



# The impact of gamification in educational settings on student learning outcomes: a meta-analysis

Rui Huang<sup>1</sup> · Albert D. Ritzhaupt<sup>1</sup> · Max Sommer<sup>1</sup> · Jiawen Zhu<sup>1</sup> · Anita Stephen<sup>1</sup> · Natercia Valle<sup>1</sup> · John Hampton<sup>1</sup> · Jingwei Li<sup>1</sup>

Published online: 20 July 2020

© Association for Educational Communications and Technology 2020

## Abstract

Gamification research in educational settings has produced mixed results on student learning outcomes. Educational researchers and practitioners both struggle with identifying when, where, and how to use gamification design concepts. The present study provides findings from a meta-analysis that integrated the empirical, quantitative research on gamification in formal educational settings on student learning outcomes. This was achieved by examining the overall effect size, identifying which gamification design elements (e.g., badges) were used, and determining under what circumstances (e.g., engineering education) gamification works. The final corpus of data included 30 independent studies and associated effect sizes comparing gamification to non-gamification conditions while accounting for  $N=3083$  participants. The overall effect size using a random-effects model is  $g=.464$  [.244 to .684] in favor of the gamification condition, which is a small to medium effect size. We examined 14 different gamification design elements (e.g., leaderboards) and showed that each leads to different effects on student learning outcomes. Further, the type of publication (e.g., journal article), student classification (e.g., undergraduate), and subject area (e.g., mathematics) are also investigated as moderators. We provide a discussion of our findings, some recommendations for future research, and some brief closing remarks.

**Keywords** Gamification · Meta-analysis · Student learning outcomes · Game design features

## Introduction

Since its inception in the early 2000s, gamification has grown in popularity and evolved along with the information and communication technologies (e.g., Internet) that support the method. Gamification has emerged across sectors (e.g., healthcare) and disciplines (e.g., computer science) as a way to engage users in non-game contexts using techniques that are derived from video game design. While there is no universally agreed upon definition, we

---

✉ Albert D. Ritzhaupt  
aritzhaupt@gmail.com

<sup>1</sup> School of Teaching and Learning, College of Education, University of Florida, 2423 Norman Hall, PO BOX 117048, Gainesville, FL 32611, USA

have adopted the definition provided by Deterding et al. (2011), which states that gamification is the “use of game design elements in non-game contexts” (p. 9). Seaborn and Fels (2015) note that “the inconsistent use of the term ‘gamification’ serves to impede attempts to define it but also exposes its multiplicity” (p. 18). In this research, we are interested in the application of gamification to educational situations, thus the “non-game context” here refers to educational settings. We have observed a steady rise in the number of empirical studies examining gamification in educational contexts.

Unfortunately, the empirical literature on gamification in education has produced mixed findings with studies showing results both in favor and not in favor of gamification (Buckley and Doyle 2016; Hanus and Fox 2015; Mekler et al. 2017; Sailer et al. 2017). Further, many of the implementations of gamification elements in educational settings have amounted to mere “pointification”, which only involves the use of points, leaderboards, and badges (Seaborn and Fels 2015) in educational studies. As noted by Landers (2014), there are several examples of how other gamification elements (e.g., storytelling) can be operationalized to engage learners in meaningful learning experiences. Thus, gamification in educational settings has more potential than that is presently examined in educational literature.

Some literature reviews of gamification have been broadly conducted in the literature (Hamari et al. 2014; Seaborn and Fels 2015); however, the empirical findings in educational settings are still largely mixed. We were able to identify seven systematic reviews and one recent meta-analysis from the literature. Some of these reviews coded samples by gamification elements and reported number or percentage of samples used each element respectively (Dicheva et al. 2015; Subhash and Cudney 2018; Ortiz et al. 2016); some reviews reported learning results of badges, leaderboards, narrative, points, and/or levels in their included samples (Alomari et al. 2019; Antonaci et al. 2019); and one used meta-analysis (Sailer and Homner 2019), yet it differed in its focus from what is examined here. Thus, the application of gamification in education is presently in disarray without conclusive evidence of the overall effects of gamification, and, more specifically, which gamification design elements and under which situations it works.

Thus, the purpose of this research study is to provide some conclusive evidence of the application of gamification in educational settings on student learning outcomes. To date, only a single, recently published meta-analysis attempts to address this concern (Sailer and Homner 2019). To address this gap, our study provides integrated evidence on the efficacy of gamification as well as how the construct should or should not be operationalized in educational settings. Our overarching research questions are:

- What effects does gamification have on students’ learning outcomes?
- Which gamification design elements are potentially more effective on facilitating the learning process?
- What other variables moderate the effects of student learning outcomes?

## Conceptual framework

Gamification in the present study is the application of the features of video games applied to formal educational settings. The intent of integrating such game design features into educational settings is to hypothetically engage students in deeper learning processes and experiences, optimistically leading to stronger student learning outcomes. While this is not

an entirely new concept in education, the advent of modern information and communication technologies (e.g., smartphones) have propelled the concept into a reality in many different types of educational settings. We adopt Landers (2014) theory of gamified learning to frame the research problem, account for the important educational variables, and operationalize gamification according to both theory and practice. This theory is visualized in Fig. 1, and elaborated in this section.

It is important to note that many different theoretical perspectives could be adopted to characterize the application of gamification to educational settings. Landers et al. (2015) review many possible theories for gamification in learning, including classical conditioning theories of learning, expectancy-based theories, goal-setting theory, and self-determination theory. We selected the theory of gamified learning as our guiding conceptual framework because it provides a context for thinking about the relationships and outcomes in the research, and it is welcoming of other theoretical perspectives to support these relationships. Landers' (2014) theory of gamified learning makes five overarching propositions:

### **Instructional content influences learning outcomes, and attitudes and behaviors**

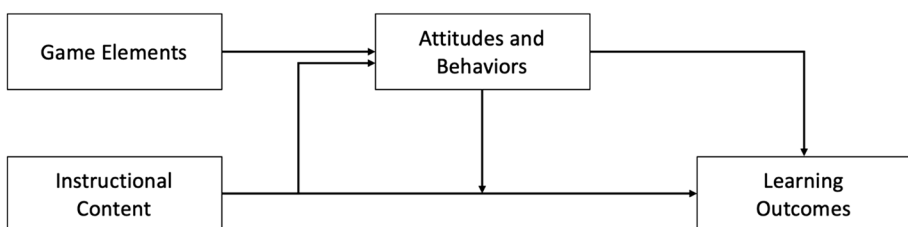
The absence of quality learning materials coupled with gamification will not have a positive influence on learning outcomes. "If the instructional content does not help students learn, gamification of that content cannot itself cause learning" (p. 760). This proposition is an important reminder to both researchers and practitioners. Quality instructional content is necessary irrespective of advancements in technology and pedagogical strategies.

### **Attitudes and behaviors influence learning outcomes**

Varying levels of attitudes and behaviors can have a substantial influence on student learning outcomes. Although the constructs (e.g., intrinsic motivation) may vary across settings, the educational literature has shown evidence of the important relationships among student attitudes and behaviors, and student learning outcomes (Hattie 2008). Students that put forth more cognitive effort, time-on-task, and hold positive attitudes towards a subject will theoretically have stronger learning outcomes (Richardson et al. 2012).

### **Game elements influence changes in attitudes and behaviors**

This proposition justifies the rationale of gamification (e.g., game elements) on influencing relevant student outcomes (e.g., motivation) that mediate student learning outcomes. Current



**Fig. 1** Landers' (2014) theory of gamified learning

literature on gamification distinguishes between both intrinsic motivation and extrinsic motivation. Intrinsic motivation is considered to produce higher-quality learning outcomes and support more creativity in students (Ryan and Deci 2000). Notably, some research on gamification suggests that extrinsic motivators may adversely affect student learning outcomes (Domínguez et al. 2013; Hanus and Fox 2015).

### **Game elements affect attitudes and behaviors, which moderate learning outcomes**

“When moderation is present, the effect of one construct on another depends upon the value of the moderating construct” (p. 761). Figure 1 shows instructional content having an influence on student learning outcomes. With the incorporation of gamification, student attitudes and behaviors should increase, which should make the relationship between the instructional content and student learning outcomes stronger, which is a moderating relationship.

### **The relationship between game elements and learning outcomes is mediated by attitudes and behaviors**

This proposition is the full cycle of the theory of gamified learning, as the influence of gamification on attitudes and behaviors (e.g., motivation) indirectly influences student learning outcomes. In the case of mediation, a mediating variable explains the causal relationship between two other variables. In this case, gamification is mediated by attitudes and behaviors on student learning outcomes.

Landers (2014) delineates among these five relationships (illustrated as lines) in Fig. 1 to operationalize the nature of gamification in educational settings. Thus, the theory of gamified learning inspired us to examine three different domains of learning aligned to existing educational theory (Bloom 1965): (1) behavioral (e.g., time-on-task), (2) affective (e.g., student satisfaction), and (3) cognitive (e.g., student performance on a final exam). The present research study presents the results from the cognitive domain of student learning outcomes. Furthermore, we seek to better understand the variables that might interact with the notion of gamification, such as the classification of the student population (e.g., K-12 versus undergraduates) and the discipline (e.g., business) under investigation. We seek to understand the conditions under which gamification works and does not work for specific types of students, disciplines, learning environments, and more. While many game elements have been described in the literature and taxonomies have been developed (Bedwell et al. 2012; Wilson et al. 2009), we propose these game design elements shown in Table 1, along with an operational definition and an example gamification application. The initial list of these game elements came from our generalized knowledge based on prior studies and our past design and development experiences. During the screening and coding process, we iteratively modified the list in order to include all game elements explicitly or implicitly mentioned in the manuscripts. This is by no means an exhaustive list; however, it does illustrate many of the approaches taken in gamification research in educational environments.

**Table 1** Game elements operationalized in the current study

Element	Description of element	Example of element in education
Quests, missions, or modules	The gamified environment is organized into meaningful “chunks” for the users to clearly understand how to complete the task	The gamified educational environment is composed of multiple units, which may be called quests, missions, tasks, or modules, etc. They are structured in sequence or in a non-linear approach. The completion of these units fulfills desired learning goals
Badges and awards	Badges and awards are presented for completing a task and are frequently displayed and shared in the system to convey status	Students earn unique, visual badges by completing educational tasks or by demonstrating desired behaviors in the environment (e.g., helping a peer)
Points and experience	Points or experience (XP) are a measure that quantifies some action or currency within a system. Often points are used to track progress within a gamified learning environment	The “points” or score earned by students will be based on how well a student completes a task according to predefined parameters, like solving a puzzle on the first attempt within a designated amount of time
Leader boards	A leaderboard is a rank-order listing of users within a gamified system based on some metric, like points	Points earned in specific areas (e.g., program efficiency or overall progress) will be used as a metric to rank the students in an environment to engender competition
Collaboration	Collaboration is manifested by interpersonal communications and joint user activity mediated by the gamified system	Collaboration in the gamified educational environment is supported by communication targeting specific tasks, modules, or quests that require action taken by multiple students
Levels and advancement	Levels convey expertise, skill, or proficiency within a gamified system. Advancement in the environment can open other pathways and experiences	Levels in the educational environment will be aligned with the completion of a quest, task, or module. Levels could be rated by difficulty or skill (e.g., easy, medium, etc.)
Avatars and customization	Users in a gamified environment can make choices about the appearance of their virtual characters and settings in the system to tailor the experience	Student in a gamified educational environment can select an avatar and control the settings for different parameters of the environment, like the display/layout of the dashboard
Timed activity	The use of time limits in the environment to encourage speedy decisions and behavior by the user	Students are tasked with answering a riddle within a two-minute time frame with a countdown clock displayed on the screen
Performance graphs	Performance graphs use the users’ performance data to provide a visual display of performance in the system at a specific point in time	Performance graphs provide a holistic review of the learners’ progress within the gamified environment by illustrating overall progress

**Table 1** (continued)

Element	Description of element	Example of element in education
Narrative or storytelling	Quests, tasks, or modules typically involve some form of narrative structure, a clear goal, and are often part of a larger progression model. The narrative refers to the way events unfold as the user "plays" in the gamified system	A gamified environment starts with an authentic context for a series of quests, tasks, or modules. It may unfold with narratives before the user progresses to a new quest, task or module
Responsive feedback	Within a gamified system, responsive feedback deals with the user experience. When certain actions are taken within the system (e.g., complete a quest), responsive feedback is the indication to the user that the benchmark was achieved	Students can take an action in the gamified educational environment and receive immediate feedback on whether the action was the "right" or "wrong" course of action
Competition	Users in a gamified environment contend for limited resources within the environment and react to other users' actions	The gamified educational environment rank orders student performance on specific tasks, quests or modules
Adaptivity and personalization	Users in the system receive "tailored" messages and feedback based on some aspect of the user or user's decisions in the system	The gamified educational environment can provide "intelligent" and "custom" feedback to the student based on the student's different behaviors
Non-linear navigation	The environment provides the user multiple pathways through the system as opposed to a prescribed sequence	The students in the gamified system can take different paths to complete the quests, tasks or modules, and advance in different ways

## Method

### Literature search strategy

We engaged in a systematic literature search across 18 academic databases (ACM Digital Library, ACS Publications, DOAJ, EBSCOhost, Gale Databases, HEINONLINE, IEEE Xplore Digital Library, LearnTechLib, NCBI Databases, Ovid, ProQuest, Sage Journals, ScienceDirect, SpringerLink, Taylor & Francis Online, Web of Science, Wiley Online Library, World Cat) to cast a wide net for publications for inclusion in the study, especially since gamification has been applied in several fields for educational purposes. We used a combination of search terms in Boolean search strings for gamification (gamification, gamify, gamified, gamifying), education (assessing, assessment, assessments, impact, impacts, effect, effects, evaluating, evaluation, comparison, comparisons, comparing, influencing, enhancing, outcome, outcomes, result, results, influence, influences), student (student, learner, user, participant, students, learners, users, participants) and effect (learning, comprehension, satisfaction, motivation, achievement, performance, understanding, outcome, outcomes, perception, perceptions, result, results, efficacy, self-efficacy, confidence, expectancy). We didn't use "\*" in search terms because some databases do not accept the wildcard character.

To narrow the search, we applied these search terms to the titles and abstracts of the publications to ensure only relevant publications were identified in the search process. Using this systematic process, we identified 4289 publications across the 18 academic databases, imported all the records into a RefWorks account, and removed duplicate entries, which resulted in a total of 2719 publications for consideration in the present study.

### Inclusion and exclusion criteria

To be included in this meta-analysis, studies had to (1) be published between January 1, 2009 and December 31, 2018 (in press included); (2) clearly identify students as the target population at any educational level in a formal educational setting; (3) compare a gamification intervention to a non-gamification intervention as a between-subject condition; (4) use a quasi-experimental or experimental research design; (5) primarily focus on students' learning outcomes, particularly on clearly defined cognitive (e.g., comprehension), affective (e.g., motivation), and behavioral measures (e.g., time on task); and (6) reported in the English language. Using these inclusion and exclusion criteria, we filtered the publications following systematic procedures outlined in the following section. Figure 2 shows the PRISMA flowchart for our search, identification, screening, coding, and extraction processes (Moher et al. 2009).

### Identification, selection, and coding

We executed the systematic process for identifying, selecting, and coding the relevant publications in stages. This process took place in two phases. After the searches, we began a screening process of the publications with eight members of the research team. First, we conducted a training meeting to prepare the team members to identify key criteria (e.g., students in the sample) for classifying a manuscript as qualified, not qualified, or unsure. As a group, we discussed the inclusion and exclusion criteria and used publications for examples. After the training session, we classified 20 manuscripts independently

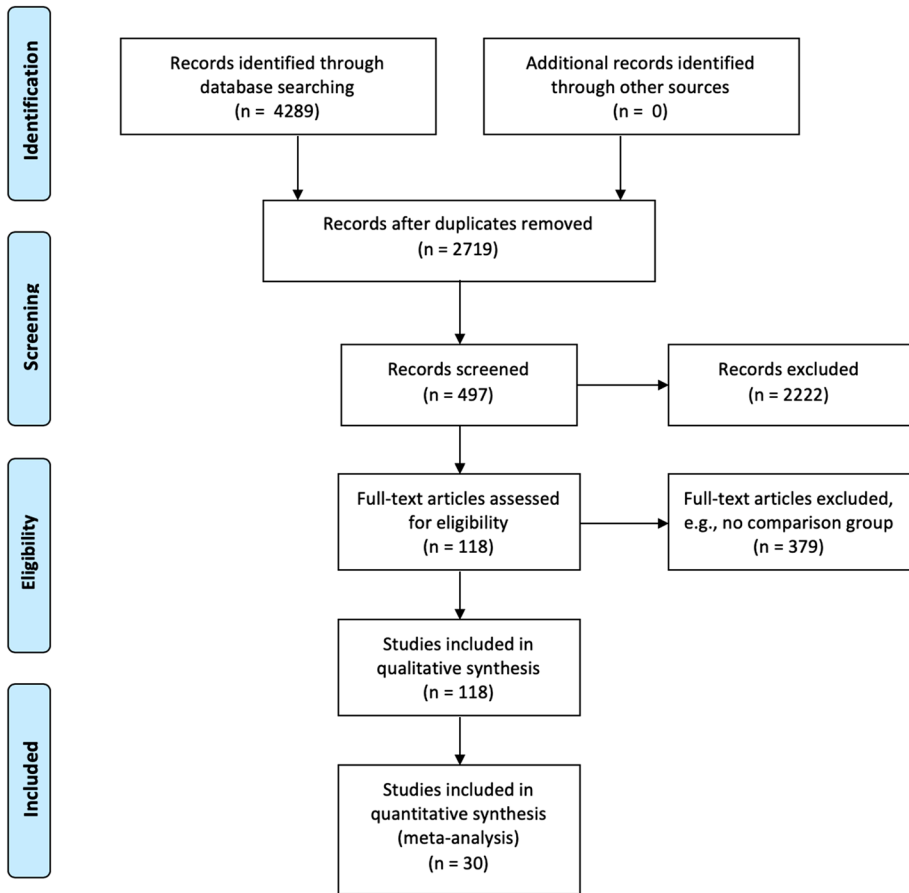


Fig. 2 PRISMA flowchart of present study process

and calculated the Fleiss' Kappa for inter-rater agreement. The Fleiss' Kappa was calculated at .62, which suggests substantial agreement (Landis and Koch 1977). Consequently, the total list of publications was evenly split among the eight team members and carefully classified independently using two possible outcomes: include and exclude. We decided to drop the unsure category and mark unsure publications as include since the subsequent process of screening the full-text publications would be the same. We selected publications that used students as research participants and indicated a comparison of at least two conditions between gamified versus non-gamified educational settings using experimental or quasi-experimental designs. We purposefully excluded publications using only qualitative methods because they are inappropriate for a meta-analysis. This phase resulted in 118 manuscripts for consideration in phase two.

In the next phase, we carefully read the full-text publications remaining in our dataset and coded and extracted information in a systematic manner. First, a codebook guide was developed based on prior literature on gamification and modeled after a meta-analysis on the flipped classroom instructional strategy (Cheng et al. 2019). The codebook extracted the following details from each manuscript:



- Author detail information including the authors' last names and year of publication.
- Title of manuscript
- Type of manuscript (1-Journal article, 2-Dissertation/thesis, 3-Conference proceeding)
- Name of proceeding, journal, etc.
- Discipline of the intervention (1-social science, including psychology, education, 2-science, including biology, physics, etc., 3-math, including statistics, 4-engineering/computing, 5-arts and humanities, 6-health, include nursing and medicine, 7-business, information systems, 8-other)
- Student level classification (1-K-12, 2-undergraduate, 3-graduate, 4-mixed higher education, 5-mixed higher education and K-12, 6-higher education (unknown of classification))
- Research design used (1-experimental, 2-quasi-experimental)
- Domain of the dependent measure (1-cognitive, 2-affective, 3-behavioral)
- Learning environment (1 = digital, 2 = physical, 3 = blended)
- Gamification design features, each coded as a separate parameter (Leadership Boards, Badges/Awards, Points/Experience, Advancement (e.g., levels), Responsive Feedback, Narrative/Storytelling, Avatars/Customization, Timed Activity, Collaboration, Competition, Non-Linear Navigation, Quests/Missions/Modules, Adaptivity/Personalization, and Performance Graphs) (1 = present, 0 = absent)
- All of the statistical information needed to calculate effect sizes (e.g.,  $n$ ,  $M$ ,  $SD$ )

Our process involved the research team examining the articles in dyads (i.e., two research team members) to improve coding and reliability. As there were eight members in our research team, we split into four dyads with each team member possessing advanced knowledge of quantitative research methods in education. We exercised reliability procedures for all team members by hosting two training sessions in which a random sample of 10 manuscripts were coded by all four dyads using a standard codebook guide. After, the coding differences were examined and discussed by team members to ensure consistency in coding practice. We then split up the 118 manuscripts among the four dyads for coding and effect size extraction. Next, we carefully extracted publication details (e.g., date published), gamification moderators (e.g., badges, leaderboards, etc.), and contextual moderators (e.g., education level, discipline, etc.) using the codebook. During the coding and data extraction process, the four dyads kept in close communication and unified decisions were made in the event that unique parameters or new questions needed to be addressed by the full team. Finally, we created a comprehensive table to record the study publication details (e.g., author and year), stated moderators, and effect size details (e.g.,  $n$ ,  $M$ ,  $SD$ ).

### Effect size extraction and calculations

Comprehensive Meta-Analysis (CMA) version 3.0 was used to calculate the effect sizes for the publications identified through our systematic procedures. We employed SPSS version 25.0 to descriptively analyze our dataset. We extracted effect sizes in three domains: (1) cognitive (e.g., student performance on a final exam), (2) affective (e.g., student satisfaction), and (3) behavioral domains (e.g., time on task). As a reminder, this current study only shows the results of the cognitive domain, or what we label as student learning outcomes. There were many effect size comparisons with multiple dependent comparisons in the same study (e.g., mid-term and final exam). In calculating the effect sizes, we computed only one effect size per study for the cognitive, affective, and behavioral models. As

noted by Lipsey and Wilson (2001), when a study contributes more than one effect size, it can lead to statistical dependence, resulting in biased overall effect sizes. Thus, we combined the effect size comparisons by study using the averages in CMA 3.0 as an independent entity (one per publication). As a reminder, the present study only presents the results from the cognitive domain, or student learning outcomes according to our conceptual framework. Our final corpus was a list of  $k=30$  manuscripts and associated effect sizes. As with any meta-analysis, effect sizes must be standardized before running the analysis. We chose to use Hedge's  $g$  as the standardized measure of effect size for continuous variables because Hedge's  $g$  is better than Cohen's  $d$  for adjusting small sample size bias (Borenstein et al. 2011). Cohen's  $d$  is the difference between outcome means with respect to the pooled standard deviations. The pooled standard deviation is calculated based on the standard deviations in both groups. Hedge's  $g$  and Cohen's  $d$  operate similarly with large sample sizes, but Hedges'  $g$  has the best properties for small samples, as it is equivalent to Cohen's  $d$  multiplied by a correction factor  $J$  that adjusts for small sample bias.

All data were assumed under random-effects models with an  $\alpha=.05$ . Borenstein et al. (2011) suggest random effects models are more appropriate when the effect sizes of the studies included in the meta-analysis differ from each other. Since the outcomes measures and gamification environments differed dramatically from study-to-study, we chose to use the random-effects model in this study. Post hoc subgroup analyses were conducted with the mixed effects analysis (MEA) as implemented in the CMA 3.0. An effect size of .2 is considered small, .5 is considered medium, and .8 is considered large (Cohen 1992). To account for the possibility that the current meta-analysis overlooked non-significant results, the fail-safe  $N$  (Rosenthal 1979), which is the number of unpublished studies needed in order to change the effect size estimate to non-significant, was calculated. Also, the Orwin's fail-safe  $N$  test was calculated to determine the number of missing null studies required to bring the existing effect size to a trivial level (Orwin 1983). Publication bias was evaluated with the fail-safe  $N$  procedure, Orwin's fail-safe  $N$  test, and by visual inspection of the funnel plot.

## Results

The final corpus accounted for  $k=30$  independent effect sizes and a total of  $N=3083$  participants. The gamification condition accounted for  $n=1524$  participants, while the non-gamification condition accounted for  $n=1559$ . The resulting sample of empirical gamification studies were published in a wide range of journals, dissertations, and conference proceedings. These journals ranged from *Computers & Education* to *Australasian Journal of Educational Technology* to *Simulation & Gaming*. Conference proceedings came from several sources as well, including *IEEE Frontiers in Education Conference* and *Learning and Teaching in Computing and Engineering*. All of these manuscripts were published from 2014 to 2018.

### Overall effect of gamification on student learning outcomes

Figure 3 illustrates the forest plot for these 30 separate studies in relation to the author(s) of the study and year, Hedge's  $g$  (sorted by effect size coefficient), the standard error, variance, confidence interval,  $Z$ -value, and  $p$ -value. The small boxes indicate the point estimates of effect size in each single study, whereas the horizontal line that crosses each box

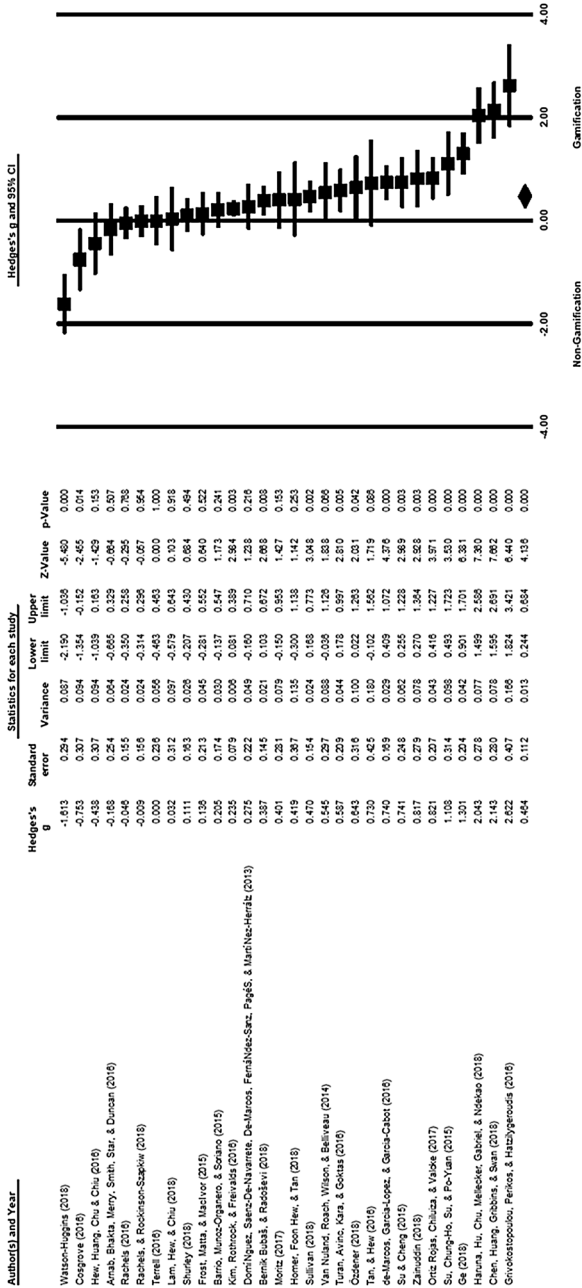


Fig. 3 Forest plot of student learning outcomes

represents the confidence interval for a study's estimate. If the confidence interval overlaps with zero, the study is easily identifiable as not statistically significant. The diamond at the bottom of the plot represents the average effect size after pooling all studies together along with a confidence interval. As can be gleaned, six of the studies resulted in negative effect sizes for the gamification condition, one zero effect size, and the remaining studies were in favor of the gamification condition. However, not all of the positive effect sizes were statistically significant (e.g., confidence intervals overlapped with zero) with 15 statistically significant Z-values in favor of gamification. The effect sizes of the studies ranged from  $g = -1.613$  to  $g = 2.622$ .

The distribution of Hedge's  $g$  was normally distributed as shown in the histogram in Fig. 4. The overall effect size using a random-effects model is  $g = .464$ , which is a small to medium effect size (Cohen 1992). This overall effect size was statistically significant at  $Z\text{-value} = 4.136$ ,  $p < .001$  with a 95% confidence interval of .244 to .684. The observed effect size varies from study to study, but a proportion of variation is expected due to sampling error. The Q-statistic provides a test of the null hypothesis that all studies in the analysis share a common effect size (Borenstein et al. 2011). The Q-value is 245.959 with 29 degrees of freedom and a p-value of  $p < .001$ . Thus, we can reject the null hypothesis that the true effect size is identical in all of the studies. The  $I^2$  is 88.209%, which informs us that the proportion of the observed variance reflects differences in the true effect sizes rather than sampling error (Borenstein et al. 2011). Higgins and Thompson (2002) suggest that values of  $I^2$  around 25%, 50%, and 75% can be interpreted as low, medium, and high heterogeneity, respectively. The  $I^2$  for the overall model shows a high level of heterogeneity and indicates one or more moderators, other than sampling error, could account for this heterogeneity in these data. Thus, exploring

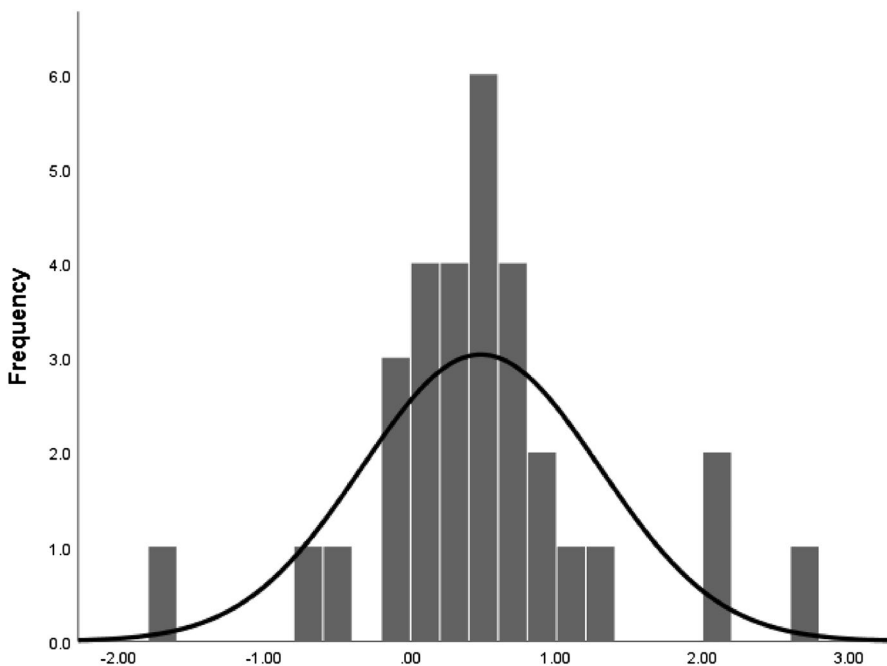


Fig. 4 Histogram of effect sizes

potential moderators in these data to predict and explain the variability is necessary. We explore possible moderators in the form of different gamification design elements and contextual factors of the studies.

## Gamification design elements

Table 2 provides the effect sizes broken down by the various gamification design elements discussed in our conceptual framework. The majority of the studies employed leaderboard, badges/awards, and points/experience. Leaderboards appear to result in a statistically significant effect size of  $g = .358$ , while those studies not employing leaderboards resulted in a statistically significant effect size of  $g = .771$ . Badges/awards, points/experience, and advancement/levels all appear to generate comparable statistically significant effect sizes irrespective of whether the design features were employed or not in the study. Responsive feedback was employed in  $k = 19$  of the studies with a statistically significant medium effect size of  $g = .509$  [.85, .833]. Narrative/storytelling was employed in few of the gamification studies, and does not appear to result in a noticeable difference. While avatars/customization was only adopted in  $k = 8$  studies, an estimated  $n = 1091$  participants experienced this design feature.

The use of timed activities in the design of learning environments as a gamification design feature appears to result in a smaller effect size than other conditions at  $g = .236$ , while the environments not using timed activities boast an effect size of  $g = .529$ . However, the timed activity condition was not statistically significant with the confidence interval overlapping zero, nor was the difference statistically significant. While collaboration was only used in  $k = 9$  of the studies, it resulted in the second highest measured effect size across the gamification design features at  $g = .609$  [.222, .997] and was statistically significant. Although competition as a design feature was used in many studies at  $k = 21$ , those studies not employing the variable resulted in a higher statistically significant effect size at  $g = .59$ . Non-linear navigation and adaptivity/personalization were only used in  $k = 5$  studies, and the differences between the conditions being present versus absent did not result in statistically significant differences. Finally, the highest noted effect size for a gamification design feature was observed in the use of quests/missions/modules in the interventions at  $g = .649$  [.279, 1.02].

## Contextual elements in gamification

### Type of publication

The type of publication was examined as a potential variable to moderate the effect size results. As noted, we collected three different types of publications, including journal articles, conference proceedings, and dissertations/theses. The results of the analyses are shown in Table 3. As shown, most effect sizes were derived from published journal articles at  $k = 18$ . Both journal articles and conference proceedings resulted in comparable effect sizes while dissertations/theses resulted in an overall negative effect size. These results were significantly different across the three conditions with both journal articles ( $Q\text{-value} = 9.555$ ,  $p = .002$ ) and conference proceedings ( $Q\text{-value} = 6.032$ ,  $p = .014$ ) being significantly different than dissertations/theses.

**Table 2** Gamification design elements effect size data

Gamification element	Effect size and 95% confidence interval						Heterogeneity				
	N	k	g	SE	Lower	Upper	Z	p	Q-value	df	p
Leaderboards	2359	23	.358	.128	.107	.608	3.115	.002			
No leaderboards	724	8	.771	.247	.286	1.256	2.797	.005			
Total between									2.199	1	.138
Badges/awards	2391	22	.458	.134	.195	.721	3.416	.001			
No badges/awards	539	8	.480	.218	.053	.907	2.205	.027			
Total between									.007	1	.932
Points/experience	2580	24	.463	.130	.208	.719	2.235	.025			
No points/experience	503	6	.479	.214	.059	.899	3.551	.000			
Total between									.004	1	.950
Advancement/levels	1831	14	.488	.181	.134	.842	2.701	.007			
No advancement/levels	1252	16	.449	.141	.173	.725	3.189	.001			
Total between									.029	1	.866
Responsive feedback	2148	19	.509	.165	.185	.833	3.076	.002			
No responsive feedback	935	11	.405	.105	.200	.610	3.877	.000			
Total between									.280	1	.597
Narrative/storytelling	519	5	.473	.162	.155	.791	2.911	.004			
No narrative/storytelling	2564	25	.464	.133	.205	.724	3.503	.000			
Total between									.002	1	.968
Avatars/customization	1091	8	.581	.264	.064	1.097	2.204	.028			
No avatars/customization	1992	22	.430	.127	.180	.679	3.378	.001			
Total between									.267	1	.605
Timed activity	710	6	.236	.222	-.199	.670	1.064	.287			
No timed activity	2373	24	.529	.133	.268	.790	3.971	.000			
Total between									1.284	1	.257

**Table 2** (continued)

Gamification element	Effect size and 95% confidence interval						Heterogeneity				
	N	k	g	SE	Lower	Upper	Z	p	Q-value	df	p
Collaboration	719	9	.609	.198	.222	.997	3.085	.002			
No collaboration	2364	21	.406	.136	.139	.673	2.978	.003			
Total between									.720	1	.396
Competition	2318	21	.412	.138	.140	.683	2.975	.003			
No competition	665	9	.590	.205	.188	.992	2.874	.004			
Total between									.517	1	.472
Non-linear navigation	729	5	.515	.293	-.058	1.088	1.761	.078			
No non-linear navigation	2354	25	.455	.125	.210	.700	3.634	.000			
Total between									.036	1	.850
Quests/missions/modules	1142	12	.649	.189	.279	1.020	3.436	.001			
No quests/missions/modules	1941	18	.343	.145	.060	.627	2.373	.018			
Total between									1.652	1	.199
Performance graphs	791	6	.309	.142	.032	.586	2.183	.029			
No performance graphs	2292	24	.513	.142	.235	.790	3.620	.000			
Total between									1.035	1	.309
Adaptivity/personalization	1080	5	.563	.257	.060	1.067	2.192	.028			
No adaptivity/personalization	2003	25	.443	.133	.182	.705	3.327	.001			
Total between									.171	1	.679

**Table 3** Type of publication effect size data

Type of publication	Effect size and 95% confidence interval								Heterogeneity		
	N	k	g	SE	Lower	Upper	Z	p	Q-value	df	p
Journal article	1430	18	.662	.150	.368	.957	4.407	.000			
Dissertation/thesis	740	7	-.170	.223	-.608	.268	-.761	.447			
Conference proceeding	913	5	.666	.257	.163	1.170	2.593	.010			
Total between									10.405	2	.006

### Subject area of the study

The subject area of the gamification study was also examined as a potential moderating variable in the present study as shown in Table 4. A great number of studies and participants are in the disciplines of the social sciences ( $k = 10$ ,  $n = 813$ ) and engineering and computing ( $k = 6$ ,  $n = 1104$ ). Several other fields, such as mathematics, healthcare, and business, appear to have been slower to adopt the use of gamification in their learning environments. While the model shows a statistically significant difference across these conditions at  $Q\text{-value} = 16.065$ ,  $p = .025$ , these results should be interpreted lightly since so many disciplines have so little representation. Notably, both the social sciences and engineering/computing have statistically significant effect sizes great than .5 or medium to large effect sizes (Cohen 1992). Also, arts and humanities, although not statistically significant, appear to have the lowest effect size for the subject areas.

### Student classification in studies

Table 5 shows the results of the analyses by the student classification of the participants involved in the gamification research studies. Most studies of gamification in formal educational settings appear to have been conducted with undergraduate students ( $k = 13$ ,  $n = 1724$ ) with a medium to large effect size at  $g = .638$  [.378, .898]. K-12 students have been studied next highest ( $k = 10$ ,  $n = 920$ ), but the effect size confidence intervals

**Table 4** Subject area effect size data

Subject area	Effect size and 95% confidence interval								Heterogeneity		
	N	k	g	SE	Lower	Upper	Z	p	Q-value	df	p
Social science	813	10	.590	.232	.135	1.045	2.540	.011			
Science	233	4	.478	.234	.019	.937	2.040	.041			
Mathematics	141	2	.266	1.878	-3.415	3.948	.142	.887			
Engineering/computing	1104	6	.621	.195	.240	1.002	3.192	.001			
Arts and humanities	561	5	.039	.085	-.128	.205	.457	.648			
Healthcare	50	1	.401	.281	-.150	.953	1.427	.153			
Business	87	1	.136	.213	-.281	.552	.640	.522			
Other	94	1	.587	.209	.178	.997	2.810	.005			
Total between									16.065	7	.025



**Table 5** Student classification effect size data

Student classification	Effect size and 95% confidence interval							Heterogeneity			
	N	k	g	SE	Lower	Upper	Z	p	Q-value	df	p
K-12 students	920	10	.306	.235	-.156	.767	1.298	.194			
Undergraduate students	1724	13	.638	.133	.378	.898	4.813	.000			
Graduate students	122	2	.856	1.291	-1.674	3.385	.663	.507			
Mixed higher education	70	1	.000	.236	-.463	.463	.000	1.000			
Mixed higher education and K-12	131	1	.205	.174	-.137	.547	1.173	.241			
Higher education (unknown level)	116	3	.102	.449	-.779	.983	.227	.821	7.999	5	.156
Total between											

overlap zero and the model for K-12 students is not statistically significant. Few graduate students appear to have been examined in gamification studies at  $n = 122$ . Notably, the effect size for K-12 students is smaller than that of the undergraduate comparison, but the difference between the two is not statistically significant. Some of the studies mixed students at various levels as shown in the Table. The  $Q$ -value shows a statistically significant difference among these conditions, but this should be interpreted lightly since only a few studies fall into several of the categories.

## Publication bias

A visual inspection of the funnel plot generated from the meta-analysis shows generally symmetrical distributions around the weighted mean effect sizes with a few outliers (see Fig. 5). A funnel plot is a scatter plot of effect sizes estimated from the studies in the meta-analysis against the standard error (Stern and Egger 2001). In the diagram, the horizontal axis is the Hedges'  $g$ , plotted alongside the standard error on the vertical axis. Generally speaking, a symmetric funnel plot suggests the absence of publication bias in the meta-analysis (Duval and Tweedie 2000), but visual inspection alone is insufficient. Thus, we also adopted the Classic fail-safe  $N$  and the Orwin's fail-safe  $N$  test. The Classic fail-safe  $N$  test showed that 869 additional studies of gamification on student learning outcomes would be required to nullify the overall effect size found in the present meta-analysis. These calculations show that the number of null or additional studies needed to nullify the overall effect sizes found in this meta-analysis is larger than the  $5k + 10$  limit (Rosenthal 1995). Finally, the results of Orwin's fail-safe  $N$  test tells us that the number of missing null studies required to bring the current overall mean effect size to a trivial level ( $g = .01$ ) was 1087. Consensus among the results of the Classic fail-safe  $N$  test, Orwin's fail-safe  $N$  test, and inspection of the funnel plot provide evidence that publication bias does not pose a serious threat to the validity of the meta-analysis.

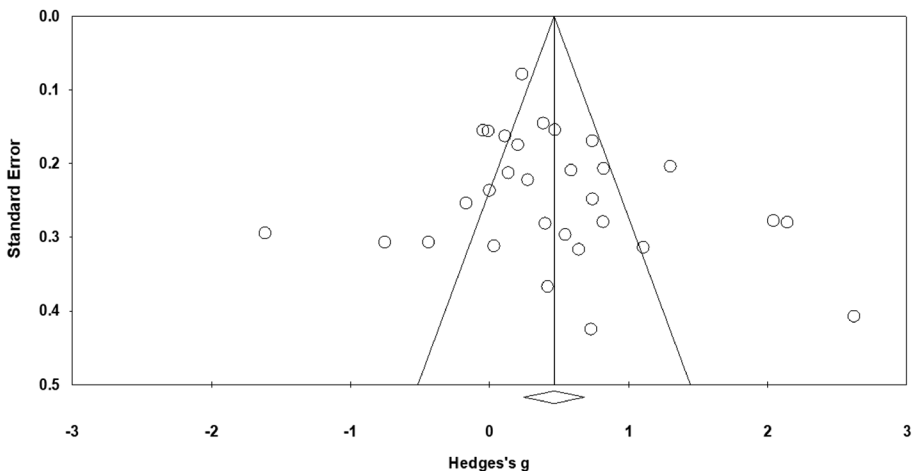


Fig. 5 Funnel plot of effect size data

## Discussion

Implementations of gamification may result in costly technology-enhanced learning solutions as several resources (e.g., graphic designers, programmers, instructional designers, subject-matter experts) must come together to fully realize the scope of the instructional approach. As many educational organizations are investing in gamification in their online and blended educational settings, it is imperative that we have evidence of the efficacy of the gamified instructional approach from the literature. In particular, it is important that we have nuanced knowledge about which gamification design features work and under which circumstances. The findings from this paper shed light on this issue by providing a comprehensive review of gamification studies in educational settings, focused on student learning outcomes. We hope the findings could contribute to the efforts of discerning the cost/value worth and the specific gamification elements which would be more prominent for future research and application.

The primary conclusion from the analyses conducted demonstrates that gamification does appear to have a positive and significant effect on student learning outcomes in formal educational settings. The effect size of  $g = .464$  is a small to medium effect according to Cohen's (1992) criteria; however, other scholars have noted that effect sizes of .40 and greater to be practically relevant to influence educational outcomes (Hattie 2008). This finding mirrors a recent meta-analysis on gamification with an effect size of  $g = .49$  on cognitive learning outcomes (Sailer and Homner 2019).

Prior to interpreting the findings, we present our limitations and delimitations so readers can consider all dimensions in making future decisions or conclusions from this study. While we did access 18 academic databases to identify our corpus, it is possible that we missed both published articles and gray literature on gamification in educational settings. Cooper (2017) further recommends citation/bibliography searches or forward reference searches, and we did not complete this step in our literature search strategy since we started with more than 2000 manuscripts in our initial search. We chose to pool effect sizes that represented student learning outcomes from across settings and participants. While this is a critique to meta-analysis in general, there was great variability in the types of outcome measures employed in the studies. While our publication bias analysis did not show strong evidence of publication bias, it is notable that the vast majority of our manuscripts and effect sizes are reported from journal articles. Finally, although we trained team members on several rounds, coded and extracted data in dyads of research team members, kept close communication during the coding and extraction process, and provide initial evidence of inter-rater agreement, we did not calculate inter-rater agreement at every stage of this research.

Consistent with the observation of Seaborn and Fels (2015), it would appear that most studies of gamification are equivalent to that of "pointification" in that most solutions involve leaderboards, points, and badges. This is evidenced by the large proportion of studies using these design features from our analyses and the large number of participants that have used these design features in the studies. However, as noted by several others (e.g., Landers) and by our own list of possible gamification design features, there is much more potential to the deployment of gamification beyond the mere "pointification" found in most studies. While we do not claim that our list of gamification design features is comprehensive by any means, we do believe that those features represent most of the incarnations of gamification in present work in formal educational settings. Thus, we encourage both researchers and practitioners in the field to expand their arsenal of gamification in our

learning designs. As such, we can now discuss several of the gamification design features we studied in greater detail.

### **Gamification design elements**

It would appear that not all gamification design elements have the same effect on student learning outcomes. To start, consider the effects of leaderboards, which are frequently employed in gamification interventions. Our results suggest that studies not using this design feature resulted in a higher statistically significant effect size than those studies that did use this feature. While the difference between the two was not statistically significant, it begs to question whether leaderboards result in the intended outcomes. While leaderboards are often intended to serve as a form of extrinsic motivation among the learners (Mekler et al. 2017), this form of motivation might undermine the intended goal of improve student learning outcomes as our conceptual framework tells us that student learning performance is moderated by attitudes and behaviors. The individualistic and competitive nature of leaderboards might serve to adversely influence students in attaining the intended educational outcomes. While more empirical research is necessary on this topic, our results minimally call into question the overuse of leaderboards in gamification solutions.

Badges/awards, points/experience, and advancement/levels appear to result in similar effect sizes both with and without the use of the gamification design feature. While points, badges, and advancement are typically intertwined (e.g., the more points earned, the greater the advancement in the environment and possible badges earned), this finding might tell both researchers and practitioners that these particularly design features may not warrant full consideration in gamification learning solutions and future research studies. Most gamification solutions must employ some form of currency (i.e., points) to show progress and completion, so points and advancement may just be necessary aspects that must be built into the learning environments without the hope of positively influencing student learning outcomes. Badges on the other hand represent a substantial amount of time to design and develop in a gamification environment. Designing the conditions upon which badges are awarded, as well as developing the graphics for the badges themselves might be deemphasized in future learning solutions.

Responsive feedback resulted in slightly higher effect size than in those solutions that did not employ the feature. Responsive feedback is not surprising finding in that we have decades of research demonstrating the effectiveness of different feedback configurations on student learning outcomes (Mory 2004; Thurlings et al. 2013). Gamified learning environments allow feedback to be presented in a unique manner, as the feedback can be carefully coordinated with the narrative and goals of the task.

One way in which a gamified environment might make use of responsive feedback is through the use of non-linear navigation within the environment. For instance, a learner could make a decision based on feedback and navigate to an advanced quest within the environment based on that feedback. While non-linear navigation was not employed often, it did result in a statistically insignificant effect size larger than the overall effect size from the meta-analysis. In many ways, non-linear navigation is indicative of the concept of learner control in the navigation of technology-enhanced learning environments (Scheiter and Gerjets 2007), like hypermedia systems. The notion of non-linear navigation is not exclusive to gamification.

While timed activities and performance graphs were not implemented widely in the gamified learning environments, the effect sizes associated with these two design features

demonstrate that both conditions may not be fruitful avenues for future gamification research or practice. Timed activities had the lowest calculated effect size at  $g = .236$  and performance graphs was at  $g = .309$ . Though the timed activity condition was not statistically significant, it still raises the question of whether timed tasks and learning activities in gamified environments are appropriate for learners, as the additional pressure might adversely influence the students' learning outcomes. As for performance graphs, these diagrams and splash screens typically show learners their overall progress and, in some cases, norm-reference performance based on other learners in the environment. While we unfortunately did not discern whether the performance graphs used norm-referencing, such a practice might have a similar effect to that of the leaderboards previously discussed. Careful attention should be placed on the information provided on a performance graph and the way the information is presented to learners, akin to a learner analytics dashboard in other types of learning environments (Verbert et al. 2013).

Two promising gamification design features are collaboration and quests/missions/modules, with statistically significant effect sizes higher than the overall effect size. In fact, the use of quests, missions, or modules resulted in the highest overall effect size at  $g = .649$  [.279, 1.02]. The application of quests, missions, or modules in gamification resembles two important educational concepts: 1) providing explicit learning goals, and 2) "chunking" the educational materials in a meaningfully engaging way. The inclusion of explicit learning goals is a hallmark of instructional design theory and practice (Dick et al. 2005). A gamified environment provides the opportunity for designers to present learning goals and activities that help manifest the educational content in a meaningful way. These goals and activities are embedded in a unit, or "chunk" of educational content that does not overwhelm a learner's working-memory while engaged in the goal. Notably, collaboration resulted in the second highest effect size across all gamification design features at  $g = .609$  [.222, .997]. Collaboration can be manifested in many ways in gamified environments, but the important takeaway is that the interaction among peers in a learning activity appears to have a durable effect on student learning outcomes (Sung et al. 2017) across educational settings.

## Contextual elements in gamification

Our results showed that the dissertations and theses on gamification in educational settings resulted in a statistically insignificant negative effect size, while both journal articles and conference proceedings produced effect sizes above .65. These findings might suggest that peer reviewers have a preference for findings with positive effects from the interventions under investigation, and makes us reflect on the "file drawer problem" (Rosenthal 1979), which results from not publishing statistically non-significant findings. Another likely explanation is that the doctoral candidates writing their dissertations on gamification did not have access to the same level of resources experienced scholars had during their implementations, and, as a consequence, had less robust research findings. Regardless, the larger effect sizes associated with the journal articles and conference proceedings does raise questions about whether other gamification studies were not published as a result of the editorial process.

As noted in the results, the largest proportion of gamification studies have been conducted in the subject areas of the social sciences and engineering/computing. These disciplines also account for the largest proportion of participants at  $n = 813$  and  $n = 1104$ , respectively. While the effect sizes for both of these disciplines were statistically significant

and greater than .50, other disciplines do not seem to favor the use of gamification, such as the arts and humanities with a statistically insignificant, trivial effect size at  $g = .039$ . Other disciplines, aside from science, only resulted in two or less studies. As a result, effect sizes for these disciplines were not statistically significant. These findings make us ask a deeper question: why does gamification appear to work in some disciplines, but not others? Are there confounding variables in these subjects, like the nature of certain subject matters and/or the nature of learning processes of these subject matters, the nature of the students or instructors, or the culture of the domain that conflict with the overall tenants of gamification? Unfortunately, this particular analysis resulted in more questions than answers, but does provide a direct path for future research endeavors.

A final consideration is for whom does gamification work? We attempted to shed light on this question by classifying the students involved in the studies to K-12, undergraduate, and graduate; however, a few of the studies mixed participants while others were unclear about the population under investigation. While most participants were classified as undergraduate students at  $n = 1724$ , a sizable portion were also in the K-12 context at  $n = 920$ . The effect size for undergraduates was statistically significant at  $g = .638$  [.378, .898], but the effect size for K-12 students was not statistically significant and nearly half the size at  $g = .306$  [-.156, .767]. Seeing more studies in undergraduate education is unsurprising as they represent convenience samples at researcher's home institutions. However, why the undergraduates' effect size nearly doubles that of K-12 students is an important concern, and again, forces us researchers to ask important questions. Why do undergraduates thrive with gamification? What supports do K-12 students need to be successful in gamified learning environments? Are K-12 students developmentally ready for the facets of gamified learning?

### Future research on gamification

Our findings call into question several areas about gamification that remain unanswered in the literature base. First, our findings call for attention to consider gamification design elements beyond "pointification" such as collaboration and quests/missions/modules when designing/implementing gamification learning environments. Second, some of the gamification design features have not been used regularly, but are worth more exploration (e.g., adaptivity/personalization). Third, we have only examined the potential of gamification in educational settings in this meta-analysis and in the primary studies in which we base our findings. More gamification elements can be designed and implemented beyond those provided in this manuscript. We encourage both researchers and practitioners to push the limits in their implementations and try alternative designs that attempt to mimic the facets of the most popular video games not presently represented in the gamification research literature.

We would like to propose researchers and practitioners to move past this current era of "pointification" (points, badges, and leaderboards) and evolve into something with more potential in facilitating learning. As we iterate through these different approaches to gamification, we will later be able to identify the features that show the most promise for influencing student learning outcomes. The results also suggest we need to fill some voids in our various fields of endeavor, such as business, mathematics, and healthcare. In carrying out gamification across settings, we must ensure that the unique aspects of the disciplines are accounted for in our research and practice. We must continue to ask questions about what factors lead to the successful integration of the instructional

method and when and where the approach is best suited for an audience. Further, our findings show an alarming difference between implementations in undergraduate and those in K-12, though the difference is not statistically significant. This aspect of the research must be more thoroughly explored in our future research, and a complete explanation as to why a gap might exist should be fully explained. Should gamification be found to be inferior for a target population, we must use theory to help explain this outcome. We know one size does not fit all, therefore, future research in this field should attempt to address the issues of what aspects of gamification to deploy with a target population.

Another clear concern is attempting to determine which combination of gamification design elements has the most potential for facilitating learning outcomes. While our sample does not presently lend itself to this form of analysis, we feel as more primary studies are generated, future systematic reviews and meta-analyses can attempt to address this important concern. Our immediate next step is to examine the data we collected in this process on the affective and behavioral outcomes since our conceptual framework calls for these variables to be systematically studied. While gamification appears to have a durable effect on student learning outcomes, it may be that the student attitudes and behaviors are inconsistent with these findings. We would eventually like to build a meta-analytic path model framed by Landers' (2014) theory of gamified learning. Unfortunately, few of the empirical studies simultaneously investigate attitudes, behaviors, and student learning outcomes in the same research paper. The realm of gamification in educational settings is ripe with opportunities for advancement.

## Closing remarks

We now have some conclusive evidence about the overall effects of gamification on student learning outcomes, suggesting the approach might offer promising outcomes for future research and practice. We would like to re-iterate the caveat from our conceptual framework which states that well-designed instructional content must be used in concert with gamification. Gamification cannot replace poor instructional designs and inadequate learning materials. While we strongly believe that “silver bullets” do not exist in educational practice, the application and integration of gamification in educational settings along with other innovations from the field (e.g., flipped classroom or learning analytics) might result in powerful combinations previously unexplored to advance educational outcomes in our domains. Further, this research provides evidence that not all gamification design elements are equal in their effects on student learning outcomes. This finding provides a more nuanced perspective of gamification and points both researchers and practitioners to other promising avenues. We hope this paper has illustrated the potential of gamification to continue to advance our shared goals of improving educational outcomes for all students.

**Funding** This study was not funded by any grants or contracts.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

### \*Meta-analysis

- Alomari, I., Al-Samarraie, H., & Yousef, R. (2019). The role of gamification techniques in promoting student learning: A review and synthesis. *Journal of Information Technology Education Research*, 18, 395–417. <https://doi.org/10.28945/4417>.
- Antonaci, A., Klemke, R., & Specht, M. (2019). The effects of gamification in online learning environments: A systematic literature review. *Informatics*, 6(3), 32. <https://doi.org/10.3390/informatics6030032>.
- \*Arnab, S., Bhakta, R., Merry, S. K., Smith, M., Star, K., & Duncan, M. (2016). Competition and collaboration using a social and gamified online learning platform. *10th European Conference on Games Based Learning: ECGBL*, 19–27.
- \*Barrio, C. M., Muñoz-Organero, M., & Soriano, J. S. (2016). Can gamification improve the benefits of student response systems in learning? An experimental study. *IEEE Transactions on Emerging Topics in Computing*, 4(3), 429–438. <https://doi.org/10.1109/TETC.2015.2497459>.
- 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.
- \*Bernik, A., Bubaš, G., & Radošević, D. (2018). Measurement of the effects of e-learning courses gamification on motivation and satisfaction of students. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 806–811. <https://doi.org/10.23919/MIPRO.2018.8400149>
- Bloom, B. S. (1965). *Taxonomy of educational objectives: The classification of educational goals*. New York: David McKay Company Inc.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Chichester: Wiley.
- Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments*, 24(6), 1162–1175.
- \*Chen, C. C., Huang, C. C., Gribbins, M., & Swan, K. (2018). Gamify online courses with tools built into your learning management system (LMS) to enhance self-determined and active learning. *Online Learning*, 22(3), 20. <https://doi.org/10.24059/olj.v22i3.1466>.
- Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Educational Technology Research and Development*, 67(4), 793–824.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>.
- Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). Thousand Oaks, CA: Sage.
- \*Cosgrove, P. J. (2016). *The Effects of Gamification on Self-Efficacy and Persistence in Virtual World Familiarization* (Doctoral dissertation). Retrieved from <https://hdl.handle.net/10355/56469>
- \*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>.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15). ACM.
- Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). *Gamification in Education: A Systematic Mapping Study*. 15.
- Dick, W., Carey, L., & Carey, J. O. (2005). *The systematic design of instruction*. 6th Ed. Pearson.
- \*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>.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463.
- \*Eunsik, K., Rothrock, L., & Freivalds, A. (2016). The effects of Gamification on engineering lab activities. *IEEE Frontiers in Education Conference (FIE)*, 2016, 1–6. <https://doi.org/10.1109/FIE.2016.7757442>.
- \*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.



- \*Ge, Z.-G. (2018). The impact of a forfeit-or-prize gamified teaching on e-learners' learning performance. *Computers & Education*, 126, 143–152. <https://doi.org/10.1016/j.compedu.2018.07.009>.
- \*Grivokostopoulou, F., Perikos, I., & Hatzilygeroudis, I. (2016). An innovative educational environment based on virtual reality and gamification for learning search algorithms. *2016 IEEE Eighth International Conference on Technology for Education (T4E)*, 110–115. <https://doi.org/10.1109/T4E.2016.029>
- Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?—a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences (HICSS)* (pp. 3025–3034). IEEE.
- Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, 80, 152–161.
- \*Haruna, H., Hu, X., Chu, S. K. W., Mellecker, R. R., Gabriel, G., & Ndekao, P. S. (2018). Improving sexual health education programs for adolescent students through game-based learning and gamification. *International Journal of Environmental Research and Public Health*, 15(9), 2027. <https://doi.org/10.3390/ijerph15092027>.
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Abington: Routledge.
- \*Hew, K. F., Huang, B., Chu, K. W. S., & Chiu, D. K. W. (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>.
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539–1558.
- \*Homer, R., Hew, K. F., & Tan, C. Y. (2018). Comparing digital badges-and-points with classroom token systems: effects on elementary school ESL students' classroom behavior and English learning. *Educational Technology & Society*, 21(1), 137–151.
- \*Lam, Y. W., Hew, K. F., & Chiu, K. F. (2018). Improving argumentative writing: Effects of a blended learning approach and gamification. *Language Learning & Technology*, 22(1), 97–118.
- Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming*, 45(6), 752–768.
- Landers, R. N., Bauer, K. N., Callan, R. C., & Armstrong, M. B. (2015). Psychological theory and the gamification of learning. In T. Reiners & L. C. Wood (Eds.), *Gamification in education and business* (pp. 165–186). Cham: Springer.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.
- 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.
- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed1000097>.
- \*Moritz, S. C. (2017). *Examination of Badges to Increase Nursing Student Engagement: A Quasi-Experimental Study* (Doctoral dissertation). Retrieved from ProQuest. (10258458)
- Mory, E. H. (2004). Feedback research revisited. *Handbook of Research on Educational Communications and Technology*, 2, 745–783.
- Ortiz, M., Chiluiza, K., & Valcke, M. (2016). *Gamification in higher education and STEM: A systematic review of Literature*. 6548–6558. <https://doi.org/10.21125/edulearn.2016.0422>
- \*Ortiz, M., Chiluiza, K., & Valcke, M. (2017). Gamification in computer programming: Effects on learning, engagement, self-efficacy and intrinsic motivation. *The 11<sup>th</sup> European Conference on Games Based Learning (ECGBL 2017)*. 507–514
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8(2), 157–159.
- \*Özdener, N. (2018). Gamification for enhancing Web 2.0 based educational activities: The case of pre-service grade school teachers using educational Wiki pages. *Telematics and Informatics*, 35(3), 564–578. <https://doi.org/10.1016/j.tele.2017.04.003>.
- \*Rachels, J. R. (2016). *The Effect of Gamification on Elementary Students' Spanish Language Achievement and Academic Self-efficacy* (Doctoral dissertation). Retrieved from <https://digitalcommons.liberty.edu/doctoral/1246/>

- \*Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer Assisted Language Learning*, 31(1–2), 72–89. <https://doi.org/10.1080/09588221.2017.1382536>.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353–387.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118(2), 183.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, 69, 371–380.
- Sailer, M., & Homner, L. (2019). The gamification of learning: A meta-analysis. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-019-09498-w>.
- Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review*, 19(3), 285–307.
- Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74, 14–31.
- \*Shurley, K. A. (2018). *Enhancing the Acquisition and Retention of the Navajo Language Using Computer-Based Instruction and the Effects of Static Pedagogical Agents and Gamification Practice* (Doctoral dissertation). Retrieved from ProQuest. (10809069)
- Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, 54(10), 1046–1055.
- \*Su, C. H., & Cheng, C. H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning*, 31(3), 268–286.
- \*Su, C. H., & Su, P. Y. (2015). Study on 3D meaningful mobile gamification learning outcome assessment: An example of blood circulation lesson. *Applied Mechanics and Materials*, 764–765, 1395–1399. <https://doi.org/10.4028/www.scientific.net/AMM.764-765.1395>.
- Subhash, S., & Cudney, E. A. (2018). Gamified learning in higher education: A systematic review of the literature. *Computers in Human Behavior*, 87, 192–206. <https://doi.org/10.1016/j.chb.2018.05.028>.
- \*Sullivan, A. L. (2018). *Open Badges and Student Motivation: A Study of Their Relationship to Student Assessment Scores* (Doctoral dissertation). Retrieved from ProQuest. (10748817)
- Sung, Y. T., Yang, J. M., & Lee, H. Y. (2017). The effects of mobile-computer-supported collaborative learning: Meta-analysis and critical synthesis. *Review of Educational Research*, 87(4), 768–805.
- \*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>.
- \*Terrell, J. E. (2016). *Instructional Methods and Engagement: The Impact of Gamification on Student Learning of APA Style* (Doctoral dissertation). Retrieved from <https://scholarworks.uark.edu/etd/1502>
- Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2013). Understanding feedback: A learning theory perspective. *Educational Research Review*, 9, 1–15.
- \*Turan, Z., Avinc, Z., Kara, K., & Goktas, Y. (2016). Gamification and education: Achievements, cognitive loads, and views of students. *International Journal of Emerging Technologies in Learning (IJET)*, 11(07), 64–69. <https://doi.org/10.3991/ijet.v11i07.5455>.
- \*Van Nuland, S. E., Roach, V. A., Wilson, T. D., & Belliveau, D. J. (2015). Head to head: The role of academic competition in undergraduate anatomical education. *Anatomical Sciences Education*, 8(5), 404–412. <https://doi.org/10.1002/ase.1498>.
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500–1509.
- \*Watson-Huggins, J. (2018). *An experimental study on the effects of a gamified software intervention mathematics achievement among sixth grade students* (Doctoral dissertation). Retrieved from ProQuest. (13820081)
- Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., ... Estock, J. L. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation & Gaming*, 40(2), 217–266.
- \*Zainuddin, Z. (2018). Students' learning performance and perceived motivation in gamified flipped-class instruction. *Computers & Education*, 126, 75–88. <https://doi.org/10.1016/j.compedu.2018.07.003>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Rui Huang** is a doctoral student in the Educational Technology program in the School of Teaching and Learning at the University of Florida. Her primary research area focuses on the design and development, implementation and measurement of immersive second-language learning environment. She has published two books as a co-translator or a co-author in China, and she serves as the Editorial Assistant for the *Journal of Research on Technology in Education*, the flagship research journal for the *International Society for Technology in Education* (ISTE).

**Albert D. Ritzhaupt** is an associate professor of educational technology and associate director for graduate studies in the School of Teaching and Learning at the University of Florida. His primary research areas focus on the design and development of technology-enhanced learning environments, and operationalizing and measuring technology integration in education. Dr. Ritzhaupt serves as the Editor for the *Journal of Research on Technology in Education*, the flagship research journal for the *International Society for Technology in Education* (ISTE).

**Max Sommer** is a doctoral student in the Educational Technology program at the University of Florida. His primary research area focuses on the design of technology-enhanced learning environments, and a new area Max is exploring is the design of instruction pertaining to information literacy education.

**Jiawen Zhu** is a research fellow and PhD student of educational technology in the School of Teaching and Learning at the University of Florida. Her primary research interests focus on informal online communities and the design of online multimedia learning environment.

**Anita Stephen** is a clinical assistant professor in College of Nursing at the University of Florida and a PhD student in the School of Teaching and Learning with a specialization in Educational Technology. Her research interest is in simulation education and its effects in clinical practice.

**Natercia Valle** is a PhD candidate in educational technology and a research assistant in the School of Teaching and Learning at the University of Florida. Her current research involves the design, development, and implementation of learning analytics dashboards to support learners' motivation and learning performance.

**John Hampton** is a third year doctoral student in Educational Technology at the University of Florida. His primary research focus is on the intersection between technology, literacy, and adolescent identity development.

**Jingwei Li** is a PhD candidate in educational technology in the School of Teaching and Learning at the University of Florida. Her primary research areas focus on individual cognitive characteristics and learning performance in game-based, adaptive learning platforms.