

Addressing boundary conditions of cognitive and motivational effects of gamified learning

Stefan Huber¹, Elizabeth Cloude², Lukas Ober¹, Moritz Edlinger¹, Antero Lindstedt³, Kristian Kiili³ & Manuel Ninaus^{1,4}

¹ University of Graz, Austria

² Michigan State University, USA

³ Tampere University, Finland

⁴ LEAD Graduate School and Research Network, University of Tübingen, Germany

Article received 24 December 2025 / Article revised 30 May 2025 / Accepted 19 August 2025/
Available online 12 September 2025

Abstract

There is a growing interest in developing gamified learning solutions to address educational challenges. However, learning is highly influenced by the conditions in which it takes place (e.g., does gamified learning in a laboratory setting replicate the outcomes of gamified learning online at home?). Hence, it is crucial to understand the boundary conditions of different learning contexts to effectively implement gamified interventions that provide optimal learner support. This work contributes to such an understanding by assessing how general contextual aspects of three studies on gamified learning influence cognitive learning and motivational outcomes. Therefore, we re-examined the results of two earlier published online studies (Study 1: n=285; Study 2: n=61) and compared the results to a recently conducted laboratory study (Study 3: n=121), all of which employed the same associative learning task. Comparing results through a Bayesian lens, we find that motivational outcomes induced by gamification differ substantially between contexts. In contrast, cognitive learning outcomes seem comparatively robust across different contextual factors, with some indication of subtle influences in agreement with cognitive learning theories. Implications are discussed for future empirical research on learning, highlighting how a better understanding of boundary conditions of gamified learning interventions could open perspectives for context-aware educational interventions.

Keywords: gamification, game-based learning, context, boundary conditions, cognition, motivation.



1. Introduction

Research on game-based pedagogies has repeatedly shown that the common factor across various forms of these pedagogies – namely, game elements – can effectively influence learning (Barz et al., 2023; Sailer & Homner, 2020; Wouters et al., 2013; Zainuddin et al., 2020). However, studies find that the efficacy of game elements can substantially differ, depending on which learning outcome is under investigation. That is, game elements might have positive effects on cognitive learning outcomes, but not on motivational or behavioral outcomes (Sailer & Homner, 2020). Under different conditions, game elements may have positive effects on motivational outcomes yet leave cognitive outcomes unaffected (Huber et al., 2023, 2024). Furthermore, these effects can depend substantially on the exact contextual features of a learning situation (Schlag et al., 2024). In particular, the efficacy of game elements may depend on education level (Arztmann et al., 2023; Sailer & Homner, 2020), gender (Conte, 2019; Mavridis et al., 2017; L. Zhang et al., 2024), subject area (Ritzhaupt et al., 2021), research context (Huang et al., 2020; Sailer & Homner, 2020), or sampled populations (Ritzhaupt et al., 2021). These findings suggest that the efficacy of game elements on learning depends on specific *boundary conditions* under which game elements become especially effective for particular learning outcomes.

With the term “boundary conditions” we refer to Mayer’s (2024) notion that each multimedia learning principle “is subject to boundary conditions including for whom the principle applies, for which kind of lesson the principle applies, and under what circumstances the principle applies” (p. 19). That is, Mayer recognizes that the efficacy of each multimedia learning principle depends on several factors: learner characteristics (“for whom the principle applies”), contextual aspects in a narrower sense (“which kind of lesson”, i.e., which subject), and in a broader sense (“under what circumstances”), the environment (e.g., online vs. classroom), or the intention (e.g., related to grading or on a voluntary basis) in which learning occurs.

In the present study, we build upon this definition of boundary conditions and the opportunity that we recently used the same associative learning task in several experiments on gamified learning under varying contextual conditions (e.g., conducting the experiments online or in the laboratory). To compare the outcomes of those experiments, we conducted secondary analyses of the findings from two previously published online studies (Huber et al., 2023, 2024), in conjunction with the results of a recent laboratory study hence allowing an assessment of how contextual features, not inherent to the learning task per se, shape the efficacy of gamified learning. As both cognitive learning and motivational outcomes were investigated in all three juxtaposed studies, the comparison yields a perspective regarding the boundary conditions of gamified learning on two learning outcomes. The implications of the findings may provide guidance for future research dedicated to further resolve the exact impact of specific contextual features on gamified learning outcomes.

Gamified learning is “the use of game design elements in non-game contexts” (Deterding et al., 2011, p. 9). We focus on this type of game-based pedagogy because game design elements are one essential and common factor among various game-based pedagogies, such as game-based learning, gamification, or playful learning (Plass et al., 2020). The focus on the efficacy of game elements on learning outcomes, as in the re-assessed previous work (Huber et al., 2023, 2024), may warrant some generalizability across game-based pedagogies.

Whereas we use the term game-based pedagogy here rather as an umbrella term subsuming game-based, gamified, and playful learning under one term, it generally refers to a teaching approach that integrates games or game design elements into the learning processes and practices to promote motivation, learning, and assessment. As such, the term includes many aspects that go beyond the mere use of game elements like teachers’ roles, integration with the curriculum, educational objectives, or classroom practices (Palha & Jukić Matic, 2025). While not considered in the present work, these aspects might reveal further boundary conditions in practice.



The remainder of the introduction section is organized as follows. In Section 1.1, we provide a theoretical account for why boundary conditions are important in game-based pedagogies. In Section 1.2, we provide an overview of related empirical work, suggesting the importance of considering boundary conditions within the framework of game-based pedagogies. In Section 1.3, we summarize the specific aims and research questions of the present study and provide short summaries of two previously published online studies (Huber et al., 2023, 2024) as we conducted secondary analyses of the data in the present work.

1.1 Theoretical key concepts

In psychology, the perspective that a person's behavior (including learning) results from the interplay between endogenous processes originating from within the person and exogenous characteristics and processes determined by the person's environment is not new. Already in 1936, Kurt Lewin coined the famous quasi-formula that behavior is a function of a person and their environment (Lewin, 1936). That is, according to Lewin (1946), human behavior needs to be regarded "as a function of the total situation" (p. 791), in which persons and their contexts constitute a unified whole (Plass & Kaplan, 2016). As a fundamental key concept, this notion is at the core of many theoretical frameworks in learning research and also applies to game-based pedagogies.

Although game-based pedagogies are a relatively recent addition to multimedia learning, Mayer's (2019) cognitive model of multimedia learning provides a useful starting point to describe their conceptual foundation. At the core of Mayer's model are three basic principles (Mayer, 2011, 2014): (i) the *dual channel principle*, according to which people have separate channels for processing visual and verbal material; (ii) the *limited capacity principle*, which states that people can only process a small amount of material in each channel at one time; and (iii) the *active processing principle*, according to which deep learning occurs when people engage in active cognitive processing during multimedia learning.

According to Mayer (2019), these core principles need to be extended beyond cognition to describe game-based pedagogies with two further elements: motivation and meta-cognition. A modern framework integrating both of these elements with Mayer's cognitive model is provided by Plass and Kaplan (2016): The integrated cognitive affective model of learning with multimedia (ICALM). A simplified version of the ICALM is depicted in Fig. 1(a). An essential ingredient of this model is that learners make meaning of multimedia material via a continuous, dynamic interaction between cognition and affect. The latter is rooted in environmental, contextual factors of the situation, and moderated by learner characteristics (e.g., meta-cognitive skills, motivation) and predispositions (e.g., expectations brought into the situation by the learner). We added the dashed arrows in Fig. 1(a) to highlight that learning is not regarded as a unidirectional cause-and-effect relation. Instead, it is viewed as a closed-loop process in which learners constantly engage in "meaning-making" (Schwartz & Plass, 2020).

Thus, in the ICALM, the high sensitivity towards specific contextual factors influencing learning is an essential feature. The ICALM is built upon the notion that learning occurs in specific contexts, driven by dynamic cognition-emotion interactions. These interactions themselves "emerge and operate in ways that are highly contextualized (and hence sensitive to contextual factors)" (Schwartz & Plass, 2020, p. 149).

Learning emerges from the interactions among cognition and affect, further providing a conceptual link to learning engagement and motivation. This is especially important as game-based pedagogies are frequently employed for their capabilities in fostering sustained engagement and motivation in learners (Greipl et al., 2020; Ryan & Rigby, 2020). As outlined by Urhahn and Wijnia (2023), many, if not all, motivational theories share the same basic motivational model depicted in Fig. 1(b), based on the general model of motivation by Heckhausen and Heckhausen (2018). Similar to the ICALM (Plass & Kaplan, 2016), the model emphasizes that motivated behavior emerges from

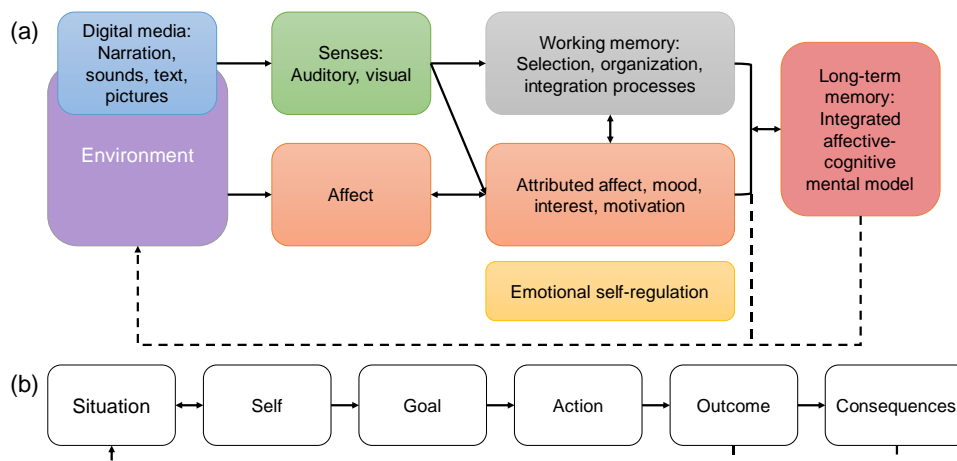


interactions between learners and the situation in which they find themselves. All subsequent concepts to describe motivated behavior [goals, actions, outcomes, consequences in Fig. 1(b)] emerge from this interaction and are evaluated in relation to it. Motivational theories like expectancy-value theory (Eccles & Wigfield, 2020), self-efficacy theory (Bandura, 1997), or self-determination theory (Ryan & Deci, 2019) then differ in how they conceptually fill, relate, and weigh the different parts of the motivation-action cycle against each other. However, they share their roots in this interaction between learners and the learning situation.

In summary, both cognition and motivation should be sensitive to the exact conditions experienced during learning. As recognized by Mayer (2024) in his implicit definition of boundary conditions quoted above, this has important implications for game-based pedagogies, as well as multimedia learning principles in general, which are expected to be sensitive to those exact conditions. This is suggested also by a plethora of empirical research as outlined next.

Figure 1

Outline of typical cognitive and motivational models



Note. (a) Simplified illustration of the Integrated Cognitive Affective Model of Learning with Multimedia (ICALM) by Plass and Kaplan (2016). (b) Basic motivational model proposed by Urhahne and Wijnia (2023). The dashed arrows were added by us in both illustrations, with the rationale for them explained in the text.

1.2 Related empirical work

While a variety of meta-analyses converge on the conclusion that game-based pedagogies can positively affect both cognitive and motivational learning outcomes, they also typically yield a high heterogeneity of effect sizes. Point estimates for cognitive learning outcomes range from medium to large effect sizes, specifically Hedge's $g = 0.46-0.85$ (Alotaibi, 2024; Arzmann et al., 2023; Bai et al., 2020; Hu et al., 2022; Huang et al., 2020; Sailer & Homner, 2020; Q. Zhang & Yu, 2022). Substantial heterogeneity between individual studies is frequently noted (e.g., Hu et al., 2022, $I^2 = 86%$; Huang et al., 2020, $I^2 = 88%$). Reported point estimates for motivational outcomes result in an even wider range of small to large effect sizes, specifically Hedge's $g = 0.25-0.92$ (Alotaibi, 2024; Arzmann et al., 2023; Fadda et al., 2022; Hu et al., 2022; Li et al., 2024; Sailer & Homner, 2020; Q. Zhang & Yu, 2022) with, again, substantial heterogeneity between individual studies, $I^2 = 83%$ (Arzmann et al., 2023; Li et al., 2024).



Effect sizes in the scope of motivation can further vary substantially depending on which aspect or construct of motivation is under investigation. While Li et al. (2024) obtained an overall effect size of $g = 0.26$ for intrinsic motivation, they obtain a large variation in effect sizes for perceptions of autonomy ($g = 0.64$, 95%-CI [0.14, 1.14]), relatedness (1.78, 95% CI [0.74, 2.81]), and competence (0.28, 95%-CI [0.00, 0.55]). The sensitivity of effect size to the exact subfactor of motivation is supported further by Fadda et al. (2022), providing additional evidence that the specific type of motivational outcome moderates the obtained effect size.

Moderator analyses provide some indications about which underlying factors are driving the heterogeneity in effect sizes for both types of learning outcomes. At large, the suggested factors can be grouped into person factors and contextual factors. Among person factors, education level (and hence, indirectly related, age) was repeatedly identified as a factor moderating learning outcomes (Arztmann et al., 2023; Sailer & Homner, 2020). While gender was not a significant moderator in a meta-analysis (Arztmann et al., 2023), the authors caution against over-generalization of that result given the limited number of studies in their meta-analysis. In fact, some empirical indications exist suggesting gender differences in game-based learning outcomes (Conte, 2019; Mavridis et al., 2017; Rodrigues et al., 2022; L. Zhang et al., 2024).

On the other hand, evidence has accumulated regarding the role of *contextual factors* on learning and its outcomes within the framework of game-based pedagogies. One such contextual factor is that the mentioned meta-analyses represent a mix of investigations into different game-based pedagogies, such as gamification, game-based learning, or serious games. Conceptual differences exist between those pedagogies (see, e.g., Deterding et al., 2011; Plass et al., 2020), potentially underlying some of the obtained empirical differences like gamification affecting extrinsic rather than intrinsic motivation, whereas the opposite holds true for game-based learning (Q. Zhang & Yu, 2022).

Another important contextual factor is subject area (Ritzhaupt et al., 2021). This is also endorsed by research in the strongly related field of learning analytics, and especially in predictive modeling of academic achievement (Alyahyan & Düşteğör, 2020; Prenkaj et al., 2021), where it has been found that predictive models produce substantially different results across different course topics (Conijn et al., 2017). That is, the transferability of predictive modeling was found to be low, which means that a model, for instance, capable of accurately predicting grades in a mathematics course could perform poorly in a language course and vice versa. Further, it was found that generalized models performed worse than course specific models.

Recent research also showed that different course-specific learning variables (e.g., learning design, student sample composition, or class size) could affect the prediction of outcomes differently (Xu et al., 2024). This is in line with the notion that learning outcomes in educational research might depend considerably on the sampled population, which is often comprised of undergraduate university students (Ritzhaupt et al., 2021).

Finally, the research context in which learning is studied has been noted as an influential factor (Huang et al., 2020; Sailer & Homner, 2020). Whereas Huang et al. (2020) found that learning outcomes reported in dissertations and theses differed significantly from outcomes reported in journal articles and conference proceedings, Sailer and Homner (2020) noted significant differences in outcomes between quasi-experimental and experimental research designs. Sailer and Homner (2020) suggested differences in methodological rigor as a possible moderator of effect sizes to explain part of those discrepancies. However, another possible explanation for this finding could be, in part, that meta-analyses typically combine studies that differ widely in many aspects (e.g., population, methodology, etc.; Borenstein et al., 2009), which are not yet identified as important determinants of learning outcomes and which can only be addressed by moderator analyses to some extent.



To give another, additional perspective, learners' engagement is considered an essential prerequisite of learning (Chi & Wylie, 2014; Fiedler & Beier, 2014; Wong & Liem, 2022). However, engagement is, like cognition and motivation, a multi-faceted psychological construct influenced strongly by context and timescale (Sinatra et al., 2015). That entails that an understanding of engagement depends also on the perspectives taken by the researcher, which can be categorized into person-oriented, context-oriented and person-in-context studies according to the taxonomy proposed by Sinatra et al. (2015). However, in comparison to person-oriented and person-in-context oriented research, context-oriented studies seem relatively scarce (Booth et al., 2023). Moreover, context-oriented studies typically consider the context only at the group level, for instance, investigating if students, on average, learn more efficiently with instructional videos in online or in blended learning courses (Seo et al., 2021). In contrast, person-in-context studies usually investigate how changes of particular features within a learning task affect learners, for instance, via emotional or gamified design of certain features (Greipl et al., 2021; Ninaus et al., 2019). That is, the term context in the person-in-context category of Sinatra et al. (2015) is rather narrowly defined. What is mostly lacking, to our best knowledge, are studies investigating how context properties (in the broader sense of context-oriented studies) affect the influence of features used typically in person-in-context studies. That is, e.g., how contextual features affect the effectiveness of gamified task design for different learning outcomes. This is precisely what Mayer's notion of boundary conditions is aiming at and what we are addressing in the present work, as outlined next.

1.3 Present work

In the present work, we take the opportunity to (re-)assess and compare the results of three studies using the same associative learning task to study the effect of gamified task design in different contexts. That is, all three studies were originally construed as individual, person-in-context studies aiming to investigate the effectiveness of gamified task design on cognitive and motivational learning outcomes. By combining studies across different contexts in the present work, we can shed some light on how cognitive and motivational outcomes are influenced by different contexts during gamified learning. By combining the different results, we end up with a context-oriented analysis to explore possible boundary conditions of gamified learning.

By reassessing and comparing the same cognitive learning and motivational outcomes under three different, contextual conditions realized in the three respective studies, we aim to improve our understanding of gamified learning at two different methodological levels:

1. To what extent do cognitive learning and motivational outcomes differ based on the contextual conditions realized in each of the compared studies?
2. Are there especially important contextual features that give rise to boundary conditions for the efficacy of gamified learning for cognitive learning and motivational outcomes?

Since the first two studies in our present work are re-analyses of already-published research (Huber et al., 2023, 2024), we provide short summaries of the respective two studies below.

Study 1. In study 1, Huber et al. (2023) explored whether gamifying an associative learning task via a specific set of game elements would affect learners' engagement more than a non-gamified learning task. The study focused on attrition, a subcomponent of behavioral engagement, in an online learning environment. By comparing the gamified with the non-gamified version of the learning task, game elements were found to reduce dropout rates and enhance engagement. While 1688 persons accessed the landing page of the study, only 685 (about 40%) proceeded to sign the online consent form, and from there, only 312 (about 18%) finished the learning task with significantly more persons ($\Delta N = 50$) dropping out with the non-gamified version than the gamified version of the task ($\Delta N = 23$), $\chi^2(1) = 12.28$, $p < .001$. Game elements also improved learning efficacy and efficiency indirectly through their ability to increase task attractiveness. Both the task's attractiveness and the perceived stimulation from the task – used to assess learners' experience of the learning task – were positively associated with the motivational subcomponents of intrinsic motivation and identified regulation. Conversely, they were negatively correlated with the subcomponents of external regulation and amotivation.



Study 2. In study 2, Huber et al. (2024) explored relations between game elements, learners' experience and motivation, and cognitive learning outcomes. The study aimed to replicate the positive, indirect association between gamified learning and cognitive learning outcomes through learner experience and motivation, as noted in study 1, using a small convenience sample (via a university course) and a slightly more difficult version of the associative learning task. The study was again conducted online and compared two versions of the learning task, a gamified one and a non-gamified one. The results showed no significant differences in cognitive learning outcomes, but medium and large positive effects on motivational outcomes and learner experience, respectively. Mediation analysis revealed that while game elements slightly hindered cognitive outcomes directly, this was offset by a stronger positive influence through increased motivation and learner experience. Learner experience was further shown to be positively associated with two subcomponents of intrinsic motivation, namely perceived interest and competence.

Study 3. Study 3 was conducted recently and will be described in detail in the present work.

2. Material and methods

Three datasets were collected from studies conducted at two different Austrian universities, guided by a value-added research design according to Mayer's taxonomy (Mayer, 2019, 2020). That is, participants (Section 2.1) were randomly assigned to one of two learning task versions in each study within an overall similar experimental procedure (Section 2.2). The two different learning task versions consisted of a control version (i.e., non-gamified = without additional game elements) and a gamified version (Section 2.3). Various learning outcome measures, including cognitive and motivational outcomes, were assessed after the learning task (Section 2.4). The results obtained in the two conditions (i.e., the two task versions) were then compared and differences were quantified using statistical analyses (Section 2.5). All studies were approved by the respective local universities' ethics committees.

Although all studies were similar to each other, they also differed in several respects. This will be clarified in the following sections. Table 1 provides an overview of features by which the compared studies differed. Except for one task-inherent feature (task difficulty and directly related to it: time-on-task) all features in Table 1 refer to the broader study context and are not inherent to the learning task common to all three studies.

2.1 Participants

Table 2 summarizes the demographic information available for the participants of the three studies. Regarding demographic information, age, gender, and whether participants were currently enrolled university students, were assessed in all three studies. Considering student status, we did not differentiate between graduate and undergraduate students. Furthermore, the information was entirely based on self-report, that is, we did not verify with the universities about participants' student or enrollment status.

Study 1. The first of the three compared studies was conducted online with a data acquisition period spanning about 7 weeks from mid-February 2021 to the beginning of April 2021. Participants were compensated by being given the option to enter a raffle for five vouchers, each worth 10 EUR for an online retailer, which were awarded at the conclusion of data collection (Huber et al., 2023). The compensation was very low on purpose to reduce the influence of incentive bias on engagement, especially dropout, which was the research focus of the study. However, some compensation was necessary to improve response rates, which can be especially low in online studies that exceed 10 minutes in duration (Sammut et al., 2021). At the same time, the study was broadly advertised using university-wide listservs of the local university to acquire a sufficient number of participants allowing a statistical analysis of attrition over the course of the study. Due to this primary aim of the study, many participants dropped out over the course of the learning task ($n = 244$ overall and significantly more



during the non-gamified than the gamified task version; for details, see Huber et al., 2023) and hence, never advanced to the questionnaires addressing motivational outcomes. As a result, in this paper we reassess only the data from those $n = 285$ participants who completed each phase of the study.

Study 2. The second of the three compared studies was conducted online. In this case, however, participants were acquired during a shorter, about 4-week period in May 2022 within the framework of a university course on empirical research at a different university than in the case of the first study, see Table 1. Furthermore, participants were compensated by course credit, whereas no course credit could be obtained in the first study. The rationale for changing compensation from Study 1 was to increase the chance that participants would stay engaged throughout the study in both gamified and non-gamified conditions, as this was important for an unbiased quantification of the motivational effect of gamification on learning. Therefore, the incentive to stay engaged in the learning task was increased by providing course credit for students enrolled in the psychology program, which reduced dropout rates overall ($n = 35$), and eliminated the difference in dropout rate between gamified and non-gamified conditions (Huber et al., 2024). Although again university-wide listservs were used for participant recruitment and course credit was provided for participating psychology students, only $n = 61$ participants completed the study, whose data will be reassessed in the present study.

Study 3. The third study was conducted on-site, in a laboratory at the same university at which study 2 was conducted. In this case, the data acquisition period lasted about 13 months from the end of April 2023 to beginning of June 2024. The reason for this long data acquisition period was related to the requirement of recruiting a minimum of 120 participants to be able to replicate the motivational effect of gamified learning found in study 2 with sufficient statistical power assuming a medium-to-large effect size. In this study, participants were compensated in the form of course credit for students enrolled in the psychology program in addition to the option to enter a raffle for five vouchers, each worth 50 EUR for an online retailer, awarded at the conclusion of data collection. A total of $n = 121$ participants were recruited, again using university-wide email broadcasts, in addition to word-of-mouth and by advertising the study across social networks.

2.2 Procedure

In all three studies, participants were administered several questionnaires before and after the learning task to obtain demographic information and motivational outcomes, in addition to other data not discussed in this work. In the present work, we focus only on the cognitive learning and motivational outcomes that were commonly used in all three studies to explore the influence of the different contextual features on the same cognitive and motivational outcomes of gamified learning. These will be described in detail in Section 2.4.

In the case of the laboratory study (study 3), we further recorded different physiological sensor data (electrodermal activity, heart rate, eye-tracking, facial expression analysis) over the course of the learning task. Such data were not recorded in the two online studies (study 1 and study 2). This also means that in the case of study 3, some additional time before and after the learning task was required to attach and detach the respective sensors.



Table 1
Different contextual features realized in the three studies

Study	Context	Incentive	Recruitment strategy	Task Difficulty	Austrian University	Data acquisition period	Time-on-task
Study 1 (Huber et al., 2023)	Online (Questionnaires and learning task)	Low monetary reward with low win chance	University-wide email broadcast	Easy (14 symbols)	Innsbruck	~ 7 weeks	10-30 min
Study 2 (Huber et al., 2024)	Online (Questionnaires and learning task)	Course credit	University-wide email broadcast	Moderate (20 symbols)	Graz	~ 4 weeks	15-45 min
Study 3	Laboratory (Questionnaires, learning task, sensors)	Course credit and monetary reward with low win chance	University-wide email broadcast, word-of-mouth, social networks	Moderate (20 symbols)	Graz	~ 13 months	15-45 min

Table 2
Demographic information on the three study samples

Demographic variable	Metric	Study 1	Study 2	Study 3
Sample size	<i>n</i>	285	61	121
Age	Mean	29.2	27.56	23.7
	Median	25	24	23
	<i>SD</i>	12.02	11.54	5.45
	MAD (robust)	5.93	4.45	2.97
	Min.	18	18	18
	Max.	74	64	64
Gender	Female (%)	62.5	72.1	61.2
	Male (%)	36.1	24.6	36.4
	Non-binary (%)	1.4	3.3	2.5
Student?	Yes (%)	59.3	85.2	90.9

Note. *SD* = standard deviation. *MAD* = Median absolute deviation adjusted for asymptotically normal consistency. All variables referring to age are given in units of years.



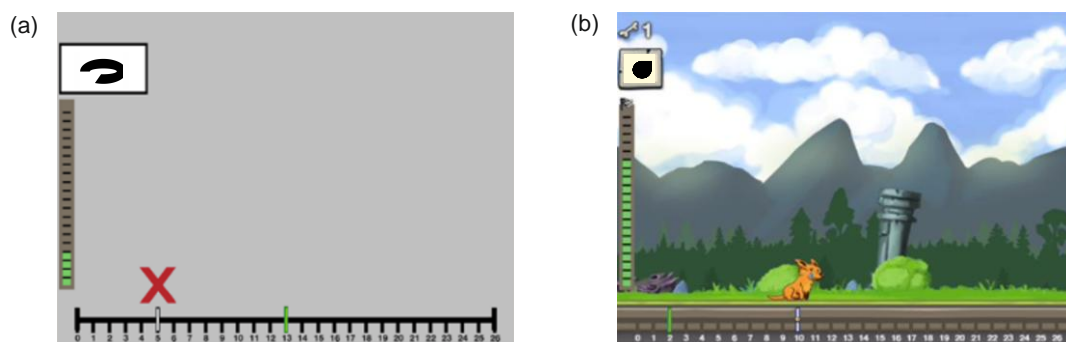
2.3 Learning task

All studies used an associative learning task in which participants needed to learn associations between symbols and numbers over the course of five levels of the task. At the beginning of the first level of the task, a symbol was presented to a participant in the upper left corner of the computer display (see Fig. 2). The participant then had to guess which number, presented on a number line at the bottom of the interface, would be associated with that symbol. The selection was made by moving a slider with the keyboard's arrow keys to the desired position on the number line and confirmed by pressing the spacebar. After participants made their selection, corrective feedback was provided. That is, the correct number was always indicated for the given symbol with a green vertical bar to allow for learning of the association between symbol and number. This procedure was repeated for all remaining symbols. Note that at the first level, participants could only guess which number would correspond to which symbol. That is, the purpose of the first level was entirely to present all the associations once to the participants. In the second level, all symbols would be shown again subsequently. At this point, the participants would typically be able to remember already some of the associations but would still receive corrective feedback for each symbol, allowing them to learn the rest. The overall goal of the learning task was to learn as many associations as possible over the course of five levels.

In each level, symbols were presented in random order. Selecting a number had to take place within 20 seconds after each symbol was displayed; otherwise, corrective feedback would be provided without having selected a number. Study 1 differed from studies 2 and 3 with respect to the total number of symbol-number associations that needed to be learned by the participants. This number was 14 in *Study 1* (Huber et al., 2023), while it was increased to 20 in *Study 2* (Huber et al., 2024) and was kept at 20 also in *Study 3* (present work). Hence, the difficulty of the learning task was also somewhat lower in the first of the three studies (see Table 1). The choice to increase the number of symbol-number associations was justified in study 2 (Huber et al., 2024) by the aim to increase the discriminatory power of the cognitive outcome measures and mitigate ceiling effects at the fifth level of the task reported in study 1 (Huber et al., 2023). However, the increase of presented symbols per level also affects the time participants engage with the task (time-on-task). In Table 1, we provide approximate minimum and maximum times participants engage with the task based on the task's inherent mechanics.

Figure 2

Non-gamified and gamified learning task versions



Note. Screenshots of the learning task in the (a) non-gamified version and (b) in the gamified version when an incorrect answer was provided. Corrective feedback was always provided, regardless of correct or incorrect responses. If responses were correct, a green check mark was provided in the non-gamified version of the task whereas the bone count was increased by 1 and the dog wagged its tail and held a bone in its mouth in the gamified version of the task (not shown here). If responses were incorrect, a red X-sign was provided in the non-gamified version of the task (see panel a) whereas the dog cried in the gamified version of the task (see panel b).



Regarding the design of the learning task, all three studies used two identical design versions, one employing some additional game elements (= gamified version), one representing a control version without those game elements (= non-gamified version). Screenshots of the two versions are provided in Fig. 2. The difference between gamified and non-gamified task versions was entirely based upon the inclusion of game elements, a core aspect also in game-based learning (Plass et al., 2015). The game elements comprised of (i) visual aesthetics (a scenery with trees, sky, mountains, etc.), (ii) a narrative of a dog taking a stroll in the woods looking for bones, and (iii) an incentive system (i.e., the “bone count” displayed in Fig. 2(b) in the upper left corner). In the gamified version, the slider’s movement and placement were also accompanied by animations of the dog walking and digging (searching for a bone), respectively. Choosing a correct number would result in the dog wagging its tail in the gamified task version, while in the non-gamified task version, only a green check mark would appear. In the gamified task version, choosing an incorrect number would result in the dog shedding tears, while a red X-sign would appear in the non-gamified task version (see also Fig. 2).

Both versions of the learning task were created using the NumberTrace game engine [see e.g., Koskinen et al. (2023)]; a short video demonstration may further be viewed at the following URL: <https://www.youtube.com/watch?v=T7s7xSILrac>] which was originally developed for fraction instruction using JavaScript.

2.4 Outcome measures

Regarding cognitive learning outcomes, we considered both learning efficacy and efficiency in all three studies. Efficacy refers to how many associations participants had learned at the end of the learning task, that is, in level 5 of the task. Because studies differed in the total number of symbol-number associations, see Section 2.2, we refer to the proportion of correctly learned associations to compare studies with each other with respect to efficacy. This proportion is given by the number of correctly learned associations in level 5 divided by the total number of symbol-number associations in the respective study (i.e., 14 in study 1, 20 in studies 2 and 3), resulting in a number between 0 and 1 for the learning efficacy of each learner.

Efficiency refers to how fast participants learned symbol-number associations. Originally, in the two online studies (i.e., studies 1 and 2), the authors fitted an exponential learning curve to the number of correct responses per level for each individual participant, and subsequently analyzed the rate constants of these learning curves (Huber et al., 2023, 2024). In the present work, we follow a simpler approach by simply adding the numbers of correct responses per level for levels 2 to 5. The rationale behind this approach is that if a participant learns very efficiently, they approach the maximum number of correct responses very fast and the sum of correct responses per level corresponds to a large number. Participants learning less efficiently approach their final performance at the fifth level slower and the sum of correct responses corresponds to a lower value. By dividing this sum score again by the maximally possible value (56 for study 1, 80 for studies 2 and 3) we again obtain a number between 0 and 1 for the learning efficiency of each learner.

Regarding motivational outcomes, studies 1 and 2 used different questionnaires to assess specific facets of motivation. However, both studies complemented these questionnaires by the subscales attractivity and stimulation of the user experience questionnaire (UEQ; Laugwitz et al., 2008; Schrepp et al., 2017). Although not targeting motivation at face value, both studies provided evidence that these subscales are associated with the construct motivation. Study 1 (Huber et al., 2023) showed that attractivity and stimulation were strongly and positively associated with intrinsic motivation ($\rho_{pb} = 0.69$ and $\rho_{pb} = 0.66$, respectively, using the percentage bend correlation described, e.g., by Wilcox (2022)). Further, identified regulation was moderately and positively associated with attractivity and stimulation ($\rho_{pb} = 0.45$ and $\rho_{pb} = 0.48$, respectively), and somewhat, negatively associated with external regulation ($\rho_{pb} = -0.25$ and $\rho_{pb} = -0.18$, respectively). A strong negative association was found with amotivation ($\rho_{pb} = -0.49$ and $\rho_{pb} = -0.50$, respectively). These facets of motivation were measured via



the situational motivation scale by Guay et al. (2000) in Study 1. Study 2 (Huber et al., 2024) further revealed that the two subscales of the UEQ were positively associated with the subscales interest ($\rho_{pb} = 0.32$ and $\rho_{pb} = 0.29$, respectively) and competence ($\rho_{pb} = 0.78$ and $\rho_{pb} = 0.82$, respectively) of the short scale for intrinsic motivation (Wilde et al., 2009). Hence, we use these two subscales of the UEQ (i.e., attractivity and stimulation) in Study 3 to compare the results of Studies 1 and 2 regarding motivational outcomes with the results obtained in Study 3 and, as such, as a proxy for motivation.

Both the attractivity and the stimulation subscale of the UEQ (Laugwitz et al., 2008; Schrepp et al., 2017) consist of several items, each presenting two opposing adjectives, forming the endpoints of a 7-point, bipolar rating scale (i.e., from -3 to 3 in steps of 1), on which participants indicate their experience of the task. In particular, attractivity aims to assess the general appeal of a task (or product) by asking, for instance, how “enjoyable” it is perceived (with the poles: “annoying”/“enjoyable”), or “pleasing” (with the poles: “unlikable”/“pleasing”). The subscale uses six items in total. Stimulation assesses the capability of a task (or product) to motivate or captivate by asking, for instance, how “interesting” (poles: “not interesting”/“interesting”), or “motivating” (poles: “not motivating”/“motivating”) it is perceived. The subscale uses four items in total.

All studies yielded good to excellent reliabilities for the two subscales attractivity and stimulation of the UEQ. In particular, we obtained Cronbach’s $\alpha = 0.92$ and $\alpha = 0.89$, respectively, for study 1, $\alpha = 0.93$ and $\alpha = 0.91$, respectively, for study 2, and $\alpha = 0.89$ and $\alpha = 0.82$, respectively, for study 3.

2.5 Statistical analyses

For statistical analyses, we employ Bayesian hypothesis testing and estimation as provided by JASP (version 0.19.1.0; JASP Team, 2024). In the present work, we apply Bayesian inference because it provides some advantages over classical inference (Wagenmakers et al., 2018) that are important for the present work. First, Bayes factor analysis can quantify the degree of evidence that the data support the null model (with a fixed parameter) compared to alternative models (with a continuous range of possible parameter values that are estimated through conditioning on the data). Second, Bayesian inference is not biased against the null model. Third, it does not depend on sampling plans varying between studies. While our research questions (Section 1.3) fall into the area of estimation, the re-assessed previous studies (Huber et al., 2023, 2024) have shown that gamified learning can have insignificant effects on cognitive learning outcomes, which would be reflected in high compatibility between a null model (assuming an effect of exactly zero) and the data. This implies that investigating the dependence of respective Bayes factors on contextual features across the three studies is informative also for investigating our present research questions. For estimating the cognitive and motivational effects of gamified learning and comparing these effects across the three studies, Bayesian estimation provides advantages over classical inference because it can gradually quantify the degree of confidence where effect sizes lie in a specific interval while conditioning on what is factually known (Wagenmakers et al., 2018).

All respective analyses and data, are available at the Open Science Framework (OSF), see <https://osf.io/snvw8> and <https://osf.io/d97bp>, respectively. The specific context of each study was included in each model as a categorical predictor (denoted as “boundary” in files provided at OSF) with three categories corresponding to studies 1-3.

To assess the effect of contextual factors on cognitive learning and motivational outcomes, we compared Bayes factors resulting from Bayesian ANOVAs for the following five models for each outcome variable (efficacy, efficiency, attractivity, stimulation): 1) a null model (containing neither task version nor context as predictors); 2) two single predictor models with either task version or context as predictor; and 3) two models including both predictors once without their interaction and once with their interaction included. These models are denoted below simply as “null”, “version”, “context”, “version



+ context” and “version * context”, respectively. Prior odds were assigned uniformly over all five models (i.e., prior odds of 20% for each model). Models were compared to the best resulting model in our analyses. All models included age, gender, and student status as random factors to provide some control for their influence because we noted that samples differed across studies (see Table 2), and multimedia learning theory suggests that these factors play a role too for cognitive learning and motivational outcomes (Mayer, 2024).

Bayesian ANOVAs were followed up by pairwise comparisons of results obtained with gamified and non-gamified task versions for each of the three studies. For these comparisons, first the assumptions about normality of outcome variables and equality of variances for gamified and non-gamified task versions were checked with Shapiro-Wilk and Brown-Forsythe tests, respectively. Based on those tests, either JASP’s Bayesian implementations of Student’s t or Mann-Whitney tests were used for pairwise comparisons. For each comparison, default Cauchy priors centered on zero and with an interquartile range of $1/\sqrt{2}$ were used. If not stated explicitly otherwise, credible intervals refer to 95% percentile intervals and are reported directly after parameters in parenthesis. For the pairwise comparisons, null hypotheses always referred to typical nil hypotheses, that is, an exact equality between outcome variables for gamified and non-gamified task versions, whereas alternatives referred to allowing outcome variables to be unequal.

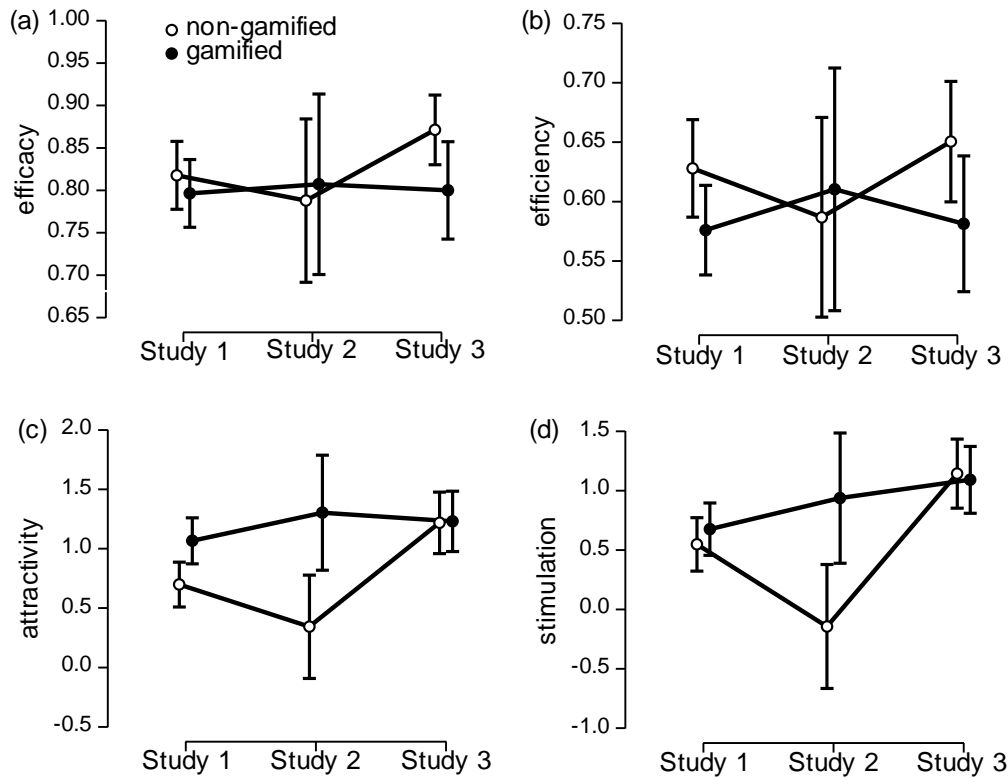
3. Results

3.1 Effects of context on cognitive learning and motivational outcomes

In Fig. 3, we depict the means and their 95% credible intervals for (a) efficacy, (b) efficiency, (c) attractivity, and (d) stimulation depending on context and task version. We find that for efficacy and efficiency, neither context nor task version seem to matter much, because all credible intervals yield substantial overlap with each other. The data does not provide any evidence for an interaction between context and task version, either for efficacy or efficiency (see Tables 3-4), respectively. In fact, for efficacy (Table 3), the best model is the null model that does not include any of the two predictors. Except for the single predictor model with task version as a predictor (2.6 times less likely than the null model according to our data), the data provides strong to decisive evidence against all other models. The same is the case for efficiency (Table 4), in which case again the single predictor model with task version as predictor and the null model are most likely from all considered models and for the given data, providing strong to decisive evidence against all other models. Overall, we conclude that for predicting efficacy and efficiency in the used learning task contextual aspects do not contribute much if anything at all. Regarding our first research question (Section 1.4), we thus conclude that cognitive learning outcomes resulting from this task do not differ much between the different contexts realized by the three experimental settings. Regarding our second research question, the results further suggest that the influence of gamification on cognitive learning outcomes does not depend much on those contextual aspects.

**Figure 3**

Means and 95% credible intervals for the considered outcome variables



Note. Means and their 95% credible intervals for (a) efficacy, (b) efficiency, (c) attractivity, and (d) stimulation for the three considered studies (from left to right in each panel) and each task version (non-gamified: open circles; gamified: full circles).

Instead, in the case of motivational outcomes, the depicted means and their 95% credible intervals in Figs. 3(c) and (d) indicate that these outcomes depend on contextual aspects. In the case of attractivity (Table 5), both two predictor models (i.e., the two models with two predictors, either including or excluding the interaction term), turn out best. In particular, the data allow no clear distinction between both two predictor models, yielding almost identical posterior probabilities. However, against all other models, moderate to strong evidence is provided by our data. In particular, the null model is more than 40 times less likely than both two predictor models. The evidence against the null model is even stronger in the case of stimulation, with the model being more than 100 times less likely than any model, including context as a predictor. Overall, this means that for motivational outcomes the data suggest that contextual aspects seem to matter substantially. Regarding our first research question (Section 1.4), we thus conclude that motivational outcomes resulting from this task differ substantially depending on the different contexts realized by the three experimental settings. Regarding our second research question, the relative importance of the interaction between task version and context for predictive quality of the resulting model further suggests that the influence of gamification on motivational outcomes also depends somewhat on those contextual aspects. To resolve this contextual dependence better, pairwise comparisons between gamified and non-gamified task versions for each study context are presented next.



Table 3

Model comparison for efficacy

Model (M)	Prior	Posterior	BF ₀₁
Null	0.2	0.683	1.000
Version	0.2	0.264	2.588
Context	0.2	0.036	19.097
Version + Context	0.2	0.015	44.211
Version * Context	0.2	0.002	405.601

Note. All models include age, gender, and student status. Models are always compared against the best model (denoted by 0 in the Bayes factor) in ascending order.

Table 4

Model comparison for efficiency

Model (M)	Prior	Posterior	BF ₀₁
Version	0.2	0.607	1.000
Null	0.2	0.351	1.732
Version + Context	0.2	0.026	23.147
Context	0.2	0.013	46.106
Version * Context	0.2	0.003	230.773

Note. All models include age, gender, and student status. Models are always compared against the best model (denoted by 0 in the Bayes factor) in ascending order.

Table 5

Model comparison for attractiveness

Model (M)	Prior	Posterior	BF ₀₁
Version + Context	0.2	0.429	1.000
Version * Context	0.2	0.410	1.047
Version	0.2	0.127	3.366
Context	0.2	0.024	18.215
Null	0.2	0.010	43.103

Note. All models include age, gender, and student status. Models are always compared against the best model (denoted by 0 in the Bayes factor) in ascending order.

Table 6

Model comparison for stimulation

Model (M)	Prior	Posterior	BF ₀₁
Context	0.2	0.472	1.000
Version * Context	0.2	0.310	1.523
Version + Context	0.2	0.216	2.188
Null	0.2	0.001	338.575
Version	0.2	0.0006	798.378

Note. All models include age, gender, and student status. Models are always compared against the best model (denoted by 0 in the Bayes factor) in ascending order.

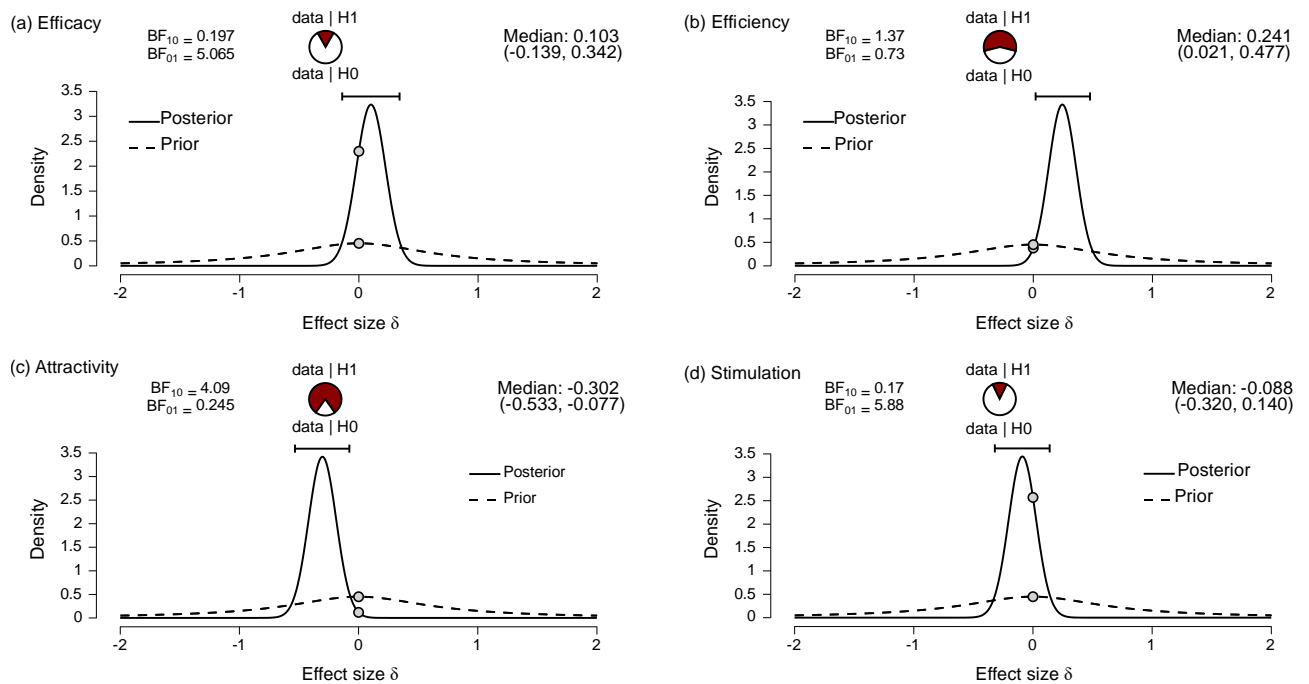


3.2 Effects of context on the influence of gamification on cognitive learning and motivational outcomes

Study 1. Regarding study 1, we depict the prior and posterior distributions for the effects of gamification on all four considered outcome measures in Fig. 4. The Bayes factor analyses provide evidence for the null hypothesis in the case of efficacy ($BF_{01} = 5.1$) and stimulation ($BF_{01} = 5.9$), yet for the alternative hypothesis in the case of attractivity ($BF_{10} = 4.1$). The median effect size in the latter case is -0.30 ($-0.53, -0.08$), where the negative sign indicates that attractivity is, on average, higher for the gamified task version than for the non-gamified one. In the case of efficiency, the data are undecisive with respect to both hypotheses ($BF_{10} = 1.4$). Still, more than 95% of the mass of the posterior distribution is located above zero, indicating that if there is some population effect on efficiency, then efficiency is probably somewhat larger for the non-gamified task version than for the gamified one. The median effect size in that case is 0.24 ($0.02, 0.48$).

Figure 4

Effects of gamification in study 1



Note. Prior and posterior distributions for effects of gamification for (a) efficiency, (b) efficacy, (c) attractivity, and (d) stimulation in the first online study. Note that a negative value of effect size means that the respective outcome is, on average, larger for the gamified than for the non-gamified task version.

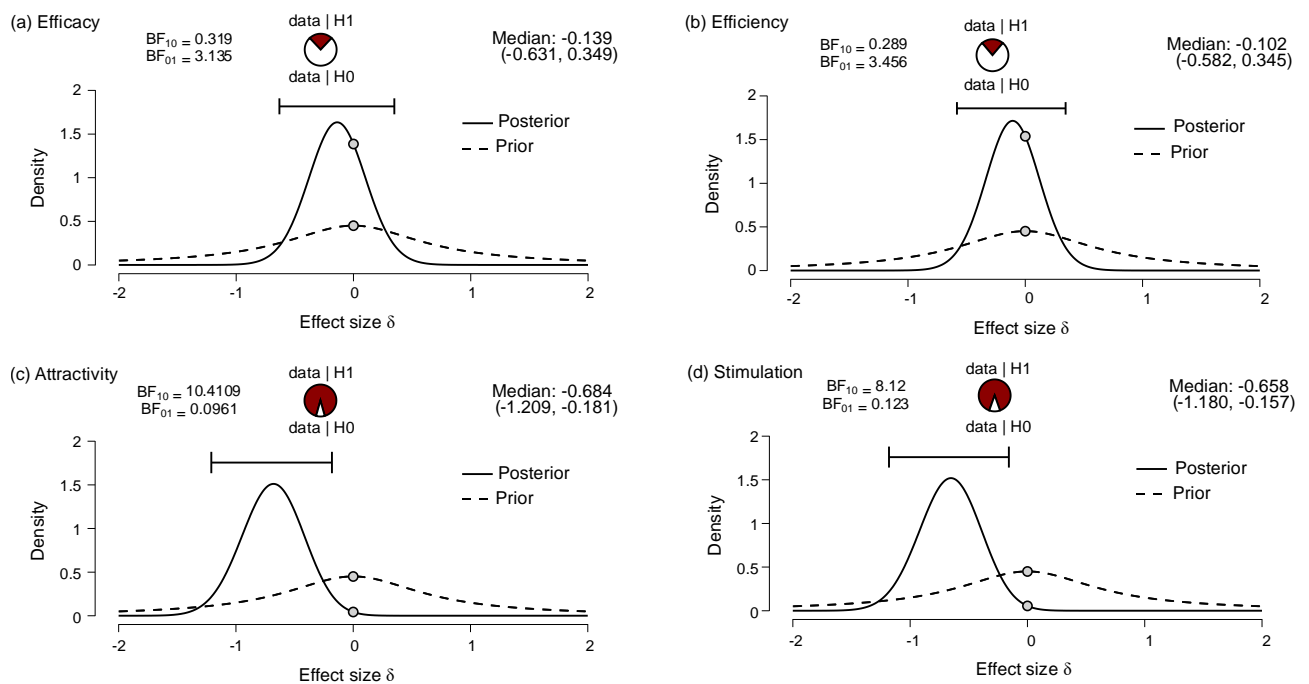


Study 2. Comparison of these results with those obtained for study 2, depicted in Fig. 5, suggests that study context might matter for the influence of gamification, especially for motivational outcomes. In particular, we find evidence for the null hypothesis in the cases of both efficacy ($BF_{01} = 3.1$) and efficiency ($BF_{01} = 3.5$). For attractivity, we obtain strong evidence in favor of the alternative hypothesis ($BF_{10} = 10.4$) with a medium median effect size of -0.68 ($-1.21, -0.18$), indicating the gamified task version to be considerably more appealing to the participants than the non-gamified task version. For stimulation, we also obtain evidence in favor of the alternative hypothesis ($BF_{10} = 8.1$), yielding a medium median effect size of -0.66 ($-1.18, -0.16$), indicating the gamified task version to raise more interest and motivation in the participants than the non-gamified task version.

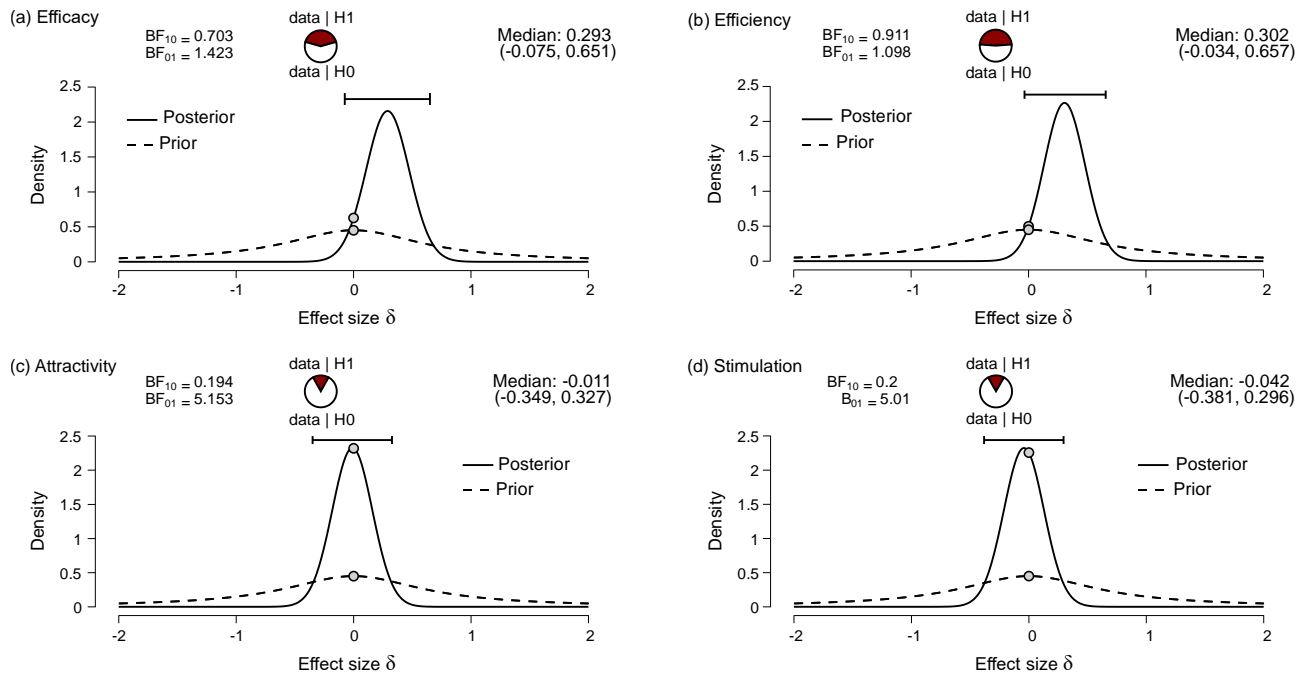
Study 3. Strong evidence for the existence of boundary conditions for the effects of gamification is found by comparing study 2 with study 3. Note that the two studies are closest to each other out of all three possible pairs of studies not only by design, but also by the resulting participant samples, see Section 2.1 and Table 1. Nevertheless, in contrast to study 2, the laboratory data are undecided regarding both hypotheses for both cognitive learning outcomes, that is, efficacy, $BF_{01} = 1.4$ and efficiency, $BF_{01} = 1.1$. In addition, the results yield posterior distributions in both cases with almost 90% of their probability mass above zero. That is, if gamification affects efficacy and efficiency at all, then, according to the data, it is more reasonable to assume that they are higher, on average, for the non-gamified task version than for the gamified one. The respective median effect sizes are 0.29 ($-0.08, 0.65$) and 0.30 ($-0.03, 0.66$). Most importantly and in stark contrast to study 1, study 3 provides evidence for the null hypothesis in the case of both motivational outcomes, attractivity, $BF_{01} = 5.2$, and stimulation, $BF_{01} = 5.0$. The median effect sizes account for practically negligible to small effects yielding -0.01 ($-0.35, 0.33$) for attractivity and -0.04 ($-0.38, 0.30$) for stimulation.

Figure 5

Effects of gamification in study 2



Note. Prior and posterior distributions for effects of gamification for (a) efficiency, (b) efficacy, (c) attractivity, and (d) stimulation in the second online study. Note that a negative value of effect size means that the respective outcome is, on average, larger for the gamified than for the non-gamified task version.

**Figure 6***Effects of gamification in study 3*

Note. Prior and posterior distributions for effects of gamification for (a) efficiency, (b) efficacy, (c) attractivity, and (d) stimulation in the laboratory study. Note that a negative value of effect size means that the respective outcome is, on average, larger for the gamified than for the non-gamified task version.

4. Discussion

4.1 Main findings

In this paper, we focused specifically on the concepts of context and environment as part of the broader conceptualization of boundary conditions in multimedia learning (Mayer, 2024). We aimed to identify the contextual features, in which the three compared studies differed, that give rise to boundary conditions influencing the effect of gamified learning on cognitive learning and motivational outcomes. Our findings indicate the existence of such boundary conditions for motivational outcomes, whereas cognitive learning outcomes are influenced by the considered contextual features, if at all, only to a substantially smaller extent. This result is supported by theory as well as meta-analyses and systematic reviews on gamification and game-based learning, consistently showing that employing game elements can have a beneficial effect on both motivational and cognitive outcomes, but revealing also a substantial heterogeneity of effect sizes across studies (Arztmann et al., 2023; Barz et al., 2023; Hu et al., 2022; Huang et al., 2020; Lampropoulos & Kinshuk, 2024; Ritzhaupt et al., 2021; Ryan & Rigby, 2020; Sailer & Homner, 2020; Schlag et al., 2024; Wouters et al., 2013). This result is further aligned with Plass and Kaplan's (Plass & Kaplan, 2016) Integrated Cognitive Affective Model of Learning with Multimedia (ICALM), which specifically describes that there is a high sensitivity towards contextual factors on cognitive processing and motivation when learning with multimedia.

The most obvious contextual difference between the three considered studies is that study 3 was carried out in the laboratory, whereas studies 1 and 2 were carried out online. This difference suggests at least two possible explanations for the substantial motivational difference between studies 1 and 2 and study 3. Both explanations are based on differences in expectations and norms between those contexts.



First, the motivational difference between the laboratory (Study 3) and online studies (Studies 1 and 2) might stem from participants coping with cognitive dissonance (Harmon-Jones & Mills, 2019) through reevaluation or self-justification. In the laboratory setting, participants invest considerably more effort compared to an equivalent online setting—they must register, schedule an appointment, travel to the lab, and navigate a social situation with stricter norms and expectations (Hoerger, 2010; Hoerger & Currell, 2012; Skitka & Sargis, 2006). Given this level of effort for completing a relatively unappealing task (i.e., memorizing number-symbol associations), participants may experience cognitive dissonance, making it difficult to justify the experience to themselves. To resolve this discomfort, they might retrospectively reassess the task in the laboratory as less unappealing than it initially seemed. In other words, to avoid cognitive dissonance (likely more prominent with the non-gamified task version) participants may reevaluate the task's attractiveness and their interest in it more favorably after its completion in the laboratory. This is indeed supported by our findings revealing that the non-gamified task is evaluated as more appealing and stimulating in the laboratory than online, whereas the gamified tasks are not evaluated considerably different between those two settings, see Fig. 3.

However, another explanation considers the possibility that both the non-gamified and gamified task versions are not merely evaluated differently post-hoc but may be genuinely experienced as equally motivating in the laboratory setting. This could be understood as a form of a figure-ground effect. Hypothetically, the controlled laboratory conditions (amplified eventually by application of sensors and being monitored), in addition to the absence of other, readily available action options, which might enhance participants' focus on the tasks, possibly reduce the distinction between the gamified and non-gamified versions in the laboratory. In other words, in the laboratory, the non-gamified task version might be experienced as appealing and interesting, given the socially- and contextually-amplified attention that it receives in that setting, in addition to the simple fact that there is nothing else to do. This is aligned with prior research which has repeatedly shown that participants underestimate the enjoyment of supposedly boring tasks if experienced in the context of a scientific experiment (Hatano et al., 2022; Kuratomi et al., 2023). In contrast, in the online environment, participants may take part in the study considerably more spontaneously, and more likely than not, also have other action options at their disposal. Under these circumstances, the non-gamified task possibly starts out with less appeal to the participants (relative to the laboratory setting) and, in addition to that, is perceived as even less appealing relative to other, alternative action options. Given such options, Bernecker and Ninaus (2021) showed that gamifying a strenuous and tedious working memory task can indeed reduce disengagement and the motivational conflict to do something else. That implies that there might be much more to gain for gamification regarding an influence on learners' motivation in an online setting than in the laboratory.

The present study also tentatively supports that the gamified and non-gamified task versions were truly equivalent regarding their motivational outcomes in the laboratory. To explicate this, the more subtle influence of contextual aspects on cognitive learning outcomes needs to be considered in conjunction with their influence on motivational outcomes. It was argued already in both earlier online studies (Huber et al., 2023, 2024), that the effects of game elements on cognitive learning outcomes are aligned with what would be expected from cognitive load theory (Sweller, 2011) or cognitive theory of multimedia learning (Mayer, 2019, 2024). In general, the additional design elements employed in the gamified task version included more stimuli requiring additional cognitive processing, but non-essential for completion of the task. In addition, the simplicity of the task may not leave much room for generative processing (or germane load). However, even if they represent merely extraneous processing demands, ignoring and discarding the game elements as irrelevant yet puts some demand on information processing in the gamified task version, which is absent in the non-gamified one.

In fact, our resulting estimates for effect sizes regarding cognitive learning outcomes show that the difference between gamified and non-gamified task versions is, descriptively, largest in the laboratory study (Fig. 6), smallest in the second online study (Fig. 5), and somewhat in between in the first online study (Fig. 4). Furthermore, the sign of the estimates indicates that in the laboratory and the first online study (Study 1), the game elements are cognitively demanding in the sense that cognitive



outcomes are slightly lower in the gamified than in the non-gamified task version, whereas the opposite is the case for the second online study (Study 2). That is, if some cognitive effect of gamification is assumed at all, which is a reasonable assumption based on cognitive learning theories (Mayer, 2019, 2024; Sweller, 2011), then our analyses suggest that the influence of gamification on cognitive learning outcomes is most detrimental in the laboratory where the two task versions are most similar in motivation, and less detrimental in online settings where some motivational difference exists between the two task versions.

However, this pattern of cognitive and motivational differences between task versions and contexts can be entirely explained by the indirect positive effect of game elements on cognitive learning outcomes through enhanced motivation, as previously suggested in both earlier online studies (Huber et al., 2023, 2024). As outlined in the ICALM framework (Plass & Kaplan, 2016), game elements can boost motivation, which in turn positively influences cognitive processing, which can explain the subtle differences in cognitive learning outcomes observed across the varying study contexts. In study 3, the motivational difference between gamified and non-gamified task versions was nearly absent. Hence, in theory, only the presumed higher cognitive demand due to the additional game elements remains, and in fact, the difference in cognitive learning outcomes turns out largest. The other way around, the motivational difference between gamified and non-gamified task versions was largest in study 2. Here, in theory, the positive indirect effect of game elements via motivation is supposedly the largest among the three studies and can compensate for the negative, direct effect due to cognitive demand. In fact, in this case, cognitive learning outcomes are, across all three studies, mostly in favor of a null difference between gamified and non-gamified task versions, descriptively favoring the gamified version over the non-gamified one, see Fig. 5.

Although the most obvious, the difference between laboratory and online settings is not the only feature in which the three studies differ, see Table 1. Hence, alternative mechanisms could underlie the noted differences in learning outcomes. Another difference between the three studies that needs to be addressed is task difficulty. Task difficulty, and in direct consequence also the time spent on the learning task, was lower in study 1, requiring participants to learn 14 symbol-number associations, than in studies 2 and 3, both requiring participants to learn 20 symbol-number associations. Clearly, task difficulty can affect learning outcomes and a quadratic relationship between difficulty and both cognitive and motivational outcomes can be expected based on conceptual grounds (Greipl et al., 2020).

According to the limited capacity principle (Mayer, 2011, 2014) cognitive resources are limited and can hence be overstressed and lead to cognitive overload. Perceiving a task as too hard may then decrease learners' engagement (Baten et al., 2020) and motivation (Sailer & Sailer, 2021). Interestingly, engagement also decreases for tasks perceived as too easy (Baten et al., 2020), perhaps related to experiences of boredom and amotivation (Engeser & Rheinberg, 2008). Hence, a part of the motivational difference between study 1 and study 2 might also be due to the difference in task difficulty between those two studies, providing more potential for the effectiveness of game elements in the case of a task that affords more engagement in the first place.

Some support for such a mechanism is provided by the fact that in both studies 2 and 3 again about a third of all participants (study 2: 34%; study 3: 32%) reached maximum efficacy at the final task level not differing much from study 1 (40%) in that respect. Moreover, in all three studies about half of the participants manage to correctly recall all, or all but one, symbol-number associations at the final level (study 1: 54%; study 2: 51%; study 3: 45%). Overall, the empirical distribution functions of all three studies are surprisingly similar given the clear difference in difficulty between study 1 and studies 2 and 3 (more details are provided in the respective analysis files provided at OSF, see Section 2.5). Hence, it seems that the increased difficulty has indeed been met with increased engagement on behalf of the participants to some extent. While this may explain some part of the motivational difference between study 1 and 2, it can hardly explain the substantial difference in motivational outcomes between studies 2 and 3 which had the same difficulty level.



Another difference between all three studies concerns the equivalence of the sampled populations. Individual differences have been noted as a potential factor influencing the effectiveness of gamification (Ritzhaupt et al., 2021). While accounting for gender, age and student status as covariates may provide some control for interindividual differences, further person characteristics that may make a difference (for instance, prior knowledge, learning strategies, or generally, meta-cognitive abilities, see e.g., Azevedo & Wiedbusch, 2023) cannot be excluded. In fact, Table 1 shows that study 1 differs substantially from studies 2 and 3 regarding the student status of the participants. Being a student is not only related to how well one is currently attuned to learning or performing well in cognitive memory tasks but is also closely related to the appeal of the incentives provided for participation (e.g., course credit). However, these incentives also differed somewhat between the studies. Incentives are a form of external regulation, which itself is a type of extrinsic motivation, and known to affect the effect of task design elements on learning (Wesenberg et al., 2025). Hence, incentives potentially influence the efficacy of gamified learning via several pathways. Together with additional contextual features like the geographic location and the data acquisition period of the study (Table 1), they first affect selection of participants and hence, the composition of the sample. Second, they may directly affect the efficacy of gamified learning. While we do not think that the differences in incentives can fully explain the reported differences in learning outcomes between the three studies, more targeted future studies will need to further disentangle the combined effect of the various contextual and individual features on the efficacy of gamified learning, particularly for motivational outcomes.

While our present study warrants further investigation, future studies should take special care about varying only one contextual feature at a time. A typical criticism of meta-analyses is that they often *compare apples with oranges* (i.e., studies that vary considerably in design, methods, contexts, etc.; Borenstein et al., 2009). Whereas our present study may represent some progress in that regard by *comparing apples with pears* instead, dedicated research that actually *compares apples with apples* will be required to sort out those contextual features that are especially relevant for the efficacy of gamified learning from those that do not matter much if at all. However, the importance of identifying such boundary conditions goes beyond the specific case of gamified learning considered in the present work. A rich body of literature exists showing that basic cognitive functions like memory (Neath & Suprenant, 2003) can be highly sensitive to subtle changes in contextual factors resulting in environmental context-dependent memory effects (Smith & Vela, 2001). Owing to the definition of learning as an “*enduring change of the mechanisms of behavior*” (Domjan, 2010, p. 17, emphasis by us), memory is a crucial factor in learning per se. Hence, at a fundamental level, learning is sensitive to the exact conditions under which it occurs, and research on boundary conditions of fundamental learning mechanisms like memory remains of utmost importance to this day (Krauspe et al., 2025). However, the findings of our present study suggest that the sensitivity on specific contextual features might be even stronger in the case of affective (including motivational) than cognitive aspects of learning. This implies that future research should delve deeper into investigating the boundary conditions of learning interventions that focus particularly on learners’ motivation.

4.2 Limitations

The scope of contextual features considered in the present work and the conclusions it allows regarding boundary conditions of gamified learning have several limitations. First, our samples were represented mostly by university students, who likely do not represent the entire population of adult learners. In addition to that, studies differed in the fraction of participants who were not currently enrolled as students at a university. Future studies will need to assess the personal background with more scrutiny to resolve the impact of sample composition or more rigorously control the population from which participants are sampled (e.g., only undergraduate university students versus primary- or secondary-school students).



Second, we solely relied on self-report instruments to measure motivational outcomes. This approach assumes that motivation was always conscious and accessible to the individual, which is sometimes not the case. Some studies find that participants report inaccurate data for several reasons, including failure to accurately monitor their internal states (Fryer & Dinsmore, 2020). In a similar vein, motivation is a multi-dimensional construct, comprising of biological, physiological, social, and cognitive factors that influence behavior (Fulmer & Frijters, 2009). However, our research only measured a single dimension of motivation, which was itself only indirectly assessed via the two aspects of learner experience, that is, task attractivity and perceived stimulation. While both are at face value obviously associated with self-report measures of intrinsic motivation (Wilde et al., 2009), reflected also in previously noted associations with intrinsic and extrinsic motivation scales (Huber et al., 2023, 2024), their use as a proxy for intrinsic motivation was solely based on the pragmatic reason of being the only motivation-related measure common to all three compared studies and may not be optimal for providing a complete picture of the potential impact of contextual features on motivational outcomes. Future studies should go beyond this limitation by at least complementing those measures with validated instruments designed to measure motivational constructs, resolving, for instance, also aspects of extrinsic motivation (e.g., Guay et al., 2000), or subcomponents of motivation like autonomy or competence experiences (e.g., Wilde et al., 2009). Assessing the task and utility values of the learning task (e.g., Gaspard et al., 2020) in its different design versions and contexts could give an additional perspective that provides a link between aspects of motivation and provided participation incentives on the basis of expectancy-value theory (Eccles & Wigfield, 2020).

It can further not be excluded that participants may respond differently to self-report items in an anonymous online setting compared to a laboratory setting (Hoerger & Currell, 2012), which is on the one hand intricately related to our research questions but, on the other hand, goes beyond the specific scope of gamified learning. Lastly, the three studies differed in more contextual features than being conducted in the laboratory or online as discussed in detail in the previous section. As outlined above, future studies will need to further disentangle the combined impact on learning outcomes of the various contextual features realized by the studies compared in the present study. Nevertheless, it remains noteworthy what difference the change from an online to a laboratory setting under the most similar study conditions (studies 2 and 3) already made for the considered self-report and performance-based outcomes alone. Informed by adjacent research domains (D'Mello & Booth, 2023), similar or even more substantial differences can be expected when going from the laboratory further into the field.

4.3 Implications for building context-aware educational interventions

Our results emphasize that for an intervention to support learning as intended the context in which learning takes place needs to be explicitly considered. The findings are relevant to the educational community and present implications for designing context-aware educational interventions that benefit both cognitive learning and motivational outcomes with gamified learning tasks (Sailer et al., 2024). Based on a comparison of three empirical studies on gamified learning under varying contextual conditions, we subsequently illustrate what challenges and difficulties existing boundary conditions can impose on a proper understanding and appropriate interpretation of outcomes under general conditions. Our results suggest that in an online setting, gamification may effectively motivate learners to stay engaged with a task. However, in a more controlled environment, perhaps closer to the general outline of a controlled laboratory environment, this influence might cease to have any effect. This suggests that learning interventions utilizing gamification should be context-specific rather than universally applied.

We envision that context-aware interventions should be designed to dynamically adapt based on real-time feedback from learners *and* the learning environment. Using data collected in-situ may provide a means to provide interventions that can be adjusted on-the-fly to maintain engagement and optimize cognitive learning and motivational outcomes, ensuring they remain effective across different boundary conditions. Furthermore, different learners may respond differently to the same interventions depending on the context, highlighting the need for personalization. Context-aware interventions should



incorporate personalization strategies, adapting the intervention based on factors such as individual learning preferences, engagement levels, and the specific learning environment (e.g., online vs. laboratory vs. classroom setting; Ninaus & Sailer, 2022; Sailer et al., 2024).

To identify patterns that can be harnessed to drive adaptive learning support across conditions, we need to understand which data are informative about which aspect of learning and under which circumstances. One research area fully dedicated to this goal is learning analytics, which seeks to optimize learning and the environments in which it occurs (Long & Siemens, 2011). In learning analytics, the learning process is envisaged as a dynamic, closed-loop system (Clow, 2012; Ninaus & Sailer, 2022; Sailer et al., 2024), similarly to the theoretical considerations outlined in Section 1.1. This loop connects learners with behavioral traces (i.e., data) they leave upon interacting with a digital learning system, the patterns that are identified within those traces that influence a learning outcome, and the interventions derived from those patterns to enhance that learning outcome. In the final step of the cycle, the learning environment adapts with respect to identified patterns in a learner's traces and provides the learner with a new, adapted learning interface, through which the learning process continues in the next iteration of the cycle. A good understanding about what learning processes and outcomes can be expected under what circumstances is of utmost importance for devising such adaptive, closed-loop learning systems.

This begins already at the data acquisition stage, because different contexts may dictate how often and what types of data can be collected during learning. For instance, it is well-known that affect expression varies over contexts (Barrett et al., 2019; D'Mello et al., 2018). If, for example, automated affect detection is then used to probe affective dynamics, for instance, for the detection of signs of confusion or frustration during learning (Cloude et al., 2022), it might well turn out that insights from studies conducted in the laboratory are hardly, if at all, transferable into the field (D'Mello & Booth, 2023). A laboratory setting is not a typical setting for learning. However, it is a typical setting for conducting highly controlled research about learning (Booth et al., 2023).

Our present study also shows that results may considerably differ between laboratory and other (here, online learning) contexts. Hence, for the development of tailored educational interventions, it seems essential to know the boundary conditions under which a certain result will be obtained. A result obtained in the laboratory may turn out practically useless in the field and vice versa (D'Mello & Booth, 2023). As such, the development of context-aware interventions requires the creation of context-specific metrics for evaluating their effectiveness and to gain a more holistic understanding of how learners interact with different learning designs. For example, metrics that work well in online environments (e.g., time-on-task, click-through rates) may not be appropriate for controlled lab settings, where other factors (e.g., effort investment or self-reported engagement) may be more relevant.

Another area of consideration for future research should be on properly capturing cognitive and affective (also including motivational) fluctuations *within and between* contexts. While much of the research has focused on studying cognitive and affective processes in situ to understand cognitive changes in single tasks and contexts, future studies should also consider how learner cognition and affect fluctuates across contexts. While notoriously difficult, a promising approach for capturing cognitive and affective fluctuations is through the use of multimodal data collected across different tasks and learning environments (Molenaar et al., 2023; Ninaus & Sailer, 2022; Sailer et al., 2024). These multimodal learning analytics (Blikstein, 2013) could include eye movements, physiological signals, facial expressions, interactions with the learning system, and repeated self-reported motivation levels. By integrating and triangulating multiple data channels, researchers may gain a deeper understanding of how cognition, affect, and motivation change in real time and how varying contexts affect their outcomes.



5. Conclusion

In the present work, we set out to investigate how different contextual study features can influence the impact of gamified learning on cognitive learning and motivational outcomes in an associative learning task. We found a substantial influence of the considered contextual features for motivational outcomes. In particular, whereas gamified learning tasks were perceived as substantially more appealing and stimulating in online settings, they exhibited no motivational benefits over non-gamified tasks in a controlled, laboratory setting under otherwise similar conditions. The influence on cognitive learning outcomes appeared relatively more nuanced but aligned well with cognitive and cognitive-affective learning theories, indicating that if cognitive outcomes are affected by gamification, these influences are best illuminated in the laboratory for the given learning task. Although only a limited scope of varying contexts, a single learning task, and motivational outcomes solely based on self-reports were considered in this study, we nevertheless think that the present work shows the importance of considering boundary conditions of gamified learning in educational research.

Keypoints

- Boundary conditions of cognitive and motivational effects of gamified learning are explored.
- Indications for boundary conditions of motivational effects are identified.
- Motivational effects differ substantially between laboratory and online settings.
- Cognitive learning outcomes are comparatively robust across contexts.
- Cognitive effects of gamification can be affected by study context to some extent via motivational effects.

Acknowledgments

This work was supported by the Strategic Research Council (Grant No. 358250) within the Research Council of Finland. The authors further acknowledge financial support by the University of Graz.

References

- Alotaibi, M. S. (2024). Game-based learning in early childhood education: A systematic review and meta-analysis. *Frontiers in Psychology, 15*, 1307881. <https://doi.org/10.3389/fpsyg.2024.1307881>
- Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education, 17*(1), 3. <https://doi.org/10.1186/s41239-020-0177-7>
- Arztmann, M., Hornstra, L., Jeuring, J., & Kester, L. (2023). Effects of games in STEM education: A meta-analysis on the moderating role of student background characteristics. *Studies in Science Education, 59*(1), 109–145. <https://doi.org/10.1080/03057267.2022.2057732>
- Azevedo, R., & Wiedbusch, M. (2023). Theories of metacognition and pedagogy applied to AIED systems. In B. Du Boulay, A. Mitrovic, & K. Yacef (Eds.), *Handbook of Artificial Intelligence in Education* (pp. 45–67). Edward Elgar Publishing. <https://doi.org/10.4337/9781800375413.00013>
- Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review, 30*, 100322. <https://doi.org/10.1016/j.edurev.2020.100322>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.



- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, 20(1), 1–68. <https://doi.org/10.1177/1529100619832930>
- Barz, N., Benick, M., Dörrenbächer-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*, 003465432311677. <https://doi.org/10.3102/00346543231167795>
- Baten, E., Vansteenkiste, M., De Muynck, G.-J., De Poortere, E., & Desoete, A. (2020). How can the blow of math difficulty on elementary school children’s motivational, cognitive, and affective experiences be dampened? The critical role of autonomy-supportive instructions. *Journal of Educational Psychology*, 112(8), 1490–1505. <https://doi.org/10.1037/edu0000444>
- 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>
- Blikstein, P. (2013). Multimodal learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 102–106. <https://doi.org/10.1145/2460296.2460316>
- Booth, B. M., Bosch, N., & D’Mello, S. K. (2023). Engagement Detection and Its Applications in Learning: A Tutorial and Selective Review. *Proceedings of the IEEE*, 111(10), 1398–1422. <https://doi.org/10.1109/JPROC.2023.3309560>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis* (1st ed.). Wiley. <https://doi.org/10.1002/9780470743386>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- 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>
- Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 134–138. <https://doi.org/10.1145/2330601.2330636>
- Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1), 17–29. <https://doi.org/10.1109/TLT.2016.2616312>
- Conte, P. D. (2019). A gender study on the effects of the “high five game” on the math learning performance of children. *International Journal of Scientific and Technology Research*, 8(12), 2063–2066.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining “gamification.” *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. <https://doi.org/10.1145/2181037.2181040>
- D’Mello, S. K., & Booth, B. M. (2023). Affect Detection From Wearables in the “Real” Wild: Fact, Fantasy, or Somewhere In between? *IEEE Intelligent Systems*, 38(1), 76–84. <https://doi.org/10.1109/MIS.2022.3221854>



- D’Mello, S. K., Kappas, A., & Gratch, J. (2018). The Affective Computing Approach to Affect Measurement. *Emotion Review*, *10*(2), 174–183. <https://doi.org/10.1177/1754073917696583>
- Domjan, M. (2010). *The Principles of Learning and Behavior*. Wadsworth Cengage Learning.
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, *32*(3), 158–172. <https://doi.org/10.1007/s11031-008-9102-4>
- Fadda, D., Pellegrini, M., Vivanet, G., & Zandonella Callegher, C. (2022). Effects of digital games on student motivation in mathematics: A meta-analysis in K-12. *Journal of Computer Assisted Learning*, *38*(1), 304–325. <https://doi.org/10.1111/jcal.12618>
- Fiedler, K., & Beier, S. (2014). Affect and Cognitive Processes in Educational Contexts. In R. Pekrun & L. Linnenbrank-Garcia (Eds.), *International Handbook of Emotions in Education* (1st ed.). Routledge.
- Fryer, L. K., & Dinsmore, D. L. (2020). The Promise and Pitfalls of Self-report. *Frontline Learning Research*, *8*(3), 1–9. <https://doi.org/10.14786/flr.v8i3.623>
- Fulmer, S. M., & Frijters, J. C. (2009). A Review of Self-Report and Alternative Approaches in the Measurement of Student Motivation. *Educational Psychology Review*, *21*(3), 219–246. <https://doi.org/10.1007/s10648-009-9107-x>
- Gaspard, H., Jiang, Y., Piesch, H., Nagengast, B., Jia, N., Lee, J., & Bong, M. (2020). Assessing students’ values and costs in three countries: Gender and age differences within countries and structural differences across countries. *Learning and Individual Differences*, *79*, 101836. <https://doi.org/10.1016/j.lindif.2020.101836>
- Greipl, S., Klein, E., Lindstedt, A., Kiili, K., Moeller, K., Karnath, H.-O., Bahnmueller, J., Bloechle, J., & 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>
- Greipl, S., Moeller, K., & Ninaus, M. (2020). Potential and limits of game-based learning. *International Journal of Technology Enhanced Learning*, *12*(4), 363. <https://doi.org/10.1504/IJTEL.2020.110047>
- 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*(3), 175–213. <https://doi.org/10.1023/A:1005614228250>
- Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology* (2nd ed.). (pp. 3–24). American Psychological Association. <https://doi.org/10.1037/0000135-001>
- Hatano, A., Ogulmus, C., Shigemasu, H., & Murayama, K. (2022). Thinking about thinking: People underestimate how enjoyable and engaging just waiting is. *Journal of Experimental Psychology: General*, *151*(12), 3213–3229. <https://doi.org/10.1037/xge0001255>
- Heckhausen, J., & Heckhausen, H. (2018). Motivation and Action: Introduction and Overview. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and Action* (pp. 1–14). Springer International Publishing. https://doi.org/10.1007/978-3-319-65094-4_1



- 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. J. Knapp, M. C. Gottlieb, M. M. Handelsman, & L. D. VandeCreek (Eds.), *APA handbook of ethics in psychology, Vol 2: Practice, teaching, and research* (pp. 385–400). American Psychological Association.
- Hu, Y., Gallagher, T., Wouters, P., Van Der Schaaf, M., & Kester, L. (2022). Game-based learning has good chemistry with chemistry education: A three-level meta-analysis. *Journal of Research in Science Teaching*, *59*(9), 1499–1543. <https://doi.org/10.1002/tea.21765>
- Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development*, *68*(4), 1875–1901. <https://doi.org/10.1007/s11423-020-09807-z>
- Huber, S. E., Cortez, R., Kiili, K., Lindstedt, A., & Ninaus, M. (2023). Game elements enhance engagement and mitigate attrition in online learning tasks. *Computers in Human Behavior*, *149*, 107948. <https://doi.org/10.1016/j.chb.2023.107948>
- Huber, S. E., Edlinger, M., Lindstedt, A., Kiili, K., & Ninaus, M. (2024). Game elements improve affect and motivation in a learning task. *International Journal of Serious Games*, *11*(4), 103–126. <https://doi.org/10.17083/ijsg.v11i4.769>
- JASP Team. (2024). *JASP* (Version 0.19.1.0) [Computer software].
- 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>
- Krauspe, J., Ebersbach, M., Ludwig, A., & Scharf, F. (2025). Do worked examples boost the spacing effect on lasting learning? *Learning and Instruction*, *97*, 102103. <https://doi.org/10.1016/j.learninstruc.2025.102103>
- Kuratomi, K., Johnsen, L., Kitagami, S., Hatano, A., & Murayama, K. (2023). People underestimate their capability to motivate themselves without performance-based extrinsic incentives. *Motivation and Emotion*, *47*(4), 509–523. <https://doi.org/10.1007/s11031-022-09996-5>
- Lampropoulos, G. & Kinshuk. (2024). Virtual reality and gamification in education: A systematic review. *Educational Technology Research and Development*, *72*(3), 1691–1785. <https://doi.org/10.1007/s11423-024-10351-3>
- Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Ed.), *HCI and Usability for Education and Work* (Vol. 5298, pp. 63–76). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-89350-9_6
- Lewin, K. (1936). *Principles of topological psychology*. McGraw-Hill.
- Lewin, K. (1946). Behavior and development as a function of the total situation. In L. Carmichael (Ed.), *Manual of child psychology*. (pp. 791–844). John Wiley & Sons Inc. <https://doi.org/10.1037/10756-016>
- Li, L., Hew, K. F., & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: A meta-analysis and systematic review. *Educational Technology Research and Development*. <https://doi.org/10.1007/s11423-023-10337-7>



- Long, P. D., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*. <https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education>
- Mavridis, A., Katmada, A., & Tsiatsos, T. (2017). Impact of online flexible games on students' attitude towards mathematics. *Educational Technology Research and Development*, 65(6), 1451–1470. <https://doi.org/10.1007/s11423-017-9522-5>
- Mayer, R. E. (2011). *Applying the Science of Learning*. Pearson.
- Mayer, R. E. (2014). *Computer Games for Learning: An Evidence-Based Approach*. MIT Press.
- Mayer, R. E. (2019). Computer Games in Education. *Annual Review of Psychology*, 70(1), 531–549. <https://doi.org/10.1146/annurev-psych-010418-102744>
- Mayer, R. E. (2020). Cognitive foundations of game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 83–110). MIT Press.
- Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, 36(1), 8. <https://doi.org/10.1007/s10648-023-09842-1>
- Molenaar, I., Mooij, S. D., Azevedo, R., Bannert, M., Järvelä, S., & Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior*, 139, 107540. <https://doi.org/10.1016/j.chb.2022.107540>
- Neath, I., & Suprenant, A. M. (2003). *Human Memory*. Wadsworth/Thomson Learning.
- Ninaus, M., Greipl, S., Kiili, K., Lindstedt, A., Huber, S., Klein, E., Karnath, H.-O., & 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., & Sailer, M. (2022). Closing the loop – The human role in artificial intelligence for education. *Frontiers in Psychology*, 13, 956798. <https://doi.org/10.3389/fpsyg.2022.956798>
- Palha, S., & Jukić Matić, L. (2025). What do teachers anticipate from education in game-based pedagogy? *Technology, Pedagogy and Education*, 1–14. <https://doi.org/10.1080/1475939X.2025.2454453>
- 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* (pp. 3–24). MIT Press.
- Plass, J. L., & Kaplan, U. (2016). Emotional Design in Digital Media for Learning. In *Emotions, Technology, Design, and Learning* (pp. 131–161). Elsevier. <https://doi.org/10.1016/B978-0-12-801856-9.00007-4>
- Prencak, B., Velardi, P., Stilo, G., Distanto, D., & Faralli, S. (2021). A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses. *ACM Computing Surveys*, 53(3), 1–34. <https://doi.org/10.1145/3388792>
- Ritzhaupt, A. D., Huang, R., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., & Li, J. (2021). A meta-analysis on the influence of gamification in formal educational settings on affective and behavioral outcomes. *Educational Technology Research and Development*, 69(5), 2493–2522. <https://doi.org/10.1007/s11423-021-10036-1>
- Rodrigues, L., Pereira, F., Toda, A., Palomino, P., Oliveira, W., Pessoa, M., Carvalho, L., Oliveira, D., Oliveira, E., Cristea, A., & Isotani, S. (2022). Are They Learning or Playing? Moderator



- Conditions of Gamification's Success in Programming Classrooms. *ACM Transactions on Computing Education*, 22(3), 1–27. <https://doi.org/10.1145/3485732>
- Ryan, R. M., & Deci, E. L. (2019). Brick by Brick: The Origins, Development, and Future of Self-Determination Theory. In *Advances in Motivation Science* (Vol. 6, pp. 111–156). Elsevier. <https://doi.org/10.1016/bs.adms.2019.01.001>
- Ryan, R. M., & Rigby, C. S. (2020). Motivational foundations of game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 153–176). MIT Press.
- Sailer, M., & Homner, L. (2020). The Gamification of Learning: A Meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Sailer, M., Ninaus, M., Huber, S. E., Bauer, E., & Greiff, S. (2024). The End is the Beginning is the End: The closed-loop learning analytics framework. *Computers in Human Behavior*, 158, 108305. <https://doi.org/10.1016/j.chb.2024.108305>
- Sailer, M., & Sailer, M. (2021). Gamification of in-class activities in flipped classroom lectures. *British Journal of Educational Technology*, 52(1), 75–90. <https://doi.org/10.1111/bjet.12948>
- 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, 104058. <https://doi.org/10.1016/j.ijnurstu.2021.104058>
- Schlag, R., Sailer, M., Tolks, D., Ninaus, M., & Sailer, M. (2024). Effectiveness of gamification in education. In A. Gegenfurtner & I. Kollar (Eds.), *Designing Effective Digital Learning Environments* (pp. 143–159). Routledge.
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). *Die UX KPI - Wunsch und Wirklichkeit*. <https://doi.org/10.18420/MUC2017-UP-0100>
- 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). MIT Press.
- Seo, K., Dodson, S., Harandi, N. M., Roberson, N., Fels, S., & Roll, I. (2021). Active learning with online video: The impact of learning context on engagement. *Computers & Education*, 165, 104132. <https://doi.org/10.1016/j.compedu.2021.104132>
- Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The Challenges of Defining and Measuring Student Engagement in Science. *Educational Psychologist*, 50(1), 1–13. <https://doi.org/10.1080/00461520.2014.1002924>
- Skitka, L. J., & Sargis, E. G. (2006). The Internet as Psychological Laboratory. *Annual Review of Psychology*, 57(1), 529–555. <https://doi.org/10.1146/annurev.psych.57.102904.190048>
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, 8(2), 203–220. <https://doi.org/10.3758/BF03196157>
- Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Elsevier. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Urhahne, D., & Wijnia, L. (2023). Theories of Motivation in Education: An Integrative Framework. *Educational Psychology Review*, 35(2), 45. <https://doi.org/10.1007/s10648-023-09767-9>
- Wesenberg, L., Jansen, S., Krieglstein, F., Schneider, S., & Rey, G. D. (2025). The influence of seductive details in learning environments with low and high extrinsic motivation. *Learning and Instruction*, 96, 102054. <https://doi.org/10.1016/j.learninstruc.2024.102054>
- Wilcox, R. R. (2022). *Introduction to robust estimation and hypothesis testing* (5th ed.). Academic Press.



- Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzsкала intrinsischer Motivation (KIM). *Zeitschrift Für Didaktik Der Naturwissenschaften*, 15, 31–45.
- Wong, Z. Y., & Liem, G. A. D. (2022). Student Engagement: Current State of the Construct, Conceptual Refinement, and Future Research Directions. *Educational Psychology Review*, 34(1), 107–138. <https://doi.org/10.1007/s10648-021-09628-3>
- 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>
- Xu, Z., Olson, J., Pochinki, N., Zheng, Z., & Yu, R. (2024). Contexts Matter but How? Course-Level Correlates of Performance and Fairness Shift in Predictive Model Transfer. *Proceedings of the 14th Learning Analytics and Knowledge Conference*, 713–724. <https://doi.org/10.1145/3636555.3636936>
- 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, 100326. <https://doi.org/10.1016/j.edurev.2020.100326>
- Zhang, L., Lei, Y., Pelton, T., Pelton, L. F., & Shang, J. (2024). An exploration of gendered differences in cognitive, motivational and emotional aspects of game-based math learning. *Journal of Computer Assisted Learning*, 40(6), 2633–2649. <https://doi.org/10.1111/jcal.12956>
- Zhang, Q., & Yu, Z. (2022). Meta-Analysis on Investigating and Comparing the Effects on Learning Achievement and Motivation for Gamification and Game-Based Learning. *Education Research International*, 2022, 1–19. <https://doi.org/10.1155/2022/1519880>