International Journal of Serious Games

ISSN: 2384-8766
https://journal.seriousgamessociety.org/

Article

Game elements improve affect and motivation in a learning task

Stefan E. Huber[1], Moritz Edlinger[1], Antero Lindstedt[2], Kristian Kiili[2], Manuel Ninaus[1,3]

[1]Department of Psychology, University of Graz, Graz, Austria; [2] Faculty of Education and Culture, Tampere University, Tampere, Finland; [3]LEAD Graduate School and Research Network, University of Tübingen, Tübingen, Germany
stefan.huber@uni-graz.at ; manuel.ninaus@uni-graz.at

Keywords:

game elements
affect
motivation
memory
game-based learning
gamification

Received: May 2024
Accepted: October 2024
Published: November 2024
DOI: 10.17083/ijsg.v11i4.769

Abstract

Earlier studies repeatedly showed increased learner motivation due to game elements, while overall cognitive effects on learning outcomes were absent. One possible explanation for this discrepancy is provided by theories integrating cognitive and affective learning processes: the beneficial effect on learner motivation eventually balancing simultaneously higher cognitive processing demands associated with game elements. In this paper, we provide results of an empirical test of this theoretical suggestion. In particular, we report results of a value-added online experiment (with $n = 61$ participants, mostly students; 44 female, 15 male, 2 diverse; median age: 24 years), comparing a more gameful with a less gameful version of a learning task. In agreement with earlier studies, we find similar cognitive learning outcomes ($\delta < 0.2$), but medium ($\delta \sim 0.5$) and large ($\delta \sim 0.9$) effects on affective and motivational outcomes, respectively. Furthermore, mediation models show that a small negative direct effect of game elements on cognitive outcomes ($\beta \sim -0.2$) is indeed effectively cancelled by an indirect path through motivational outcomes ($\beta \sim +0.4$). Overall, our results corroborate the tenability of the integrated cognitive affective model of learning with multimedia. This implies its feasibility in purposefully designing learning environments with specific motivational or cognitive aims in mind.

1. Introduction

Using game elements or even full-fledged games in educational contexts is often based on the aim to leverage their capabilities in capturing and holding people's attention and fostering sustained engagement and motivation [1]. Meta-analyses on game-based learning interventions indeed support their effectiveness concerning cognitive and motivational outcomes for learning in school [2] and higher education [3]. Meta-analyses on gamification have further shown that already the inclusion of specific, separable game features in digital tasks can enhance engagement [4] and motivation [5]. While game-based learning and gamification are certainly distinct approaches [6], the motivational capabilities of their common feature of game elements have been corroborated for both game-related pedagogies [7].

However, the exact mechanisms by which game elements exert their effects during learning are not yet fully elaborated [8]. The present work contributes to clarifying two unresolved issues: (i) the interrelation of motivational aspects of game elements with cognitive learning outcomes; (ii) the interrelation between motivational aspects of game elements and affective dynamics in learners.

1.1 Characterization of the current knowledge gap

The first addressed issue is how motivational aspects may be interrelated with cognitive learning outcomes. Numerous studies [9], [10], [11], [12], [13] report significant effects of game elements on learner motivation or the closely related construct engagement [14], whereas, simultaneously, cognitive learning outcomes remained largely unaffected. This absence of cognitive effects of game elements on learning outcomes remains in need of explanation, because increased motivation and engagement would suggest enhanced cognitive outcomes [15], [16]. These previous findings thus imply that sometimes game elements are associated with both motivational and cognitive benefits only in a number of cases. However, knowing under which circumstances game elements are associated with what effects is an imperative prerequisite for designing tailored, personalized learning environments. Hence, we aim to shed further empirical light on the interplay between motivational and cognitive aspects during learning.

The second issue addressed in the present work is how motivation is interrelated with affective dynamics during learning. Previous research revealed that emotional design elements (such as characters in task illustrations exhibiting facial expressions, or appealing colors or graphical elements) can improve learning [17], [18], [19], [20], showing that emotional design features can be comparably effective in facilitating learning as instructional features based on purely cognitive principles of multimedia learning [21]. By aiming at generating positive emotional responses [22], emotional design features link the respective learning tasks inevitably to motivation. Intrinsic motivation occurs for activities that hold some intrinsic appeal for an individual, which can be in the form of aesthetic value or enjoyment associated with that activity [23]. According to the self-determination theory [24] interest and enjoyment are actually *the* self-report measures of intrinsic motivation. Furthermore, positive emotions are considered to strengthen both intrinsic and extrinsic motivation [25]. Thus, the motivational capabilities of game elements are diversely linked to their emotional aspects and their effect on affective dynamics in learners. Hence, to clarify the interplay between motivational and cognitive aspects during learning, the interrelations between motivational and affective aspects also need to be considered.

1.2 Theoretical background

Aiming for a holistic understanding of the mechanisms by which game elements can influence learning, it seems advisable to consider the possible interrelations between all three, i.e., cognitive, affective, and motivational components in the learning process. A theoretical framework that integrates these components is provided by the integrated cognitive affective model of learning with multimedia (ICALM) developed by Plass and Kaplan [26].

Essentially, the ICALM combines Mayer's cognitive theory of multimedia learning [21] with Russell's concepts of core and attributed affect [27], and Izard's concept of emotion schemas [28]. A crucial point in the ICALM is that cognitive processes are inseparably intertwined with affective processes inducing the emergence of highly context-sensitive emotion schemas that serve as motivators for learning in multimedia environments. Due to the inevitable interrelatedness of cognition and affect, affective processes influence cognitive processes and vice versa [26].

Hence, on the one hand, emotional design features may induce positive affect or enjoyment in learners. Positive affect involving appraisal is subjectively experienced in the form of interest or motivation [26]. By facilitating engagement in generative cognitive processing, motivation can support the utilization of free cognitive capacity [29]. Or in other words: Design features – such as game elements – may induce positive affect, which may subsequently increase motivation and engagement, which can subsequently enhance cognitive learning outcomes.

On the other hand, affective processes also require cognitive resources [26]. If unessential for the task at hand, they can be associated with extraneous cognitive processes, occupying a portion of the generally limited cognitive capacity [21], [30]. Affective processes that distract learners' attention from learning objectives can indeed hinder learning, also known as seductive detail effect [31], [32].

The ICALM thus provides the following potential, theoretical explanation for the absence of cognitive learning outcomes in the above noted game-based learning studies [9], [10], [11], [12], [13], even if effects on motivation or engagement were clearly present: While game elements indeed acted as motivators enhancing cognitive learning outcomes, they simultaneously demanded additional cognitive resources, reducing cognitive learning outcomes. This would finally result in an approximately zero net effect on cognitive learning outcomes.

1.3 Specific aims of the present study – Hypotheses

The present study aims to empirically test the outlined explanatory framework by testing the following hypotheses. Supposing that the framework captures the essentials of the learning process, we would expect the following:

- Hypothesis 1: Cognitive learning outcomes should be similar in more and less gameful versions of the same task, if the difference in used game elements between task versions is also similar as in the above noted earlier studies [10], [11], [12] (see Section 2.3 for details on task versions and game elements). Note that the hypothesis is thus conditional on the use of a specific set of game elements, which are described in detail in Section 2.3 below.
- Hypothesis 2: The more and less gameful task versions should differ with respect to the change in positive affect from before to after the learning task.
- Hypothesis 3: The more and less gameful task versions should differ with respect to motivation.
- Hypothesis 4: The effects of game elements on cognitive outcomes should be partially mediated by motivational effects. That is, there should be a positive indirect association between game elements and cognitive outcomes via the

mediator motivation. Furthermore, to result in an overall similar cognitive effect (hypothesis 1), this indirect pathway should be (partially) cancelled by a negative direct effect. Figure 1(a) illustrates these expected interrelations between game elements, cognitive and motivational outcomes.

- Hypothesis 5: According to the considerations from an emotional design perspective outlined above, the motivational effects of game elements should further be mediated (partially or completely) by the increase of positive affect from before to after the task. That is, we expect that, at least partially, game elements induce positive affect which in turn enhances motivation. Figure 1(b) illustrates these expected interrelations between game elements, positive affect and motivation.

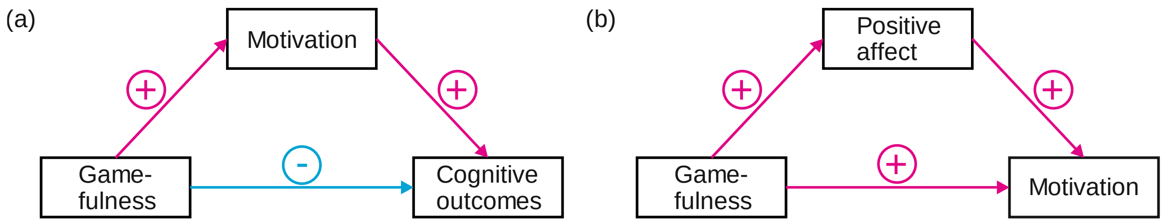

Figure 1. Illustrations of the expected interrelations between (a) gamefulness of the learning task, motivation and cognitive outcomes (hypothesis 4) and (b) gamefulness, positive affect and motivation (hypothesis 5).

2. Methods and Material

2.1 Participants, data collection and sampling considerations

In total, 61 participants (44 female, 15 male, 2 diverse) completed the study. The participants' age ranged from 18 to 64 years (*M* = 27.56, *Mdn* = 24, *SD* = 11.54, *Median Absolute Deviation (MAD)* = 4.45; all in years). Most of the participants were students (52 of 61). Apart from their student status no further information regarding the exact occupation or field of study was obtained. Psychology students were compensated for study participation by course credit. All study participants provided informed consent. The study was approved by the local university's ethics committee.

Beyond the 61 participants completing the study, 35 additional participants had at least started the learning task, but disengaged from it later. Apart from assessing attrition in dependence of task version for the sake of exploring condition equivalence in Section 3.1, their data were not analyzed further.

Data collection took place in May 2022 in the framework of a university course on empirical research. A university-wide e-mail broadcast was used to invite participants to take part in the study. Hence, the sample is a typical example of a relatively small opportunity sample. However, earlier studies [11], [13] have shown that motivational effects of game elements can be substantial (Cohen's *d* ~ 0.8).

2.2 Study design

We conducted a typical value-added [22] online experiment to empirically test the hypothesized (indirect) cognitive, affective, and motivational effects of game elements. In particular, two experimental conditions were implemented using two versions of a learning task differing solely in the use of specific, separable game elements (described in detail in Section 2.3). Participants were randomly assigned to one of these two experimental conditions. Cognitive, affective, and motivational outcome measures (described in Section 2.4) were assessed and compared between the two conditions.

For assessing outcome measures (besides socio-demographic data), participants were administered questionnaires before (pre-task survey) and after (post-task survey) the learning task. Both surveys were implemented in LimeSurvey. The two versions of the learning task were based on the NumberTrace game engine (see e.g., Ref. [33]; 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. This information was communicated in the e-mail broadcast through which participants were invited to the study. Apart from this, we had no control over the exact conditions in which participants took part in the study, as with any online study. Particularly, we had no control over when exactly participants would join the experiment, in what exact environment, and how long it would take them exactly to finish the experiment (including the pre- and post-task surveys). To some extent, participation would require conducting all tasks (pre- and post-task surveys and the learning task) in one session, as closing the browser window would lead to an incomplete data set and exclusion of that data from analysis.

2.3 Learning task

Both more and less gameful learning task versions were the same (with one exception outlined in the following) as in an earlier study [10]. In both versions, participants needed to learn the associations between 20 pairs of symbols and numbers.

In both task versions, the symbols were displayed consecutively in the left upper corner of the screen. For each shown symbol, the participants had to select a number visualized on a number line with 27 possible numbers (from 0 to 26) in ascending order. The selection required the movement of a slider with arrow keys and confirmation of the selection via pressing the space bar of the keyboard. The confirmation had to occur within 20 seconds, otherwise no number would be registered as selected. In the first level of the task, participants did not know to which number a shown symbol would belong and would thus need to guess some number for each symbol. After selecting a number, corrective feedback showed if the number was correct and always provided also the correct number for the given symbol (regardless of the correctness of the response) by displaying a green vertical bar over the correct number, see Fig. 2. This corrective feedback would also be provided after exceeding the maximum response time of 20 seconds. Neither of the two task versions involved any audio or sound effects.

After having seen all 20 symbols and their numbers once in the first level of the task, the whole procedure was repeated in four further task levels. In those levels, the participants could use their obtained knowledge about the associations in the previous levels to respond to the displayed symbols and to learn the remaining, not yet memorized associations. To exclude order effects, symbols were displayed in random order in each task level.

This core mechanic was exactly the same in both task versions. The only differences between the less and the more gameful task versions consisted of a narrative, visual aesthetics, and a virtual incentive system. An illustration of the differences between the more and less gameful task versions is provided in Fig. 2.

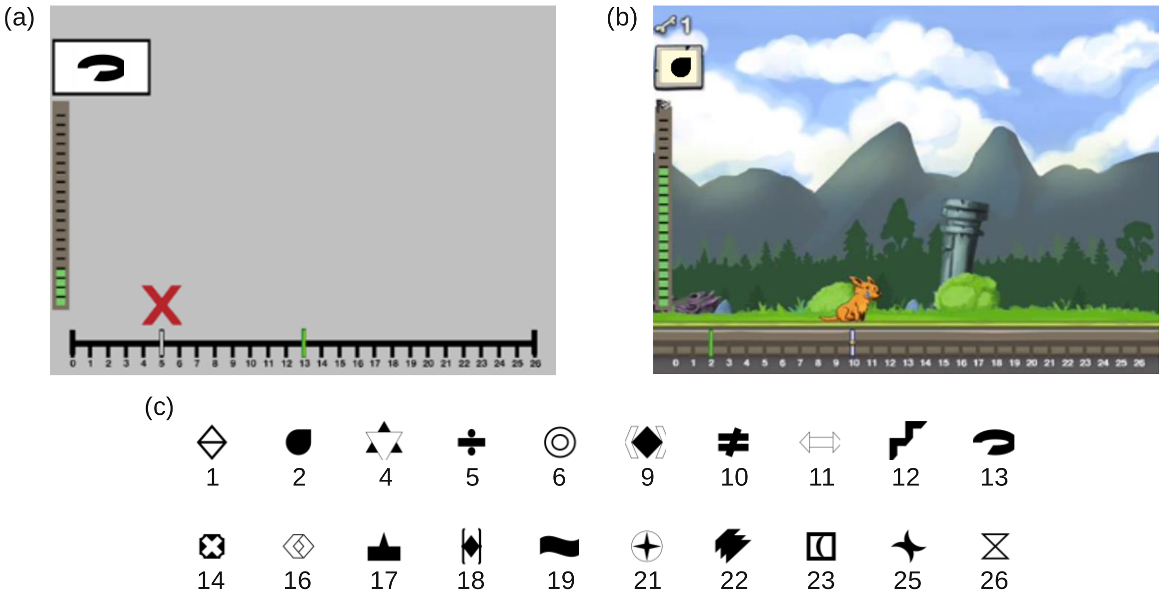

Figure 2. Illustrations of the (a) less and (b) more gameful task implementations during the corrective feedback phase and (c) the 20 pairs of symbols and numbers that are required to be learned over the course of the task.

The narrative in the more gameful task version was provided in the framework of the task instruction and consisted of a dog walking in a forest searching for hidden bones. The only hint the dog would have for the location of the bones was a set of symbols, each associated with a certain position in the ground. By memorizing the associations between symbols and numbers, the participants could assist the dog in finding the hidden bones. In the more gameful task version, the cursor's movement was accompanied by a walking animation of the dog. The placement of the cursor (by pressing the spacebar) would initiate a digging animation. Correct positioning resulted in the dog wagging its tail and the bone count displayed above the display of the current symbol increased by one (incentive system). In case that the position was incorrect, the dog would cry instead, see Fig. 2(b). In the less gameful task version, a green check mark and a red X-symbol would indicate correct and incorrect responses, respectively. The less gameful task version would also lack any fictional narrative accompanying the task instruction and all of the described visual aesthetics. Instead, a constant, grey background was presented. There was also no count of correct responses provided in the less gameful task version.

The total number of symbol-number pairs was increased from 14 to 20 with respect to said previous investigation [10] to increase the discriminatory power of our cognitive outcome measures (see Section 2.4) and mitigate ceiling effects. Using 14 symbol-number pairs had resulted in about a third of the participants remembering all of them at the fifth task level (see below), and some of the participants remembering them even already at the second, third or fourth level [10].

2.4 Outcome measures

Overall, we differentiate between cognitive, affective, and motivational outcome measures presented in separate subsections below. Whereas cognitive outcomes were measured using performance indices, affective and motivational outcomes were measured using self-report questionnaires.

2.4.1 Cognitive outcomes

Regarding cognitive outcomes, we discern between learning efficacy and learning efficiency. Learning efficacy refers to the questions if and how much participants learned over the course of the learning task. This is operationalized by the number of correct responses given in the final level (i.e., level 5).

Learning efficiency refers to how fast participants learned new symbol-number associations over the course of the task, i.e., how fast they arrive at the learning efficacy provided by the number of correct responses at level 5. To operationalize this construct, we modeled the individual, temporal increase of learned material using an exponential learning curve [34]:

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

In Eq. (1), $N_{\mathrm{corr},i}(L)$ denotes the number of correct responses of the i-th participant at task level $L = 1, \ldots, 5$, $N_{\max} = 20$ denotes the maximum number of correct responses, and the coefficient c_i denotes the rate constant quantifying the learning efficiency of the i-th participant. The higher the rate constant, the faster participants approach the maximum number of correctly memorized symbol-number associations.

Note that according to Eq. (1) at task level 1, $N_{\mathrm{corr},i}(1) = 0$, because no associations could have been memorized prior to completing level 1. Note further that the computational model for the temporal course of learning provided by Eq. (1) does not take into account more complex dynamic processes such as forgetting already learned material or omission errors due to inattention or carelessness. However, we opted for such a simple, one-parameter model because only four data points (i.e., the numbers of correct responses for levels 2-5) were available for fitting Eq. (1) for each participant. The latter was obtained using a non-linear least squares fit provided by the nls-function in R [35].

2.4.2 Affective outcomes

Affective outcomes were assessed using the German version of the positive and negative affect schedule (PANAS) [36] which was developed based on the original English version provided by Watson et al. [37]. Because we were particularly interested in the change of positive affect from before to after the learning task (see hypotheses 2 and 5 above), the PANAS was administered directly before and after the learning task.

For positive affect, the PANAS provides ten adjectives [37] such as “attentive”, “interested”, or “enthusiastic”. For each of those adjectives, participants are asked to indicate the intensity with which they were experiencing these emotions in the immediate past on a 5-point scale ranging from “not at all” to “very much”. By averaging responses to those ten items an overall scale for positive affect is provided.

As a measure of internal consistency, we provide McDonald’s ω which has been recently suggested as a more reliable estimate for a scale’s reliability taking into account the possibility of unequal factor loadings [38] (Cronbach’s α assumes equal factor loadings instead). We obtained McDonald’s ω = 0.91 and ω = 0.90 for positive affect before and after the learning task, respectively.

According to our hypothesis we were not particularly interested in negative affect and it is less clear than for positive affect how the implemented game elements might affect negative affect [25]. However, because PANAS was administered in full and provides a scale for negative affect too, we report also results for changes in negative affect from before to after the learning task for the sake of completeness. The procedure for negative affect is analogous to the one for positive affect, but obviously using a different list of ten adjectives [37] such as “distressed”, “upset”, or “afraid”. For negative affect, we obtained McDonald’s ω = 90 and ω = 0.88 after the learning task, respectively.

2.4.3 Motivational outcomes

To assess motivational learning outcomes, we used several subscales of two distinct self-report questionnaires administered after completion of the learning task directly subsequent to the PANAS post-test described above.

In particular, we used the two subscales *interest* and *perceived competence* of the German short scale to measure intrinsic motivation developed by Wilde et al. [39]. The subscales *perceived choice* and *pressure/tension* were not included since the mechanics of the learning task did not allow for customization of the learning activity.

The subscale *interest* includes three items addressing the general appeal of or interest in the activity within the learning task (like “The activity in the learning task was fun”, or “I found the activity in the learning task very interesting”). The subscale *perceived competence* includes three items addressing the satisfaction with the own performance in the learning task (like “I think I was pretty good at what I did in the learning task”). All items were answered on a 5-point rating scale ranging from “does not apply at all” to “applies completely”. Total scores for each of the two subscales were computed by averaging responses to the respective three individual items. Regarding internal consistency, we obtained McDonalds $\omega = 0.90$ for the subscale *interest*, and $\omega = 0.94$ for the subscale *perceived competence*.

Motivational outcomes were complemented by the *attractivity* and *stimulation* subscales of the German version of the user experience questionnaire [40], [41]. Both subscales consist of several items, each presenting two opposing adjectives, forming the endpoints of a 7-point, bipolar rating scale, on which participants indicate their experience of the task.

In particular, *attractivity* aims to assess (using six items in total) the general appeal of a task (or product) by asking, for instance, how “enjoyable” (with the poles: “annoying”/”enjoyable”), or “pleasing” (with the poles: “unlikable”/”pleasing”) it is perceived. *Stimulation* aims to assess (using four items in total) the capability of a task (or product) to motivate or captivate by asking, for instance, how “interesting” (poles: “not interesting”/”interesting”), or “motivating” it is perceived. As outlined above and noted also by Wilde et al. [39], the interest in and enjoyment of an activity is *the* self-report measure of intrinsic motivation [24]. As the subscales *attractivity* and *stimulation* overlap substantially with those two constructs, i.e., enjoyment and interest, we use them as additional measures of intrinsic motivation. Regarding internal consistency, we obtained McDonalds $\omega = 0.93$ for the subscale *attractivity*, and McDonalds $\omega = 0.91$ for the subscale *stimulation*.

2.5 Statistical analyses and hypothesis testing

Empirical tests of the hypotheses listed in Section 1.3 are all assuming the comparability of the two experimental conditions in any respect apart from the difference in the used game elements explicated in Section 2.3. Hence, before statistical tests of any hypothesis were conducted, we assessed the similarity between conditions with respect to gender, status of being a student, attrition over the course of the task, typical age of the participants and positive and negative affect before the task.

To do so, count data (gender, student status, attrition) were analyzed for statistical significance of differences between expected and observed frequencies using Fisher’s exact test. Metric variables (age, positive and negative affect scales) were assessed for statistically significant differences between either the means or the trimmed means with a trim factor of 0.2 of the two empirical distributions for each metric variable. Means, denoted as M, or trimmed means, denoted as M_t, were chosen for representing typical values of the respective metric variable in each empirical distribution. Standard deviations, SD, and winsorized standard deviations, s_w were taken as appropriate measures of scale, respectively. For normally distributed random variables means and trimmed means coincide. In the presence of considerable skew, outliers, or generally substantial deviations from normality, trimmed means provide a more precise estimate of what can be considered a typical value for the respective variable [42], [43].

Hence, for each metric variable, we first explored the existence of outliers via the boxplot-rule and tested deviations from normality via Shapiro-Wilk tests with $\alpha = 0.20$ (increasing the type I error probability to be more sensitive to deviations from normality). Only if no considerable deviations from normality were obtained based on these criteria, the equality of means was tested for statistical significance using Welch's t-test (to take into account heteroskedasticity). Otherwise, the equality of trimmed means was tested using the bootstrap version of Yuen's test provided by the WRS2 package [44] using 2000 bootstrap samples. By this procedure, always the difference between *typical* values of the respective variable in the less and more gameful task condition was tested for statistical significance. Like with our checks of normality, conditions were deemed reasonably similar, if we failed to reject the null hypothesis of equal (trimmed) means with $\alpha = 0.20$.

This procedure was exactly the same for the metric cognitive, affective, motivational outcomes involved in hypotheses 1-3. That is, first their conformity with normality was explored, and based on this investigation either means or trimmed means were tested for statistically significant differences between conditions. According to hypothesis 1, we expected similar cognitive outcomes. Also in that case, conditions were deemed reasonably similar if we failed to reject the null hypothesis of equal (trimmed) means with $\alpha = 0.20$ for both learning efficacy and efficiency.

In addition to the comparison of typical values of learning efficacy and efficiency, we also computed a robust between-within subjects analysis of variance on the trimmed means of correct responses given by participants in levels 2-5. To do so, we used the bwtrim-function provided by the WRS2 package [44]. Assessing the interaction between the within-subjects factor level and the between-subjects factor condition provides an alternative perspective on differences between conditions regarding how participants learn associations over the course of the task not captured by individual learning curves modeled using Eq. (1).

From hypotheses 2 onwards, we interpreted results of statistical inference tests according to Tukey's three decision rule [45]. Regarding hypotheses 2 and 3 that is, if the sample (trimmed) mean of the considered variable was significantly smaller/larger in the less than in the more gameful condition with $\alpha = 0.05$, then we deemed it reasonable to decide that the population (trimmed) mean of the considered variable is smaller/larger in the less than in the more gameful condition. Instead, for $p > \alpha = 0.05$, no, in such a way, reasonable decision can be made about which population (trimmed) mean is larger. So the goal was not to test for exact equality, but to assess the empirical evidence that a decision can be made about which population (trimmed) mean is larger [43]. Note that the p-value neither reflects the probability that a correct decision is made nor informs about the importance or size of a difference between population values.

Effect sizes of differences between conditions are reported in the form of Cohen's *d* [46] in the case of comparing means, and its robust generalization, δ_t, suggested by Algina et al. [47] in the case of comparing trimmed means. Cohen's [46] rules of thumb apply for both effect size measures, that is 0.2, 0.5, and 0.8 correspond to small, medium, and large effects, respectively, for both measures. If we refer to any of these effect size measures, we simply use the generic symbol δ. As a measure of the precision of obtained point estimates and effect sizes we provide 95%-confidence intervals, based on 2000 bootstrap samples and a percentile bootstrap method. A bootstrap-t method is used in the case of point estimates for trimmed means. The confidence intervals are always reported directly following the corresponding point estimate in squared brackets. For interpretation, lower and upper limits of confidence intervals are also discussed as reasonably compatible with data (with $\alpha = 0.05$) besides point estimates of effect sizes [48].

In the case of motivational outcomes, we had four different measures to assess for significant differences between conditions (see Section 2.4.3). Because one or more significant differences with respect to any of those measures would suffice to conclude that the conditions differ regarding motivational outcomes at the population level, a correction for multiple comparisons is required, for which we used Hochberg's method (see, e.g., Ref. [42]) with a family-wise error rate of 5%. However, instead of reporting modified, comparison-wise *p*-

values, we provide unmodified p-values in conjunction with the respective critical values, d_k, obtained via Hochberg's procedure for the sake of reproducibility.

For testing hypotheses 4 and 5, we first explored associations between all cognitive, affective and motivational outcomes using percentage bend correlations, denoted as ρ_{pb}, as a robust measure of association, which gives essentially the same values as Pearson's correlation coefficient for bivariate normal data [44]. For testing mediation of the effect of game elements on cognitive outcomes via motivation (hypothesis 4) and mediation of the effect of game elements on motivation via the change of positive affect from before to after the task (hypothesis 5), we tested the corresponding indirect effects for statistical significance using Zu and Yuan's [49] robust approach. For interpretation, we again employed Tukey's three decision rule. That is, if an indirect effect was significant with $\alpha = 0.05$ and smaller/larger than zero, we concluded that, given these data, it appears reasonable to decide that the indirect effect is smaller/larger than zero at the population level. If an indirect effect was insignificant, no such reasonable decision can be made about the population effect based on the given data.

To provide information about the coefficients for individual paths in the corresponding mediation models, we further fitted robust, linear regression models using M-estimators as implemented in the rlm-function of the MASS package [50]. All mediation models are based on studentized metric variables to allow comparison of the effects of individual paths.

All statistical analyses were conducted using R [35] and RStudio [51] complemented by the following packages: tidyverse [52], XML [53], ggpubr [54], rstatix [55], WRS2 [44], MASS [50], ltm [56], pwr [57], MBESS [58].

2.6 Availability of data and computational analyses

Data and analyses supporting this study are openly available from the Open Science Framework (OSF) at https://osf.io/cq3aj/ and https://osf.io/p4hvw/, respectively.

3. Results

3.1 Comparability of conditions

Of the 61 participants who completed the study, 33 were part of the less gameful condition, and 28 were part of the more gameful condition. The conditions were similar regarding gender distribution, $p > 0.999$, and regarding counts of student and non-student participants, $p = 0.488$. In the non-game condition, the participants' age ranged from 18 to 64 years ($M_t = 23.81$, $s_w = 4.08$), while in the game condition, it ranged from 19 to 61 years ($M_t = 24.33$, $s_w = 4.02$), yielding no significant difference (with $\alpha = 0.20$) either, $Y_t = 0.30$, $p = 0.769$.

The conditions were also similar regarding participant attrition during the learning task, $p = 0.526$. In the non-game condition, 49 persons started the task and 33 completed it. In the game condition, 47 started the task and 28 completed it. The conditions were further similar regarding positive affect before the task (non-game: $M = 3.02$, $SD = 0.80$; game: $M = 2.90$, $SD = 0.72$), $t(58.76) = 0.61$, $p = 0.547$, and negative affect before the task (non-game: $M_t = 1.25$, $s_w = 0.19$; game: $M_t = 1.32$, $s_w = 0.27$), $Y_t = 0.73$, $p = 0.452$.

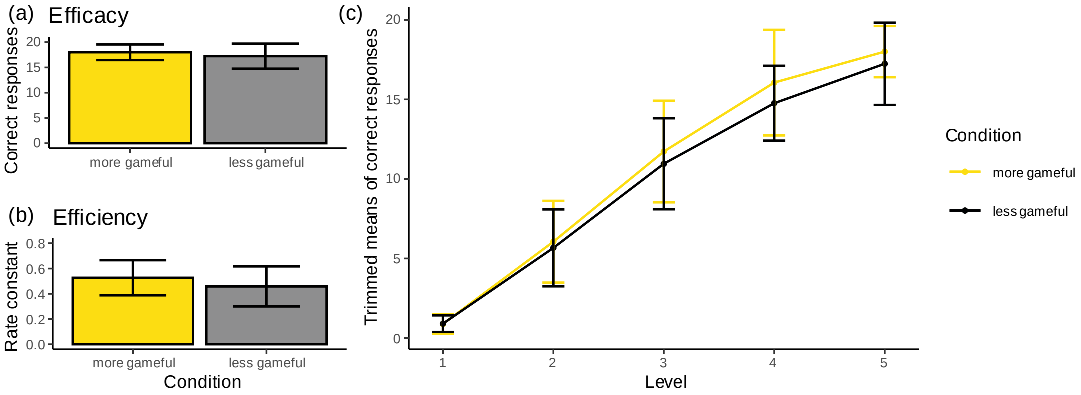

Figure 3. Trimmed means of (a) learning efficacy (i.e., the number of correct responses at level 5), (b) learning efficiency (i.e., the rate constant in Eq. (1)), and (c) correct responses at each task level for the more gameful (yellow) and less gameful (grey/black) task versions. Error bars represent 95%-confidence intervals of trimmed means computed via a bootstrap-t method.

3.2 Cognitive outcomes (hypothesis 1)

Neither learning efficacy (more gameful: M_t = 18.00, s_w = 2.61; less gameful: M_t = 17.24, s_w = 3.65; ΔM_t = 0.76 [-1.84, 3.37], Y_t = 0.57, p = 0.559) nor learning efficiency (more gameful: M_t = 0.53, s_w = 0.23; less gameful: M_t = 0.46, s_w = 0.27; ΔM_t = 0.07 [-0.13, 0.27], Y_t = 0.65, p = 0.503) differed significantly between the two conditions, see also Figs. 3(a) and (b). Hence, we cannot reject the hypothesis of equal trimmed means, but note that small negative (δ_t ~-0.4) to moderate positive (δ_t ~0.7) effect sizes are also reasonably compatible with our data given our assumptions and the small size of our sample. Test statistics and effect sizes are provided in Table 1.

The number of correct responses did not significantly differ between conditions in task levels 2-5. In particular, a robust two-way between-within subjects ANOVA yielded a significant main effect of task level, $Q_{level}(3, 28.19)$ = 167.76, p < 0.001, but no significant main effect of condition, $Q_{condition}(1, 33.87)$ = 0.30, p = 0.590, and also no significant interaction, $Q_{level,condition}(3, 28.19)$ = 0.52, p = 0.673. The trimmed means and their 95% CIs are provided for each task level in Fig. 3(c).

Table 1. Test statistics and effect sizes of differences between task conditions regarding the considered cognitive, affective and motivational learning outcomes. In addition to change scores, (trimmed) means of positive and negative affect are also provided before and after the task. Outcome measures differing significantly between conditions are marked with an asterisk. In the case of motivational outcomes, multiple comparisons were taken into account by Hochberg's procedure. Critical values to which (uncorrected) p-values were compared are denoted by d_k. The given 95%-confidence intervals are based on a percentile bootstrap method.

Outcome category	Outcome measure	Test statistics	Effect size
Cognitive	Efficacy	Y_t = 0.57, p = 0.556	δ_t = 0.15 [-0.41, 0.65]
	Efficiency	Y_t = 0.65, p = 0.503	δ_t = 0.18 [-0.33, 0.73]
Affective	Change in pos. affect*	$t(53.78)$ = 2.01, p = 0.049	d = 0.52 [0.04, 1.03]
	Change in neg. affect	Y_t = 1.67, p = 0.084	δ_t = 0.48 [-0.10, 1.10]
Motivational	Interest	Y_t = -0.56, p = 0.544 > d_k = 0.05	δ_t = -0.16 [-0.66, 0.37]
	Competence*	Y_t = 2.97, p = 0.005 < d_k = 0.025	δ_t = 0.79 [0.31, 1.47]
	Attractivity*	Y_t = 3.00, p = 0.003 < d_k = 0.025	δ_t = 0.82 [0.31, 1.30]
	Stimulation*	Y_t = 3.17, p = 0.002 < d_k = 0.025	δ_t = 0.87 [0.26, 1.56]

3.3 Affective outcomes (hypothesis 2)

In line with hypothesis 2, we obtained a significantly larger increase in positive affect from before to after the task in the more gameful as compared to the less gameful condition, yielding a medium effect, see Table 1 and Figure 4(a). Negative affect decreased slightly from before to after the task in the less gameful condition, while it remained almost unchanged in the more gameful condition, see Table 1 and Figure 4(b). The difference between conditions was not significant.

3.4 Motivational outcomes (hypothesis 3)

Three out of four motivational outcomes differed significantly between conditions, see Table 1 and Figure 4(c)-(f). Hence, in line with hypothesis 3, task versions differ with respect to motivation. Trimmed means of perceived competence, attractivity, and stimulation were significantly larger in the more gameful than in the less gameful version of the learning task, accounting for similarly large effects in all three cases. No significant difference between task versions was obtained for self-reported interest.

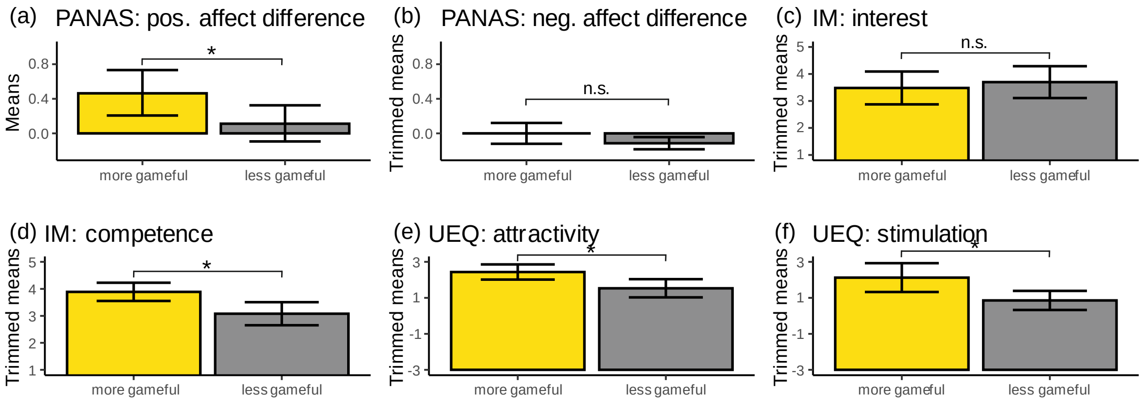

Figure 4. (Trimmed) Means of (a) change in positive affect from before to after task, (b) change in negative affect from before to after task, (c) self-reported interest, (d) perceived competence, (e) task attractivity, and (f) perceived stimulation by the task. Error bars represent 95%-confidence intervals computed via a bootstrap-t method for trimmed means and a percentile bootstrap method for means.

3.5 Interrelations between cognitive, affective, and motivational outcomes

3.5.1 Pairwise associations

Table 2 lists all pairwise percentage-bend correlations obtained for all considered cognitive, affective, and motivational learning outcomes. We note that both cognitive outcomes are highly correlated with each other. Both cognitive outcomes are further highly correlated with self-reported interest in the activity within the learning task. That is, the more participants were interested in the activity, the better their cognitive outcomes. Learning efficacy was moderately correlated with the change in positive affect from before to after the learning task, and also with the three other motivational outcomes, i.e., perceived competence, task attractivity and stimulation by the task. Learning efficiency was also positively associated with these outcomes, but to a lesser extent and significantly only in the case of attractivity. The change in positive affect was positively associated with all motivational outcomes. Interest was moderately correlated with the three motivational outcomes, which were highly correlated among each other. Change in negative affect was not significantly correlated with any of the other outcome measures, although we note a slight (but not significant) correlation with stimulation.

Table 2. Pairwise associations between all considered outcome measures. Changes in positive and negative affect from before to after the learning task are denoted as ΔPA and ΔNA, respectively. Percentage bend correlations, ρ_{pb}, are provided above the diagonal. Corresponding p-values are provided below the diagonal.

Outcome	1.	2.	3.	4.	5.	6.	7.	8.
1. Efficacy	1	0.85	0.35	-0.14	0.78	0.42	0.43	0.36
2. Efficiency	< 0.001	1	0.20	-0.14	0.72	0.23	0.27	0.21
3. ΔPA	0.006	0.121	1	0.06	0.38	0.44	0.38	0.35
4. ΔNA	0.272	0.276	0.673	1	-0.16	0.17	0.13	0.24
5. Interest	< 0.001	< 0.001	0.003	0.233	1	0.33	0.32	0.29
6. Competence	< 0.001	0.070	< 0.001	0.196	0.009	1	0.78	0.82
7. Attractivity	< 0.001	0.034	0.002	0.322	0.011	< 0.001	1	0.86
8. Stimulation	< 0.001	0.100	0.006	0.066	0.023	< 0.001	< 0.001	1

3.5.2 Motivation mediating effects of game elements on cognitive outcomes (hypothesis 4)

Tables 3 and 4 provide information on regression models considering motivational outcomes as mediators of the effect of the considered game elements on cognitive outcomes. The results for interest as a mediator suggest an effect of interest on cognitive outcomes in line with the above noted strong, positive association between those two variables (β ~ 0.7 and β ~ 0.5 for efficacy and efficiency in Tables 3 and 4, respectively). However, game elements do not significantly affect interest. Hence, interest is not a significant mediator of the cognitive effect of game elements neither regarding learning efficacy nor efficiency, although the data suggest a slightly negative association between game elements and cognitive outcomes via interest (β ~ -0.2). That is, descriptively, game elements reduce interest in the activity within the learning task, which in turn reduces cognitive outcomes. Hence, the direct effect of game elements (β ~ 0.2-0.3) is somewhat, although insignificantly larger than their total effect (β ~ 0.1).

In contrast, perceived competence, task attractivity, and stimulation by the task are significant mediators of the effect of game elements on learning efficacy, see Table 3. That is, the considered game elements are positively associated with the motivational outcomes (with β ~ 0.7-0.8), and the motivational outcomes are again moderately positively associated with learning efficacy (with β ~ 0.3-0.4). In line with hypothesis 4, this results in a mediation of the effect of game elements on learning efficacy by these motivational outcomes (β ~ 0.3-0.5). Furthermore, the data suggest a slightly negative, direct effect of game elements on cognitive outcomes. However, the regression coefficients regarding the direct paths are not significantly different from zero in any of the considered regression models. Hence, no decision can be made about the population coefficients of the direct paths being slightly positive or negative. In any case, the direct effect of game elements (β ~ -0.1/-0.2) is significantly reduced relative to their total effect (β ~ 0.1) by their indirect effect via these three motivational outcomes.

The regression coefficients quantifying the effect of competence, attractivity, and stimulation on learning efficiency are qualitatively similar to their effect on efficacy, but smaller (β ~ 0.2-0.3), see Table 4, and significantly different from zero only in the case of attractivity (β ~ 0.2). Also in the case of learning efficiency, the data suggest a slightly negative, yet insignificant direct effect of game elements on cognitive outcomes and a reduction of their total effect (β ~ 0.1) via the indirect pathway to a slightly negative, direct effect (β ~ -0.1).

Table 3. (Standardized) regression coefficients, β, 95%-confidence intervals and p-values obtained for the robust regression models considering mediation of the effect of game elements on learning efficacy via motivational outcomes. In the case of indirect effects, confidence intervals are computed via bootstrap according to Zu and Yuan [49], whereas in the case of all other paths, confidence intervals are approximated based on asymptotic normality.

Mediator	Type	Effect	β	*p*
Interest	Component	Condition → Interest	-0.19 [-0.80, 0.41]	0.524
		Interest → Efficacy	0.70 [0.55, 0.85]	< 0.001
	Indirect	Cond. → Interest → Efficacy	-0.19 [-0.61, 0.20]	0.334
	Direct	Condition → Efficacy	0.21 [-0.08, 0.51]	0.150
	Total	Condition → Efficacy	0.11 [-0.31, 0.53]	0.600
Competence	Component	Condition → Competence	0.81 [0.32, 1.29]	0.001
		Competence → Efficacy	0.43 [0.21, 0.65]	< 0.001
	Indirect	Cond. → Competence → Efficacy	0.39 [0.12, 0.73]	< 0.001
	Direct	Condition → Efficacy	-0.22 [-0.66, 0.22]	0.320
	Total	Condition → Efficacy	0.11 [-0.31, 0.53]	0.600
Attractivity	Component	Condition → Attractivity	0.73 [0.31, 1.15]	< 0.001
		Attractivity → Efficacy	0.44 [0.23, 0.65]	< 0.001
	Indirect	Cond. → Attractivity → Efficacy	0.45 [0.15, 0.85]	< 0.001
	Direct	Condition → Efficacy	-0.20 [-0.62, 0.22]	0.350
	Total	Condition → Efficacy	0.11 [-0.31, 0.53]	0.600
Stimulation	Component	Condition → Stimulation	0.78 [0.25, 1.31]	0.004
		Stimulation → Efficacy	0.30 [0.08, 0.53]	0.010
	Indirect	Cond. → Stimulation → Efficacy	0.26 [0.06, 0.60]	0.003
	Direct	Condition → Efficacy	-0.13 [-0.58, 0.33]	0.583
	Total	Condition → Efficacy	0.11 [-0.31, 0.53]	0.600

Table 4. (Standardized) regression coefficients, β, 95%-confidence intervals and p-values obtained for the robust regression models considering mediation of the effect of game elements on learning efficiency via motivational outcomes. In the case of indirect effects, confidence intervals are computed via bootstrap according to Zu and Yuan [49], whereas in the case of all other paths, confidence intervals are approximated based on asymptotic normality.

Mediator	Type	Effect	β	*p*
Interest	Component	Condition → Interest	-0.19 [-0.80, 0.41]	0.524
		Interest → Efficiency	0.53 [0.39, 0.67]	< 0.001
	Indirect	Cond. → Interest → Efficiency	-0.14 [-0.48, 0.15]	0.316
	Direct	Condition → Efficiency	0.27 [-0.02, 0.55]	0.063
	Total	Condition → Efficiency	0.10 [-0.34, 0.53]	0.659
Competence	Component	Condition → Competence	0.81 [0.32, 1.29]	0.001
		Competence → Efficiency	0.22 [-0.02, 0.45]	0.072
	Indirect	Cond. → Competence → Efficiency	0.18 [-0.01, 0.47]	0.072
	Direct	Condition → Efficiency	-0.08 [-0.55, 0.39]	0.738
	Total	Condition → Efficiency	0.10 [-0.34, 0.53]	0.659
Attractivity	Component	Condition → Attractivity	0.73 [0.31, 1.15]	< 0.001
		Attractivity → Efficiency	0.26 [0.04, 0.48]	0.020
	Indirect	Cond. → Attractivity → Efficiency	0.24 [0.03, 0.55]	0.021
	Direct	Condition → Efficiency	-0.12 [-0.56, 0.31]	0.572
	Total	Condition → Efficiency	0.10 [-0.34, 0.53]	0.659
Stimulation	Component	Condition → Stimulation	0.78 [0.25, 1.31]	0.004
		Stimulation → Efficiency	0.18 [-0.05, 0.41]	0.129
	Indirect	Cond. → Stimulation → Efficiency	0.12 [-0.03, 0.37]	0.141
	Direct	Condition → Efficiency	-0.02 [-0.48, 0.44]	0.940
	Total	Condition → Efficiency	0.10 [-0.34, 0.53]	0.659

3.5.3 Positive affect mediating effects of game elements on motivation (hypothesis 5)

In line with hypothesis 5, we find that all considered motivational outcomes are positively associated with game elements via an indirect path through the change in positive affect from before to after the task, see Table 5. In line with our correlational analysis in Section 3.5.1, the change in positive affect is positively associated with all four motivational outcomes (β ~ 0.3-0.4).

In the case of self-reported interest in the activity in the learning task, the data suggest overall a small negative, but insignificant association with game elements (β ~ -0.2), which decreases further (direct pathway; β ~ -0.4), but remains insignificant, if controlling for the indirect, positive association via the change in positive affect. That indicates that, in our sample, game elements may have to some extent undermined interest in the learning activity while simultaneously increasing interest by inducing positive affect.

This is further supported by the inspection of the corresponding effect sizes and confidence intervals. That is, point estimates and confidence intervals for the indirect effect are very similar as for the other three motivational outcomes, whereas point estimates and confidence intervals for the direct effects diverge between interest and the other three outcomes (i.e., confidence intervals do not overlap). That means, while we cannot exclude (with α = 0.05) a negligibly small positive direct effect of game elements on interest, we can conclude (with α = 0.05, comparison-wise) that direct effects are larger for the other motivational outcomes. Moreover, in the case of competence, attractivity, and stimulation, both direct and indirect pathways contribute significantly to an overall large, positive effect of game elements on these motivational outcomes (β ~ 0.7-0.8). While a significant, small portion stems from the indirect effect through positive affect (β ~ 0.2), the dominant portion remains in the direct effect (β ~ 0.6).

Table 5. (Standardized) regression coefficients, β, 95%-confidence intervals and p-values obtained for the robust regression models considering mediation of the effect of game elements on learning motivational outcomes via the change in positive affect over the course of the learning task (denoted in the table as ΔPA). In the case of indirect effects, confidence intervals are computed via bootstrap according to Zu and Yuan [49], whereas in the case of all other paths, confidence intervals are approximated based on asymptotic normality.

Outcome	Type	Effect	β	*p*
Interest	Component	Condition → ΔPA	0.46 [-0.06, 0.98]	0.080
		ΔPA → Interest	0.42 [0.16, 0.68]	0.002
	Indirect	Cond. → ΔPA → Interest	0.22 [0.01, 0.48]	0.037
	Direct	Condition → Interest	-0.42 [-0.94, 0.09]	0.104
	Total	Condition → Interest	-0.19 [-0.80, 0.41]	0.524
Competence	Component	Condition → ΔPA	0.46 [-0.06, 0.98]	0.080
		ΔPA → Competence	0.40 [0.16, 0.64]	0.001
	Indirect	Cond. → ΔPA → Competence	0.21 [0.01, 0.48]	0.037
	Direct	Condition → Competence	0.62 [0.14, 1.10]	0.013
	Total	Condition → Competence	0.81 [0.32, 1.29]	0.001
Attractivity	Component	Condition → ΔPA	0.46 [-0.06, 0.98]	0.080
		ΔPA → Attractivity	0.32 [0.11, 0.53]	0.004
	Indirect	Cond. → ΔPA → Attractivity	0.17 [0.01, 0.42]	0.038
	Direct	Condition → Attractivity	0.58 [0.16, 0.99]	0.008
	Total	Condition → Attractivity	0.73 [0.31, 1.15]	< 0.001
Stimulation	Component	Condition → ΔPA	0.46 [-0.06, 0.98]	0.080
		ΔPA → Stimulation	0.35 [0.10, 0.59]	0.006
	Indirect	Cond. → ΔPA → Stimulation	0.18 [0.01, 0.44]	0.036
	Direct	Condition → Stimulation	0.61 [0.13, 1.10]	0.014
	Total	Condition → Stimulation	0.78 [0.25, 1.31]	0.004

4. Discussion

4.1 Main results

Altogether, our results provide substantial empirical support for the theoretical considerations based on the ICALM [26]. That is, the obtained point estimates of effects size indeed yield negligible effect sizes in the case of cognitive outcomes ($\delta < 0.2$; hypothesis 1), a moderate effect in the case of the change of positive affect from before to after the task ($\delta \sim 0.5$; hypothesis 2), and large effect sizes in the case of motivational outcomes ($\delta \sim 0.8$; hypothesis 3), with the noteworthy exception of the subcomponent interest which will be discussed further below. While the ranges of effect sizes comparably compatible with our data (i.e., the confidence intervals) show that small to moderate effect sizes for cognitive outcomes in either direction (i.e., in favor of either the less or the more gameful condition) cannot be excluded, we can conclude (with $\alpha = 0.05$) that effects of positive affect and motivation are larger in the more gameful condition. We further obtained evidence that three of the four considered motivational outcomes are associated with increased cognitive outcomes (hypothesis 4) and all motivational outcomes are partially mediated by increased change in positive affect over the course of the learning task (hypothesis 5).

Thus, in line with our theoretical considerations, our results corroborate a beneficial effect of game elements on cognitive outcomes, which is mediated by their beneficial effect on motivation, which is, in turn, partially mediated by the beneficial effect of game elements on positive affect. To arrive then at overall similar cognitive outcomes in both task versions, game elements must be associated with a detrimental effect on cognitive outcomes at the same time, in line with the known additional demands they pose on cognitive processing [59].

A surprising aspect revealed by our mediation analysis is that although interest in the learning activity is clearly positively associated with cognitive learning outcomes and the change in positive affect is clearly positively related with interest, the relation between game elements and interest differs considerably from the relation between game elements and the other three motivational outcomes. While the direct effect (i.e., not mediated by induced positive affect) of game elements on competence, attractivity, or stimulation, is clearly positive (i.e., adding game elements is associated with increases in those motivational outcomes), the direct effect of game elements on interest is negligibly positive at best, and according to the estimate most compatible with our data, likely rather negative (i.e., interest decreases when adding game elements). This is unexpected also in comparison to literature suggesting definitely a facilitative link between the used game elements (narrative, visual aesthetics, incentive system) and motivation and engagement [60], [61].

A hint to resolve this apparently contradictory finding may lie in the exact wording of the self-report items used to assess different aspects of motivation. Whereas the items used to investigate the interest in the learning activity referred explicitly to the "activity *in* the learning task", the items used for the three other motivational outcomes were broadly referring to the appeal "of the task". However, the activity *within* the learning task can be conceived as the concrete mental activity of memorizing the given symbol-number pairs. The learning task as a whole certainly incorporates this mental activity, but, in addition, also the task's design features, the environment in which the symbol-number pairs are presented. That is, the less gameful learning task consists of the activity of memorizing some symbol-number pairs in a sensorially scarce environment, whereas the more gameful task consists of the *identical activity* but in a sensorially more stimulating environment. Or stating it more provocatively: the participants could probably differentiate between the chocolate (appealing design features) and the broccoli (concrete mental activity) in the more gameful task version, with the little intrinsic appeal of the learning activity eventually emphasized by its contrast to the environment in which it was presented [62]. And although being inherent to the task (as part of its design), the used game elements are extrinsic to the concrete mental activity within the task. However, by the (visual) appeal and stimulation they provide, they may yet facilitate internal regulation [23], eventually on the cost of intrinsic interest in the learning activity [63], [64], [65].

Nevertheless, and staying in the metaphor, the chocolate still did its job. Not only was motivation, taking into account all of its considered facets, positively associated with cognitive outcomes, but further was the interest in the learning activity fueled by enhanced positive affect just in the same amount as the other motivational outcomes.

To summarize, three aspects appear especially noteworthy in light of these findings:

a. Using self-report questionnaires, which allow to differentiate between core and contextual aspects of tasks, may to some extent allow to empirically illuminate and further disentangle the subtleties of the multi-faceted construct of motivation [66].
b. The differential impact of game elements on different aspects of motivation further underscores the well-substantiated notion that enhancing a learning activity by adding some game elements may miss out on substantial parts of the motivational power of game-based learning by missing the opportunity to redesign the learning activity into an intrinsically rewarding experience [6].
c. Nevertheless, and mostly relevant perhaps regarding the general state-of-the-art of game-based or gamified learning, our results also indicate that broccoli and chocolate may yet be preferred over broccoli alone by yielding overall higher scores in motivation and positive affect. Various meta-analyses also confirm that gamification typically increases motivation [5], [67], [68], [69]. For an early formation of motivation and hence, engagement with a learning task, an appealing and stimulating learning environment might just do well enough. The induced positive affect might suffice to “start up the positive feedback loop of internally rewarding learning experiences” [65].

4.2 Comparison with previous research

The importance of identifying factors for the formation of affect supporting motivation and the sustainment of engagement has been noted earlier [70]. Here, we found that game elements seem to facilitate especially the formation of positive affect, which, in turn, appear to act as one source of motivation. In contrast, negative affect was neither associated with positive affect nor with cognitive nor motivational outcomes. On the one hand, this resonates well with the more complex interplay between negative affect and learning processes, depending besides arousal characteristics (see, e.g., Ref. [25]) also on temporal aspects (see, e.g., [71]). On the other hand, our findings agree with the capability of game-based learning to elicit especially positive epistemic emotions being linked to higher levels of engagement [72], [73], [74]. Positive epistemic emotions like curiosity and enjoyment were also found to be associated with higher learner performance [75].

Earlier results further showed that high engagement in a learning game does not presuppose the learners’ interest in the topic of learning material [73] suggesting some motivational effect of game elements independent of prior interest in the learning topic. Also in the present study, game elements were associated with substantial motivational effects although being only loosely integrated with the learning activity in comparison with earlier studies on fraction estimation [11], [12], [76], employing the same game elements but interweaving the game mechanics and narrative neatly with the fraction estimation task (i.e., intrinsically integrated design; see, e.g., Ref. [77]). Yet the more noteworthy seem the effects of the employed game elements regarding affective and motivational outcomes in the present case.

Using the same experimental paradigm, but a slightly easier task (less symbol-number pairs), a previous online study [10] indicated that these affective and motivational effects of game elements translate into pragmatically important implications. That is, a significantly higher propensity of learners staying engaged with the learning task in contrast to an otherwise substantial number of learners simply disengaging from the task [10]. This further emphasizes on how important an initial sparkle of motivation, as potentially provided by such game elements, could turn out in practice.

4.3 Limitations

The main limitation of the present study is its sample size (n = 61). Although the substantial motivational impact of the considered game elements already allowed reasonable decisions about the existence of the hypothesized interrelations, a more precise estimation of their effect size requires larger samples. Furthermore, in the case of learning efficiency (i.e., the speed with which participants learned), no conclusions could be drawn at the population level regarding motivational mediation of cognitive effects of game elements except for the aspect of task attractivity. However, it is unreasonable to assume that the same interrelations are absent, which are clearly present in the case of efficacy, not the least due to the fact that both efficiency and efficacy are strongly associated with each other. In this case, it is more reasonable to assume that the effect on efficiency is a little weaker than the effect on efficacy, requiring higher statistical power to be resolved.

Apart from this limitation, the limited characterization of the assessed participant sample does not allow us to determine if and how our results, particularly positive and negative affect baselines, may depend on comfort or familiarity with (serious) games. Besides investigating these relations, future studies are also required to scrutinize how the effects of game elements are related to prior knowledge or existing expertise with memorization tasks by extending the study beyond the assessment of (mainly) university students. Although primarily owed to the intended high levels of experimental control, another limitation of our study could be seen in its focus on an associative learning task. From an educational perspective, the task relies mostly on the first level of Bloom's taxonomy (i.e., the category of knowledge/remembering) [78], [79] involving mainly cognitive functions like recalling or recognizing. Future studies will be required to investigate the effects of game elements on other (more complex) learning goals or cognitive domains beyond knowledge/remembering [78], [79].

Finally, our correlational analysis cannot confirm causal relations. Causation also implies temporal directionality and exclusion of alternative causation pathways. Regarding temporal directionality, our results cannot confirm that attractive design elements cause a positive, affective response, which, in turn, leads to enhanced motivation. Illuminating this requires digging deeper into the microstructure of the learning process. One way (among others) to approach this would be to temporally resolve the development of affective dynamics by, for instance, leveraging the potential of multimodal assessment of physiological correlates in future studies besides longitudinal, repeated sampling of situational interest, motivation, and emotion.

5. Conclusions

The results of our value-added online experiment corroborate the theoretical implication that game elements can have simultaneously antagonistic effects. They can improve cognitive outcomes via motivation, while, simultaneously, they pose additional cognitive demands which may reduce cognitive outcomes. They further can improve motivation via positive affect. At the same time, they may reduce to some extent the intrinsic interest in the activity within the learning task. These findings have several implications for future research. First, they further corroborate the tenability of the ICALM highlighting its potential for the design of tailored learning systems taking into account not only cognitive, affective, and motivational processes but also their manifold and sometimes antagonistic interrelations. Second, they show how relatively simple and easy-to-apply self-report instruments and their analyses can already allow to illuminate and disentangle some of those interrelations. Third, they suggest the formation of positive affect as one source of motivational processes which upon reliable identification could be utilized in adaptive learning systems allowing a dynamical support of learners. In such a framework, positive affect induced by an appealing and stimulating learning environment could be seen as an affective scaffold, supporting an early formation of motivation and hence, engagement with the learning material.

Acknowledgments

S.E.H. cordially thanks Hanna Weber for the numerous insightful and instructive discussions on all kinds of statistical matters, but in the scope of this work especially on the importance of methodologically proper specification of confidence intervals and the various measures to estimate internal consistency. S.E.H. and M.N. thank Mea-Maria Leinonen for her committed support during data acquisition. The authors acknowledge financial support by the University of Graz. This work was also supported by the Strategic Research Council (SRG) established within the Research Council of Finland (Grants: 335625, 358250).

Conflicts of 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.

References

[1] R. M. Ryan and C. S. Rigby, "Motivational foundations of game-based learning," in *Handbook of game-based learning*, J. L. Plass, R. E. Mayer, and B. D. Homer, Eds., MIT Press, 2020, pp. 153–176.

[2] N. Barz, M. Benick, L. Dörrenbächer-Ulrich, and F. Perels, "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*, p. 003465432311677, May 2023, doi:10.3102/00346543231167795

[3] Y. Hu, T. Gallagher, P. Wouters, M. Van Der Schaaf, and L. Kester, "Game-based learning has good chemistry with chemistry education: A three-level meta-analysis," *J Res Sci Teach*, vol. 59, no. 9, pp. 1499–1543, Nov. 2022, doi:10.1002/tea.21765

[4] J. Looyestyn, J. Kernot, K. Boshoff, J. Ryan, S. Edney, and C. Maher, "Does gamification increase engagement with online programs? A systematic review," *PLoS ONE*, vol. 12, no. 3, p. e0173403, Mar. 2017, doi:10.1371/journal.pone.0173403

[5] M. Sailer and L. Homner, "The Gamification of Learning: a Meta-analysis," *Educ Psychol Rev*, vol. 32, no. 1, pp. 77–112, Mar. 2020, doi:10.1007/s10648-019-09498-w

[6] J. L. Plass, B. D. Homer, R. E. Mayer, and C. K. Kinzer, "Theoretical foundations of game-based and playful learning," in *Handbook of game-based learning*, J. L. Plass, R. E. Mayer, and B. D. Homer, Eds., MIT Press, 2020, pp. 3–24.

[7] Q. Zhang and Z. Yu, "Meta-Analysis on Investigating and Comparing the Effects on Learning Achievement and Motivation for Gamification and Game-Based Learning," *Education Research International*, vol. 2022, pp. 1–19, Aug. 2022, doi:10.1155/2022/1519880

[8] K. Loderer, R. Pekrun, and J. L. Plass, "Emotional foundations of game-based learning," in *Handbook of game-based learning*, J. L. Plass, R. E. Mayer, and B. D. Homer, Eds., MIT Press, 2020, pp. 111–151.

[9] K. Bernecker and M. Ninaus, “No Pain, no Gain? Investigating motivational mechanisms of game elements in cognitive tasks,” *Computers in Human Behavior*, vol. 114, p. 106542, Jan. 2021, doi:10.1016/j.chb.2020.106542
[10] S. E. Huber, R. Cortez, K. Kiili, A. Lindstedt, and M. Ninaus, “Game elements enhance engagement and mitigate attrition in online learning tasks,” *Computers in Human Behavior*, vol. 149, p. 107948, Dec. 2023, doi:10.1016/j.chb.2023.107948
[11] M. De Nicolò *et al.*, “Engaging learners with games–Insights from functional near-infrared spectroscopy,” *PLoS ONE*, vol. 18, no. 6, p. e0286450, Jun. 2023, doi:10.1371/journal.pone.0286450
[12] M. Ninaus *et al.*, “The added value of game elements: better training performance but comparable learning gains,” *Education Tech Research Dev*, vol. 71, no. 5, pp. 1917–1939, Oct. 2023, doi:10.1007/s11423-023-10263-8
[13] M. Ninaus *et al.*, “Increased emotional engagement in game-based learning – A machine learning approach on facial emotion detection data,” *Computers & Education*, vol. 142, p. 103641, Dec. 2019, doi:10.1016/j.compedu.2019.103641
[14] R. N. Schwartz and J. L. Plass, “Types of engagement in learning with games,” in *Handbook of game-based learning*, J. L. Plass, R. E. Mayer, and B. D. Homer, Eds., MIT Press, 2020, pp. 53–80.
[15] B. M. McLaren, D. M. Adams, R. E. Mayer, and J. Forlizzi, “A Computer-Based Game that Promotes Mathematics Learning More than a Conventional Approach:,” *International Journal of Game-Based Learning*, vol. 7, no. 1, pp. 36–56, Jan. 2017, doi:10.4018/IJGBL.2017010103
[16] X. Huang and R. E. Mayer, “Benefits of adding anxiety-reducing features to a computer-based multimedia lesson on statistics,” *Computers in Human Behavior*, vol. 63, pp. 293–303, Oct. 2016, doi:10.1016/j.chb.2016.05.034
[17] J. L. Plass, S. Heidig, E. O. Hayward, B. D. Homer, and E. Um, “Emotional design in multimedia learning: Effects of shape and color on affect and learning,” *Learning and Instruction*, vol. 29, pp. 128–140, Feb. 2014, doi:10.1016/j.learninstruc.2013.02.006
[18] E. “Rachel” Um, J. L. Plass, E. O. Hayward, and B. D. Homer, “Emotional design in multimedia learning.,” *Journal of Educational Psychology*, vol. 104, no. 2, pp. 485–498, May 2012, doi:10.1037/a0026609
[19] R. E. Mayer and G. Estrella, “Benefits of emotional design in multimedia instruction,” *Learning and Instruction*, vol. 33, pp. 12–18, Oct. 2014, doi:10.1016/j.learninstruc.2014.02.004
[20] R. M. Wong and O. O. Adesope, “Meta-Analysis of Emotional Designs in Multimedia Learning: A Replication and Extension Study,” *Educ Psychol Rev*, vol. 33, no. 2, pp. 357–385, Jun. 2021, doi:10.1007/s10648-020-09545-x
[21] R. E. Mayer, “Cognitive Theory of Multimedia Learning,” in *The Cambridge Handbook of Multimedia Learning*, 2nd ed., R. E. Mayer, Ed., Cambridge University Press, 2014, pp. 43–71. doi:10.1017/CBO9781139547369.005
[22] R. E. Mayer, “Cognitive foundations of game-based learning,” in *Handbook of game-based learning*, J. L. Plass, R. E. Mayer, and B. D. Homer, Eds., MIT Press, 2020, pp. 83–110.
[23] R. M. Ryan and E. L. Deci, “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions,” *Contemporary Educational Psychology*, vol. 25, no. 1, pp. 54–67, Jan. 2000, doi:10.1006/ceps.1999.1020
[24] “Intrinsic Motivation Inventory (IMI) – selfdeterminationtheory.org.” Accessed: Apr. 03, 2024. [Online]. Available: https://selfdeterminationtheory.org/intrinsic-motivation-inventory/
[25] R. Pekrun, “The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice,” *Educ Psychol Rev*, vol. 18, no. 4, pp. 315–341, Nov. 2006, doi:10.1007/s10648-006-9029-9
[26] J. L. Plass and U. Kaplan, “Emotional design in digital media for learning,” in *Emotions, Technology, Design, and Learning*, S. Y. Tettegah and M. Gartmeier, Eds., Academic Press, 2016, pp. 131–161. doi:10.1016/B978-0-12-801856-9.00007-4
[27] J. A. Russell, “Core affect and the psychological construction of emotion.,” *Psychological Review*, vol. 110, no. 1, pp. 145–172, 2003, doi:10.1037/0033-295X.110.1.145
[28] C. E. Izard, “Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues,” *Annu. Rev. Psychol.*, vol. 60, no. 1, pp. 1–25, Jan. 2009, doi:10.1146/annurev.psych.60.110707.163539

[29] R. Moreno and R. Mayer, "Interactive Multimodal Learning Environments: Special Issue on Interactive Learning Environments: Contemporary Issues and Trends," *Educ Psychol Rev*, vol. 19, no. 3, pp. 309–326, Sep. 2007, doi:10.1007/s10648-007-9047-2
[30] J. Sweller, "Cognitive Load Theory," in *Psychology of Learning and Motivation*, vol. 55, Elsevier, 2011, pp. 37–76. doi:10.1016/B978-0-12-387691-1.00002-8
[31] G. D. Rey, "A review of research and a meta-analysis of the seductive detail effect," *Educational Research Review*, vol. 7, no. 3, pp. 216–237, Dec. 2012, doi:10.1016/j.edurev.2012.05.003
[32] L. Bender, A. Renkl, and A. Eitel, "When and how seductive details harm learning. A study using cued retrospective reporting," *Applied Cognitive Psychology*, vol. 35, no. 4, pp. 948–959, Jul. 2021, doi:10.1002/acp.3822
[33] A. Koskinen, J. McMullen, M. Ninaus, and K. Kiili, "Does the emotional design of scaffolds enhance learning and motivational outcomes in game-based learning?," *Computer Assisted Learning*, vol. 39, no. 1, pp. 77–93, Feb. 2023, doi:10.1111/jcal.12728
[34] N. Leibowitz, B. Baum, G. Enden, and A. Karniel, "The exponential learning equation as a function of successful trials results in sigmoid performance," *Journal of Mathematical Psychology*, vol. 54, no. 3, pp. 338–340, Jun. 2010, doi:10.1016/j.jmp.2010.01.006
[35] R Core Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, 2022. [Online]. Available: https://www.R-project.org
[36] B. Breyer and M. Bluemke, "Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel)," *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*, 2016, doi:10.6102/ZIS242
[37] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales.," *Journal of Personality and Social Psychology*, vol. 54, no. 6, pp. 1063–1070, 1988, doi:10.1037/0022-3514.54.6.1063
[38] A. F. Hayes and J. J. Coutts, "Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But…," *Communication Methods and Measures*, vol. 14, no. 1, pp. 1–24, Jan. 2020, doi:10.1080/19312458.2020.1718629
[39] M. Wilde, K. Bätz, A. Kovaleva, and D. Urhahne, "Überprüfung einer Kurzskala intrinsicher Motivation (KIM)," *Zeitschrift für Didaktik der Naturwissenschaften*, vol. 15, pp. 31–45, 2009.
[40] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," in *HCI and Usability for Education and Work*, vol. 5298, A. Holzinger, Ed., in Lecture Notes in Computer Science, vol. 5298. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 63–76. doi:10.1007/978-3-540-89350-9_6
[41] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Die UX KPI - Wunsch und Wirklichkeit," 2017, doi:10.18420/MUC2017-UP-0100
[42] R. R. Wilcox, *Understanding and applying basic statistical methods using R*. Wiley, 2016.
[43] R. R. Wilcox, *Introduction to robust estimation and hypothesis testing*, 5th ed. Academic Press, 2022. doi:10.1016/B978-0-12-820098-8.00007-5
[44] P. Mair and R. Wilcox, "Robust statistical methods in R using the WRS2 package," *Behav Res*, vol. 52, no. 2, pp. 464–488, Apr. 2020, doi:10.3758/s13428-019-01246-w
[45] L. V. Jones and J. W. Tukey, "A sensible formulation of the significance test.," *Psychological Methods*, vol. 5, no. 4, pp. 411–414, 2000, doi: 10.1037/1082-989X.5.4.411.
[46] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Academic Press, 1988.
[47] J. Algina, H. J. Keselman, and R. D. Penfield, "An Alternative to Cohen's Standardized Mean Difference Effect Size: A Robust Parameter and Confidence Interval in the Two Independent Groups Case.," *Psychological Methods*, vol. 10, no. 3, pp. 317–328, Sep. 2005, doi:10.1037/1082-989X.10.3.317
[48] V. Amrhein, S. Greenland, and B. McShane, "Scientists rise up against statistical significance," *Nature*, vol. 567, no. 7748, pp. 305–307, Mar. 2019, doi:10.1038/d41586-019-00857-9
[49] J. Zu and K.-H. Yuan, "Local Influence and Robust Procedures for Mediation Analysis," *Multivariate Behavioral Research*, vol. 45, no. 1, pp. 1–44, Jan. 2010, doi:10.1080/00273170903504695
[50] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. Springer, 2002. doi:10.1007/978-0-387-21706-2
[51] RStudio Team, "RStudio. Boston, MA: Integrated Development Environment for R." RStudio, PBC, 2022. [Online]. Available: http://www.rstudio.com/

[52] H. Wickham *et al.*, "Welcome to the Tidyverse," *JOSS*, vol. 4, no. 43, p. 1686, Nov. 2019, doi:10.21105/joss.01686
[53] D. Temple Lang, "_XML: Tools for Parsing and Generating XML Within R and S-Plus_. R package version 3.99-0.13." 2022. [Online]. Available: https://CRAN.R-project.org/package=XML
[54] A. Kassambara, "Ggpubr: 'ggplot2' based Publication Ready Plots. R package version 0.6.0." 2023. [Online]. Available: https://CRAN.R-project.org/package=ggpubr
[55] A. Kassambara, "rstatix: Pipe-Friendly framework for basic statistical tests." 2023. [Online]. Available: https://CRAN.R-project.org/package=rstatix
[56] D. Rizopoulos, "**ltm** : An *R* Package for Latent Variable Modeling and Item Response Theory Analyses," *J. Stat. Soft.*, vol. 17, no. 5, 2006, doi:10.18637/jss.v017.i05
[57] S. Champely, "_pwr: Basic Functions for Power Analysis_. R package version 1.3-0." 2020. [Online]. Available: https://CRAN.R-project.org/package=pwr
[58] K. Kelley, "_MBESS: The MBESS R package_. R package version 4.9.3." 2023. [Online]. Available: https://CRAN.R-project.org/package=MBESS
[59] B. Park, T. Flowerday, and R. Brünken, "Cognitive and affective effects of seductive details in multimedia learning," *Computers in Human Behavior*, vol. 44, pp. 267–278, Mar. 2015, doi:10.1016/j.chb.2014.10.061
[60] W. L. Bedwell, D. Pavlas, K. Heyne, E. H. Lazzara, and E. Salas, "Toward a Taxonomy Linking Game Attributes to Learning: An Empirical Study," *Simulation & Gaming*, vol. 43, no. 6, pp. 729–760, Dec. 2012, doi:10.1177/1046878112439444
[61] A. M. Toda *et al.*, "A Taxonomy of Game Elements for Gamification in Educational Contexts: Proposal and Evaluation," in *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, Maceió, Brazil: IEEE, Jul. 2019, pp. 84–88. doi:10.1109/ICALT.2019.00028
[62] R. R. Hunt, "The subtlety of distinctiveness: What von Restorff really did," *Psychonomic Bulletin & Review*, vol. 2, no. 1, pp. 105–112, Mar. 1995, doi:10.3758/BF03214414
[63] W. S. Grolnick, "Extrinsic Incentives/Rewards: Short-Term Fix That Can Undermine Long-Term Motivation," in *Motivation Science*, 1st ed., M. Bong, J. Reeve, and S. Kim, Eds., Oxford University PressNew York, 2023, pp. 154-C4.2T1. doi:10.1093/oso/9780197662359.003.0026
[64] S. E. Hidi and K. Ann Renninger, "Interest and Its Relation to Rewards, Reward Expectations, and Incentives," in *Motivation Science*, 1st ed., M. Bong, J. Reeve, and S. Kim, Eds., Oxford University PressNew York, 2023, pp. 159–164. doi:10.1093/oso/9780197662359.003.0027
[65] L. Bardach and K. Murayama, "The Role of Rewards in Motivation--Beyond dichotomies." Nov. 28, 2023. doi:10.31219/osf.io/hgbw4
[66] M. Bong, J. Reeve, and S. Kim, Eds., *Motivation Science: Controversies and Insights*, 1st ed. Oxford University PressNew York, 2023. doi:10.1093/oso/9780197662359.001.0001
[67] L. Li, K. F. Hew, and J. Du, "Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: a meta-analysis and systematic review," *Education Tech Research Dev*, Jan. 2024, doi:10.1007/s11423-023-10337-7
[68] J. Mula-Falcón, I. Moya-Roselló, and A. Ruiz-Ariza, "The Active Methodology of Gamification to Improve Motivation and Academic Performance in Educational Context: A Meta-Analysis," *RES*, vol. 14, no. 2, p. 32, Apr. 2022, doi:10.5539/res.v14n2p32
[69] Q. Zhang, L. Yu, and Z. Yu, "A Content Analysis and Meta-Analysis on the Effects of Classcraft on Gamification Learning Experiences in terms of Learning Achievement and Motivation," *Education Research International*, vol. 2021, pp. 1–21, Nov. 2021, doi:10.1155/2021/9429112
[70] A. Hamrouni and F. Bendella, "A taxonomy of learner-player's emotions in serious games," *IJSG*, vol. 10, no. 4, pp. 17–32, Nov. 2023, doi:10.17083/ijsg.v10i4.637
[71] E. B. Cloude, D. A. Dever, D. L. Hahs-Vaughn, A. J. Emerson, R. Azevedo, and J. Lester, "Affective Dynamics and Cognition During Game-Based Learning," *IEEE Trans. Affective Comput.*, vol. 13, no. 4, pp. 1705–1717, Oct. 2022, doi:10.1109/TAFFC.2022.3210755
[72] K. Kiili, J. Siuko, E. Cloude, and M. Dindar, "Motivation and Emotions in a Health Literacy Game: Insights from Co-occurrence Network Analysis," in *Games and Learning Alliance*, vol. 13647, K. Kiili, K. Antti, F. De Rosa, M. Dindar, M. Kickmeier-Rust, and F. Bellotti, Eds., in Lecture Notes in Computer Science, vol. 13647. , Cham: Springer International Publishing, 2022, pp. 149–159. doi:10.1007/978-3-031-22124-8_15

[73] K. Kiili, S. Juho, E. Cloude, and M. Dindar, “Demystifying the Relations of Motivation and Emotions in Game-Based Learning: Insights from Co-Occurrence Network Analysis,” *IJSG*, vol. 10, no. 4, pp. 93–112, Nov. 2023, doi:10.17083/ijsg.v10i4.629

[74] J. Siuko, K. Kiili, L. Paakkari, E. Cloude, and M. Dindar, “Health Literacy Game vs. Textbook: Gameplay Boosts Motivation and Emotional Engagement.” Jan. 05, 2023. doi:10.31219/osf.io/gef6h

[75] J. Zheng, S. P. Lajoie, S. Li, and H. Wu, “Temporal change of emotions: Identifying academic emotion trajectories and profiles in problem-solving,” *Metacognition Learning*, vol. 18, no. 2, pp. 315–345, Aug. 2023, doi:10.1007/s11409-022-09330-x

[76] K. Kiili, K. Moeller, and M. Ninaus, “Evaluating the effectiveness of a game-based rational number training - In-game metrics as learning indicators,” *Computers & Education*, vol. 120, pp. 13–28, May 2018, doi:10.1016/j.compedu.2018.01.012

[77] K. Kiili, A. Koskinen, and M. Ninaus, “Intrinsic integration in rational number games – A systematic literature review,” in *GamiFIN 2019: Proceedings of the 3rd International GamiFIN Conference*, CEUR-WS, 2019, pp. 35–46. [Online]. Available: http://urn.fi/urn:nbn:de:0074-2359-5

[78] B. S. Bloom, “Taxonomy of educational objectives: the classification of educational goals,” in *Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive domain*, M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, Eds., New York: David McKay.

[79] D. R. Krathwohl, “A Revision of Bloom’s Taxonomy: An Overview,” *Theory Into Practice*, vol. 41, no. 4, pp. 212–218, Nov. 2002, doi:10.1207/s15430421tip4104_2