirements (i.e., complete data), on participants staying engaged with the task through its end. It hence neglects the clear difference between the two conditions in terms of attrition. Accounting for attrition, however, provides an entirely different picture. Participants, who completely disengage from the task, can have learned at most what they learned until the point at which they disengaged. Taking this fact into account results in a significantly higher proportion of correct responses in the game as compared to the non-game condition (i.e., attrition-corrected estimate of learning efficacy; see Fig. 5) which may be better suited to illustrate the potential of game-based learning.

4.3. Game elements may benefit rather learners with lower self-efficacy

Besides differences in learning performance, our results indicate that game elements may interact with the self-efficacy beliefs of participants (Hypothesis H3). Participants scoring low on self-efficacy were more likely to disengage from the task and drop out of the study in the non-game condition, whereas the opposite appeared to be the case in the game condition. Self-efficacy has been shown to be associated with task performance (Bouffard, 1990; Hunsu et al., 2023; Stajkovic & Luthans, 1998). Interestingly, we found a significant, positive association between self-efficacy and task performance (encompassing both efficacy and efficiency) only in the case of the non-game condition, whereas the association in the game condition is still positive, yet notably lower and not statistically significant.

It seems unreasonable to assume that the association between self-efficacy and performance ceases to be in effect in the presence of game elements. The results may rather indicate that, in the presence of game elements, other effects are coming into play, mitigating the influence of self-efficacy. That is, in the non-game condition, it is rather what the participants bring themselves to the task, that determines the outcome, whereas in the game condition, it does not depend so much on the abilities participants have acquired – or belief to have acquired – so far. A recent outcome that situational self-efficacy can be improved in learning tasks by replacing abstract scaffolds with scaffolds based on (non-player) game characters seems to point in the same direction (Koskinen et al., 2023).

From an instructional design perspective, this represents an intriguing implication, because it could mean that, although game-based learning might be less beneficial, or even hindering, for learners already doing very well on a task, it might eventually serve rather those more in need of support (learners with low self-efficacy). This aligns neatly also with the result from our performance analysis: Participants with the highest learning performance might indeed suffer from seductive detail effects (Rey, 2012), i.e., they could be distracted rather than motivated by the game elements. For the other participants, however, the increased cognitive demand posed by game elements appears effectively balanced by the enhanced engagement they induce.

Hence, we conclude that the eventual relation between game elements and self-efficacy regarding learners' inclination to drop out should be scrutinized further in future research. We cannot rule out the possibility of a false positive result, especially given the relatively weak evidence for making a decision about an interaction between attrition and task condition in this case ($p = 0.031$, see Section 3.3 and Fig. 6) and the numerous comparisons made regarding personality dispositions (Hypothesis 3). Yet we argue that this result calls for dedicated replication attempts. As outlined above, the cost of discarding an eventual relation between game elements, task engagement, and general self-efficacy could be severe from an educational viewpoint. In contrast, the cost of scrutinizing an eventual relation boils down to administering general self-efficacy questionnaires [like Engeser's (2005) three-item scale] in game-based learning research. In line with McDonald's argumentation (2014, p. 260), we would argue that in this case, the cost of an eventual false negative outweighs the cost of an eventual false positive.

4.4. Game elements affect learning performance via task attractivity

Lastly, we could shed some light on the mechanisms by which game elements enhance engagement with a learning task. We found that participants not only rated task attractivity considerably higher in the game than in the non-game condition (hypothesis H4), but that task attractivity also partially mediated the effect of game elements on task performance (hypothesis H5), encompassing both efficacy and efficiency. In fact, the mediation analyses revealed that task attractivity provides a counter-measure against an otherwise inverse association between game elements and task performance such that the overall net effect nearly cancels. This aligns well with the discussion provided so far and may help to resolve some aspects of the still somewhat ongoing debate about the utility of game-based learning (Zainuddin et al., 2020).

Controlling for task attractivity indeed reveals a detrimental association between game elements and learning outcomes in line with earlier findings such as lower scores in exams (de-Marcos et al., 2014) or distraction of learners from actual learning material (Kocadere & Çağlar, 2015). The positive association between task attractivity and learning outcomes, however, almost entirely remedies that effect, and provides an explanatory link between objective design properties, enhanced engagement, and learning outcomes in agreement with a solid body of empirical work consolidating a causal pathway from game elements over emotional and cognitive engagement (Greipl, et al., 2021; Ninaus et al., 2019; Ninaus et al., 2020; Plass et al., 2015) to (learning) performance (Bernecker & Ninaus, 2021; Ninaus et al., 2015 ; Plass et al., 2015).

Unsurprisingly, task attractivity was closely related to aspects of situational motivation and especially to intrinsic motivation. Mediation models with either task attractivity or intrinsic motivation basically reflect the same characteristics. While game elements are associated with a direct, impeding effect on cognitive outcomes, they enhance attractivity and intrinsic motivation at the same time. These motivational outcomes are, however, associated with an increase in cognitive outcomes. Hence, overall, cognitive outcomes are comparable to the task condition without game elements. The same holds for external regulation and amotivation, but with an opposite sign. In this case, game elements reduce external regulation and amotivation. Reduced external regulation and amotivation in turn increase cognitive outcomes and thus counteract the higher cognitive demand by game elements. Also, similar to attractivity and intrinsic motivation, identified regulation and the other considered user experience dimension, stimulation, appear to positively influence cognitive outcomes. In contrast to attractivity and intrinsic motivation, however, identified regulation and stimulation appear to be relatively unaffected by game elements.

Engagement finally appears as the behavioral link integrating the learners' motivation and interest in the task with visual information processing to arrive at the overall learning outcome. Overall, our results thus provide further empirical support for the mechanisms suggested by the ICALM model (Plass & Kaplan, 2016). They particularly highlight a pathway by which affective and cognitive dynamics may be linked by game elements in human-computer interaction. While placing a higher cognitive burden on (visual) information processing, they provide enhanced task attractivity at the same time. Induced positive affect may then provide a link to subjectively perceived enhanced motivation and interest in the digital task manifesting in behaviorally quantifiable increased levels of engagement. Visual aesthetic design is known to influence learners' emotions (Loderer et al., 2020) and the same likely holds for a more generalized notion of attractivity.

Task attractivity, as assessed by the user experience questionnaire (Laugwitz et al., 2008), refers to a general property of how appealing a task (or product) is perceived by a user. Importantly, the general appeal or pleasantness of a task is a core characteristic of intrinsic motivation (Ryan & Deci, 2020). In the present case, the general appeal of the task also encompasses how narrative elements introduced in the task description are translated or become active during gameplay. We found that attrition was strongest after having engaged for a single level with the task [Section 3.1, Fig. 3(b)]. The especially elevated attrition in the case of complete absence of game narrative in the non-game condition could thus also be an indication that the early experience of how the game narrative translates into gameplay is of importance for keeping learners engaged. While earlier studies provide some empirical support for the effectiveness of narration regarding cognitive (Jackson et al., 2018; Lester et al., 2014), affective and motivational (Dickey, 2020) outcomes in game-based learning, results from meta-analyses remain inconclusive (Barz et al., 2023; Wouters et al., 2013). Neither Wouters et al. (2013) nor Barz et al. (2023) could find evidence for a difference between games with or without a narrative regarding learning, although Barz et al. (2023) report at least a trend.

4.5. Limitations and future directions

Scrutinizing the specific effect of narration on learning outcomes further by disentangling the combined effect of narrative, aesthetic, and incentivizing elements in the present work represents another promising avenue for future research. On the one hand, the integration of several game features represents a prerequisite for providing a full-fledged game-based learning task (Deterding et al., 2011). On the other hand, it represents an obvious limitation for the isolated study of particular game elements. However, future studies can effectively extend our present understanding by implementing scarcer versions of the same task by subtracting individual elements.

As already outlined above, future studies will be required to further illuminate or discard the eventual relation between self-efficacy, game elements, and attrition. In the ICALM (Plass & Kaplan, 2016), the dependence of learning outcomes on personality dispositions is generally considered by the dependence of cognitive and affective processes on emotional, self-regulative capabilities. Particularly for its educational implications regarding the development of individually-tailored learning environments, a finer resolution of this dimension definitely represents an important avenue for further research.

Our results are limited regarding their generalizability due to the limited demographic information we could provide for our initial sample. Due to assessing demographic information at the end of the online study, we cannot exclude that behavioral engagement and especially attrition, but also cognitive, affective, and motivational outcomes are influenced to some extent by demographic variables. Particularly concerning attrition, dedicated future studies will be needed to resolve the sociodemographic dimension of game-based learning, as the systematic dropout of participants sharing common characteristics like, e.g., gender, occupation, or socioeconomic status, directly affects generalizability and can lead to biased results (Jankovsky & Schroeders, 2022).

While our implementation of the online study was certainly less restrained than typical laboratory settings (Hoerger, 2010), it was also not entirely free of constraints (i.e., compensation by the option to enter a draw for low cash prizes with low winning probability). Even without any incentive for study participation, self-selection of participants in online studies would provide a source of bias regarding sampling (Khazaal et al., 2014). A solution may again only come in the form of a systematic investigation of the effect of contextual restraints on game-based learning research by repeated experiments.

Finally, complementing self-report questionnaires and performance data by psychophysiological measures could further disentangle the intricate relations between game elements, engagement, cognitive, affective, and motivational outcomes, especially by making unattributed affect methodologically accessible. Although this would require to temporarily fall back on laboratory research, it would allow a focused investigation of the dynamics of psychophysiological correlates of affect over the course of the task. As representing different starting conditions, participants with, e.g., varying self-efficacy, would give rise to an associated varying task experience and in consequence of that, development of differing affective patterns over time. This could not only provide an empirical test for some of the presumptions raised above, but also allow to aid the development of tailored learning systems providing optimal support based on current, individual needs.

5. Conclusion

The current study goes beyond previous research on behavioral engagement in game-based learning by utilizing objective metrics of engagement over multiple phases of a digital learning study in an online environment. With our current approach, we demonstrate that a largely unbiased investigation into how behavioral engagement is affected by some intervention requires experimental design choices that allow for rather natural learning behavior to occur in the first place. We further show that game elements increase behavioral engagement (i.e., reduce learner attrition) in online learning environments and seem to benefit especially learners with lower self-efficacy. While negative effects of game elements on cognitive learning outcomes might subtly affect high-performing learners, for others, cognitive costs appear closely balanced by increased engagement eventually mediating motivation and interest in the task. Overall, our study thus provides further support for an account integrating both cognitive and affective processes to conceptualize learning in human-computer interaction as provided by the integrated cognitive-affective model of learning with multimedia (Plass & Kaplan, 2016). On the other hand, it goes beyond an established body of knowledge by providing an empirical lens on the mechanisms by which the intricate relations between emotional, motivational, and cognitive aspects may be behaviorally implemented during learning in digital environments.

CRediT authorship contribution statement

Stefan E. Huber: Methodology, Validation, Formal analysis, Writing – original draft, Visualization. **Rodolpho Cortez:** Validation, Formal analysis, Writing – original draft. **Kristian Kiili:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Antero Lindstedt:** Software, Resources, Data curation, Writing – review & editing. **Manuel Ninaus:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

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.

Acknowledgments

The authors acknowledge the financial support by the University of Graz. This work was also supported by the Strategic Research Council (SRC) established within the ResearchCouncil of Finland (No. 335625).

References

de-Marcos, L., Domínguez, A., Saenz-de-Navarrete, J., & Pagés, C. (2014). An empirical study comparing gamification and social networking on e-learning. *Computers & Education, 75*, 82–91. https://doi.org/10.1016/j.compedu.2014.01.012

Abdul Jabbar, A. I., & Felicia, P. (2015). Gameplay engagement and learning in game-based learning: A systematic review. *Review of Educational Research, 85*(4), 740–779. https://doi.org/10.3102/0034654315577210

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. In R. E. Mayer, & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 522–560). Routledge. https://doi.org/10.4324/9781315736419.ch24.

Algina, J., Keselman, H. J., & Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods, 10*, 317–328. https://doi.org/10.1037/1082-989X.10.3.317

Axelson, R. D., & Flick, A. (2010). Defining student engagement. *Change: The Magazine of Higher Learning, 43*(1), 38–43. https://doi.org/10.1080/00091383.2011.533096

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Barz, N., Benick, M., Dörrenbacher-Ulrich, L., & Perels, F. (2023). The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: A meta-analysis. *Review of Educational Research.* https://doi.org/10.3102/00346543231167795

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning: An empirical study. *Simulation & Gaming, 43*(6), 729–760. https://doi.org/10.1177/1046878112439444

Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2013). Kurzskala zur Erfassung allgemeiner Selbstwirksamkeitserwartungen (ASKU). *Methoden, Daten und Analyse, 7*(2), 251–278.

Beißert, H., Köhler, M., Rempel, M., & Kruyen, P. (2019). Ein Vergleich traditioneller und computergestützter Methoden zur Erstellung einer deutschsprachigen Need for Cognition Kurzskala. *Diagnostica, 66*(1), 37–49. https://doi.org/10.1026/0012-1924/a000242

Bender, L., Renkl, A., & Eitel, A. (2021). When and how seductive details harm learning. A study using cued retrospective reporting. *Applied Cognitive Psychology, 35*, 948–959. https://doi.org/10.1002/acp.3822

Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose(s)? *Educational Psychology Review, 33*, 1675–1715. https://doi.org/10.1007/s10648-021-09615-8

Bernecker, K., & Ninaus, M. (2021). No Pain, no Gain? Investigating motivational mechanisms of game elements in cognitive tasks. *Computers in Human Behavior, 114*, 106542. https://doi.org/10.1016/j.chb.2020.106542

Bouffard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *The Journal of Social Psychology, 130*(3), 353–363. https://doi.org/10.1080/00224545.1990.9924591

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., … Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education, 94*, 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Breyer, B., & Bluemke, M. (2016). Deutsche version der Positive and negative affect schedule PANAS (GESIS panel). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. https://doi.org/10.6102/zis242

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

Cheng, C., & Su, C. (2012). A Game-based learning system for improving student's learning effectiveness in system analysis course. *Procedia - Social and Behavioral Sciences, 31*, 669–675. https://doi.org/10.1016/j.sbspro.2011.12.122

Christenson, S., Reschly, A., & Wylie, C. (2012). *Handbook of research on student engagement*. Springer. https://doi.org/10.1007/978-1-4614-2018-7

Cloude, E. B., Dever, D. A., Hahs-Vaughn, D. L., Emerson, A. J., Azevedo, R., & Lester, J. (2022). Affective dynamics and cognition during game-based learning. *IEEE Transactions on Affective Computing, 13*(4), 1705–1717. https://doi.org/10.1109/TAFFC.2022.3210755

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. *MindTrek 2011 Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. https://doi.org/10.1145/2181037.2181040

Dickey, M. D. (2020). Narrative in game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 283–306). MIT Press.

Engeser, S. (2005). Messung des expliziten Leistungsmotivs: Kurzform der Achievement Motives Scale. URL: http://www.uni-trier.de/fileadmin/fb1/prof/PSY/PGA/bilder/Engeser__2005_Kurzform_der_AMS.pdf. (Accessed 12 March 2023).

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7*(1), e11. https://doi.org/10.2196/jmir.7.1.e11

Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy, 98*, 19–38. https://doi.org/10.1016/j.brat.2017.05.013

Fox, J., & Weisberg, S. (2019). *An {R} Companion to applied regression* (3rd ed.). Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763–782). Springer US. http://link.springer.com/10.1007/978-1-4614-2018-7.

Ge, X., & Ifenthaler, D. (2017). Designing engaging educational games and assessing engagement in game-based learning. In R. Zheng, & M. Gardner (Eds.), *Handbook of research on serious Games for educational Applications* (pp. 253–270). https://doi.org/10.4018/978-1-5225-0513-6.ch012. IGI Global.

Greipl, S., Bernecker, K., & Ninaus, M. (2021). Facial and bodily expressions of emotional engagement: How dynamic measures reflect the use of game elements and subjective experience of emotions and effort. *Proceedings of the ACM on Human-Computer Interaction, 5(CHI, PLAY),*, 1–25. https://doi.org/10.1145/3474667

Greipl, S., Klein, E., Lindstedt, A., Kiili, K., Moeller, K., Karnath, H.-O., … Ninaus, M. (2021). When the brain comes into play: Neurofunctional correlates of emotions and reward in game-based learning. *Computers in Human Behavior, 125*, 106946. https://doi.org/10.1016/j.chb.2021.106946

Grosjean, P., & Ibanez, F. (2018). *pastecs: Package for analysis of space-time Ecological Series*. R package version 1.3.21. URL: https://CRAN.R-project.org/package=pastecs.

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The situational motivation scale (SIMS). *Motivation and Emotion, 24*, 175–213. https://doi.org/10.1023/A:1005614228250

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. https://doi.org/10.1016/S0166-4115(08)62386-9

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112. https://doi.org/10.3102/003465430298487

Hayes, A. F. (2022). *Introduction to mediation, Moderation, and conditional process analysis* (3rd ed.). New York: Guilford.

Hoerger, M. (2010). Participant dropout as a function of survey length in internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking, 13*(6), 697–700. https://doi.org/10.1089/cyber.2009.0445

Hoerger, M., & Currell, C. (2012). Ethical issues in Internet research. In S. Knapp, M. Gottlieb, M. Handelsman, & L. VandeCreek (Eds.), *APA handbook of ethics in psychology: Vol. 2. Practice, teaching, and research* (pp. 385–400). Washington, DC: APA.

Huang, X., & Mayer, R. E. (2016). Benefits of adding anxiety-reducing features to a computer-based multimedia lesson. *Computers in Human Behavior, 63*, 293–303. https://doi.org/10.1016/j.chb.2016.05.034

Huang, B.-G., Yang, J. C., & Chen, S. Y. (2021). An investigation of the approaches for integrating learning materials and digital games: A prior ability perspective. *Universal Access in the Information Society, 20*(1), 57–68. https://doi.org/10.1007/s10209-020-00710-1

Hunsu, N. J., Olaogun, O. P., Oje, A. V., Carnell, P. H., & Morkos, B. (2023). Investigating students' motivational goals and self-efficacy and task beliefs in relationship to course attendance and prior knowledge in an undergraduate statics course. *The Research Journal For Engineering Education, 112*(1), 108–124. https://doi.org/10.1002/jee.20500

Jackson, L. C., O'Mara, J., Moss, J., & Jackson, A. C. (2018). A critical review of the effectiveness of narrative-driven digital educational games. *International Journal of Game-Based Learning, 8*(4), 32–49. https://doi.org/10.4018/IJGBL.2018100103

Jankovsky, K., & Schroeders, U. (2022). Validation and generalizability of machine learning prediction models on attrition in longitudinal studies. *International Journal of Behavioral Development, 46*(2), 169–176. https://doi.org/10.1177/01650254221075034

Javora, O., Hannemann, T., Stárková, T., Volná, K., & Brom, C. (2019). Children like it more but don't learn more: Effects of esthetic visual design in educational games. *British Journal of Educational Technology, 50*(4), 1942–1960. https://doi.org/10.1111/bjet.12701

Kassambara, A. (2023a). *Ggpubr: 'ggplot2' based Publication Ready Plots*. R package version 0.6.0. URL: https://CRAN.R-project.org/package=ggpubr.

Kassambara, A. (2023b). rstatix: Pipe-Friendly framework for basic statistical tests. R package version 0.7.2. URL: https://CRAN.R-project.org/package=rstatix.

Kelders, S. M., Kok, R. N., Ossebaard, H. C., & Van Gemert-Pijnen, J. E. (2012). Persuasive system design does matter: A systematic review of adherence to web-based interventions. *Journal of Medical Internet Research, 14*(6), e152. https://doi.org/10.2196/jmir.2104

Ke, F., Xie, K., & Xie, Y. (2016). Game-based learning engagement: A theory- and data-driven exploration: Game-based learning engagement. *British Journal of Educational Technology, 47*(6), 1183–1201. https://doi.org/10.1111/bjet.12314

Khazaal, Y., van Singer, M., Chatton, A., Achab, S., Zullino, D., Rothen, S., Khan, R., Billieux, J., & Thorens, G. (2014). Does self-selection affect samples' representativeness in online surveys? An investigation in online video game research. *Journal of Medical Internet Research, 16*(7), e164. https://doi.org/10.2196/jmir.2759

Kocadere, S. A., & Çağlar, Ş. (2015). The design and implementation of a gamified assessment. *Journal of e-Learning and Knowledge Society, 11*(3), 85–99. https://doi.org/10.20368/1971-8829/1070

Koskinen, A., McMullen, J., Ninaus, M., & Kiili, K. (2023). Does the emotional design of scaffolds enhance learning and motivational outcomes in game-based learning? *Journal of Computer Assisted Learning, 39*(1), 77–93. https://doi.org/10.1111/jcal.12728

Kurzban, R., Duckworth, A., Kable, J., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences, 36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196

Laguilles, J. S., Williams, E. A., & Saunders, D. B. (2011). Can lottery incentives boost web survey response rates? Findings from four experiments. *Research in Higher Education, 52*, 537–553. https://doi.org/10.1007/s11162-010-9203-2

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In A. Holzinger (Ed.), *Lecture Notes in computer Science: Vol. 5298. HCI and Usability for education and work. USAB 2008*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-89350-9_6.

Lavoué, É., Ju, Q., Hallifax, S., & Serna, A. (2021). Analyzing the relationships between learners' motivation and observable engaged behaviors in a gamified learning environment. *International Journal of Human-Computer Studies, 154*, Article 102670. https://doi.org/10.1016/j.ijhcs.2021.102670

Leibowitz, N., Baum, B., Enden, G., & Karniel, A. (2013). The exponential learning equation as a function of successful trials results in sigmoid performance. *Journal of Mathematical Psychology, 54*, 338–340. https://doi.org/10.1016/j.jmp.2010.01.006

Lester, J. C., Spires, H. A., Nietfeld, J. L., Minogue, J., Mott, B. W., & Lobene, E. V. (2014). Designing game-based learning environments for elementary science education: A narrative-centered learning perspective. *Information Sciences, 264*, 4–18. https://doi.org/10.1016/j.ins.2013.09.005

Litvin, S., Saunders, R., Maier, M. A., & Lüttke, S. (2020). Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: A randomized controlled trial. *PLoS One, 15*(9), Article e0237220. https://doi.org/10.1371/journal.pone.0237220

Loderer, K., Pekrun, R., & Plass, J. L. (2020). Emotional foundations of game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (1st ed., pp. 111–151). The MIT Press.

Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does gamification increase engagement with online programs? A systematic review. *PLoS One, 12*(3), Article e0173403. https://doi.org/10.1371/journal.pone.0173403

Lumsden, J., Skinner, A., Coyle, D., Lawrence, N., & Munafo, M. (2017). Attrition from web-based cognitive testing: A repeated measures comparison of gamification techniques. *Journal of Medical Internet Research, 19*(11), e395. https://doi.org/10.2196/jmir.8473

Mair, P., & Wilcox, R. R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods, 52*, 464–488. https://doi.org/10.3758/s13428-019-01246-w

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of multimedia learning* (pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005.

McDonald, J. H. (2014). *Handbook of Biological Statistics* (3rd ed.). Sparky House Publishing. URL: http://www.biostathandbook.com.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Medina-Labrador, M., Martinez Quintero, S., Escobar Suarez, D., & Sarmiento Rincón, A. (2022). Dropout reduction in MOOCs through gamification and length of videos. In S. Hosseini, D. H. Peluffo, J. Nganji, & A. Arrona-Palacios (Eds.), *Transactions on computer systems and NetworksTechnology-enabled Innovations in education* (pp. 255–267). Singapore: Springer. https://doi.org/10.1007/978-981-19-3383-7_20.

Miller, E. M., Walton, G. M., Dweck, C. S., Job, V., Trzesniewski, K. H., & McClure, S. M. (2012). Theories of willpower affect sustained learning. *PLoS One, 7*(6), Article e38680. https://doi.org/10.1371/journal.pone.0038680

Ninaus, M., Greipl, S., Kiili, K., Lindstedt, A., Huber, S., Klein, E., … Moeller, K. (2019). Increased emotional engagement in game-based learning – A machine learning approach on facial emotion detection data. *Computers & Education, 142*, 103641. https://doi.org/10.1016/j.compedu.2019.103641

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/10.1109/TLT.2020.3031644

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), 3–16. https://doi.org/10.17083/ijsg.v2i1.60

Ninaus, M., & Sailer, M. (2022). Closing the loop – The human role in artificial intelligence for education. *Frontiers in Psychology, 13*, 956798. https://doi.org/fpsyg.2022.956798.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Plass, J. L., Homer, B. D., Mayer, R. E., & Kinzer, C. K. (2020). Theoretical foundations of game-based and playful learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (1st ed., pp. 3–24). The MIT Press.

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review, 31*, 339–359. https://doi.org/10.1007/s10648-019-09473-5

Plass, J. L., & Kaplan, U. (2016). Chapter 7 – emotional design in digital media for learning. In S. Y. Tettegah, & M. Gartmeier (Eds.), *Emotions, Technology, design, and learning* (pp. 131–161). Academic Press. https://doi.org/10.1016/B978-0-12-801856-9.00007-4.

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education, 52*, 275–300. https://doi.org/10.1080/15391523.2020.1719943

R Core Team. (2022). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/.

Radkowitsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., & Fischer, F. (2021). Learning to diagnose collaboratively – effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction, 75*, Article 101487. https://doi.org/10.1016/j.learninstruc.2021.101487

Revelle, W. (2022). *psych: Procedures for personality and psychological research*. Evanston, Illinois, USA: Northwestern University. URL: https://CRAN.R-project.org/package=psych.

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/10.1016/j.edurev.2012.05.003

RStudio Team. (2022). *RStudio*. Boston, MA: Integrated Development Environment for R. RStudio, PBC. URL http://www.rstudio.com/.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*, 77–112. https://doi.org/10.1007/s10648-019-09498-w

Sammut, R., Griscti, O., & Norman, I. J. (2021). Strategies to improve response rates to web surveys: A literature review. *International Journal of Nursing Studies, 123*, Article 104058. https://doi.org/10.1016/j.ijnurstu.2021.104058

Schwartz, R. N., & Plass, J. L. (2020). Types of engagement in learning with games. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 53–80). The MIT Press.

Skinner, E. A., Marchand, G., Furrer, C., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic. *Journal of Educational Psychology, 100*(4), 765–781. https://doi.org/10.1037/a0012840

Skitka, L. J., & Sargis, E. G. (2006). The internet as psychological laboratory. *Annual Review of Psychology, 57*, 529–555. https://doi.org/10.1146/annurev.psych.57.102904.190048

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261. https://doi.org/10.1037/0033-2909.124.2.240

Suárez, N., Regueiro, B., Estévez, I., del Mar Ferradás, M., Guisande, M. A., & Rodríguez, S. (2019). Individual precursors of student homework behavioral engagement: The role of intrinsic motivation, perceived homework utility and homework attitude. *Frontiers in Psychology, 10*, 941. https://doi.org/10.3389/fpsyg.2019.00941

Sweller, J. (2011). Chapter two - cognitive load theory. *Psychology of Learning and Motivation, 55*, 37–76. https://doi.org/10.1016/b978-0-12-387691-1.00002-8

Venables, W. N., & Ripley, B. D. (2002). *Modern applied Statistics with S* (4th ed.). New York: Springer.

Wangberg, S. C., Bergmo, T. S., & Johnsen, J. K. (2008). Adherence in internet-based interventions. *Patient Preference and Adherence, 2*, 57–65.

Wang, J. H., Chen, S. Y., Chang, B., & Chan, T. W. (2016). From integrative to game-based integrative peer response: High ability versus low ability: IPR versus G-IPR. *Journal of Computer Assisted Learning, 32*(2), 170–185. https://doi.org/10.1111/jcal.12125

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wong, R. M., & Adesope, O. O. (2021). Meta-analysis of emotional designs in multimedia learning: A replication and extension study. *Educational Psychology Review, 33*, 357–385. https://doi.org/10.1007/s10648-020-09545-x

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*(2), 249–265. https://doi.org/10.1037/a0031311

Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education, 60*(1), 412–425. https://doi.org/10.1016/j.compedu.2012.07.018

Wouters, P., & van Oostendorp, H. (2017). Overview of instructional techniques to facilitate learning and motivation of serious games. In P. Wouters, & H. van Oostendorp (Eds.), *Instructional Techniques to facilitate learning and motivation of serious Games. Advances in game-based learning.* Cham: Springer. https://doi.org/10.1007/978-3-319-39298-1_1.

Yeo, J.-H., Cho, I.-H., Hwang, G.-H., & Yang, H.-H. (2022). Impact of gender and prior knowledge on learning performance and motivation in a digital game-based learning biology course. *Educational Technology Research & Development, 70*(3), 989–1008. https://doi.org/10.1007/s11423-022-10099-8

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review, 30*, Article 100326. https://doi.org/10.1016/j.edurev.2020.100326

Zu, J., & Yuan, K. (2010). Local influence and robust procedures for mediation analysis. *Multivariate Behavioral Research, 45*(1), 1–44. https://doi.org/10.1080/00273170903504695