META-ANALYSIS

The Gamification of Learning: a Meta-analysis

Michael Sailer [1] · Lisa Homner [1]

Abstract

This meta-analysis was conducted to systematically synthesize research findings on effects of gamification on cognitive, motivational, and behavioral learning outcomes. Results from random effects models showed significant small effects of gamification on cognitive ($g = .49$, 95% CI [0.30, 0.69], $k = 19$, $N = 1686$), motivational ($g = .36$, 95% CI [0.18, 0.54], $k = 16$, $N = 2246$), and behavioral learning outcomes ($g = .25$, 95% CI [0.04, 0.46], $k = 9$, $N = 951$). Whereas the effect of gamification on cognitive learning outcomes was stable in a subsplit analysis of studies employing high methodological rigor, effects on motivational and behavioral outcomes were less stable. Given the heterogeneity of effect sizes, moderator analyses were conducted to examine *inclusion of game fiction*, *social interaction*, *learning arrangement of the comparison group*, as well as situational, contextual, and methodological moderators, namely, *period of time*, *research context*, *randomization*, *design*, and *instruments*. Inclusion of game fiction and social interaction were significant moderators of the effect of gamification on behavioral learning outcomes. Inclusion of game fiction and combining competition with collaboration were particularly effective within gamification for fostering behavioral learning outcomes. Results of the subsplit analysis indicated that effects of competition augmented with collaboration might also be valid for motivational learning outcomes. The results suggest that gamification as it is currently operationalized in empirical studies is an effective method for instruction, even though factors contributing to successful gamification are still somewhat unresolved, especially for cognitive learning outcomes.

✉ Michael Sailer
Michael.Sailer@psy.lmu.de

Lisa Homner
lisahomner@gmail.com

[1] Chair of Education and Educational Psychology, LMU Munich, Munich, Germany

In recent years, the concept of gamification, defined as "the use of game design elements in non-game contexts" (Deterding et al. 2011, p. 9), has received increased attention and interest in academia and practice, with education among the top fields of gamification research (Dichev and Dicheva 2017; Hamari et al. 2014; Seaborn and Fels 2015). Its hypothesized motivational power has made gamification an especially promising method for instructional contexts. However, as the popularity of gamification has increased, so have critical voices describing gamification as "the latest buzzword and the next fad" (Boulet 2012, p. 1) or "Pavlovication" (Klabbers 2018, p. 232). But how effective is gamification when it comes to learning, and what factors contribute to successful gamification?

Even though considerable research efforts have been made in this field (Hamari et al. 2014; Seaborn and Fels 2015), conclusive meta-analytic evidence on the effectiveness of gamification in the context of learning and education has yet to be provided. Therefore, the aim of this analysis was to statistically synthesize the state of research on the effects of gamification on cognitive, motivational, and behavioral learning outcomes in an exploratory manner. Furthermore, with this meta-analysis, not only did we try to answer the question of *whether* learning should be gamified but also *how*. Thus, we also investigated potential moderating factors of successful gamification to account for conceptual heterogeneity in gamification (Sailer et al. 2017a). What is more, we included contextual, situational, and methodological moderators to account for different research contexts and study setups as well as methodological rigor in primary studies. To further investigate the stability and robustness of results, we assessed publication bias and computed subsplit analyses, which included studies with high methodological rigor.

Effects of Gamification on Learning

Gamification in the context of learning can be referred to as gamified learning (see Armstrong and Landers 2017; Landers 2014). Even though gamified learning and game-based learning have overlapping research literatures, a common game design element toolkit (Landers et al. 2018), and the same focus on adding value beyond entertainment, that is, using the entertaining quality of gamification interventions or (serious) games for learning (see Deterding et al. 2011; Zyda 2005), they are different in nature. Whereas game-based learning approaches imply the design of fully fledged (serious) games (see Deterding et al. 2011), gamified learning approaches focus on augmenting or altering an existing learning process to create a revised version of this process that users experience as game-like (Landers et al. 2018). Thus, gamification is not a product in the way that a (serious) game is; gamification in the context of learning is a design process of adding game elements in order to change existing learning processes (see Deterding et al. 2011; Landers et al. 2018).

Although many studies that have examined gamification have lacked a theoretical foundation (Hamari et al. 2014; Seaborn and Fels 2015), some authors have attempted to explain the relationship between gamification and learning by providing frameworks such as the theory of gamified learning (Landers 2014). This theory defines four components: instructional content, behaviors and attitudes, game characteristics, and learning outcomes. The theory proposes that instructional content directly influences learning outcomes as well as learners' behavior. Because gamification is usually not used to replace instruction, but rather to improve it, effective instructional content is a prerequisite for successful gamification (Landers 2014). The aim of gamification is to directly affect behaviors and attitudes relevant to learning. In turn, these behaviors and attitudes are

hypothesized to affect the relationship between the instructional content and learning outcomes via either moderation or mediation, depending on the nature of the behaviors and attitudes targeted by gamification (Landers 2014). The theory of gamified learning proposes a positive, indirect effect of gamification on learning outcomes. However, it is important to note that this theory provides no information about effective learning mechanisms triggered by game design elements. Such mechanisms can be found in well-established psychological theories such as self-determination theory (Ryan and Deci 2002).

Self-determination theory has already been successfully applied in the contexts of games (see Rigby and Ryan 2011) and gamification (see Mekler et al. 2017; Sailer et al. 2017a). It postulates psychological needs for competence, autonomy, and social relatedness. The satisfaction of these needs is central for intrinsic motivation and subsequently for high-quality learning (see Ryan and Deci 2000), with self-determination theory emphasizing the importance of the environment in satisfying these psychological needs (Ryan and Deci 2002). Enriching learning environments with game design elements modifies these environments and potentially affects learning outcomes. From the self-determination perspective, different types of feedback can be central learning mechanisms triggered by game design elements. Constantly providing learners with feedback is a central characteristic of serious games (Wouters et al. 2013; Prensky 2001) and of gamification (Werbach and Hunter 2012). Although the effectiveness of feedback has been shown to vary depending on a variety of criteria such as the timing of feedback (immediate, delayed), frame of reference (criterial, individual, social), or level of feedback (task, process, self-regulation, self), feedback is among the most powerful factors in the relationship between educational interventions and learning in general (Hattie and Timperley 2007). Based on the theory of gamified learning as well as self-determination theory, gamification might influence learning outcomes in a positive way.

Previous research that has attempted to synthesize effects of gamification on learning outcomes has done this almost exclusively with (systematic) reviews. An exception is an early meta-analysis on gamification in educational contexts that was based on data from 14 studies published between 2013 and 2015 (Garland 2015). Under a random effects model, this meta-analysis identified a positive, medium-sized correlation between the use of gamification and learning outcomes ($r = .31$, 95% CI [0.11, 0.47]). It should be noted that learning outcomes in this case almost exclusively pertained to motivational outcomes, except for one study that investigated knowledge retention. Though promising, these results should be viewed with caution because they were methodologically limited. The analysis was highly likely to be underpowered and had only limited generalizability due to the small sample size. Furthermore, r as an estimator of effect size cannot correct for the bias caused by small samples, which can lead to an overestimation of the effect. Also, no evaluation of publication bias was attempted.

Apart from this analysis, a series of reviews providing at least the general tendency of research findings in the field have been conducted. A mutlidisciplinary literature review identified 24 peer-reviewed empirical studies published between 2008 and 2013 (Hamari et al. 2014). Nine of the 24 studies were conducted in the context of education and learning. All in all, the results were mixed, despite the presence of a positive tendency suggesting the existence of confounding variables. Seaborn and Fels (2015) returned concordant findings in their review of eight empirical studies (published between 2011 and 2013) from the domain of education; they also identified a positive tendency but predominantly found mixed results. Results from these reviews were attributed to differences in gamified contexts and participant characteristics, but effects of novelty and publication bias were also discussed as possible reasons for these mixed findings.

A series of reviews specifically focusing on educational contexts was conducted by Dicheva et al. (2015), Dicheva and Dichev (2015), and Dichev and Dicheva (2017). The review by Dicheva et al. (2015) consisted of 34 empirical studies published between 2010 and 2014. The majority of experiments ($k = 18$) reported positive results for gamification on various motivational, behavioral, and cognitive variables. Using the same search strategies as Dicheva et al. (2015), a subsequent review by Dicheva and Dichev (2015) resulted in 41 reports published between July 2014 and June 2015. The majority of studies were identified as inconclusive due to methodological inadequacies. Only 10 reports provided conclusive positive evidence, and three studies showed negative effects. Dichev and Dicheva (2017) followed up by conducting another subsequent literature search focusing on studies published between July 2014 and December 2015. They identified 51 additional studies, of which 41 experiments investigated the effect of gamification on motivational, behavioral, and cognitive outcomes. Whereas 12 experiments reported positive results, three reported negative results. Again, the majority of the experiments ($k = 26$) were inconclusive. The large numbers of inconclusive studies in these reviews point toward a general problem: Gamification research lacks methodological rigor.

According to meta-analytic results and several reviews, gamification tends to have positive effects on different kinds of learning outcomes, albeit with mixed results. Thus, in this meta-analysis, we statistically synthesized the state of current research on gamification to investigate the extent to which gamification affects cognitive, motivational, and behavioral learning outcomes compared with conventional instructional methods.

Moderating Factors

Gamification applications can be very diverse, and research has often failed to acknowledge that there are many different game design elements at work that can result in different affordances for learners, modes of social interactions, and learning arrangements (Sailer et al. 2017a). Thus, we included different moderating factors to account for conceptual heterogeneity in gamification. Further, because contextual and situational factors might influence the effects of gamification on learning outcomes (Hamari et al. 2014), and gamification research lacks methodological rigor (Dicheva and Dichev 2015; Dichev and Dicheva 2017), we also included contextual, situational, and methodological moderators. The process of choosing potential moderating factors for the effects of gamification on learning outcomes was iterative in nature. We included moderating factors that were both theoretically interesting and which the literature was large enough to support their inclusion.

Inclusion of Game Fiction

From a self-determination theory perspective, game contexts can potentially satisfy the need for autonomy and relatedness by including choices, volitional engagement, sense of relevance, and shared goals (Rigby and Ryan 2011). These are assumed to be triggered by meaningful stories, avatars, nonplayer characters, or (fictional) teammates (Rigby and Ryan 2011; Sailer et al. 2017a). A shared attribute of these elements is that they provide narrative characteristics or introduce a game world, both of which include elements of fantasy (Bedwell et al. 2012; Garris et al. 2002). In general, they focus on *game fiction*, which is defined as the inclusion of a

fictional game world or story (Armstrong and Landers 2017). Inclusion of game fiction is closely related to the use of narrative anchors and has been shown to be effective for learning as it situates and anchors learning in a context (Clark et al. 2016) and can further serve as a cognitive framework for problem solving (Dickey 2006). In the context of games, results on the effectiveness of the inclusion of game fiction have been mixed (Armstrong and Landers 2017). Whereas Bedwell et al. (2012) found positive effects of the inclusion of game fiction on knowledge and motivation in their review, meta-analyses by others such as Wouters et al. (2013) found that serious games with narrative elements are not more effective than serious games without narrative elements. Thus, we investigated whether the use of game fiction moderates the effects of gamification on cognitive, motivational, and behavioral learning outcomes.

Social Interaction

The impact of relatedness in interpersonal activities can be crucial (Ryan and Deci 2002). Therefore, the type of social interaction that is likely to occur as a result of gamification could affect its relationship with learning outcomes. Collaboration and competition can be regarded as particularly important in this context (Rigby and Ryan 2011).

The term *collaboration* in this meta-analysis subsumes both collaborative and cooperative learning arrangements (i.e., situations in which learners work together in groups to achieve a shared goal), while being assessed either as a group (collaborative) or individually (cooperative; Prince 2004). In the broader context of games, collaboration has the potential to affect the needs for both relatedness and competence (Rigby and Ryan 2011). Collaboration not only allows for team work and thus the experience of being important to others, but it also enables learners to master challenges they otherwise might not be able to overcome on their own, which can result in feelings of competence.

Competition can cause social pressure to increase learners' level of engagement and can have a constructive effect on participation and learning (Burguillo 2010). However, it also has the potential to either enhance or undermine intrinsic motivation (Rigby and Ryan 2011). In this context, two types of competition can be distinguished. On the one hand, *destructive* competition occurs if succeeding by tearing others down is required, resulting in feelings of irrelevance and oppression. On the other hand, *constructive* competition occurs if it is good-natured and encourages cooperation and mutual support (i.e., if competition is aimed at improving everyone's skills instead of defeating someone). In this sense, constructive competition has the potential to foster feelings of relatedness, thereby enhancing intrinsic motivation (Rigby and Ryan 2011).

Collaboration—as well as competition augmented by aspects of collaboration (i.e., constructive competition)—can have additional beneficial effects on intrinsic motivation when compared with solitary engagement in an activity; as in cases of collaboration and competition augmented by aspects of collaboration, the need for relatedness is fostered additionally. Mere competition, however, can thwart feelings of relatedness when the goal is to defeat each other rather than to improve skills together (see Rigby and Ryan 2011).

Findings from the context of games have shown that collaborative gameplay can be more effective than individual gameplay (Wouters et al. 2013). Clark et al. (2016) included competition in their meta-analysis on digital games and found that combinations of competition and collaboration as well as single-player games without competitive elements can outperform games with mere competition. In this meta-analysis, we investigated whether

different types of social interaction moderate the effects of gamification on cognitive, motivational, and behavioral learning outcomes.

Learning Arrangement of the Comparison Group

Active engagement in cognitive processes is necessary for effective and sustainable learning as well as deep levels of understanding (see Wouters et al. 2008). This emphasis on active learning in educational psychology is aligned with the (inter)active nature of games (Wouters et al. 2013). Similar to games, gamification also has high potential to create instructional affordances for learners to engage in active learning. Based on the theory of gamified learning, gamification is assumed to affect learning outcomes by enhancing the attitudes and behaviors that are relevant for learning (e.g., when rewards for taking high-quality notes are provided in gamification; Landers 2014). A prerequisite is that the behavior or attitude that is targeted by gamification must itself influence learning (Landers 2014) and thus create instructional affordances for learners to actively engage in cognitive processes with the learning material. However, how learners interact with the environment has to be considered because learners may interact with the environment in different ways and carry out certain learning activities whether or not they are intended by gamification designers or researchers (see Chi and Wylie 2014; Young et al. 2012).

Further, in between-subject studies, which were included in this meta-analysis, the learning arrangement of the comparison condition, against which gamification was contrasted, is crucial (see Chi and Wylie 2014). Learners in a comparison condition can receive different prompts or instructions to engage in different learning activities and thus bias the effects of gamification. Therefore, it is important to differentiate between the passive and active instructions of comparison groups. Whereas passive instruction includes listening to lectures, watching instructional videos, and reading textbooks, active instruction involves explicitly prompting the learners to engage in learning activities (e.g., assignments, exercises, laboratory experiments; Sitzmann 2011; Wouters et al. 2013). Similar to the approach used by Wouters et al. (2013) in the context of games and Sitzmann (2011) in the context of simulation games, we included the comparison group's learning arrangement as a potential moderator of effects of gamification on cognitive, motivational, and behavioral learning outcomes.

Apart from these moderators, situational and contextual moderators were also included in the analysis. Thus, we included the period of time in which gamification was applied and the research context as potential moderators.

Period of Time

Previous reviews have indicated that the period of time during which gamification was used and investigated in primary studies has shown substantial variance (Dichev and Dicheva 2017; Seaborn and Fels 2015). One the one hand, reviews have raised the question of whether effects of gamification persist in the long run (Dichev and Dicheva 2017; Hamari et al. 2014; Seaborn and Fels 2015). On the other hand, research in the context of games has indicated that the effects of games are larger when players engage in multiple sessions and thus play over longer periods of time (Wouters et al. 2013). Thus, we included the period of time gamification was used as a potential moderator of its effects on cognitive, motivational, and behavioral learning outcomes.

Research Context

Gamification has been studied in different research contexts. Whereas the majority of studies found in reviews focusing on education were conducted in higher education settings, some of them were performed in primary and secondary school settings (see Dichev and Dicheva 2017). Further, some studies have been deployed in the context of further education or work-oriented learning (e.g., Lombriser et al. 2016) or informal settings with no reference to formal, higher, or work-related education (e.g., Sailer et al. 2017a). Therefore, this meta-analysis includes the context of research as a moderating factor for effects of gamification.

As meta-analytic methods allow for a synthesis of studies using different designs and instruments, the degree of methodological rigor can vary between primary studies and thus jeopardize the conclusions drawn from meta-analyses (Wouters et al. 2013). Because studies in the context of gamification have often lacked methodological rigor (Dicheva and Dichev 2015, 2017), we included methodological factors to account for possible differences in methodological study design and rigor across primary studies.

Randomization

Randomly assigning learners to experimental conditions allows researchers to rule out alternative explanations for differences in learning outcomes between different conditions. However, quasi-experimental studies do not allow alternative explanations to be ruled out (Wouters et al. 2013; Sitzmann 2011). For this reason, we included randomization as a moderating factor to account for methodological rigor.

Design

Besides randomization, the design of the primary studies can indicate methodological rigor. Primary studies using posttest-only designs cannot account for prior knowledge or initial motivation. The administration of pretests is particularly relevant for quasi-experimental studies because effects can be biased first by not randomly assigning learners to conditions and second by not controlling for learners' prior knowledge and motivation. Thus, we included design as a moderating factor to further account for methodological rigor in primary studies.

Instruments

Previous reviews have indicated that primary studies investigating gamification have often not used properly validated psychometric measurements to assess relevant outcomes (Dichev and Dicheva 2017; Hamari et al. 2014; Seaborn and Fels 2015). The use of standardized instruments can help to ensure the comparability of study results and further ensure the reliable measurement of variables (see Hamari et al. 2014). Thus, we included the type of instruments used in primary studies as a moderating factor of effects of gamification on cognitive, motivational, and behavioral learning outcomes.

To sum up, in this meta-analysis, we aimed to statistically synthesize the state of current research on gamification on cognitive, motivational, and behavioral learning outcomes. Further, we included the moderating factors *inclusion of game fiction*, *social interaction*, *learning arrangement of the comparison group*, *period of time*, *research context*, *randomization*, *design*, and *instruments* that potentially affect the relationships between gamification and

cognitive, motivational, and behavioral learning outcomes. Further, we investigated publication bias and the stability of the findings by performing subsplits that included only primary studies that applied high methodological rigor.

Method

Literature Search

In order to maximize the sensitivity of the search and adopting the search criteria used in the review by Seaborn and Fels (2015), the terms *gamification* and *gamif** were used in the academic literature search in all subject areas. Specific game design elements were not used as search terms because gamification research lacks consistent terms, definitions, and taxonomies for game design elements and agreed upon lists of game design elements that claim to be exhaustive are lacking (see Landers et al. 2018; Sailer et al. 2017a). Including specific game design elements as search terms would put the analysis in danger of being biased toward specific game design elements while leaving out others that are less common and thus potentially not part of the search terms. The years of publication were not restricted. The literature search was conducted on March 3, 2017. We searched the following academic data bases: ACM Digital Library, ERIC, IEEE Xplore, JSTOR, PubMed, ScienceDirect, and SpringerLink. Citations from these data bases were directly exported from the websites. Aside from this search in academic data bases, we conducted a Google Scholar search with the same terms to further maximize the scope of our literature search. To retrieve citations from Google Scholar, we used Publish or Perish (version 5), a software tool that allows all accessible search results to be automatically downloaded. We also screened the reference lists from Garland's (2015) meta-analysis and the reviews by Hamari et al. (2014), Seaborn and Fels (2015), Dicheva et al. (2015), Dicheva and Dichev (2015), and Dichev and Dicheva (2017). After removing duplicates, the literature search resulted in a total of 5548 possibly eligible records. An overview of the total number of search results per data base is found in Fig. 1. Google Scholar only shows a maximum of 1000 search results per search query. Therefore, 1100 denotes the number of records identified and retrieved from separate queries using *gamification* and *gamif**.

Inclusion and Exclusion Criteria

Gamification Studies were required to include at least one condition in which gamification, defined as "the use of game design elements in non-game contexts" (Deterding et al. 2011, p. 9), was used as an independent variable. Accordingly, studies describing interventions using the term gamification were excluded if the definition of gamification used in this meta-analysis did not apply to the intervention.

Learning Outcomes Eligible studies were required to assess at least one learning outcome. Learning outcomes were divided into three categories: cognitive, motivational, and behavioral learning outcomes. *Cognitive learning outcomes* refer to conceptual knowledge or application-oriented knowledge. Conceptual knowledge contains knowledge of facts, principles, and concepts, whereas application-oriented knowledge comprises procedural knowledge, strategic knowledge, and situational knowledge (de Jong and Ferguson-Hessler 1996). Adopting a
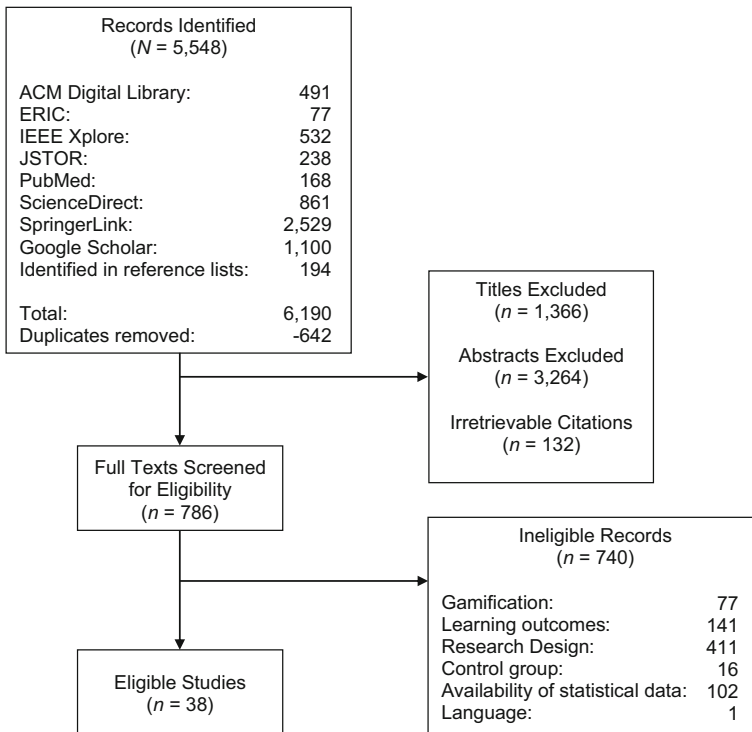
Fig. 1 Study flow diagram

broad view of motivation (see Wouters et al. 2013), m*otivational learning outcomes* encompass (intrinsic) motivation, dispositions, preferences, attitudes, engagement, as well as feelings of confidence and self-efficacy. *Behavioral learning outcomes* refer to technical skills, motor skills, or competences, such as learners' performance on a specific task, for example, a test flight after aviation training (Garris et al. 2002).

Language Eligible studies were required to be published in English.

Research Design Only primary studies applying quantitative statistical methods to examine samples of human participants were eligible. Furthermore, descriptive studies that did not compare different groups were excluded because the data obtained from such studies does not allow effect sizes to be calculated.

Control Group Studies were required to use a between-subject design and to compare at least one gamification condition with at least one condition involving another instructional approach. As the goal of this meta-analysis was to investigate the addition of game design elements to systems that did not already contain them, studies comparing gamification with fully fledged games were ineligible.

Availability of Statistical Data Studies were required to report sufficient statistical data to allow for the application of meta-analytic techniques.

Coding Procedure

First, we screened all titles for clearly ineligible publications (e.g., publications published in languages other than English). Next, we coded all remaining abstracts for eligibility. Finally, the remaining publications were retrieved and coded at the full-text level. An overview of the number of search results excluded per step is found in Fig. 1.

Moderator coding for eligible studies was then performed by two independent coders on a random selection of 10 studies (approximately 25%), with interrater reliability ranging from $\kappa = .76$ to perfect agreement. Remaining coding discrepancies were discussed between the authors until mutual agreement was reached before coding the remaining studies. Furthermore, all statistical data were double-coded, and differing results were recalculated by the authors. Finally, data on moderator variables were extracted. The previously introduced moderators, which potentially influence the effectiveness of gamification on learning outcomes, were coded as follows.

Inclusion of Game Fiction Studies using game fiction by providing a narrative context or introducing a game world (e.g., meaningful stories or avatars) were coded *yes*, whereas studies that did not use game fiction were coded *no*. For this moderator, interrater reliability was $\kappa = .76$.

Social Interaction Studies in which learners competed against each other or nonplayer characters during gamified interventions were coded as *competitive*, whereas studies in which learners collaborated with each other or nonplayer characters were coded as *collaborative*. If a study included both competitive and collaborative elements, it was assigned the code *competitive-collaborative*. Studies in which learners engaged in a learning activity entirely on their own were coded as *none*. For this moderator, interrater reliability was $\kappa = .83$.

Learning Arrangement of the Comparison Group *Passive* instructional methods of the comparison group include listening to lectures, watching instructional videos, and reading textbooks. *Active* instruction refers to learning arrangements that explicitly prompt learners to engage in learning activities (e.g., assignments, exercises, laboratory experiments; see Sitzmann 2011). *Mixed* instruction refers to a combination of passive and active instructional methods (see Wouters et al. 2013). Studies using a waitlist condition as a control group were coded as *untreated*. For this moderator, interrater reliability was $\kappa = .86$.

Period of Time The duration of the intervention was operationalized as the period of time over which the intervention took place. Studies were assigned to one of the categories: *1 day or less*, *1 week or less* (but longer than 1 day), *1 month or less* (but longer than 1 week), *half a year or less* (but longer than 1 month), or *more than half a year*. For this moderator, coders achieved perfect agreement.

Research Context Depending on the research context, studies were coded as *school setting* or *higher education setting*. Studies in the context of further education or work-oriented learning were coded as *work-related learning setting*. Studies with no reference to formal, higher, or work-related education were coded as *informal training setting*. For this moderator, coders achieved perfect agreement.

Randomization If participants were assigned to the experimental and control groups randomly, the study was coded as *experimental*, whereas publications using nonrandom assignment were assigned the value *quasi-experimental*. For this moderator, coders reached perfect agreement.

Design Furthermore, studies were coded on the basis of whether they exclusively used posttest measures (*posttest only*) or also administered a pretest (*pre- and posttest*). For this moderator, coders achieved perfect agreement.

Instruments Studies using preexisting, standardized instruments to measure the variables of interest were coded as *standardized*, whereas studies using adapted versions of standardized measures were coded as *adapted*. If the authors developed a new measure, studies were assigned the value *self-developed*. For this moderator, coders reached perfect agreement.

Final Sample

Application of the exclusion criteria detailed above resulted in a final sample of 38 publications reporting 40 experiments. Three of the studies (Wasson et al. 2013; Grivokostopoulou et al. 2016; Sanmugam et al. 2016) were excluded from the analysis to avoid bias after the forest plot; displaying the effect sizes for the analysis of cognitive learning outcomes showed that the corresponding effect sizes were extraordinarily large (see Fig. 2). An examination of the forest plots for motivational and behavioral learning outcomes revealed no remarkable outliers (see Fig. 2). Furthermore, the studies by de-Marcos et al. (2014) and Su and Cheng (2014) plus the first of two studies reported by Hew et al. (2016) were excluded from the analysis because they reported the same experiments as previously published or more extensive studies included in this sample.

The final sample of studies reporting cognitive learning outcomes comprised 19 primary studies reporting 19 independent experiments. They were published between 2013 and 2017 and examined a total of 1686 participants. Considering motivational learning outcomes, the final sample of studies consisted of 16 primary studies reporting 16 independent experiments that were published between 2013 and 2017 and examined a total of 2246 participants. Finally, the final sample of studies examining behavioral learning outcomes consisted of nine primary studies reporting 10 independent experiments. These experiments were published between 2014 and 2017 and examined a total of 951 participants.

Statistical Analysis

Effect sizes were estimated using the formulas provided by Borenstein et al. (2009). First, Cohen's *d* was determined by dividing the difference between the means of the experimental and control groups by the pooled standard deviation. If no means and/or standard deviations were available, effect sizes were estimated on the basis of the *t*, *F*, or *r* statistics. To correct for possible bias caused by small samples, Cohen's *d* was then used to calculate Hedges' *g* (Hedges 1981).

For studies with pretests, effect sizes were adjusted to allow for a more accurate estimation of the effect by controlling for pretest effects. In these cases, Hedges' *g* was calculated for both pre- and posttest comparisons between groups. Posttest values
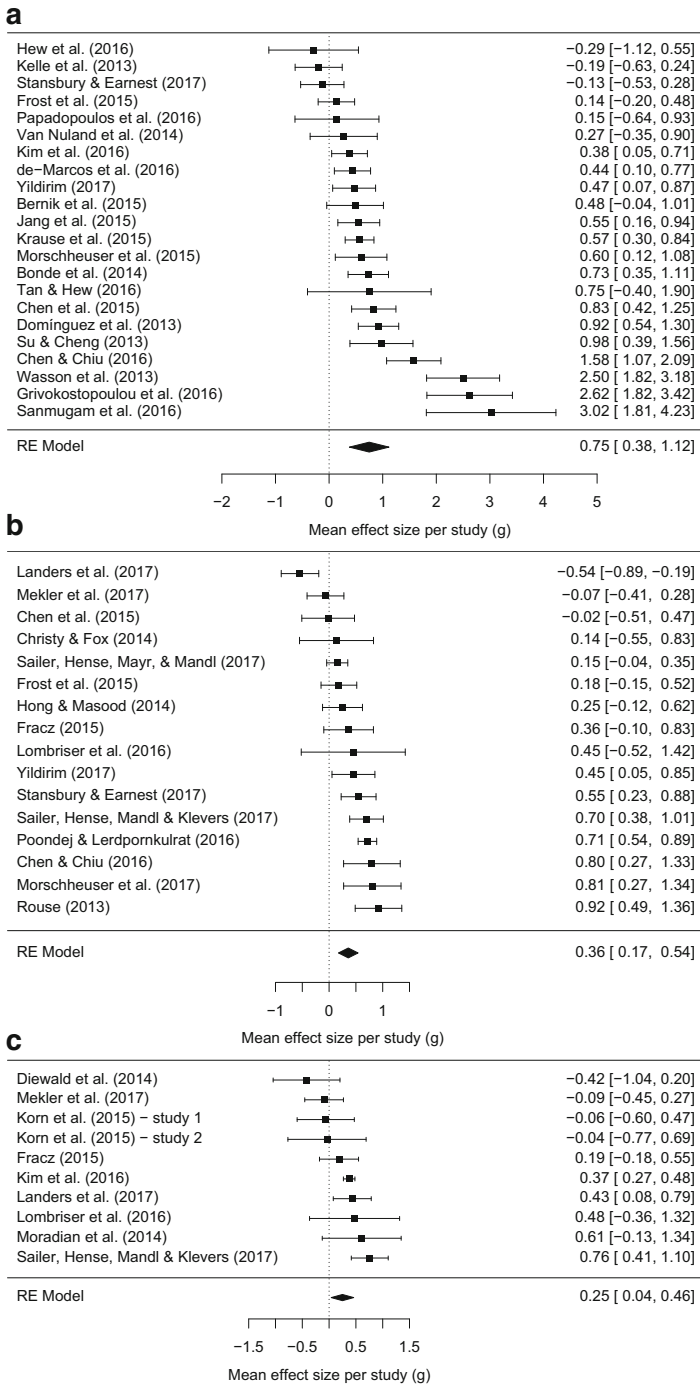
Fig. 2 Forest plots showing the distribution of effect sizes for cognitive (a), motivational (b), and behavioral (c) learning outcomes. Points and lines display effect sizes (Hedges' *g*) and confidence intervals, respectively. The overall effect size is shown by the diamond on the bottom, with its width reflecting the corresponding confidence interval

were then adjusted by subtracting pretest effect sizes from posttest effect sizes, while the respective variances were added up (see Borenstein et al. 2009).

Another issue concerns studies reporting multiple effect sizes. This can occur when studies use multiple outcome measures or compare the same control group with more than one experimental group. To avoid the bias caused by dependent effect sizes, they were synthesized by calculating the mean effect size per study and the respective variance, taking into account the correlation between outcomes as recommended by Borenstein et al. (2009). The aggregation was performed using the R statistical environment (version 3.4.1) and Rstudio (version 1.0.143) with the MAd package.

For some studies that reported data on different experimental groups or outcomes, even though they belonged to the same study, they required assignment to different moderator levels. In these cases, aggregation would lead to a loss of data and not allow these studies to be included in the respective subgroup analyses because they could not be unambiguously assigned to a single moderator level. Therefore, groups were treated separately and not aggregated for the moderator analysis but were aggregated for all analyses not affected by this problem.

Because heterogeneity among the participant samples and the experimental conditions could be presumed, we used a random effects model for the main analysis (Borenstein et al. 2009), which was also conducted using R and Rstudio with the metafor and matrix packages. Because true effect sizes can vary under a random effects model, it is important to identify and quantify this heterogeneity. Using the Q statistic (Borenstein et al. 2009), for which the p value indicates the presence or absence of heterogeneity, the degrees of homogeneity were assessed for the effect sizes. Additionally, I^2 was used to quantify the degree of inconsistency for the results of the included studies on a scale ranging from 0 to 100%.

Meta-regressions under a mixed effects model (Borenstein et al. 2009) were conducted for the moderator analyses using the metafor and matrix packages in R and R studio. This analysis uses a Q test as an omnibus test to estimate the heterogeneity among studies explained by the moderator, as well as residual heterogeneity (i.e., whether unexplained between-study variance remained). If a moderator level had a sample size smaller than two, it was excluded from the analysis. For categorical moderators, post hoc comparisons under a random effects model were calculated with the MAd package if the omnibus test indicated significance.

Because the reporting of a sufficient amount of statistical data on at least one outcome of interest was used as an eligibility criterion, there were no missing data in the summary effect analyses. However, there were missing data for the moderation analyses because not all studies reported sufficient information about all variables, and thus, they could not be unambiguously coded. These cases were excluded from the respective moderator analyses.

Furthermore, bivariate correlations between the presence of learning outcomes and moderator levels were computed for all studies. Therefore, all learning outcomes and moderator levels were dummy coded and included in a correlation matrix. The moderators *inclusion of game fiction*, *randomization*, and *design* were already initially dichotomously coded, and thus, only the moderator levels that indicated the presence of game fiction and randomization and the application of pre-posttest designs were included in the matrix. The database for this analysis is the final sample detailed above, excluding studies reporting the same experiment as other studies and the outliers described above.

Publication Bias

The absence of publication bias is highly unlikely unless a set of preregistered reports is examined (Carter et al. 2019). Because this was not the case for the primary studies included in this meta-analysis, the possibility of publication bias was explored by using a funnel plot, a selection model, and the fail-safe number. Evaluations of the funnel plots showed no obvious asymmetries for the samples of studies reporting cognitive, motivational, and behavioral learning outcomes (see Fig. 3), initially indicating the absence of strong publication bias.

The selection model indicated that there was no publication bias for the subsets of studies reporting cognitive, $\chi^2(1) = .45$, $p = .50$, motivational, $\chi^2(1) = 1.04$, $p = .31$, and behavioral learning outcomes, $\chi^2(1) = .44$, $p = .50$. For this reason, estimates from the initial random effects model were computed for the cognitive, motivational, and behavioral learning outcomes.

Further, Rosenberg's (2005) fail-safe number was used as an estimate of the degree to which publication bias existed in the sample. The fail-safe number indicates the number of nonsignificant studies that would need to be added to reduce a significant effect to a nonsignificant effect. It was computed for all significant summary effects as well as for every moderator level that showed a statistically significant effect. The fail-safe number can be considered robust when it is greater than $5n + 10$, with n standing for the number of primary studies (Rosenthal 1991).

To further investigate the stability of the effects, subsplits were computed for the summary effect and moderator analyses for conceptual, situational, and contextual moderators. In these subsplits for cognitive, motivational, and behavioral learning outcomes, only studies applying experimental designs or using quasi-experimental designs with pre- and posttests were included. Thus, we assumed that the studies included in these subsplit analyses were characterized by high methodological rigor.

Results

Before reporting the summary effect and moderator analyses, we report correlations of the presence of different learning outcomes and moderator levels in gamification studies included in the final sample to illustrate possible co-occurrences of outcomes and moderator levels. The resulting correlation matrix, including the number of studies for each cell, is shown in Table 1. All of the following results refer to significant correlations. Studies investigating cognitive learning outcomes were less likely to also observe motivational ($r = - .47$) and behavioral learning outcomes ($r = - .60$). Further, studies observing cognitive learning outcomes were more likely to use a mixed instruction comparison group ($r = .41$), including self-developed instruments ($r = .35$), and less likely to last 1 day ($r = - .53$), indicating that these studies often took longer periods of time. Motivational learning outcomes were more likely to use standardized ($r = .44$) or adapted instruments ($r = .54$). Investigating behavioral learning outcomes was positively correlated with using an active comparison group ($r = .35$).

Studies showed some specific patterns for their respective research context: Studies performed in school settings were more likely to use passive instruction for the comparison group ($r = .43$). Higher education studies were more likely to investigate cognitive learning outcomes

a

b

c
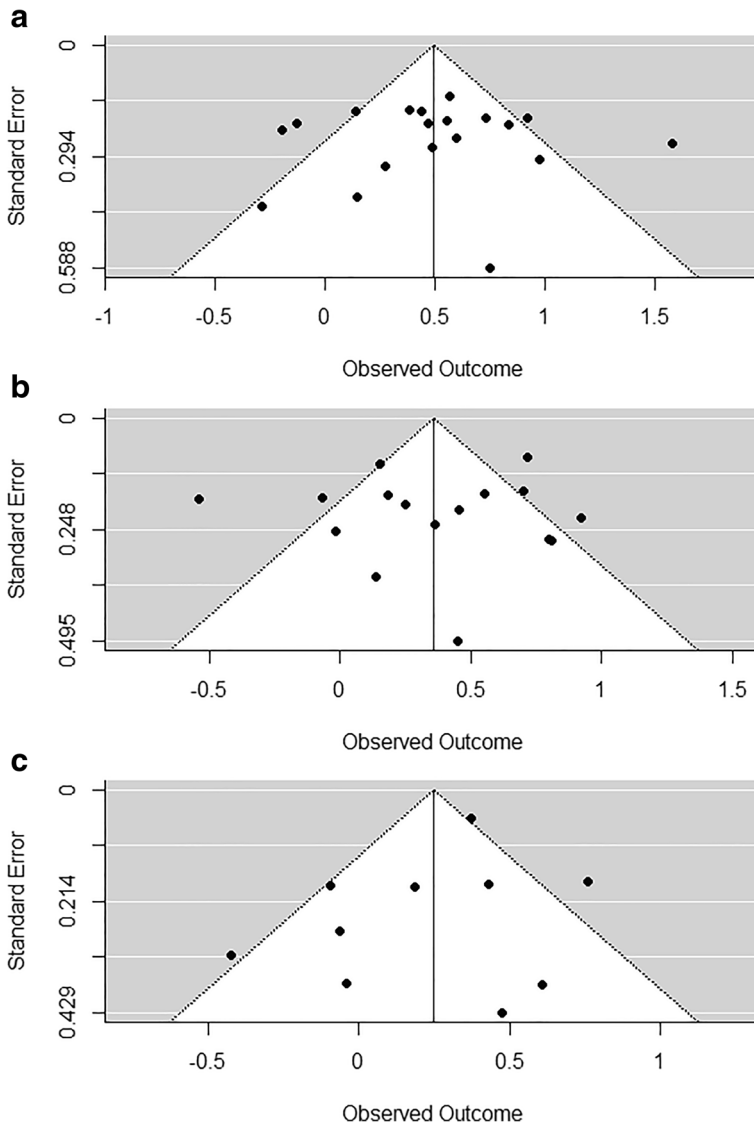
Fig. 3 Funnel plot for cognitive (a), motivational (b), and behavioral (c) learning outcomes. Black dots indicate studies from the present sample positioned by their respective estimated effect size and standard error

($r = .35$) that were compared with a mixed instruction control group ($r = .45$) over a period of half a year ($r = .48$). Some kind of social interaction was more likely to occur in higher education contexts (correlation with *no social interaction* $r = -.39$). In work-related training settings, studies were more likely to observe behavioral learning outcomes ($r = .37$), and results were likely to be compared with a comparison group receiving active instruction ($r = .34$). The use of no social interaction was positively correlated with work settings ($r = .49$), and no study investigated competitive or collaborative gamification implementation in this setting. Studies in informal training settings were likely to apply randomization ($r = .35$).

Further significant correlations were shown for the inclusion of game fiction: Studies including game fiction were less likely to apply randomization ($r = -.41$) and were more likely to include collaborative modes of social interaction ($r = .45$).

Summary Effect Analyses

Cognitive Learning Outcomes The random effects model yielded a significant, small effect of gamification on cognitive learning outcomes ($g = .49$, $SE = .10$, $p < .01$, 95% CI [0.30, 0.69]). Homogeneity estimates showed a significant and substantial amount of heterogeneity for cognitive learning outcomes, $Q(18) = 57.97$, $p < .01$, $I^2 = 72.21\%$. The fail-safe number could be considered robust for cognitive learning outcomes (fail-safe $N = 469$).

Motivational Learning Outcomes The results of the random effects model showed a significant, small effect of gamification on motivational learning outcomes ($g = .36$, $SE = .09$, $p < .01$, 95% CI [0.18, 0.54]) and a significant amount of heterogeneity, $Q(15) = 73.54$, $p < .01$, $I^2 = 75.13\%$. The fail-safe number indicated a robust effect for motivational learning outcomes (fail-safe $N = 316$).

Behavioral Learning Outcomes The random effects model showed a significant, small effect of gamification on behavioral learning outcomes ($g = .25$, $SE = .11$, $p < .05$, 95% CI [0.04, 0.46]). Results showed a significant and substantial amount of heterogeneity for behavioral learning outcomes, $Q(9) = 22.10$, $p < .01$, $I^2 = 63.80\%$. The fail-safe number for behavioral learning outcomes could be interpreted as robust (fail-safe $N = 136$).

Moderator Analyses

As the homogeneity estimates showed a significant and substantial amount of heterogeneity for cognitive, motivational, and behavioral learning outcomes, moderator analyses were conducted to determine whether additional factors could account for the variance observed in the samples. Not all of the following comparisons contained all possible levels of the respective moderator because levels with $k \leq 1$ were excluded from these analyses. Summaries of the moderator analysis are shown in Table 2 for cognitive, Table 3 for motivational, and Table 4 for behavioral learning outcomes.

Inclusion of Game Fiction The mixed effects analysis concerning game fiction resulted in no significant effect size differences for cognitive learning outcomes, $Q(1) = 0.04$, $p = .85$, and the residual variance was significant ($p < .01$). For motivational learning outcomes, effect size magnitude did not vary significantly, $Q(1) = 0.13$, $p = .72$, with significant residual variance remaining ($p < .01$). Finally, the results of the mixed effects analysis for behavioral learning outcomes showed a significant difference in the magnitudes of the effect sizes for the game fiction moderator, $Q(1) = 5.45$, $p < .05$, and no residual variance was left ($p = .08$). An evaluation of the separate levels showed that inclusion of game fiction yielded a significant, small effect on behavioral learning outcomes in contrast to a nonsignificant result of not including game fiction.

Table 1 Observed learning outcomes and moderator level presence correlation matrix including number of studies for each cell

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. Cognitive learning outcome													
2. Motivational learning outcome	-.47** (5)												
3. Behavioral learning outcome	-.60** (1)	.04 (5)											
4. Inclusion of game fiction	.11 (7)	.10 (6)	-.03 (3)										
5. Competitive	-.13 (9)	.18 (10)	-.04 (5)	-.10 (5)									
6. Collaborative	.07 (2)	.12 (2)	n/a (0)	.45** (3)	-.12 (1)								
7. Competitive-collaborative	.07 (5)	.03 (4)	.10 (3)	-.09 (2)	n/a (0)	n/a (0)							
8. No social interaction	-.13 (3)	-.19 (2)	.15 (3)	-.04 (2)	-.39* (1)	n/a (0)	n/a (0)						
9. Passive instruction comparison group	-.04 (2)	.02 (2)	n/a (0)	.14 (2)	-.02 (2)	.21 (1)	n/a (0)	.04 (1)					
10. Active instruction comparison group	-.24 (8)	.18 (10)	.35* (8)	.02 (6)	-.18 (8)	.29 (3)	-.03 (4)	.19 (5)	-.20 (1)				
11. Mixed instruction comparison group	.41* (10)	-.20 (4)	-.34* (1)	.02 (4)	.08 (7)	n/a (0)	.17 (4)	-.22 (1)	n/a (0)	n/a (0)			
12. Untreated comparison group	n/a (0)	n/a (0)	.27 (1)	n/a (0)	.16 (1)	n/a (0)	n/a (0)	n/a (0)	n/a (0)	n/a (0)	n/a (0)		
13. Duration of 1 day	-.53** (5)	.12 (9)	.39* (8)	-.31 (3)	0 (9)	-.10 (1)	0 (4)	.22 (5)	0 (2)	.12 (10)	-.25 (4)	.17 (1)	
14. Duration of 1 month	.28 (3)	n/a (0)	n/a (0)	.01 (1)	.09 (2)	n/a (0)	.07 (1)	n/a (0)	n/a (0)	-.12 (1)	.20 (2)	n/a (0)	n/a (0)
15. Duration of half a year	.31 (8)	.17 (6)	-.28 (1)	.11 (4)	.22 (7)	.03 (1)	-.05 (2)	n/a (0)	-.04 (1)	-.17 (4)	.20 (5)	n/a (0)	n/a (0)
16. School setting	.14 (3)	.2 (3)	n/a (0)	-.06 (1)	-.02 (2)	.21 (1)	.01 (1)	n/a (0)	.43* (2)	.16 (3)	n/a (0)	n/a (0)	0 (2)
17. Higher education setting	.35* (13)	-.17 (7)	-.30 (3)	.02 (6)	.17 (11)	-.12 (1)	.11 (5)	-.39* (1)	-.02 (2)	-.42* (6)	.45** (10)	n/a (0)	-.35* (6)
18. Work-related learning setting	n/a (0)	.02 (2)	.37* (3)	-.06 (1)	n/a (0)	n/a (0)	.01 (1)	.49** (3)	n/a (0)	.34* (4)	n/a (0)	n/a (0)	.18 (3)
19. Informal training setting	-.28 (2)	.10 (4)	.31 (4)	.11 (3)	.04 (4)	.10 (1)	-.11 (1)	.28 (3)	n/a (0)	.19 (5)	-.22 (1)	.34* (1)	.22 (5)
20. Randomization	-.15 (6)	-.01 (6)	.02 (4)	-.41* (1)	.14 (8)	-.03 (1)	-.01 (3)	.05 (3)	n/a (0)	.26 (9)	-.20 (3)	.22 (1)	.30 (9)
21. Pre- and posttest	.31 (6)	-.33 (1)	n/a (0)	-.20 (1)	.04 (4)	.10 (1)	.06 (2)	n/a (0)	.04 (1)	.04 (4)	.08 (3)	n/a (0)	-.07 (3)
22. Standardized instrument	-.13 (2)	.44** (5)	-.09 (1)	-.11 (1)	.06 (3)	n/a (0)	.16 (2)	-.01 (1)	n/a (0)	.06 (3)	.04 (2)	n/a (0)	-.08 (2)
23. Adapted instrument	-.28 (2)	.54** (7)	-.01 (2)	.11 (3)	.33 (6)	.10 (1)	-.11 (1)	n/a (0)	.27 (2)	.04 (4)	-.22 (1)	n/a (0)	.22 (5)
24. Self-developed instrument	.35* (17)	-.59** (8)	.21 (9)	-.06 (8)	-.38* (11)	-.07 (2)	.14 (7)	.28 (7)	-.23 (2)	.03 (14)	.12 (10)	.10 (1)	0 (13)

Table 1 (continued)

	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.
1. Cognitive learning outcome										
2. Motivational learning outcome										
3. Behavioral learning outcome										
4. Inclusion of game fiction										
5. Competitive										
6. Collaborative										
7. Competitive-collaborative										
8. No social interaction										
9. Passive instruction comparison group										
10. Active instruction comparison group										
11. Mixed instruction comparison group										
12. Untreated comparison group										
13. Duration of 1 day										
14. Duration of 1 month										
15. Duration of half a year	n/a (0)									
16. School setting	n/a (0)	−.04 (1)								
17. Higher education setting	.29 (3)	.48** (9)	n/a (0)							
18. Work-related learning setting	n/a (0)	n/a (0)	n/a (0)	n/a (0)						
19. Informal training setting	n/a (0)	n/a (0)	n/a (0)	n/a (0)	n/a (0)					
20. Randomization	−.03 (1)	−.24 (2)	n/a (0)	−.11 (6)	−.10 (1)	.35* (5)				
21. Pre- and posttest	.10 (1)	−.01 (2)	.04 (1)	.33 (6)	n/a (0)	n/a (0)	.20 (4)			
22. Standardized instrument	n/a (0)	.28 (3)	.11 (1)	.06 (3)	n/a (0)	−.01 (1)	.19 (3)	−.01 (1)		
23. Adapted instrument	n/a (0)	−.01 (2)	.27 (2)	−.25 (2)		.28 (3)	−.10 (2)	n/a (0)	n/a (0)	
24. Self-developed instrument	.17 (3)	−.25 (6)	−.23 (2)	.03 (14)	.20 (4)	−.06 (5)	.01 (10)	.11 (6)	−.36* (2)	−.57** (2)

Pearson correlations between each cell with number of study samples (*k*) in parentheses. Moderator levels 1 week or less and more than half a year were omitted because the sample size was 0

n/a not applicable because *k* = 0

*$p < .05$; **$p < .01$

Table 2 Results of the moderator analyses for cognitive learning outcomes

Moderator variable	p	g	95% CI	k	N	Q	df	I^2 (%)	Fail-safe N
Inclusion of game fiction									
Yes		.47**	[0.19, 0.76]	7	762	21.06**	6	74.89	72*
No		.51**	[0.26, 0.76]	13	1027	40.19**	12	71.96	212*
Social interaction									
Competitive		.50**	[0.34, 0.67]	9	922	14.03	8	35.36	134*
Collaborative		.40	[− 0.67, 1.48]	2	195	9.20**	1	89.13	–
Competitive-collaborative		.63*	[0.06, 1.20]	5	275	18.87**	4	77.34	33
None		.37	[− 0.18, 0.93]	3	326	10.59**	2	82.61	–
Learning arrangement comparison group									
Passive instruction		.81**	[0.48, 1.14]	2	182	0.80	1	0.00	11
Active instrution		.47*	[0.06, 0.88]	8	629	40.92**	7	82.32	45
Mixed instruction		.49**	[0.32, 0.46]	10	941	13.45**	9	36.62	141*
Untreated[a]		–	–	–	–	–	–	–	–
Period of time									
1 day or less		.60**	[0.24, 0.96]	5	344	5.85	4	51.70	37*
1 week or less[a]		–	–	–	–	–	–	–	–
1 month or less		.52**	[0.29, 0.75]	3	297	1.01	2	0.00	13
Half a year of less		.50**	[0.14, 0.86]	8	763	36.68**	7	85.01	80*
More than half a year[a]		–	–	–	–	–	–	–	–
Research context	**								
School setting		1.12**	[0.68, 1.57]	3	257	5.21	2	58.17	43*
Higher education setting		.41**	[0.25, 0.57]	13	1181	24.22*	12	42.15	171*
Work-related learning setting[a]		–	–	–	–	–	–	–	–
Informal training setting		.19	[− 0.55, 0.92]	2	212	6.23*	1	83.95	–
Randomization									
Experimental		.29**	[0.07, 0.51]	6	388	7.73	5	0.00	5
Quasi-experimental		.56**	[0.26, 0.88]	11	1161	46.08**	10	85.49	233*
Design									
Posttest only		.41*	[0.08, 0.73]	14	1445	128.25**	13	89.97	163*
Pre- and posttest		.51**	[0.16, 0.87]	6	423	9.76	5	43.69	31
Instruments									
Standardized[a]		1.84**	[1.23, 2.44]	1	58	–	–	–	–
Adapted[a]		.14	[− 0.20, 0.48]	1	88	–	–	–	–
Self-developed		.50**	[0.32, 0.69]	18	1630	44.38**	17	65.90	430*

g posttest effect sizes adjusted for pretest scores between groups, k number of study samples, N total sample size, Q homogeneity statistic, df degrees of freedom for the Q test, I^2 between-study variance component

[a] Not estimated due to a small number of studies. An asterisk (*) in the fail-safe N column indicates the fail-safe number is robust ($> 5n + 10$)

*$p < .05$; **$p < .01$

Social Interaction Concerning social interaction, the results of the mixed effects analysis showed no significant difference in effect sizes between different forms of social interaction for cognitive learning outcomes, $Q(3) = 0.80$, $p = .85$, or for motivational learning outcomes, $Q(3) = 3.85$, $p = .28$. For both outcomes, significant residual variance remained ($p < .01$). However, for behavioral learning outcomes, there was a significant difference between competitive, competitive-collaborative, and no social interaction, $Q(2) = 12.80$, $p < .01$, with no significant residual variance remaining ($p = .05$). Post hoc comparisons under a random effects model showed a significant difference between competitive-collaborative interaction and no social interaction ($p < .05$), with the former outperforming the latter.

Table 3 Results of the moderator analyses for motivational learning outcomes

Moderator variable	p	g	95% CI	k	N	Q	df	I^2 (%)	Fail-safe N
Inclusion of game fiction									
Yes		.34**	[0.20, 0.48]	5	515	10.50*	4	0.00	26
No		.32**	[0.08, 0.57]	12	1751	66.72**	11	81.95	159*
Social interaction									
Competitive		.18	[− 0.05, 0.42]	11	1737	66.36**	10	81.25	–
Collaborative		.33	[− 0.06, 0.72]	2	321	4.05*	1	75.34	–
Competitive-collaborative		.63**	[0.41, 0.85]	4	269	1.45	3	0.00	30
None		.38	[− 0.42, 1.17]	2	175	6.05*	1	83.48	–
Learning arrangement comparison group									
Passive instruction		.51*	[0.06, 0.96]	2	606	4.92*	1	79.68	31*
Active instruction		.27*	[0.01, 0.54]	10	1202	44.63**	9	76.75	31
Mixed instruction		.47**	[0.16, 0.77]	4	308	7.11	3	56.63	16
Untreated[a]		–	–	–	–	–	–	–	–
Period of time	*								
1 day or less		.19	[− 0.07, 0.45]	9	1173	35.52**	8	76.99	-
1 week or less[a]		–	–	–	–	–	–	–	–
1 month or less[a]		–	–	–	–	–	–	–	–
Half a year of less		.59**	[0.39, 0.79]	6	932	10.89	5	53.98	147*
More than half a year[a]		–	–	–	–	–	–	–	–
Research context									
School setting		.33	[− 0.14, 0.80]	3	191	5.09	2	67.78	-
Higher education setting		.52**	[0.33, 0.71]	7	1025	12.98*	6	50.17	157*
Work-related learning setting		.72**	[0.25, 1.19]	2	53	0.40	1	0.00	3
Informal training setting		.07	[− 0.44, 0.57]	4	847	28.03**	3	91.59	–
Randomization	*								
Experimental		.13	[− 0.23, 0.50]	6	961	23.83**	5	82.35	–
Quasi-experimental		.51**	[0.35, 0.68]	10	1155	20.89*	9	46.12	269*
Design									
Posttest only		.37**	[0.17, 0.56]	15	2038	73.06**	14	78.07	289*
Pre- and posttest[a]		–	–	–	–	–	–	–	–
Instruments									
Standardized		.45*	[0.08, 0.81]	5	464	15.46**	4	68.16	16
Adapted		.13	[− 0.26, 0.54]	5	999	45.38**	4	87.15	–
Self-developed		.39**	[0.25, 0.52]	6	653	12.62*	5	0.00	44*

g posttest effect sizes adjusted for pretest scores between groups, k number of study samples, N total sample size, Q homogeneity statistic, df degrees of freedom for the Q test, I^2 between studies variance component

[a] Not estimated due to a small number of studies. An asterisk (*) in the fail-safe N column indicates the fail-safe number is robust ($> 5n + 10$)

*$p < .05$; ** $p < .01$

Learning Arrangement of the Comparison Group As for the learning arrangement of the comparison group, no difference in effect sizes could be found regarding cognitive learning outcomes, $Q(2) = 1.49$, $p = .48$, or for motivational learning outcomes, $Q(2) = 1.17$, $p = .56$. Significant residual variance remained for both outcomes ($p < .01$). We could not assess the effect of the learning arrangement of the comparison group for behavioral learning outcomes because there was only one subgroup with more than one study.

Period of Time There was no significant difference between interventions with different durations for cognitive, $Q(2) = 0.17$, $p = .92$, or behavioral learning outcomes, $Q(1) = 0.28$, $p = .59$. For both outcomes, significant residual variance remained ($p < .01$). For motivational

Table 4 Results of the moderator analyses for behavioral learning outcomes

Moderator variable	p	g	95% CI	k	N	Q	df	I² (%)	Fail-safe N
Inclusion of game fiction	**								
Yes		.41**	[0.31, 0.51]	3	254	4.39	2	0.00	46*
No		.10	[−0.14, 0.34]	7	697	11.61	6	41.92	–
Social interaction	**								
Competitive		.12	[−0.14, 0.38]	6	759	16.24**	5	73.10	–
Collaborative[a]		–	–	–	–	–	–	–	–
Competitive-collaborative		.70**	[0.41, 0.99]	3	135	0.45	2	0.00	15
None		−.06	[−0.36, 0.24]	3	193	< 0.01	2	0.00	–
Learning arrangement comparison group									
Passive instruction[a]		–	–	–	–	–	–	–	–
Active instrution		.33**	[0.14, 0.52]	8	848	15.51*	7	46.42	117*
Mixed instruction[a]		.19	[−0.18, 0.55]	1	73	–	–	–	–
Untreated[a]		−.42	[−1.04, 0.20]	1	30	–	–	–	–
Period of time									
1 day or less		.20	[−0.07, 0.47]	8	800	20.27**	7	62.45	–
1 week or less[a]		–	–	–	–	–	–	–	–
1 month or less[a]		–	–	–	–	–	–	–	–
Half a year of less[a]		.37**	[0.27, 0.48]	1	140	–	–	–	–
More than half a year[a]		–	–	–	–	–	–	–	–
Research context									
School setting[a]		–	–	–	–	–	–	–	–
Higher education setting		.36**	[0.26, 0.47]	3	234	1.39	2	0.00	36*
Work-related learning setting		.06	[−0.33, 0.44]	3	71	1.21	2	0.00	–
Informal training setting		.20	[−0.31, 0.72]	4	646	17.04**	3	84.83	–
Randomization	*								
Experimental		.12	[−0.33, 0.58]	4	564	8.88*	3	71.74	–
Quasi-experimental		.39**	[0.30, 0.49]	4	327	5.74	3	0.00	61*
Design									
Posttest only		.25*	[0.04, 0.46]	10	951	22.10**	9	63.80	136*
Pre- and posttest[a]		–	–	–	–	–	–	–	–
Instruments									
Standardized[a]		–	–	–	–	–	–	–	–
Adapted[a]		.43*	[0.08, 0.79]	1	240	–	–	–	–
Self-developed		.22	[−0.01, 0.45]	9	711	21.80**	8	64.59	110*

g posttest effect sizes adjusted for pretest scores between groups, *k* number of study samples, *N* total sample size, *Q* homogeneity statistic, *df* degrees of freedom for the *Q* test, *I²* between studies variance component

[a] Not estimated due to a small number of studies. An asterisk (*) in the fail-safe *N* column indicates the fail-safe number is robust ($> 5n + 10$)

*$p < .05$; **$p < .01$

learning outcomes, there was a significant difference between gamification interventions lasting 1 day or less and interventions lasting half a year or less, $Q(1) = 4.93$, $p < .05$. Gamification interventions lasting half a year or less showed significantly larger effects on motivational learning outcomes than interventions lasting 1 day or less. Significant residual variance remained ($p < .01$).

Research Context Mixed effects analyses showed no significant difference between different research contexts for motivational, $Q(3) = 5.09$, $p = .17$, or behavioral learning outcomes, $Q(2) = 0.67$, $p = .71$. For both outcomes, significant residual variance remained ($p < .01$). For cognitive learning outcomes, there was a significant difference between school, higher

education, and informal training settings, $Q(2) = 12.48$, $p < .01$, with significant residual variance remaining ($p < .01$). Post hoc comparisons under a random effects model showed a significant difference between studies performed in school settings and studies performed in higher education settings ($p < .01$) or informal training settings ($p < .05$). The effects found in school settings were significantly larger than those found in either higher education settings or informal education settings.

Randomization Results from the mixed effects analyses showed no significant difference between experimental and quasi-experimental studies regarding cognitive, $Q(1) = 1.18$, $p = .28$, or behavioral learning outcomes, $Q(1) = 1.63$, $p = .20$. The residual variance was significant ($p < .01$). However, for motivational learning outcomes, there was a significant difference in the magnitude of the effect size between experimental and quasi-experimental studies, $Q(1) = 4.67$, $p < .05$. Quasi-experimental studies showed a significant medium-sized effect on motivational learning outcomes, whereas experimental studies showed a nonsignificant effect. The residual variance was significant ($p < .01$).

Design Mixed effects analyses showed no significant effects of applied designs for cognitive learning outcomes, $Q(1) = 0.05$, $p = .82$, and the residual variance was significant ($p < .01$). An evaluation of the influence of the applied design was not possible for the motivational and behavioral learning outcomes because there was only one subgroup with more than one study.

Instrument An evaluation of the influence of the instrument that was used was not possible for cognitive or behavioral learning outcomes because there was only one subgroup with more than one study in each of these areas. Moreover, there was no significant difference in effect sizes between studies using standardized, adapted, or self-developed instruments for motivational learning outcomes, $Q(2) = 2.85$, $p = .24$, with significant residual variance remaining ($p < .01$).

Subsplit Analyses Including Studies with High Methodological Rigor

Subsplits were performed for studies with high methodological rigor. Thus, in the following analyses, only studies with experimental designs or quasi-experimental designs that used pre- and posttests were included. In line with the summary effect analysis above, the subsplit for cognitive learning outcomes showed a small effect of gamification on cognitive learning outcomes ($g = .42$, $SE = .14$, $p < .01$, 95% CI [0.14, 0.68], $k = 9$, $N = 686$, fail-safe $N = 51$), with homogeneity estimates showing a significant and substantial amount of heterogeneity, $Q(8) = 19.90$, $p < .05$, $I^2 = 51.33\%$. In contrast to the summary effects analysis above, results of the subsplit for motivational learning outcomes showed no significant summary effect of gamification on motivational learning outcomes ($g = .22$, $SE = .17$, $p = .20$, 95% CI [−0.11, 0.56], $k = 7$, $N = 1063$). Significant residual variance was left, $Q(6) = 35.99$, $p < .01$, $I^2 = 84.23\%$. Further, the subsplit summary effect of gamification on behavioral learning outcomes was not significant ($g = .27$, $SE = .22$, $p = .22$, 95% CI [−0.16, 0.70], $k = 5$, $N = 667$). The residual variance was significant, $Q(4) = 17.66$, $p < .01$, $I^2 = 78.59\%$.

Moderator analyses for the subsplit for cognitive learning outcomes could only be performed for *social interaction*, *learning arrangement of the comparison group*, and *period of time* because the subgroups were too small for several other moderator

levels (see Table 5). In line with the moderator analysis described above (see Table 2), social interaction, learning arrangement of the comparison group, and period of time did not significantly moderate the effects of gamification on cognitive learning outcomes in the subsplit analysis.

For motivational learning outcomes, moderator analyses were conducted for *inclusion of game fiction*, *social interaction*, and *research context*. Again, other moderator analyses were not possible because the subgroups were too small (see Table 6). In line with the initial moderator analyses for motivational learning outcomes (see Table 3), game fiction and research context did not significantly moderate the effects of gamification on motivational learning outcomes. Contrary to the initial moderator analysis, the subsplit analysis for social interaction showed significant differences in the magnitude of the effect size, $Q(2) = 7.20$, $p < .05$, with significant residual variance remaining ($p < .01$). Gamification with combinations of competition and collaboration showed a medium-sized effect on motivational learning outcomes and thus outperformed the gamification environments that solely used competition.

Subsplit moderator analyses for behavioral learning outcomes were only possible for the *social interaction* moderator (see Table 7). The results of this analysis were in line with the

Table 5 Results of the subsplit moderator analyses for cognitive learning outcomes including high methodological rigor studies only

Moderator variable	p g	95% CI	k	N	Q	df	I^2 (%)	Fail-safe N
Inclusion of game fiction								
Yes[a]	.98**	[0.39, 1.56]	1	102	–	–	–	–
No	.29**	[0.09, 0.50]	8	584	10.26	7	15.97	15
Social interaction								
Competitive	.59**	[0.33, 0.86]	4	324	3.77	3	6.86	20
Collaborative[a]	.98**	[0.39, 1.56]	1	102	–	–	–	–
Competitive-collaborative	.30	[−0.24, 0.85]	3	162	3.02	2	39.68	–
None[a]	−.19	[−0.63, 0.24]	1	98	–	–	–	–
Learning arrangement comparison group								
Passive instruction[a]	1.09**	[0.41, 1.77]	1	68	–	–	–	–
Active instrution	.23	[−0.20, 0.65]	4	269	6.83	3	47.53	–
Mixed instruction	.42**	[0.15, 0.71]	5	383	3.62	4	29.03	16
Untreated[a]	–	–	–	–	–	–	–	–
Period of time								
1 day or less	.36	[−0.22, 0.95]	3	133	3.59	2	42.27	–
1 week or less[a]	–	–	–	–	–	–	–	–
1 month or less[a]	.15	[−0.64, 0.93]	1	36	–	–	–	–
Half a year of less	.58**	[0.29, 0.88]	3	317	2.91	2	26.20	15
More than half a year[a]	–	–	–	–	–	–	–	–
Research context								
School setting[a]	.98**	[0.39, 1.56]	1	102	–	–	–	–
Higher education setting	.46**	[0.19, 0.73]	6	418	7.93	5	22.70	26
Work-related learning setting[a]	–	–	–	–	–	–	–	–
Informal training setting[a]	−.19	[−0.63, 0.24]	1	98	–	–	–	–

g posttest effect sizes adjusted for pretest scores between groups, k number of study samples, N total sample size, Q homogeneity statistic, df degrees of freedom for the Q test, I^2 between studies variance component

[a] Not estimated due to a small number of studies. An asterisk (*) in the fail-safe N column indicates the fail-safe number is robust ($> 5n + 10$)

*$p < .05$; **$p < .01$

Table 6 Results of the subsplit moderator analyses for motivational learning outcomes including high methodological rigor studies only

Moderator variable	p	g	95% CI	k	N	Q	df	I^2 (%)	Fail-safe N
Inclusion of game fiction									
Yes		.41	[− 0.13, 0.95]	2	883	7.97**	1	87.45	–
No		.13	[− 0.23, 0.50]	6	331	23.86**	5	82.29	–
Social interaction	*								
Competitive		− .09	[− 0.32, 0.14]	5	847	12.40*	4	54.12	–
Collaborative[a]		.15	[− 0.07, 0.37]	1	228	–	–	–	–
Competitive-collaborative		.61**	[0.36, 0.85]	2	200	0.88	1	0.00	10
None		.38	[− 0.42, 1.17]	2	175	6.05*	1	83.48	–
Learning arrangement comparison group									
Passive instruction[a]		–	–	–	–	–	–	–	–
Active instrution		.19	[− 0.21, 0.58]	6	967	34.07**	5	86.65	3
Mixed instruction[a]		.45*	[0.05, 0.85]	1	97	–	–	–	–
Untreated[a]		–	–	–	–	–	–	–	–
Period of time									
1 day or less		.19	[− 0.21, 0.58]	6	967	34.07**	5	86.65	–
1 week or less[a]		–	–	–	–	–	–	–	–
1 month or less[a]		–	–	–	–	–	–	–	–
Half a year of less[a]		.45*	[0.05, 0.85]	1	97	–	–	–	–
More than half a year[a]		–	–	–	–	–	–	–	–
Research context									
School setting[a]		–	–	–	–	–	–	–	–
Higher education setting		.37*	[0.03, 0.72]	2	175	0.60	1	0.00	1
Work-related learning setting[a]		.81**	[0.27, 1.34]	1	42	–	–	–	–
Informal training setting		.07	[− 0.44, 0.57]	4	847	28.03**	3	91.59	–

g posttest effect sizes adjusted for pretest scores between groups, k number of study samples, N total sample size, Q homogeneity statistic, df degrees of freedom for the Q test, I^2 between studies variance component

[a] Not estimated due to a small number of studies. An asterisk (*) in the fail-safe N column indicates the fail-safe number is robust ($> 5n + 10$)

*$p < .05$; **$p < .01$

initial moderator analysis (see Table 4) in showing a significant moderating effect of social interaction on the relationship between gamification and behavioral learning outcomes, $Q(1) = 6.87$, $p < .01$. Gamification with competitive-collaborative modes of social interaction outperformed gamification that solely used competitive modes of social interaction.

Discussion

The aim of this meta-analysis was to statistically synthesize the current state of research on the effects of gamification on cognitive, motivational, and behavioral learning outcomes, taking into account potential moderating factors. Overall, the results indicated significant, small positive effects of gamification on cognitive, motivational, and behavioral learning outcomes. These findings provide evidence that gamification benefits learning, and they are in line with the theory of gamified learning (Landers 2014) and self-determination theory (Ryan and Deci 2002). In addition, the results of the summary effect analyses were similar to results from meta-analyses conducted in the context of games (see Clark et al. 2016; Wouters et al. 2013), indicating that the power of games can be transferred to non-game contexts by using game

Table 7 Results of the subsplit moderator analyses for behavioral learning outcomes including high methodological rigor studies only

Moderator variable	p	g	95% CI	k	N	Q	df	I^2 (%)	Fail-safe N
Inclusion of game fiction									
Yes[a]		.76**	[0.41, 1.10]	1	103	–	–	–	–
No		.12	[− 0.33, 0.58]	4	564	8.88*	3	71.74	–
Social interaction	**								
Competitive		− .01	[− 0.36, 0.34]	4	546	8.39*	3	61.97	–
Collaborative[a]		–	–	–	–	–	–	–	–
Competitive-collaborative		.73**	[0.41, 0.99]	2	124	0.13	1	0.00	10
None[a]		− .07	[− 0.48, 0.35]	1	133	–	–	–	–
Learning arrangement comparison group									
Passive instruction[a]		–	–	–	–	–	–	–	–
Active instrution		.40*	[0.06, 0.75]	4	637	11.68**	3	64.74	13
Mixed instruction[a]		–	–	–	–	–	–	–	–
Untreated[a]		− .42	[− 1.04, 0.20]	1	30	–	–	–	–
Period of time									
1 day or less		.27	[− 0.16, 0.70]	5	646	17.66**	4	78.59	–
1 week or less[a]		–	–	–	–	–	–	–	–
1 month or less[a]		–	–	–	–	–	–	–	–
Half a year of less[a]		–	–	–	–	–	–	–	–
More than half a year[a]		–	–	–	–	–	–	–	–
Research context									
School setting[a]		–	–	–	–	–	–	–	–
Higher education setting[a]		.61	[− 0.13, 1.34]	1	21	–	–	–	–
Work-related learning setting[a]		–	–	–	–	–	–	–	–
Informal training setting		.20	[− 0.31, 0.72]	4	646	17.04**	3	84.83	–

g posttest effect sizes adjusted for pretest scores between groups, k number of study samples, N total sample size, Q homogeneity statistic, df degrees of freedom for the Q test, I^2 between studies variance component

[a] Not estimated due to a small number of studies. An asterisk (*) in the fail-safe N column indicates the fail-safe number is robust ($> 5n + 10$)

*$p < .05$; **$p < .01$

design elements. Given that gamification research is an emerging field of study, the number of primary studies eligible for this meta-analysis was rather small, and effects found in this analysis were in danger of being unstable. Therefore, we investigated the stability of the summary effects. Fail-safe numbers as an estimate of the degree to which publication bias may exist in the samples were applied and indicated that the summary effects for cognitive, motivational, and behavioral outcomes were stable. However, subsplits exclusively including studies with high methodological rigor only supported the summary effect of gamification on cognitive learning outcomes. Although the analyses were underpowered, the summary effects for motivational and behavioral learning outcomes were not significant. Thus, summary effects of gamification on motivational and behavioral learning outcomes are not robust according to the subsplit analyses. For both motivational and behavioral learning outcomes, the subsplits indicate effects of gamification depend on the mode of social interaction. In a nutshell, gamification of motivational and behavioral learning outcomes can be effective when applied in competitive-collaborative settings in contrast to mere competitive settings.

The significant, substantial amount of heterogeneity identified in all three subsamples was in line with the positive, albeit mixed findings of previous reviews. For this reason, moderator analyses were computed to determine which factors were most likely to contribute to the observed variance.

Moderators of the Effectiveness of Learning with Gamification

Inclusion of Game Fiction For both cognitive and motivational learning outcomes, there was no significant difference in effect sizes between the inclusion and the exclusion of game fiction. The mechanisms that are supposed to be at work according to self-determination theory (Rigby and Ryan 2011) were not fully supported in this analysis. However, the results on cognitive and motivational learning outcomes were in line with meta-analytic evidence from the context of games, which found that including game fiction was not more effective than excluding it (Wouters et al. 2013). Nevertheless, for behavioral learning outcomes, the effects of including game fiction were significantly larger than the effects without game fiction. The fail-safe number for the effect of game fiction on behavioral learning outcomes indicated a stable effect that did not suffer from publication bias. However, studies including game fiction were less likely to use experimental designs, which can be a confounding factor.

These results raise the question as to why the use of game fiction seems to matter only for behavioral learning outcomes. A striking divergence between the data representing behavioral learning outcomes and cognitive or motivational learning outcomes is the point of time in which the data were collected: Whereas cognitive and motivational learning outcomes were almost exclusively measured *after* interventions, behavioral learning outcomes were almost exclusively measured *during* interventions (i.e., the measurement was to some extent confounded with the intervention itself). Therefore, it makes sense to ask whether the significant difference in behavioral learning outcomes regarding the use of game fiction really reflects a difference in effectiveness for learning or rather points toward an effect of gamification with game fiction on assessment. Previous research has in fact shown that gamification can affect the data (e.g., the number of questions completed or the length of answers produced; Cechanowicz et al. 2013). Even though behavioral learning outcomes were typically measured by participants completing specific tasks such as assembling Lego® cars (Korn et al. 2015), it may well be the case that these findings also transfer to the number of completed tasks or the time and effort invested in completing a task. From a theoretical point of view, the present results might merely reflect the idea that including game fiction was more effective in getting learners to invest more effort in completing tasks than not including game fiction. It remains uncertain whether these differences would also appear if performance was assessed without gamification.

We attempted to include the time of assessment as a moderator in this analysis; however, there were not enough studies in the subgroups to allow for a conclusive evaluation. Future studies should consider assessing behavioral learning outcomes not only as process data during the intervention but also after the intervention to avoid confounding effects.

Additionally, as gamification can be described as a design process in which game elements are added, the nonsignificant result of this moderator for cognitive and motivational learning outcomes could be explained by the quality with which the (game) design methods were applied: Most learning designers who apply and investigate gamification in the context of learning are not trained as writers and are probably, on average, not successful at applying game fiction effectively. Further, the findings could also be affected by how the moderator was coded. For example, the effectiveness of gamification might be affected by whether game fiction was only used at the beginning of an invervention to provide initial motivation, but it might not be relevant afterwards (e.g., avatars that cannot be developed), or they might continue to be relevant throughout the intervention (e.g., meaningful stories that continue). These possible qualitative differences in the design and use of game fiction could have contributed to the mixed results found in the present analysis. Due to the small sample size,

a more fine-grained coding was not possible because it would have led to subgroups that were too small to be used to conduct any conclusive comparisons. Further, subsplit analyses regarding the moderator *inclusion of game fiction* were not possible for behavioral learning outcomes because the subgroups were too small.

Social Interaction For the cognitive and motivational learning outcomes, no significant difference in effect sizes were found between the different types of social interaction. For behavioral outcomes, a significant difference was found between competitive-collaborative and no interaction in favor of the former. As mentioned previously, behavioral learning outcomes, as opposed to cognitive and motivational learning outcomes, were almost exclusively comprised of process data. Therefore, the results of this analysis suggested that different types of social interactions affect learners' behavior within gamification. Evoking social interactions via gamification in the form of combinations of collaboration and competition was most promising for behavioral learning outcomes. This result for behavioral learning outcomes is in line with evidence from the context of games, showing that combinations of competition and collaboration in games are promising for learning (Clark et al. 2016).

Although the positive effect of competitive-collaborative modes of interaction on behavioral learning outcomes was in danger of being unstable with respect to its fail-safe number, the subsplit of studies with high methodological rigor confirmed the advantage of the combination of competition and collaboration, but here, the advantage was over competition by itself. Interestingly, the findings from the subsplits of motivational and behavioral learning outcomes showed parallels as both showed that combinations of competition and collaboration outperformed mere competition. These results can be interpreted to mean that mere competition might be problematic for fostering learners' motivation and performance—at least for some learners under certain circumstances.

A factor that might have contributed to these results is learners' subjective perceptions. As Rigby and Ryan (2011) pointed out, satisfying the need for relatedness, especially in competitive situations, largely depends on how competition is experienced. Perceiving competition as constructive or destructive is "a function of the people involved rather than the activity itself" (Rigby and Ryan 2011, p. 79). The interaction between learner characteristics and gamification should therefore be investigated in primary studies. A closely related problem concerns whether learners engaged in the interventions in the manner in which they were intended. Most of the studies did not provide information about whether or not learners took advantage of the affordances to engage in collaborative and/or competitive interactions. These issues can only be resolved by conducting future primary studies with data on learning processes, which will allow for investigations of these aspects on a level more closely related to what actually occurs during interventions and checks for implementation fidelity. Besides observations of learners' interactions with learning environments, gamification research should take advantage of log-file analyses and learning analytics to analyze the human-environment interaction in gamified environments (see Young et al. 2012).

Additionally, differences in learners' skill levels may have also contributed to the mixed results found in this analysis. Competition may be problematic if it occurs between learners with widely different skill levels because succeeding can be unattainable for learners with lower skill levels (Slavin 1980; Werbach and Hunter 2012). Only a few primary studies have considered participants' skill levels. Future studies applying competitive or competitive-

collaborative interventions should take prior skill level into account and report it. Further, the effectiveness of different types of social interaction could be influenced by the way the gamification environment and the specific game design elements were designed. Landers et al. (2017) showed how leaderbords that are implemented to foster competition can be interpreted as goals, suggesting that leaderbords with easy goals would likely be less effective than leaderboards with difficult goals. This thereby suggests that also within the single moderator levels of our meta-analysis, the effects could vary depending on their specific design.

Learning Arrangement of the Comparison Group Results of the moderator *learning arrangement of the comparison group* did not show significant differences between gamification when compared with different types of instruction (i.e., passive, active, or mixed instruction) in the comparison group for cognitive and motivational learning outcomes. For behavioral learning outcomes, the analysis was not possible because the moderator subgroups were too small. On the one hand, these results may indicate that gamification is not yet used in a way that focuses on fostering high-quality learning activities and, thus, does not take full advantage of the possibilities that gamification might have. As proposed by the theory of gamified learning, gamification can affect learning outcomes by enhancing activities that are relevant for learning and might thus create instructional affordances for learners to actively engage in cognitive processes with the learning material. Several primary studies included in this analysis could have failed to specifically provide affordances for high-quality learning activities.

On the other hand, learners might not actually take advantage of the instructional affordances provided by the gamified system. Similar to the social interaction moderator, learners probably did not engage in certain (high-quality) learning activities, as intended by the gamification environment (see Chi and Wylie 2014). Primary studies in this meta-analysis did not report sufficient data on the learning processes to clarify this issue. Therefore, future primary studies should use the human-environment interaction as the unit of analysis to account for learners interacting in different ways with the gamified environment (see Young et al. 2012). This would allow researchers to investigate different levels of learning activities that are fostered by certain game design elements while taking into account different situational factors and learner characteristics. Differentiating actual learning activities on a more fine-grained level into passive, active, constructive, and interactive learning activities, as suggested by Chi and Wylie (2014), would enable gamification research to find answers about how, for whom, and under which conditions gamification might work best.

Situational, Contextual, and Methodological Moderators Besides the moderators discussed above, *period of time*, *research context*, *randomization*, *design*, and *instruments* were included as moderators to account for situational and contextual factors as well as methodological rigor. Results regarding period of time for cognitive and behavioral learning outcomes indicated that gamification can be effective in both the short and long term. For motivational learning outcomes, interventions lasting half a year or less (but more than 1 month) showed a medium-sized effect, whereas interventions lasting 1 day or less showed a nonsignificant result. The results of this meta-analysis can weaken the fear that effects of gamification might not persist in the long run and might thus contradict the interpretations presented in reviews (see Hamari et al. 2014; Seaborn and Fels 2015). For motivational outcomes, it might even take longer to affect motivation. However, this does not allow for any conclusions about how enduring the obtained effects will be. An attempt to include the time of the posttest as an

additional methodological moderator in this analysis was not possible across all three sub-samples because only one study reported a delayed posttest. Future primary studies should therefore focus on conducting follow-up tests so that conclusions about the endurance of the effects of gamification can be drawn.

Regarding the research context, a significant difference was found between school settings and higher education settings as well as informal settings in favor of the school setting for cognitive learning outcomes. Therefore, gamification works especially well for school students' knowledge acquisition. No significant differences were found for motivational and behavioral learning outcomes. It should be noted that studies in school contexts were more likely to compare gamification groups with groups receiving passive instruction as opposed to studies from higher education, which were more likely to use control groups that received mixed instructions. This could have led to biased effects.

Randomization did not affect the relationship between gamification and cognitive learning outcomes but did so for motivational and behavioral learning outcomes, indicating that methodological rigor might moderate the size of the effect. For both outcomes, effect sizes for quasi-experimental studies were significantly larger than for experimental studies. In cases in which evaluation of the other methodological moderators, namely *design* and *instruments*, was possible, no influence on the size of the effect for the respective learning outcomes could be detected.

Limitations

One limitation of our study was that the sample size was rather small, especially for behavioral learning outcomes and all subsplit analyses. This limits the generalizability of the results and is also problematic for statistical power because, for random effects models, power depends on the total number of participants across all studies and on the number of primary studies (Borenstein et al. 2009). If there is substantial between-study variance, as in this meta-analysis, power is likely to be low. For this reason, nonsignificant results do not necessarily indicate the absence of an effect but could be explained by a lack of statistical power, especially if effects are rather small.

A common criticism of meta-analytic techniques is that they combine studies that differ widely in certain characteristics, often referred to as *comparing apples and oranges* (Borenstein et al. 2009). This issue can be addressed, to some extent, by including moderators that can account for such differences. In the present meta-analysis, we therefore investigated a set of moderating factors, even though only some of them were shown to be significant moderators of the relationship between gamification and learning. It is important for future research not to disregard the factors that did not show any significant influence, given the issues discussed above regarding power. Furthermore, this set of factors is not exhaustive, and other variables could also account for the observed heterogeneity. A few such factors (e.g., the more fine-grained distinction of game fiction or the time of the posttest) were discussed above, but other factors such as participant characteristics (e.g., familiarity with gaming or differences in the individual perception of game design elements, personality traits, or player types; see Hamari et al. 2014; Seaborn and Fels 2015) or different types of feedback addressed by certain game design elements may also account for heterogeneity in effect sizes. These aspects could not be investigated in the present analysis because most primary studies did not examine or report them. Further, some of the

aspects mentioned above refer to design decisions made by the developers of gamification interventions. Such design decisions can lead to variance in the effect sizes of certain game design elements as shown in the context of leaderboards (Landers et al. 2017) and in the context of achievements, showing that effectivity of achievements strongly depends on the design (i.e., the quantity and difficulty of achievements; Groening and Binnewies 2019). Our meta-analysis did not include such aspects of design and rather synthesized effects of gamification as it is currently operationalized in the literature. By doing so, a mean design effect was assumed. Nevertheless, our meta-analysis offers an interpretive frame for applying and further investigating gamification. Such a frame is especially needed in such a young field. However, research investigating different aspects of certain game design elements is valuable for fully capturing the role of design in the process of applying gamification.

Another limitation concerns the quality of primary studies in the present analysis, a problem often described with the metaphor *garbage in*, *garbage out* (Borenstein et al. 2009). As mentioned earlier, gamification research suffers from a lack of methodological rigor, which, from the perspective of meta-analyses, can be addressed by either assessing methodological differences as moderators or excluding studies with insufficient methodological rigor. In this analysis, both approaches were applied: methodological factors were included as moderators, and subsplits involving studies that applied high methodological rigor were performed. For motivational and behavioral learning outcomes, the results showed that quasi-experimental studies found significant effects, whereas experimental study showed nonsignificant results, emphasizing the need for more rigorous primary study designs that allow alternative explanations for differences in learning outcomes between different conditions to be ruled out. The subsplit analyses showed that the summary effects for the motivational and behavioral outcomes were not robust. However, given the small sample sizes in the subgroup analyses, these findings were highly likely to be underpowered and should be viewed with caution.

Outcome measures are a particular aspect affected by both these problems: A lack of uniformity in measurement leads to differences in reliability and validity that are not considered in the calculation of the mean effect sizes (Walker et al. 2008). The use of agreed-upon measurement instruments with good psychometric properties is therefore needed to increase both the comparability of studies and methodological rigor.

Conclusion

Gamification in the context of learning has received increased attention and interest over the last decade for its hypothesized benefits on motivation and learning. However, some researchers doubt that effects of games can be transferred to non-game contexts (see Boulet 2012; Klabbers 2018). The present meta-analysis supports the claim that gamification of learning works because we found significant, positive effects of gamification on cognitive, motivational, and behavioral learning outcomes. Whereas the positive effect of gamification on cognitive learning outcomes can be interpreted as stable, results on motivational and behavioral learning outcomes have been shown to be less stable. Further, the substantial amount of heterogeneity identified in the subsamples could not be accounted for by several moderating factors investigated in this analysis, leaving partly unresolved the question of which factors contribute to successful gamification. More theory-guided empirical research is needed to work toward a comprehensive theoretical framework with clearly defined components that describes precise mechanisms by which gamification can affect specific learning processes and

outcomes. Future research should therefore explore possible theoretical avenues in order to construct a comprehensive framework that can be empirically tested and refined.

The theory of gamified learning offers a suitable framework for doing so (see Landers 2014; Landers et al. 2018). Combined with evidence from well-established educational and psychological theories that provide clear starting points for effective gamification, primary studies should work toward an evidence-based understanding of how gamification works: focusing on specific game design elements affecting specific psychological needs postulated by self-determination theory (see Ryan and Deci 2002) and exploring ways to create affordances for high-quality learning activities, namely, constructive and interactive learning activities (see Chi and Wylie 2014). Future research has to take into account how psychological needs and high-quality learning activities can be fostered by gamification and to what extent as well as under what conditions learners actually take advantage of these affordances by focusing on the human-environment interaction of learners in gamified interventions (see Young et al. 2012).

In general, more high-quality research, applying experimental designs or quasi-experimental designs with elaborated controls for prior knowledge and motivation, is needed to enable a more conclusive investigation of the relationship between gamification and learning, as well as possible moderating factors. As the abovementioned issues concerning the moderators are merely post hoc explanations, future research should specifically investigate issues such as learners' experiences and perceptions of gamification, their actual activities in the interventions, the role of learners' skill level in competition, the influence of learners' initial motivation, the adaptiveness of gamified systems, other individual characteristics, and the endurance of effects after interventions. Nevertheless, the moderators considered in this analysis should also not be disregarded yet due to the abovementioned limitations of this meta-analysis.

The fact that the significant results in the analyses of the conceptual moderators were mostly found for behavioral learning outcomes is of particular interest for future investigations. Because these outcomes, in contrast to the cognitive variables, were almost exclusively measured during the interventions, these results could indicate that certain variables affect behavior and performance in the immediate situation, which does not necessarily transfer to situations outside the gamified context. The transparent reporting of study characteristics, control group arrangements, and combinations of learning processes and learning outcome data as well as investigating several outcome types in single studies would provide a more comprehensive (meta-analytic) investigation of the factors that contribute to the effectiveness of gamification.

The results of the present meta-analysis offer an interpretive frame for applying gamification and the following practical implications. First, the results suggest that, in general, gamification has the potential to serve as an effective instructional approach for interventions focusing on cognitive, motivational, and behavioral learning outcomes. Second, when considering interventions focusing on behavioral learning outcomes, including game fiction is promising; for example, introducing a fictional game world, which is relevant throughout the gamified intervention, combined with an avatar system, which allows for developing an avatar over time, can help to foster learners' skills (e.g., Sailer et al. 2017b). Further, creating a gamification environment that allows learners to engage in both competitive and collaborative interaction can be beneficial: Letting learners work together in teams, while competing with other teams, can help to improve learners' quality of performance and skills (e.g., Sailer et al. 2017b). There is evidence from the subsplit analysis of studies using high methodological rigor

that this also holds true for motivational learning outcomes. For example, implementing a transparent badge system that allows peers to challenge themselves as well as awarding badges for helping and collaborating with others can help to foster motivation (e.g., Yildirim 2017). Both subsplits for motivational and behavioral learning outcomes further indicate that mere competition within gamification might be suboptimal, but competition augmented with collaboration can be effective. These aspects should be considered when designing gamified interventions, particularly when targeting behavioral learning outcomes, although these effects may especially apply to experiences and performance during the gamified intervention and might not affect performance in contexts that are devoid of gamification.

The initial question was whether gamification is effective for learning. The results from the present meta-analysis suggest that, *yes*, gamification might in fact be effective when it comes to learning. However, the question of which factors contribute most to successful gamification remains partly unresolved, at least for cognitive learning outcomes.

References

References marked with an asterisk indicate studies included in the meta-analysis. References marked with two asterisks indicate studies included in the subsplit analysis.

Armstrong, M. B., & Landers, R. N. (2017). An evaluation of gamified training: using narrative to improve reactions and learning. *Simulation & Gaming, 48*(4), 513–538. https://doi.org/10.1177/1046878117703749.

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

*Bernik, A., Bubaš, G., & Radošević, D. (2015). A pilot study of the influence of gamification on the effectiveness of an e-learning course. In T. Hunjak, V. Kirinić, & M. Konecki (Eds.), *Central european conference on information and intelligent systems. CECIIS 26th International Conference* (pp. 73–79). Varaždin: Faculty of Organization and Informatics, University of Zagreb.

*Bonde, M. T., Makransky, G., Wandall, J., Larsen, M. V., Morsing, M., Jarmer, H., & Sommer, M. O. (2014). Improving biotech education through gamified laboratory simulations. *Nature Biotechnology, 32*(7), 694–697. https://doi.org/10.1038/nbt.2955.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley. https://doi.org/10.1002/9780470743386.

Boulet, G. (2012). Gamification: the latest buzzword and the next fad. *ELearn Magazine, 2012*(12). https://doi.org/10.1145/2407138.2421596.

Burguillo, J. C. (2010). Using game theory and competition-based learning to stimulate student motivation and performance. *Computers & Education, 55*(2), 566–575. https://doi.org/10.1016/j.compedu.2010.02.018.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: a comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115–144. https://doi.org/10.1177/2515245919847196.

Cechanowicz, J., Gutwin, C., Brownell, B., & Goodfellow, L. (2013). Effects of gamification on participation and data quality in a real-world market research domain. In L. E. Nacke, K. Harrigan, & N. Randall (Eds.), *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (pp. 58–65). New York: ACM. https://doi.org/10.1145/2583008.2583016.

*Chen, C.-H., & Chiu, C.-H. (2016). Employing intergroup competition in multitouch design-based learning to foster student engagement, learning achievement, and creativity. *Computers & Education, 103*, 99–113. https://doi.org/10.1016/j.compedu.2016.09.007.

*Chen, C.-H., Liu, G.-Z., & Hwang, G.-J. (2015). Interaction between gaming and multistage guiding strategies on students' field trip mobile learning performance and motivation. *British Journal of Educational Technology, 47*(6), 1032–1050. https://doi.org/10.1111/bjet.12270.

Chi, M. T., & 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.

**Christy, K. R., & Fox, J. (2014). Leaderboards in a virtual classroom: a test of stereotype threat and social comparison explanations for women's math performance. *Computers & Education, 78*, 66–77. https://doi.org/10.1016/j.compedu.2014.05.005.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: a systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79–122. https://doi.org/10.3102/0034654315582065.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". In A. Lugmayr (Ed.), *Proceedings of the 15th International Academic Mindtrek Conference: Envisioning Future Media Environments* (pp. 9–15). New York: ACM. https://doi.org/10.1145/2181037.2181040.

Dichev, C., & Dicheva, D. (2017). Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International Journal of Educational Technology in Higher Education, 14*(9), 1–36. https://doi.org/10.1186/s41239-017-0042-5.

Dicheva, D., & Dichev, C. (2015). Gamification in education: where are we in 2015? In C. Ho & G. Lin (Eds.), *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 1445–1454). Waynesville: Association for the Advancement of Computing in Education.

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: a systematic mapping study. *Educational Technology & Society, 18*(3), 75–88.

Dickey, M. D. (2006). Game design narrative for learning: appropriating adventure game design narrative devices and techniques for the design of interactive learning environments. *Educational Technology Research and Development, 54*(3), 245–263. https://doi.org/10.1007/s11423-006-8806-y.

**Diewald, S., Lindemann, P., Möller, A., Stockinger, T., Koelle, M., & Kranz, M. (2014). Gamified training for vehicular user interfaces - effects on drivers' behavior. In R. Pfliegl (Ed.), *2014 International Conference on Connected Vehicles and Expo (ICCVE) Proceedings* (pp. 250–257). Piscataway: IEEE. https://doi.org/10.1109/iccve.2014.7297551.

*Domínguez, A., Saenz-de-Navarrete, J., de-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: practical implications and outcomes. *Computers & Education, 63*, 380–392. https://doi.org/10.1016/j.compedu.2012.12.020.

*Frącz, W. (2015). An empirical study inspecting the benefits of gamification applied to university classes. In *2015 7th Computer Science and Electronic Engineering Conference (CEEC). Conference Proceedings* (pp. 135–139). Piscataway: IEEE. https://doi.org/10.1109/ceec.2015.7332713.

*Frost, R. D., Matta, V., & MacIvor, E. (2015). Assessing the efficacy of incorporating game dynamics in a learning management system. *Journal of Information Systems Education, 26*(1), 59–70.

Garland, C. M. (2015). *Gamification and implications for second language education: a meta analysis* (Master's thesis). St. Cloud State University, St. Cloud.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simulation & Gaming, 33*(4), 441–467. https://doi.org/10.1177/1046878102238607.

Grivokostopoulou, F., Perikos, I., & Hatzilygeroudis, I. (2016). An innovative educational environment based on virtual reality and gamification for learning search algorithms. In V. Kumar, S. Murthy, & Kinshuk (Eds.), *Proceedings IEEE Eighth International Conference on Technology for Education. T4E 2016* (pp. 110–115). IEEE: Los Alamitos. https://doi.org/10.1109/t4e.2016.029.

Groening, C., & Binnewies, C. (2019). "Achievement unlocked!" - the impact of digital achievements as a gamification element on motivation and performance. *Computers in Human Behavior, 97*, 151–166. https://doi.org/10.1016/j.chb.2019.02.026.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - a literature review of empirical studies on gamification. In R. H. Sprague Jr. (Ed.), *Proceedings of the 47th Annual Hawaii International Conference on System Sciences* (pp. 3025–3034). Washington, DC: IEEE. https://doi.org/10.1109/hicss.2014.377.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. https://doi.org/10.3102/10769986006002107.

**Hew, K. F., Huang, B., Chu, K. W., & Chiu, D. K. (2016). Engaging Asian students through game mechanics: findings from two experiment studies. *Computers & Education, 92-93*, 221–236. https://doi.org/10.1016/j.compedu.2015.10.010.

*Hong, G. Y., & Masood, M. (2014). Effects of gamification on lower secondary school students' motivation and engagement. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 8*(12), 3448–3455.

*Jang, J., Park, J. J., & Yi, M. Y. (2015). Gamification of online learning. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial intelligence in education. AIED 2015. Lecture notes in computer science, vol. 9112* (pp. 646–649). Cham: Springer. https://doi.org/10.1007/978-3-319-19773-9_82.

de Jong, T., & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105–113. https://doi.org/10.1207/s15326985ep3102_2.

**Kelle, S., Klemke, R., & Specht, M. (2013). Effects of game design patterns on basic life support training content. *Educational Technology & Society, 16*(1), 275–285.

*Kim, E., Rothrock, L., & Freivalds, A. (2016). *The effects of gamification on engineering lab activities. In 2016 IEEE Frontiers in Education Conference Proceedings*. Piscataway: IEEE. https://doi.org/10.1109/fie.2016.7757442.

Klabbers, J. H. (2018). On the architecture of game science. *Simulation & Gaming, 49*(3), 207–245. https://doi.org/10.1177/1046878118762534.

*Korn, O., Funk, M., & Schmidt, A. (2015). Towards a gamification of industrial production. A comparative study in sheltered work environments. In J. Ziegler (Ed.), *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (pp. 84–93). New York: ACM. https://doi.org/10.1145/2774225.2774834.

*Krause, M., Mogalle, M., Pohl, H., & Williams, J. J. (2015). A playful game changer: fostering student retention in online education with social gamification. In G. Kiczales (Ed.), *Proceedings of the 2nd (2015) ACM Conference on Learning @ Scale* (pp. 95–102). New York: ACM. https://doi.org/10.1145/2724660.2724665.

Landers, R. N. (2014). Developing a theory of gamified learning: linking serious games and gamification of learning. *Simulation & Gaming, 45*(6), 752–768. https://doi.org/10.1177/1046878114563660.

**Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: a goal setting experiment. *Computers in Human Behavior, 71*, 508–515. https://doi.org/10.1016/j.chb.2015.08.008.

Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science, its history and future: definitions and a research agenda. *Simulation & Gaming, 49*(3), 315–337. https://doi.org/10.1177/1046878118774385.

*Lombriser, P., Dalpiaz, F., Lucassen, G., & Brinkkemper, S. (2016). Gamified requirements engineering: model and experimentation. In M. Daneva & O. Pastor (Eds.), *Requirements engineering: foundation for software quality. REFSQ 2016. Lecture notes in computer science, vol. 9619* (pp. 171–187). Cham: Springer. https://doi.org/10.1007/978-3-319-30282-9_12.

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

**de-Marcos, L., Garcia-Lopez, E., & Garcia-Cabot, A. (2016). On the effectiveness of game-like and social approaches in learning: comparing educational gaming, gamification & social networking. *Computers & Education, 95*, 99–113. https://doi.org/10.1016/j.compedu.2015.12.008.

**Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior, 71*, 525–534. https://doi.org/10.1016/j.chb.2015.08.048.

**Moradian, A., Nasir, M., Lyons, K., Leung, R., & Sim, S. E. (2014). Gamification of collaborative idea generation and convergence. In M. Jones & P. Palanque (Eds.), *CHI '14 extended abstracts on human factors in computing systems* (pp. 1459–1464). New York, NY: ACM. https://doi.org/10.1145/2559206.2581253.

**Morschheuser, B., Henzi, C., & Alt, R. (2015). Increasing intranet usage through gamification – insights from an experiment in the banking industry. In T. X. Bui & R. H. Sprague Jr. (Eds.), *Proceedings of the 48th Annual Hawaii International Conference on System Sciences. HICSS 2015* (pp. 635–642). IEEE: Los Alamitos, CA. https://doi.org/10.1109/hicss.2015.83.

**Morschheuser, B., Maedche, A., & Walter, D. (2017). Designing cooperative gamification: conceptualization and prototypical implementation. In C. P. Lee & S. Poltrock (Eds.), *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing* (pp. 2410–2421). New York: ACM. https://doi.org/10.1145/2998181.2998272.

**Papadopoulos, P. M., Lagkas, T., & Demetriadis, S. N. (2016). How revealing rankings affects student attitude and performance in a peer review learning environment. In S. Zvacek, M. T. Restivo, J. Uhomoibhi, & M. Helfert (Eds.), *Computer supported education. CSEDU 2015. Communications in computer and information science, vol. 583* (pp. 225–240). Cham: Springer. https://doi.org/10.1007/978-3-319-29585-5_13.

*Poondej, C., & Lerdpornkulrat, T. (2016). The development of gamified learning activities to increase student engagement in learning. *Australian Educational Computing, 31*(2), 1–16.

Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x.

Rigby, S., & Ryan, R. M. (2011). *Glued to games: how video games draw us in and hold us spellbound*. Santa Barbara: Praeger.

Rosenberg, M. S. (2005). The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution, 59*(2), 464–468. https://doi.org/10.1111/j.0014-3820.2005.tb01004.x.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park: Sage. https://doi.org/10.4135/9781412984997.

*Rouse, K. E. (2013). *Gamification in science education: the relationship of educational games to motivation and achievement* (Doctoral dissertation). The University of Southern Mississippi, Hattiesburg, MS.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020.

Ryan, R. M., & Deci, E. L. (2002). Overview of self-determination theory: an organismic dialectical perspective. In R. M. Ryan & E. L. Deci (Eds.), *Handbook of self-determination research* (pp. 3–33). Rochester, NY: University of Rochester Press.

**Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017a). How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033.

**Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2017b). Fostering development of work competencies. In M. Mulder (Ed.), *Competence-based vocational and professional education – bridging the world of work and education* (pp. 795–818). Cham: Springer. https://doi.org/10.1007/978-3-319-41713-4_37.

Sanmugam, M., Abdullah, Z., Mohamed, H., Aris, B., Zaid, N. M., & Suhadi, S. M. (2016). The affiliation between student achievement and elements of gamification in learning science. In Y. Rusmawati & T. A. B. Wirayuda (Eds.), *2016 4th International Conference on Information and Communication Technology (ICoICT)* (pp. 1–4). Piscataway: IEEE. https://doi.org/10.1109/icoict.2016.7571962.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: a survey. *International Journal of Human-Computer Studies, 74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*(2), 489–528. https://doi.org/10.1111/j.1744-6570.2011.01190.x.

Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research, 50*(2), 315–342. https://doi.org/10.3102/00346543050002315.

*Stansbury, J. A., & Earnest, D. R. (2017). Meaningful gamification in an industrial/organizational psychology course. *Teaching of Psychology, 44*(1), 38–45. https://doi.org/10.1177/0098628316677645.

**Su, C.-H., & Cheng, C.-H. (2013). A mobile game-based insect learning system for improving the learning achievements. *Procedia - Social and Behavioral Sciences, 103*, 42–50. https://doi.org/10.1016/j.sbspro.2013.10.305.

Su, C.-H., & Cheng, C.-H. (2014). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning, 31*(3), 268–286. https://doi.org/10.1111/jcal.12088.

**Tan, M., & Hew, K. F. (2016). Incorporating meaningful gamification in a blended learning research methods class: examining student learning, engagement, and affective outcomes. *Australasian Journal of Educational Technology, 32*(5), 19–34. https://doi.org/10.14742/ajet.2232.

**Van Nuland, S. E., Roach, V. A., Wilson, T. D., & Belliveau, D. J. (2014). Head to head: the role of academic competition in undergraduate anatomical education. *Anatomical Sciences Education, 8*, 404–412. https://doi.org/10.1002/ase.1498.

Walker, E., Hernandez, A. V., & Kattan, M. W. (2008). Meta-analysis: its strengths and limitations. *Cleveland Clinic Journal of Medicine, 75*(6), 431–439.

Wasson, R., Mould, D., Biddle, R., & Martinez, C. S. (2013). A sketching game for art history instruction. In S. N. Spencer (Ed.), *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling* (pp. 23–31). New York: ACM. https://doi.org/10.1145/2487381.2487384.

Werbach, K., & Hunter, D. (2012). *For the win: how game thinking can revolutionize your business*. Philadelphia: Wharton Digital Press.

Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2008). How to optimize learning from animated models: a review of guidelines based on cognitive load. *Review of Educational Research, 78*(3), 645–675. https://doi.org/10.3102/0034654308320320.

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.

**Yildirim, I. (2017). The effects of gamification-based teaching practices on student achievement and students' attitudes toward lessons. *The Internet and Higher Education, 33*, 86–92. https://doi.org/10.1016/j. iheduc.2017.02.002.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: a review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61–89. https://doi.org/10.3102/0034654312436980.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 38*(9), 25–32. https://doi. org/10.1109/MC.2005.297.