The impact of peer competition and collaboration on gamified learning performance in educational settings: a Meta-analytical study

Jana Chi-San Ho[1] · Yu-Sheng Hung[2] · Letty Y.-Y. Kwan[3]

Abstract

This study is a meta-analytical study that examines the effectiveness of gamification in learning performance in educational settings (n = 29; year-span = 2011–2019). Specifically, it aimed to investigate (a) whether gamification could improve learning performance, and (b) whether peer interaction (i.e., peer competition and peer collaboration) moderated the effectiveness of gamification in learning performance. Results from random-effects models showed significant effects of gamification in learning performance ($g = .595$, 95% CI [.432, .758], $N = 3515$). This effect remained robust after excluding outliers and was stable in a sub-split analysis that excludes studies with low methodological rigor (i.e., studies with pre-post test design). Subgroup analyses revealed a moderating effect of peer competition in gamification in learning, suggesting that competitive games were better than non-competitive games for promoting learning performance in educational settings. However, this effect was not robust and no evidence of subgroup differences were found in the sub-split analysis. Peer collaboration did not moderate the effectiveness of gamification in learning as no subgroup differences were found between collaborative games and non-collaborative games. The effectiveness of games that were both competitive and collaborative did not differ from those that were only competitive. Other moderators such as education level and research design were also investigated. No subgroup differences were found for these two moderators. Educational implications and limitations were further discussed.

✉ Jana Chi-San Ho
janahochisan@gmail.com

[1] Department of Psychology, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR

[2] Hong Kong, Hong Kong SAR

[3] Department of Psychology, The University of Macau, Macau SAR, China

1 Introduction

Gamification has gained rapid popularity since the last decade due to its potential to foster motivation, behavioral changes, friendly users' interactions such as competition and collaboration in various contexts, especially in education (Dicheva & Dichev, 2015). Students who have used gamification in their curriculum praised that gamification integration can improve learning outcomes, such as learning engagement and motivation (Kim et al., 2018) and ample evidence has demonstrated increased students' learning performance in gamified learning (e.g., Eseryel et al., 2014a; Zainuddin, 2018a). Recent research has shifted its focus to investigating different game designs and their effects on learning outcomes. Among a variety of game designs, peer interaction in learning games has been a popular topic. In traditional learning settings, many researchers advocated the positive effects of peer collaboration over peer competition. However, in the context of gamified learning, positive effects of peer competition have been found, e.g., competitive gamified learning motivated students to achieve better performance (Karakostas & Demetriadis, 2011). As the effects of peer interaction on learning performance seem to vary depending on the learning contexts, this study aimed to investigate the role of peer interaction in determining the effectiveness of gamification in educational settings. More specifically, this meta-analytical study investigated whether gamification could improve learning performance and how peer collaboration and peer competition played a role in moderating the effectiveness of gamification in students' learning performance.

In this paper, existing research on the effects of gamification and peer interaction in learning, as well as previous meta-analyses on this topic were reviewed in the literature review section. The process of study selection and data analyses using CMA were illustrated in the method section. Results of the main analysis, subgroup analyses, sub-split analyses and sensitivity analyses were reported in the results section. Limitations and educational implications were further discussed.

2 Literature review

Gamification, as defined by Deterding et al. (2011, p. 9), refers to "the use of game design elements in non-game contexts", usually with the aim of engaging people in various tasks. Different from serious games, the purpose of gamification or gamified learning is to engage users using game design elements rather than full-fledged games whereas serious games are designed to "convey learning material in being played through" (Deterding et al., 2011, p. 10). Typical game designs include but are not limited to the sole use or combination of the following game elements: quests, levels, progress, points, leaderboards, badges, virtual goods, teams, etc., with a purpose to make learning more enjoyable (Buckley & Doyle, 2017). As gamification has been widely applied in educational settings, in this current meta-analysis, we would like focus on the context of education and investigate the effectiveness of gamification in educational settings.

In 2006, the Federation of American Scientists (FAS) issued a public report stating that games offer powerful affordances as a medium for education (Clark et al., 2016). Among motivational theories, the motivational affordances of game elements could be best explained by Deci and Ryan's Self-determination Theory (Deci & Ryan, 2002). This theory proposes a continuum of motivation with two ends – extrinsic motivation (when learners are motivated by extrinsic parameters such as rewards or school grades) and intrinsic motivation (when learners are motivated by the nature of learning and gaining knowledge itself). Long-term positive effects are associated with intrinsic motivation. Gamified learning might promote students' intrinsic motivation as it is designed to turn learning material more enjoyable. In addition, the Need Satisfaction Theory (Deci & Ryan, 2002) suggests that every individual has the need to be in control, connected with the environment and feeling competent. Thus, when game designs can satisfy students' sense of autonomy, relatedness and competence, students are more motivated. In Rigby and Ryan (2011), it was believed that the success of video games in engaging students was because the nature of games fulfilled these three intrinsic human needs.

Furthermore, a theoretical model has been proposed by Deterding et al. (2011) to explain the motivational pull of game mechanics. This model considers the interaction of artifactual and situational motivational affordances of game elements in explaining the underlying mechanism of gamification in learning.

"Situated motivational affordances describe the opportunities to satisfy motivational needs provided by the relation between the features of an artifact and the abilities of a subject in a given situation, comprising of the situation itself (situational affordances) and the artifact in its situation-specific meaning and use (artifactual affordances)." (Deterding et al., 2011, p. 3)

To give an example for illustration in the context of gamified learning: the use of leaderboard to motivate students' learning. The artifactual motivational affordances of leaderboards are that they can ignite social comparison that leads to a competitive environment among involved individuals, fueled by the need for achievement (Deterding et al., 2011). On the other hand, the situational motivational affordances are that the rankings on leaderboards used in games have no actual consequence to the students' grades, which gives students a sense of autonomy. The situation present here not only has its motivational affordance of autonomy but also further shapes the situation-specific meaning of leaderboards – artifacts that allow competition with no negative consequences. The motivational affordances from the artifact and the context are closely intertwined to satisfy students' motivational needs.

While much meta-analytical evidence has supported the effectiveness of gamification in learning (e.g., Bai et al., 2020; Briffa et al., 2020; Huang et al., 2020), some research has shredded some light on the downside of gamification, for instance, the motivational elements in games such as real-time scoring during play may distract trainers from the training task, resulting in weakened improvement, especially in the early learning stage (Katz et al., 2014). Distraction by

game mechanics was shown to be harmful in organizational settings as well, with losses in productivity as a consequence (Blohm & Leimeister, 2013; Thiebes et al., 2014). Nevertheless, the majority of research findings have found positive effects of gamification in learning.

2.1 The role of peer interaction in gamification

One of the components of the Need Satisfaction Theory is relatedness (Deci & Ryan, 2002), which contributes greatly to the motivational affordances of gamified learning. Individuals are driven to experience a sense of connection to the others. Thus, in the context of educational settings, the peer interaction induced by games is likely to play a pivotal role in determining the effectiveness of gamified learning. From a sociocultural perspective, social interaction and play are important for learning. "Play" has been viewed as one of the most significant leading activities in childhood that helps to develop children's cognitive, social and emotional development (Verenikina et al., 2003; Vygotsky, 1977). The Sociocultural Theory proposes that the broader cultural, historical and institutional context shapes individuals' mental functioning via social interaction, thus interaction with others plays an important role in psychological development. Learning occurs when the child is socially interacting with others and their peers (Scott & Palincsar, 2013). Vygotsky (Cole et al., 1978) further introduced the concept of the *zone of proximal development* (ZPD), which proposes that children's learning is associated with two levels of development: the actual and the potential levels of development where the latter is determined through problem-solving in collaboration with more capable peers (Cole et al., 1978).

Two common modes of peer interaction in learning have been widely researched ─ peer competition and peer collaboration (e.g., Johnson et al., 1981; Pareto et al., 2012; Plass et al., 2013). Research findings have shown that both collaborative and competitive learning activities seemed to carry a strong motivational effect for students to play games (Pareto et al., 2012). Yet, students' goal structures are different in the two configurations. Michaels (1977) referred to collaboration as positive reward interdependence and competition as negative reward interdependence. To be more specific, collaborative learning is when students' achievements are positively correlated, that is, when one student achieves their goal, others who collaborate with this student achieve their goals too. By contrast, competitive learning is when students' achievements are negatively correlated. When one student achieves his goal, the others whom he competes against fail to achieve their goals (Deutsch, 1962).

2.1.1 Peer competition

Competition has been regarded as an effective way to stimulate individuals' progress (Cagiltay et al., 2015). In the context of gamified learning, competition occurs when individuals or teams compete for finite resources such as levels, badges, and points. The use of leaderboards can further ignite competition by displaying students' rankings, fostering social comparison among peers (de Byl, 2013). Frequency analysis

(Kim et al., 2018) demonstrated that 80% of the students were motivated by competitive game mechanics (i.e., rankings and scores).

Research has mixed findings on the effect of competition on learning. On the one hand, studies have shown the benefits of competition in learning participation, engagement (Burguillo, 2010) and learning performance (Ames, 1984). In gamified learning, peer competition offered a better balance of learning and gaming, where students could learn the materials with an aim to win the game (Chen, 2014). Competitive gamified learning motivated students to achieve better performance (Karakostas & Demetriadis, 2011). On the other hand, the negative effects of competition in learning should be considered. For example, students might feel anxious and have low self-esteem when they fail in competitive games (Lam et al., 2004). Controversies and critiques regarding the reward mechanism in games such as badges have also raised concern as they do not raise students' intrinsic motivation (Facey-Shaw et al., 2020). While some researchers were in favor of the use of badges in games (Immorlica et al., 2015), some others believed that badges might devalue the learning experience when they were simply viewed as external rewards rather than a performance assessment (Reid et al., 2015).

2.1.2 Peer collaboration

While some research (Dillenbourg, 1999; Prince, 2004) has made a distinctive difference between collaboration and cooperation (i.e., students in collaboration work together to achieve a shared goal whereas students in cooperation exchange resources in support of each other's individual goals), peer collaboration in this meta-analysis refers to the situation where peers work together to achieve common goals. Generally speaking, peer collaboration was found to be positively related to academic achievement across a variety of content areas (Slavin, 1980, 1983; Slavin et al., 1984). In traditional classroom settings, plenty of research on the effects of collaboration has been conducted in second language learning. Empirical evidence has shown that working collaboratively to produce a written piece is effective in second language learning (Storch, 2011).

However, findings have suggested that the effect of peer collaboration depended on other factors as well. Peer collaboration was shown to have a stronger positive effect on group task performance (mean effect size = .31) than on individual achievement (mean effect size = .15) (Lou et al., 2001). In addition, Mullins et al. (2011) found that the type of knowledge moderated the relationship between collaboration and knowledge acquisition. More specifically, collaboration was found to be positively associated with conceptual knowledge gain but had no such effect on procedural knowledge gain. Dillenbourg and Fischer (2007, p. 122) argued that "collaborative learning per se is not effective since productive social interactions often do not occur spontaneously". In order to attain positive learning outcomes, learning environments must be purposely designed to trigger collaboration. With the advancement of technologies, Dillenbourg and Fischer believed that interaction could be designed and fostered in computer-supported collaborative learning. In terms of learning with digital games, meta-analytical evidence has demonstrated that collaborative gameplay was more effective than individual gameplay (Wouters

et al., 2013). An additive effect was seen in digital games with both competition and collaboration, which had a greater effect than games with competition only (Clark et al., 2016).

2.2 Objectives

The topic of gamification in learning in assorted contexts (including school setting, higher education, and informal training setting) has been widely researched (Clark et al., 2016; Sailer & Homner, 2020) and reviewed (Faiella & Ricciardi, 2015). Although some scholars have made the effort to specifically review and analyze the effectiveness of gamification in educational contexts (Bai et al., 2020; Zainuddin et al., 2020), the moderating roles of peer competition and peer collaboration in gamified learning, specifically in formal educational settings, have not been examined in detail. For example, Bai et al. (2020) collected studies from 2010 to 2018 and included participants in K-12 or higher education setting but didn't examine the moderating effect of peer interaction. Since literature on learning have revealed different merits of peer competition and peer collaboration in learning performance, with the additional affordances by gamification, it is important to understand how these two types of peer interaction affect learning performance in the context of gamified learning.

Thus, this meta-analytical study focused on education settings and aimed to investigate (a) whether gamification could improve learning performance, and (b) how each type of peer interaction (i.e., collaboration and competition) played a role in moderating the effectiveness of gamification in students' learning performance. Learning performance was focused on the academic learning outcomes, e.g., academic performance, knowledge and skills gained in class (measures of learning performance in the included studies were shown in Appendix 3). Typically in academic contexts, the general mode of assessment for learning performance is final test scores. All included studies for this meta-analytical study were published from 2011 to 2019, covering more studies than previous meta-analysis research did.

Moreover, existing meta-analyses on investigating the effectiveness of gamification in learning performance included only studies with a between-subject design and discarded studies with a within-subject design as effects from those studies were inflated due to measuring the learning effects along with gamification effects. Nevertheless, studies with a within-subject design could provide insight on how effective gamification was in promoting improvement in learning performance in educational settings. Hence, studies with a within-subject design were included in this meta-analysis and a further sub-split analysis on studies with a between-subject design and studies with a within-subject design was conducted to examine the robustness of the findings. Differences between the inclusion and exclusion of these studies were further discussed. Studies with a between-subject design provided evidence with much stronger internal validity regarding the average effectiveness of gamification vs. non-gamification while studies with a within-subject design provided suggestive

evidence about improvements associated with gamification but with a weaker internal validity given the lack of comparison conditions.

3 Method

3.1 Protocol and registration

No review protocol was registered for this meta-analysis.

3.2 Eligibility criteria

3.2.1 Gamification

Eligible studies must include at least one comparison of a game condition versus a non-game condition. The only criterion for the control condition or pre-test condition was the absence of gamification. Studies aimed to investigate the effectiveness of sports games were not included in this current meta-analysis. Interventions that focused on the effectiveness of designing or programming games for learning purposes were also excluded as they were considered to be in closer alignment with design-based learning than gamified learning (Clark et al., 2016). Full-fledged games were not included as they fell in the category of serious games but not gamification (Deterding et al., 2011). Although the majority of games used in educational settings are digital games, the analysis in this paper includes non-digital games, as well as the aim of this research, is to investigate the impact of gamification in learning.

3.2.2 Participants

Participants were students in educational institutions or enrolled in schools (i.e., elementary, secondary, tertiary and post-graduate schools). Learners from pre-school learning centers were not included as pre-school education varied across different institutions and was usually considered as informal educational settings. Studies focusing on samples from specific clinical populations of students (e.g., autism spectrum, individuals with learning difficulties) were also excluded.

3.2.3 Research designs

The current meta-analysis included studies that adopted the experimental design and quasi-experimental design as both designs can be used to examine the effectiveness of a construct. Both within-subject and between-subject designs were included. As a result, studies included were required to have either a control condition or a pre-test. Studies that had both pre-test scores (baseline control) and a control group were also included and the pre-test score was controlled to yield more accurate results.

3.2.4 Learning outcomes

To be included, the gamification intervention had to target participants' learning performance in educational settings, which contains academic achievement as well as conceptual knowledge and skills of related subjects (e.g., knowledge tests and problem-solving skills). Only objective data of performance (test performance, task performance, etc.) were included as self-report data are likely to be biased when participants were asked to evaluate their performance due to some reasons (e.g., social desirability and subjective evaluation).

3.2.5 Publication type

As gamification in education had only emerged as a popular phenomenon over the past decade, this meta-analysis aimed to investigate studies that were published between January 2009 and September 2019. Eligible studies were those published in peer-reviewed journal articles. Books were excluded in this current meta-analysis.

3.2.6 Study site and language

Studies were required to be published in the English language (but not necessarily conducted in an English-speaking country).

3.2.7 Effect sizes

Eligible studies were required to report sufficient information needed to calculate Hedges' *g* and 95% confidence intervals. While pre-test scores were a requirement for studies that adopted pre-post test design, they were optional for studies using between-subject experimental design. In order to have a bigger pool of studies for the meta-analysis, studies that failed to report pre-post test correlations but used the same measures in both tests were also included in this study. Sensitivity analysis of different pre-post test correlation estimates was conducted and showed no substantial change in effect sizes (see Appendix 1). Thus, missing pre-post test correlations were imputed to be .60 (drawn from the average of other similar studies in this analysis, for more details, see Footnote[1]).

---

[1] As the pre-post test correlation was required to compute the effect sizes of studies with pre-test measures (e.g., see Borenstein, 2009, p. 227, Equation 12.21), missing values of the correlation in some studies would lead to missing effect sizes of those studies. As the pre-post test correlations reported were inconsistent across studies (the average estimate is .634, range from .387 to .8), it was inappropriate to impute a common value from them (Higgins & Green, 2011). Therefore, a sensitivity analysis was conducted to evaluate the robustness of the results. Different estimates drawn from the aforementioned range of pre-post test correlations (.40, .60 and .8 0) were imputed for missing pre-post test correlations and no substantial change in results was found . Thus, we reported findings of an imputed estimate of .60 in this meta-analysis as this was approximate to the average

3.3 Information sources

Key terms were searched in the following databases and platforms: Web of Science (Index: SCI-EXPANDED, SSCI), Ovid (includes PsycINFO, Inspec, Medline) and ERIC.

This search was conducted in Sept 2019.

3.4 Search

The following terms *"gamif*"* or *"game*"* and *"performance"* or *"achievement"* and *"participation"*, *"involvement"* or *"engagement"* in the study abstract or title were included in our search criteria. The mentioned search terms were deemed likely to identify potentially eligible studies and were searched in the databases mentioned in the previous section.

3.5 Study selection

Eligible coding started at the title and abstract level, where the first author screened and excluded studies of unrelated topics and studies that used different methodology or measured outcomes other than specified (if given in abstracts). Secondly, full-texts of eligible studies were retrieved from databases, online resources and corresponding researchers. Full-text screening of each article was then conducted to further exclude studies that did not fit the inclusion criteria. Thirdly, the study characteristics and information regarding effect sizes of eligible studies were coded in an excel file.

In terms of effect size metric, several studies have reported more than one effect size (e.g., multiple measures of learning performance or multiple experimental groups using different games). In this meta-analysis, effect sizes of all experimental groups with gamification conditions and different measures of learning performance were included as long as they met the inclusion criteria. As these effect sizes were not independent, treating them as if they were individual studies would yield biased results. Based on Borenstein et al. (2009)'s suggestion on handling complex data structure, effect sizes of the same studies were averaged and combined for statistical data analysis. Nevertheless, effect sizes of different experimental groups were treated separately in subgroup analyses since different experimental groups used different games, whose features might be different and were assigned to different subgroups in potential moderators. Aggregation in these cases would have excluded these studies in subgroup analyses and have led to a loss of data.

3.6 Data collection process

During the coding process, corresponding researchers were contacted via emails when eligible studies did not provide sufficient data for calculating effect sizes.

Four eligible studies in this meta-analysis had failed to retrieved data of effect sizes from online resources and corresponding researchers, therefore, could not be further analyzed statistically. No duplication of data was suspected in the analysis. Coding for all eligible studies was then performed by two independent coders (i.e., the first and the second author) with the interrater reliability ranging from $\kappa=.74$ to perfect agreement. The data on moderator variables were extracted and coded as mentioned in the following paragraph. Any coding discrepancies were discussed between the coders until a mutual agreement was reached.

3.6.1 Peer competition

Studies where participants competed against their peers for finite resources (such as rankings, levels, badges, and points) were coded yes, whereas studies that did not use such game design were coded no. For this moderator, interrater reliability was $\kappa=.83$.

3.6.2 Peer collaboration

Studies where participants collaborated with their peers were coded yes (e.g., playing the game in teams), whereas studies that did not use such game design were coded no. For this moderator, interrater reliability was $\kappa=.85$.

3.6.3 Research design

Studies were categorized in the following research designs: experimental, quasi-experimental and pre-post test design. For this moderator, interrater reliability was $\kappa=.74$.

3.6.4 Education level

Studies were categorized in the following educational levels: elementary, secondary and tertiary. For this moderator, interrater reliability was $\kappa=1$.

3.7 Data items

The following variables were coded and used for descriptive purposes or examined as potential effect size moderators: publication year, study design, participants' educational levels, media of the games (i.e. the channel that the game was played, e.g., non-digital, digital, virtual reality, etc.) and measures of learning performance in the primary studies.

3.8 Risk of Bias in individual studies

As recommended by Cochrane Handbook for Systematic reviews of interventions (Higgins & Green, 2011), the common classification scheme in Newcastle Ottawa

Scale (Wells et al., 2014) was used for evaluating the quality of evidence in this meta-analysis since it included non-randomized studies. After thoughtful consideration of all issues specifically related to the effectiveness of gamification in learning, the scale was adapted to the context of this meta-analysis. The quality of evidence was assessed in three categories: selection, comparability and outcome. Two independent raters scrutinized each study regarding these three categories. The criteria for judging risk of bias were illustrated in Appendix 2. The raters computed scores representing the total number of criteria on the list and assessed whether the study is at high risk of bias (score from 1 to 5) or low risk of bias (score from 6 to 9). Risk of bias in individual studies was shown in Table 4 (in Appendix 2), with the score ranging from 3 to 8. The majority of the studies were classified as low risk of bias.

3.9 Summary measures

Meta-analysis was conducted using Comprehensive Meta-Analysis, Version 3.3.070 (CMA; Borenstein et al., 2014). All effect sizes were estimated using the formulas provided by the computer software CMA. Effect sizes were computed using the available statistics provided by researchers, e.g., means, standard deviations, sample sizes of experimental groups and control groups, differences in means between groups, the pre-post difference in means or other statistics. In order to correct the biased results due to the small sample sizes of some included studies, the statistical estimate Hedges' g, its 95% confidence interval, the associated z and p values were computed in this meta-analysis to estimate the effect sizes of all studies (Hedges, 1981).

3.10 Synthesis of results

Random-effects models were assumed and used in this meta-analysis as the true effect sizes in each study might vary due to different sampling populations and other factors. Random-effects models take account of the difference in variances caused by the difference among studies. The Q-test, quantified by the index I^2, was used to assess the presence of heterogeneity in the pooled studies. A significant Q value indicates that there exists heterogeneity in the data that goes beyond random errors while I^2 indicates the percentage of variance caused by the true heterogeneity among studies (Borenstein et al., 2009).

3.11 Risk of Bias across studies

Since one of the inclusion criteria of this meta-analysis was peer-reviewed journal articles, findings could be subject to publication bias resulting from the exclusion of unpublished studies with non-significant results (Rothstein et al., 2005). Some non-significant findings might have been missing in our search. As a consequence, publication bias was assessed to examine the extent of validity of findings in this meta-analysis. Because the interpretation of a funnel plot is subjective, several tests were conducted to quantify the bias or test the relationship between sample size and

effect size (Borenstein et al., 2009). Publication bias was assessed in five ways: funnel plots, Egger's regression test (Egger et al., 1997), Begg and Mazumdar rank correlation test (Begg & Mazumdar, 1994) and trim and fill analysis (Duval & Tweedie, 2000).

3.12 Additional analyses

As existing literature suggested that peer competition and peer collaboration could influence students' learning performance in educational settings, subgroup analyses of potential moderators were performed. Moderator variables, e.g., competition and collaboration were coded and effect sizes in different categories (absence or presence) were further investigated. *Q*-tests for heterogeneity were performed for each subgroup analysis.

Furthermore, sensitivity analyses were conducted in both main analysis and subgroup analyses in order to test the robustness of the effect sizes. Studies with extreme effects (i.e., studies with residuals greater/lower than two standard deviations or above) were identified as outliers and removed.

4 Results

4.1 Study selection

As shown in Fig. 1, a total of 1066 titles and abstracts were screened. Then 992 studies were excluded at the level of title and abstract, and 41 studies were further excluded after screening the full texts. Only 33 studies met the eligibility criteria and four of them had irretrievable data for calculating the effect sizes. As a result, a final set of 29 studies was entered for this meta-analysis. A total of 50 effect sizes were retrieved from the study pool.

4.2 Study characteristics

This meta-analysis included 15 experimental studies with a control group, 7 quasi-experimental studies with a control group and 7 within-subject design studies, with a total of 3515 participants (1522 in gamification condition, 1396 in the control condition and 597 in within-subject test design). Appendix 3 demonstrated descriptive statistics of the study sample size, participants' education level, study design and the medium of the game.

The majority of the studies ($n = 13$) included in the meta-analysis were students in tertiary education or above, followed by elementary education ($n = 8$) and secondary education ($n = 8$).

In terms of measures for learning performance, no existing standardized tests and scales were used. Test scores and exam scores were measured for learning performance in the majority of the studies.
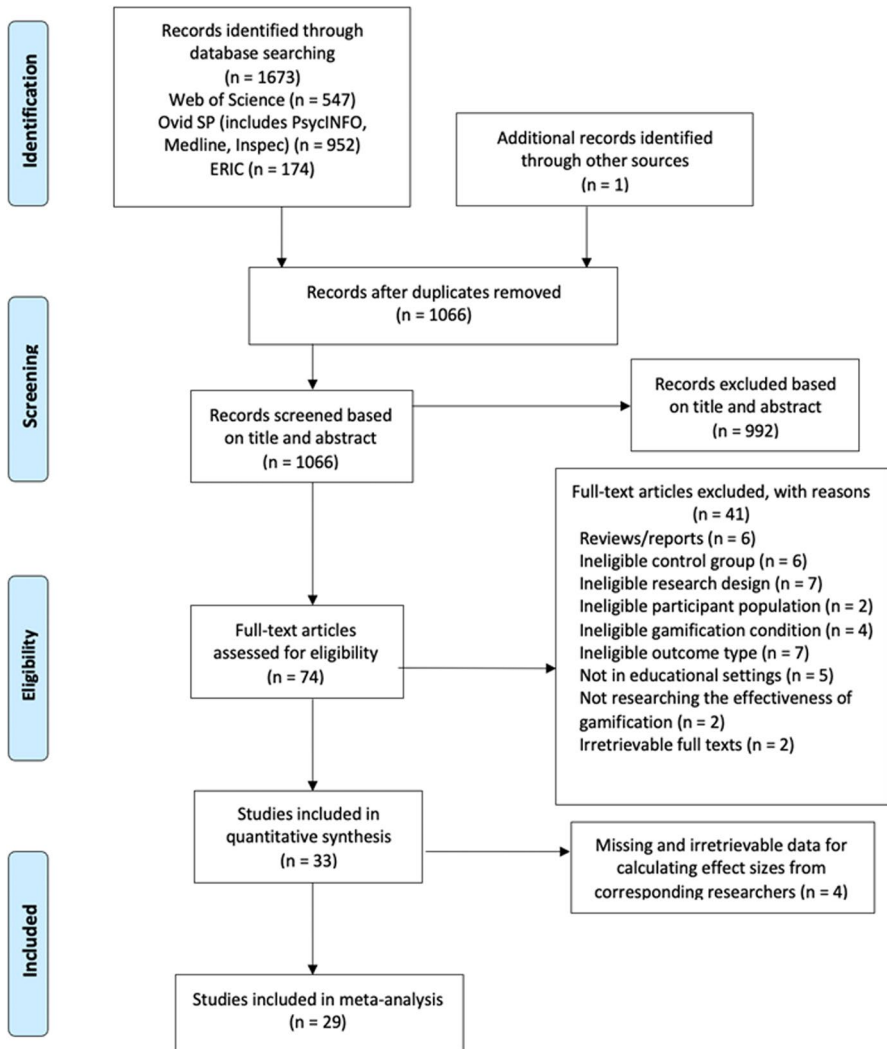
Fig. 1 The PRISMA Flowchart of Study Selection

4.3 Results of individual studies

Figure 2 shows summary data (effect sizes and corresponding 95% confidence intervals of individual studies) and a forest plot (sorting by the standard errors from small to large) for the random-effects meta-analysis of the effectiveness of gamification in learning performance in educational settings. The combined effect size (g) was .595, 95% CI [.432, .758], $p < .001$ (as shown in the last row).
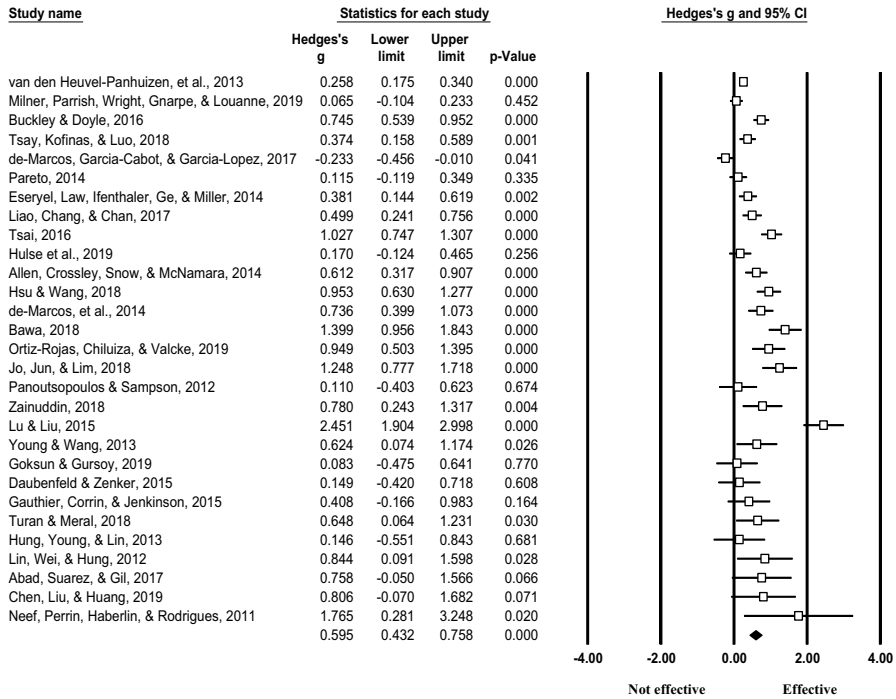
Study name	Statistics for each study				Hedges's g and 95% CI
	Hedges's g	Lower limit	Upper limit	p-Value	
van den Heuvel-Panhuizen, et al., 2013	0.258	0.175	0.340	0.000	
Milner, Parrish, Wright, Gnarpe, & Louanne, 2019	0.065	-0.104	0.233	0.452	
Buckley & Doyle, 2016	0.745	0.539	0.952	0.000	
Tsay, Kofinas, & Luo, 2018	0.374	0.158	0.589	0.001	
de-Marcos, Garcia-Cabot, & Garcia-Lopez, 2017	-0.233	-0.456	-0.010	0.041	
Pareto, 2014	0.115	-0.119	0.349	0.335	
Eseryel, Law, Ifenthaler, Ge, & Miller, 2014	0.381	0.144	0.619	0.002	
Liao, Chang, & Chan, 2017	0.499	0.241	0.756	0.000	
Tsai, 2016	1.027	0.747	1.307	0.000	
Hulse et al., 2019	0.170	-0.124	0.465	0.256	
Allen, Crossley, Snow, & McNamara, 2014	0.612	0.317	0.907	0.000	
Hsu & Wang, 2018	0.953	0.630	1.277	0.000	
de-Marcos, et al., 2014	0.736	0.399	1.073	0.000	
Bawa, 2018	1.399	0.956	1.843	0.000	
Ortiz-Rojas, Chiluiza, & Valcke, 2019	0.949	0.503	1.395	0.000	
Jo, Jun, & Lim, 2018	1.248	0.777	1.718	0.000	
Panoutsopoulos & Sampson, 2012	0.110	-0.403	0.623	0.674	
Zainuddin, 2018	0.780	0.243	1.317	0.004	
Lu & Liu, 2015	2.451	1.904	2.998	0.000	
Young & Wang, 2013	0.624	0.074	1.174	0.026	
Goksun & Gursoy, 2019	0.083	-0.475	0.641	0.770	
Daubenfeld & Zenker, 2015	0.149	-0.420	0.718	0.608	
Gauthier, Corrin, & Jenkinson, 2015	0.408	-0.166	0.983	0.164	
Turan & Meral, 2018	0.648	0.064	1.231	0.030	
Hung, Young, & Lin, 2013	0.146	-0.551	0.843	0.681	
Lin, Wei, & Hung, 2012	0.844	0.091	1.598	0.028	
Abad, Suarez, & Gil, 2017	0.758	-0.050	1.566	0.066	
Chen, Liu, & Huang, 2019	0.806	-0.070	1.682	0.071	
Neef, Perrin, Haberlin, & Rodrigues, 2011	1.765	0.281	3.248	0.020	
	0.595	0.432	0.758	0.000	

-4.00   -2.00   0.00   2.00   4.00

Not effective   Effective

Fig. 2 Forest Plot of the Effectiveness of Gamification

4.4 Interpretation of results

Gamified learning was found to be effective in students' learning in educational settings when compared with non-game conditions. The random-effects model yielded a significant effect size of gamification in students' learning performance in educational settings, $g = .595$, SE $= .083$, 95% CI [.432, .758], $p < .001$. Based on Higgins, Thompson, Deeks and Altman's rule of thumb for I^2 (2003), heterogeneity was significant and high among these studies, $Q(28) = 218.055$, $p < .001$, $I^2 = 87.159\%$, $\tau^2 = .147$.

4.5 Risk of Bias across studies

4.5.1 Publication Bias

Initial evaluation of the funnel plot showed no obvious asymmetries for the pool of studies in this meta-analysis (see Fig. 3) and no imputed studies were added, indicating no presence of strong publication bias. However, at least one obvious outlier was observed.
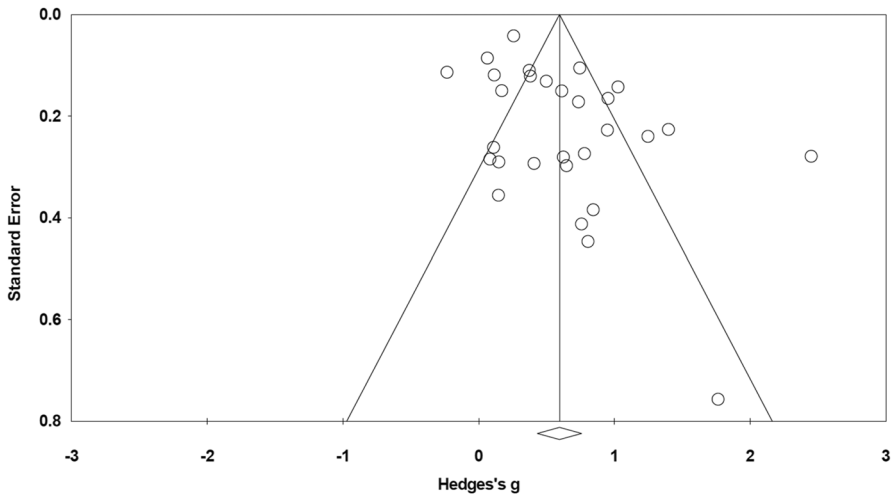
Fig. 3 Funnel Plot of Standard Error by Hedge's g

Egger's test revealed a significant association between effect size and result precision, $b_0 = 2.253$, SE $= .798$, $t = 2.824$, $p = .004$, one-tailed. Nonetheless, Begg and Mazumdar rank correlation test revealed no significant association between standard error and effect size, Tau $= .182$, $z = 1.388$, $p = .083$, one-tailed, but this could be due to the low power of the test (Borenstein et al., 2009). Additionally, no adjusted values for the random-effects model in the Trim and Fill method were suggested.

The results of the publication bias analysis could be classified into three categories, that are, a) the impact of bias is trivial; b) the impact is not trivial but the major finding is still valid; c) the validity of major finding is questioned (Borenstein et al., 2009). In this meta-analysis, based on the various indices mentioned above, the impact of publication bias seemed to be present as Egger's test suggested larger effects in smaller studies. Nevertheless, with consideration of other indicators, the effectiveness of gamification in learning performance in educational settings should still be considered valid.

4.6 Additional analyses

4.6.1 Sensitivity analysis for extreme values

As the funnel plot shows that there appear to be some outliers, extreme values were identified and removed in order to test the robustness of the combined effect size. Studies being removed were de-Marcos et al. (2017) and Lu and Liu (2015). Possible reasons for the extreme values found in these studies could be the distinctive gamification condition (social gamification with the focus on social aspect) in de-Marcos et al.'s study (De-Marcos et al., 2017) and the different medium of the game design used (virtual reality) in Lu and Liu's study (Lu & Liu, 2015).

After removing the outliers, assuming a random-effects model, the result remained almost the same as before, with a slightly smaller combined effect size than the result with outliers. A significant combined effect size was reported, $g = .563$, SE $= .072$, 95% CI $= [.422, .704]$, $p < .001$. The Q-test still showed significant and high heterogeneity among studies although it was lower than the results including the outliers, $Q(26) = 133.405$, $I^2 = 80.511\%$, $p < .001$, $\tau^2 = .090$.

4.6.2 Subgroup analysis

As the Q-tests and I^2 showed a significant and substantial amount of heterogeneity (I^2 was larger than 60%) for the main analysis (i.e., the effectiveness of gamification), subgroup analyses were performed to examine the proposed moderators and their accountability for the variance observed. Mixed effect models, i.e., a random-effects model within subgroups and a fixed-effect model across subgroups (the approach generally advocated), were used for the subgroup analyses (Borenstein et al., 2009) (for more details, see Footnote[2]). As estimates of subgroups that contain five or fewer effect sizes are likely to be imprecise (Borenstein et al., 2009), subgroups that contained five or fewer effect sizes were not evaluated in subgroup analyses.

For studies that contained more than one game condition comparing with the non-game condition (i.e., Göksün & Gürsoy, 2019; Hsu & Wang, 2018), the effect sizes of different game conditions were evaluated separately as different games might have different configurations of peer interaction. After averaging effect sizes of multiple measures within the same studies (when applicable), a total of 31 effect sizes from 29 studies were evaluated in subgroup analyses.

Table 1 shows the effect size estimates for group comparisons of peer competition, peer collaboration and the additive effect of peer collaboration in competition.

Peer competition In terms of peer competition, competitive games were used in the experimental condition in 24 out of 31 effect sizes. The effect of gamification on learning performance was larger in conditions using competitive games ($g = .685$, SE $= .112$, 95% CI $= [.465, .906]$, $p < .001$) than non-competitive games ($g = .345$, SE $= .074$, 95% CI $= [.200, .491]$, $p < .001$). Games featuring peer competition demonstrated a significantly bigger effect size than games without peer competition features on students' learning performance when compared with non-game conditions. The subgroup difference of peer competition in the effects of gamification on learning performance in educational settings was significant, $Q(1) = 6.387$, $p = .011$.

One outlier was detected in this subgroup analysis (i.e., Lu & Liu, 2015). After the removal of the outlier, a significant difference in effect sizes was still found

---

[2] In subgroup analysis, CMA offers two options in terms of computational models: a pooled estimate of τ^2 and with separate estimates of τ^2. In this study, separate estimates of τ^2 were used for all moderators since a) the true study-to-study dispersion was expected to differ across subgroups; b) there were unequal numbers of studies in each subgroup; and c) the variances in subgroups were likely to vary for moderators with intervention effects (e.g., competitive games versus non-competitive games) (Borenstein et al., 2009)

Table 1 Effect Size Estimates for Group Comparisons on Each Subgroup Analysis

Moderator	Subgroup	n[a]	g	SE	95% CI	p	Subgroup difference
Competition	Competition	24	.685	.112	[.465, .906]	< .001	.011
	No competition	7	.345	.074	[.200, .491]	< .001	
Collaboration	Collaboration	11	.748	.205	[.346,1.151]	< .001	.336
	No collaboration	20	.534	.088	[.361, .706]	< .001	
Joint effect of competition and collaboration	Competition-only	15	.584	.129	[.330, .838]	< .001	.250
	Competition & Collaboration	9	.897	.240	[.428, 1.367]	< .001	
After the removal of outliers							
Competition	Competition	23	.607	.102	[.407, .806]	< .001	.038
	No competition	7	.345	.074	[.200, .491]	< .001	
Collaboration	Collaboration	10	.542	.150	[.248, .836]	< .001	.822
	No collaboration	19	.581	.085	[.415, .747]	< .001	
Joint effect of competition and collaboration	Competition-only	15	.584	.129	[.330, .838]	< .001	.747
	Competition & Collaboration	8	.655	.178	[.307, 1.003]	< .001	

[a]n is the number of effect sizes in the group

between the subgroups, $Q(1) = 4.303$, $p = .038$. When comparing with non-game conditions, the effect of gamification on learning performance was larger in conditions using competitive games ($g = .607$, SE $= .102$, 95% CI $= [.407, .806]$, $p < .001$) than non-competitive games ($g = .345$, SE $= .074$, 95% CI $= [.200, .491]$, $p < .001$). Games featuring peer competition demonstrated a significantly bigger effect size than games without peer competition on students' learning performance when compared with non-game conditions.

Peer collaboration In terms of peer collaboration, collaborative games were used in the experimental condition in 11 out of 31 effect sizes. No significant differences were found in the two subgroups, i.e., games with peer collaboration and without peer collaboration, in terms of the effect of gamification on learning performance, $Q(1) = .926$, $p = .336$. Both subgroups demonstrated a significant effect size (collaborative: $g = .748$, SE $= .205$, 95% CI $= [.346, 1.151]$, $p < .001$; non-collaborative: $g = .534$, SE $= .088$, 95% CI $= [.361, .706]$, $p < .001$).

After the removal of two outliers (i.e., de-Marcos et al., 2017; Lu & Liu, 2015), still no evidence of subgroup differences were found between the subgroups, $Q(1) = .051$, $p = .822$. When comparing with non-game conditions, both groups reported a significant effect size (collaborative: $g = .542$, SE $= .150$, 95% CI $= [.248, .836]$, $p < .001$; non-collaborative: $g = .581$, SE $= .085$, 95% CI $= [.415, .747]$, $p < .001$), indicating that games featuring peer collaboration or not had the same, significant effect size on students' learning performance.

The joint effect of peer competition and peer collaboration As existing research (e.g., Hung et al., 2015b; Clark et al., 2016) has found that games featuring both peer competition and peer collaboration also had a positive impact on students' learning performance, the joint effect of peer competition and peer collaboration was further examined. Since subgroups that contained five or fewer effect sizes would yield imprecise estimates (Borenstein et al., 2009), a subgroup analysis (containing 24 effect sizes) was performed to compare the two subgroups only: games featuring both peer competition and peer collaboration (9 effect sizes) and games featuring peer competition only (15 effect sizes). The effect sizes of the subgroups: games featuring neither peer competition nor peer collaboration (only 5 effect sizes) and games featuring peer collaboration (only 2 effect sizes) only were not evaluated.

No evidence of differences were reported between the subgroups, $Q(1) = 1.322$, $p = .250$. Findings suggested no significant difference in the effectiveness of between the games with the joint effects and with competition only. Although no subgroup differences were found, when comparing with non-game conditions, a significant effect size of was reported in games with competition only ($g = .584$, SE $= .129$, 95% CI $= [.330, .838]$, $p < .001$) whilst a significant effect size was reported in games with joint effects ($g = .897$, SE $= .240$, 95% CI $= [.428, 1.367]$, $p < .001$).

After removing one outlier (i.e., Lu & Liu, 2015), still no evidence of differences were found between the subgroups, $Q(1) = .104$, $p = .747$. Significant effect sizes were seen in both subgroups when compared with non-game conditions (peer competition only: $g = .584$, SE $= .129$, 95% CI $= [.330, .838]$, $p < .001$; both peer competition and collaboration: $g = .655$, SE $= .178$, 95% CI $= [.307, 1.003]$, $p < .001$). Results indicated that there were no additive effects of peer collaboration for competitive games in terms of learning performance.

Education level To investigate whether the effectiveness of gamification differed in different age groups, a subgroup analysis on the education level of students was conducted. No evidence of subgroup differences were found in the effectiveness of gamification in learning performance among participants from elementary, secondary and tertiary education or higher, $Q(2) = .782$, $p = .676$. The result remained nonsignificant after removing one outlier (i.e., Lu & Liu, 2015), $Q(2) = 3.59$, $p = .166$.

4.6.3 Sub-split analyses excluding studies with low methodological rigor

In terms of research design, this meta-analysis included experimental, quasi-experimental and pre-post test designs as they all examined the effectiveness of gamification and compared the differences in learning performance between the presence and the absence of gamification. However, the post-tests in the studies with the pre-post test design not only reflected the effectiveness of gamification but also the general learning effect. Therefore, the effect sizes might have been inflated. A sub-group analysis on study design showed that there existed a significant difference between studies with a between-subject design and with a within-subject design, $Q(1) = 3.891$, $p = .049$. The effect of gamification on learning performance was larger in studies with a within-subject design ($g = .903$, SE $= .191$, 95% CI $= [.528, 1.278]$, $p < .001$) than studies with a between-subject design ($g = .482$, SE $= .095$,

Table 2 Effect Size Estimates in the Sub-split Analysis on the Subgroup: Between-subject

Moderator	Subgroup	n^{a}	g	SE	95% CI	p	Subgroup difference
Competition	Competition	19	.550	.116	[.324, .777]	< .001	.145
	No competition	5	.329	.098	[.136, .522]	< .01	
Collaboration	Collaboration	9	.580	.187	[.214, .947]	< .01	.560
	No collaboration	15	.454	.111	[.236, .671]	< .001	
Joint effect of competition and collaboration	Competition-only	12	.469	.139	[.197, .741]	< .01	.333
	Competition & Collaboration	7	.732	.233	[.275, 1.188]	< .01	
Education level	Elementary	7	.503	.150	[.209, .798]	< .01	.929
	Secondary	4	.558	.177	[.211, .904]	< .01	
	Tertiary and above	13	.471	.140	[.196, .746]	< .01	

[a]n is the number of effect sizes in the group

95% CI = [.296, .667], $p < .001$). To further test the rigor of the findings, a sub-split analysis was performed on the studies with the two designs separately.

Effect size estimates in the sub-split analysis on the subgroup between-subject were demonstrated in the Table 2. No outliers were found in all the subgroup analyses after excluding the pre-post test design studies. Results showed that there existed a significant effect size of the effectiveness of gamification on learning performance, $g = .482$, SE = .095, 95% CI = [.296, .667], $p < .001$. Subgroup analyses on peer collaboration [$Q(1) = .340$, $p = .560$], the joint effect of peer competition and collaboration [$Q(1) = .937$, $p = .333$], and education level [$Q(2) = .147$, $p = .929$] remained non-significant. However, the subgroup analysis on peer competition became non-significant after excluding the studies with the pre-post test design, $Q(1) = 2.12$, $p = .145$.

For studies with a within-subject design, one outlier was detected and removed for further subgroup analyses (i.e., Lu & Liu, 2015). Effect size estimates for the studies with a within-subject design were not reported as there were fewer than 5 studies in each subgroup.

5 Discussion

5.1 Summary of evidence

Overall, regarding the effectiveness of gamification, similar to the results in the previous meta-analysis in the context of digital games and learning outcomes (Clark et al., 2016), our findings supported a significant and robust positive effect size in learning performance in educational settings. Moreover, our findings suggested no subgroup differences in the effectiveness of gamification in learning performance in terms of the education level, which differed from the findings in Huang et al.'s (2020) study. In their study, the effectiveness of gamification

in learning outcomes (including behavioral, affective and cognitive outcomes in educational settings) was investigated and results indicated that undergraduates' effect size nearly doubled that of K-12 students. This current meta-analysis only examined cognitive outcomes in terms of learning, thus, it is possible that subgroup differences of various education levels differed depending on the types of learning outcomes.

In terms of peer competition in gamified learning effectiveness, the results suggested that competitive games promoted better learning performance (specifically in terms of performance improvements) than non-competitive games, supporting the previous findings based on self-determination theoretical perspective (Bai et al., 2020) – competitive games satisfy learner's need for competence and thus lead to improved learning performance (Landers et al., 2015; Sailer et al., 2017). These findings are also in line with The Sociocultural Theory (Vygotsky, 1977) regarding the positive impact of "play" with peers on learning. However, this finding differed from Sailer and Homner's meta-analytical study (Sailer & Homner, 2020) on the effectiveness of gamification in learning in the context of general learning settings (including work-related, educational and informal training settings) where peer interaction was not found to moderate the effectiveness of gamification on cognitive learning outcomes. The differences might be attributed to the different contexts in these two meta-analyses. Although subgroup analysis suggested a difference between competitive games and non-competitive games, this finding was not robust. After excluding the studies with pre-post test design in a sub-split analysis, the subgroup difference between competitive games and non-competitive games became non-significant. Subgroup difference between competitive games and non-competitive games was only found in studies with pre-post test design, $Q(1) = 5.35$, $p = .021$.

One possible explanation of the significant subgroup difference between competitive games and non-competitive games found in pre-post test design but not in experimental or quasi-experimental design might be that peer competition only promotes learning effect (i.e., improvement in performance after learning regardless of the application of games) but it does not amplify the gamification effect. The subgroup difference in studies using pre-post test design indicates the learning effect and gamification effect while the subgroup difference in studies using experimental or quasi-experimental design indicates only the gamification effect. This suggests that peer competition might be beneficial to learning in general and might not have additional benefits in gamified learning contexts. Previous research findings have already demonstrated the positive effect of competitive learning environments: students in competitive conditions were more likely to adopt performance-oriented goals, which predicts better short-term and long-term academic performance (Lam et al., 2004). Another possible explanation could be that the subgroup difference between competitive games and non-competitive games is small, therefore, after excluding the studies with pre-post test design, the number of remaining studies ($n = 22$) did not provide enough power to reject the null hypothesis of no significant subgroup difference (for more details, please

refer to the first paragraph in the limitation section). Although findings in the sub-group analysis indicated that competitive games were better than non-competitive games for improving students' learning performance, according to the finding in the sub-split analysis, further research shall be conducted to fully understand the role of peer competition in gamified learning and whether its impact is similar to that in traditional learning.

As for peer collaboration, the moderating effect of peer collaboration was not seen in our study. One explanation might be that learning performance in educational settings is usually measured by test and exam results, which are likely to reflect individual performance instead of group performance. Collaboration in groups might not necessarily affect the learning performance measured individually (Lou et al., 2001).

Most evidence favoring the positive effects of peer collaboration suggests that collaborative learning promotes cognitive gains (e.g., Skon et al., 1981). Mullins et al. (2011) found that collaborative learning had positive effects on gaining conceptual knowledge but not procedural knowledge. On the other hand, some other studies (Hwang et al., 2019; Matsuda et al., 2013) suggested a link between peer competition and learning gain in procedural knowledge. Future research can further distinguish learning performance in educational settings in terms of the types of measures and explore whether or not collaboration moderates the effectiveness of gamification in specific types of learning outcomes in educational settings.

This meta-analysis focused on examining the role of peer interaction in gamified learning and how different types of peer interaction affected learning performance in terms of two aspects of effectiveness (learning improvement and learning performance). Nevertheless, there still existed a great amount of significant within-group variances unaccounted for after considering the effect of peer competition. Future research shall have a closer look at the dynamics of peer competition and account for students' skill levels in order to have a broader picture of how the different factors together have an impact on the effectiveness of gamification in educational settings.

6 Limitations

One limitation of this meta-analysis was that the number of included studies was quite small, especially for the subgroup analysis of the joint effect of peer competition and peer collaboration versus competition only ($n = 23$). When testing for categorical moderators in a meta-analysis, fewer than 40 studies included might be underpowered (Rubio-Aparicio et al., 2017), which might have limited the generalizability of results, especially for random-effect models in subgroup analyses. Borenstein et al. (2009) reported that the number of primary studies affected the precision of estimate in random-effects models. Non-significant results in subgroup analyses should never be immediately interpreted as an absence of group difference without considering the possibility of the failure to detect a small effect

due to poor power. Therefore, results regarding the non-significant moderating effect of peer collaboration in this meta-analysis should be interpreted with caution. Furthermore, as there were not enough studies ($n = 2$) using games featuring peer collaboration only in the gaming condition, we could not further examine the effect of peer collaboration when partialling out the effect of peer competition. Considering learning theories (e.g., Vygotsky's Zone of Proximal Development, Cole et al., 1978) have suggested the positive role of peer collaboration in learning and the aforementioned issues regarding statistical power, future research should still consider the effect of peer collaboration in gamified learning. Further investigation on this factor should be conducted before jumping to a conclusion, for example, examining the effects of collaborative games versus non-collaborative games or the effects of collaborative games on different types of learning outcomes in terms of learning performance.

Despite the small number of studies included, findings from this meta-analysis could still provide valuable insights into the impact of peer competition and peer collaboration on the effectiveness of gamified learning in educational settings. The included studies revealed that competitive learning games were frequently applied in educational settings than collaborative games (as reflected in the higher number of studies being analyzed in the subgroup analysis of peer competition). Moreover, despite the concerns of the negative effects (i.e., worries of losing, lower intrinsic motivation, etc.) brought by peer competition in learning, our findings illustrated the positive effects of peer competition in the context of gamified learning, which could inspire future research on investigating the role of peer competition in gamified learning and how the affordances of gamification might influence its impact on learning outcomes in gamified learning contexts.

The second limitation concerns the ability to generalize findings to different learning games. In this study, the majority of games were played digitally (from simply digital games to advanced simulations, virtual reality games and massively multiplayer online games) and only two studies used non-digital games. As a consequence, the findings of the effectiveness of gamification might not able to generalize to all types of games. In indeed, research on using virtual reality (VR) games in learning (Virvou & Katsionis, 2008) illustrated that VR games could increase distraction in learning. Students might benefit less than they could from the game. Hence, future research can explore whether gamification using different media has different effects on students' learning performance. On that note, we advise simplistic quotations of findings suggesting all types of gamification is more effective than no gamification in educational settings ought not to be made as it fails to acknowledge the representativeness of games as well as the complex composition of the non-game conditions included in this meta-analysis.

7 Conclusion

Along with the evidence from well-established learning theories and models (i.e., Self-determination Theory, Sociocultural Learning Theory, situated motivational affordances, social comparison theory, etc.), our findings suggest that gamification

is effective in learning performance in educational settings and competitive games seemed to be more effective than non-competitive games and stimulated improvement in learning performance. However, results should be interpreted carefully with the limitations mentioned above. Despite the non-significance, peer collaboration should still be further examined in order to fully understand its role in gamification in learning.

In terms of peer competition, previous literature shows that it has both positive and negative effects. In the scope of this current meta-analysis, a positive effect of peer competition in games was found in students' learning performance, e.g., increased test scores or corresponding knowledge gains. However, the effect was not robust and no evidence of subgroup differences were suggested after controlling for the research rigor. Thus, more research is needed to understand the complexity of the effects of peer competition. As students' learning performance is associated with many other factors such as learning interest and students' prior achievement levels (Abrantes et al., 2007) and previous research has also shown that competition might harm learning interest and low-achieving students might feel defeated and discouraged in a learning environment with peer competition (Slavin, 1980; Werbach & Hunter, 2012), future research should work on factors that can distinguish constructive peer competition from destructive peer competition.

This current meta-analysis offers a preliminary framework for building an effective gamified learning environment. Due to rapid advancement in technology, digital games have become increasingly popular for learning in educational settings. This trend has also been shown in the characteristics of our included studies. Therefore, further investigation on gamified learning is in desperate need. Future studies should further investigate the dynamics of peer competition in order to maximize the effectiveness of gamification featuring peer competition.

Appendix 1

See Table 3.

Table 3 Sensitivity Analysis

Imputation of pre-post test score	Effect size and Q-statistic
.40	$g = .585$, SE $= .084$, 95% CI [.420, .749], $p < .001$ $Q(28) = 214.235$, $p < .001$, $I^2 = 86.930$, $\tau^2 = .148$
.60	$g = .595$, SE $= .083$, 95% CI [.432, .758], $p < .001$ $Q(28) = 218.055$, $p < .001$, $I^2 = 87.159\%$, $\tau^2 =$
.80	$g = .633$, SE $= .086$, 95% CI [.464, .802], $p < .001$ $Q(28) = 237.023$, $p < .001$, $I^2 = 88.187$, $\tau^2 = .159$

Table 4 Risk of Bias in Individual Studies

Study	Total Score	Risk of bias
Abad et al., 2017	6	Low
Allen et al., 2014	7	Low
Bawa, 2019	4	High
Buckley & Doyle, 2016	6	Low
Hwang et al., 2019	7	Low
Daubenfeld & Zenker, 2014	5	High
De-Marcos et al., 2014	8	Low
De-Marcos et al., 2017	5	High
Eseryel et al., 2014b	5	High
Gauthier et al., 2015	6	Low
Göksün & Gürsoy, 2019	7	Low
Hsu & Wang, 2018	7	Low
Hulse et al., 2019	5	High
Hung, Young, & Lin, 2013	6	Low
Jo et al., 2018	7	Low
Liao et al., 2018	6	Low
Lin et al., 2012	5	High
Lu & Liu, 2015	4	High
Milner et al., 2015	7	Low
Neef et al., 2011	4	High
Ortiz-Rojas et al., 2019	7	Low
Panoutsopoulos & Sampson, 2012	3	High
Pareto, 2014	6	Low
Tsai, 2017	7	Low
Tsay et al., 2018	5	High
Turan & Meral, 2018	5	High
van den Heuvel-Panhuizen et al., 2013	7	Low
Young & Wang, 2014	6	Low
Zainuddin, 2018b	5	High

Appendix 2

Risk of Bias

<u>Note</u>: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories.

Selection

1) <u>Representativeness of the gaming condition</u>

a) truly representative of the average students in educational settings*
b) somewhat representative of the average students in educational settings*
c) selected group of users
d) no description of the derivation of the condition

2) <u>Selection of the non-gaming condition</u>

a) drawn from the same community as the gaming condition *
b) drawn from a different source
c) no description of the derivation of the non-gaming condition

3) <u>Ascertainment of gamification</u>

a) features of gamification were demonstrated *
b) game designs were illustrated *
c) no description

4) <u>Demonstration that learning performance was equal at start of study</u>

a) yes *
b) no

Comparability.

Comparability of conditions on the basis of the design or analysis

1) study controls for pre-condition scores for learning performance *
2) participation in the gaming condition does not directly affect learning performance (e.g., no extra points are rewarded toward the test score for participation in games)*

Outcome

1) Assessment of learning performance

a) record linkage (e.g., test score) *
b) knowledge or skills measured *
c) others

2) Was follow-up long enough (> two weeks) for the effect of gamification to occur

a) yes *
b) no
c) unclear

3) Adequacy of follow up of conditions

a) complete follow up - all subjects accounted for *
b) subjects lost to follow up unlikely to introduce bias - small number lost - > 90%
   follow up, or description provided of those lost) *
c) follow up rate < 90% and no description of those lost
d) no statement

Appendix 3

See Table 5.

Table 5 Study Characteristics

Study	Sample size of gaming condition (studies with a between-subject design)	Sample size of non-gaming condition (studies with a between-subject design)	Sample size (studies with a within-subject design)	Participants' education level	Study design	Medium of game	Measure of learning performance
Abad et al., 2017	11	13	/	tertiary	experimental	non-digital	final exam
Allen et al., 2014	/	/	42	secondary	pre-post	digital	holistic essay scores
Bawa, 2019	48	48	/	tertiary	experimental	digital	final exam
Buckley & Doyle, 2016	/	/	80	tertiary	pre-post	digital	knowledge test
Hwang et al., 2019	10	10	/	tertiary	experimental	digital	vocabulary tests
Daubenfeld & Zenker, 2014	23	23	/	tertiary	experimental	digital	final exam
De-Marcos et al., 2014	108	66	/	tertiary	quasi-experimental	digital	final exam
De-Marcos et al., 2017	175	139	/	tertiary	experimental	digital	final exam
Eseryel et al., 2014a	/	/	88	secondary	pre-post	massively multiplayer online game	problem representation score
Gauthier et al., 2015	24	22	/	tertiary	experimental	simulation	anatomy test scores
Göksün & Gürsoy, 2019	46	25	/	tertiary	experimental	digital	academic achievement test

Table 5 (continued)

Study	Sample size of gaming condition (studies with a between-subject design)	Sample size of non-gaming condition (studies with a between-subject design)	Sample size (studies with a within-subject design)	Participants' education level	Study design	Medium of game	Measure of learning performance
Hsu & Wang, 2018	80	82	/	elementary	experimental	digital	algorithmic thinking skills
Hulse et al., 2019	185 (two conditions together)		/	elementary	experimental	digital	math performan-ce
Hung, Young, & Lin, 2013	15	15	/	elementary	experimental	digital	English vocabulary achieve-ment
Jo et al., 2018	/	/	30	secondary	pre-post	digital	summative assess-ment
Liao et al., 2018	139	106	/	elementary	quasi-experimental	digital	written expression task
Lin et al., 2012	19	11	/	secondary	experimental	digital	achieve-ment
Lu & Liu, 2015	/	/	51	elementary	pre-post	virtual reality	learning achieve-ment
Milner, Parrish, Wright, Gnarpe, & Louanne, 2019	252	297	/	tertiary	experimental	digital	final exam
Neef et al., 2011	4	4	/	post-graduate	experimental	non-digital	weekly quizzes
Ortiz-Rojas et al., 2019	55	34	/	tertiary	quasi-experimental	digital	knowledge test
Panoutsopoulos & Sampson, 2012	29	28	/	secondary	experimental	simulation	standard curriculum math
Pareto, 2014	154	129	/	elementary	quasi-experimental	digital	learning gain in math
Tsai, 2017	/	/	74	secondary	pre-post	digital	energy knowledge test
Tsay et al., 2018	160	175	/	tertiary	quasi-experimental	digital	final exam

Table 5 (continued)

Study	Sample size of gaming condition (studies with a between-subject design)	Sample size of non-gaming condition (studies with a between-subject design)	Sample size (studies with a within-sub-ject design)	Participants' education level	Study design	Medium of game	Measure of learning performance
Turan & Meral, 2018	23	23	/	secondary	quasi-experimental design	digital	achieve-ment test
van den Heuvel-Panhuizen et al., 2013	/	/	232	elementary	pre-post	simulation	algebraic test
Young & Wang, 2014	27	25	/	elementary	experimental	digital	pronuncial-tion test
Zainuddin, 2018a	27	29	/	secondary	quasi-experimental	digital	academic achieve-ment test

Declarations

References

Abrantes, J. L., Seabra, C., & Lages, L. F. (2007). Pedagogical affect, student interest, and learning performance. *Journal of Business Research, 60*(9), 960–964.

Ames, C. (1984). Achievement attributions and self-instructions under competitive and individualistic goal structures. *Journal of Educational Psychology, 76*(3), 478–487.

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review, 30*, 100322–100341.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088–1101.

Blohm, I., & Leimeister, J. M. (2013). Gamification. *Wirtschaftsinformatik, 55*(4), 275–278.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2014). Comprehensive meta-analysis (CMA) software. Version 3.0.

Briffa, M., Jaftha, N., Loreto, G., Pinto, F. C. M., Chircop, T., & Hill, C. (2020). Improved students' performance within gamified learning environment: A meta-analysis study. *International Journal of Education and Research, 8*(1), 223–244.

Buckley, P., & Doyle, E. (2017). Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. *Computers & Education, 106*, 43–55.

Burguillo, J. C. (2010). Using game theory and competition-based learning to stimulate student motivation and performance. *Computers & Education, 55*(2), 566–575.

Cagiltay, N. E., Ozcelik, E., & Ozcelik, N. S. (2015). The effect of competition on learning in games. *Computers & Education, 87*, 35–41.

Chen, Z. H. (2014). Exploring students' behaviors in a competition-driven educational game. *Computers in Human Behavior, 35*, 68–74.

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.

Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (Eds.). (1978). *L.S. Vygotsky: Mind in society: The development of higher psychological processes*. Harvard University Press.

De Byl, P. (2013). Factors at play in tertiary curriculum gamification. *International Journal of Game-Based Learning (IJGBL), 3*(2), 1–21.

de Sousa Borges, S., Durelli, V. H., Reis, H. M., & Isotani, S. (2014, March). A systematic mapping on gamification applied to education. In *proceedings of the 29th annual ACM symposium on applied computing* (pp. 216-222). ACM.

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

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.

Deutsch, M. (1962). Cooperation and trust: Some theoretical notes.

Dicheva, D., & Dichev, C. (2015, October). Gamification in education: Where are we in 2015?. In *E-learn: World conference on E-learning in corporate, government, healthcare, and higher education* (pp. 1445-1454). Association for the Advancement of computing in education (AACE).

Dillenbourg, P. (1999). What do you mean by 'collaborative learning'? *Elsevier Science*, 1–16.

Dillenbourg, P., & Fischer, F. (2007). Basics of computer-supported collaborative learning. *Zeitschrift für Berufs- und Wirtschaftspädagogik., 21*, 111–130.

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.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634.

Eseryel, D., Law, V., Ifenthaler, D., Ge, X., & Miller, R. (2014a). An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning. *Journal of Educational Technology & Society, 17*(1), 42–53.

Facey-Shaw, L., Specht, M., van Rosmalen, P., & Bartley-Bryan, J. (2020). Do badges affect intrinsic motivation in introductory programming students? *Simulation & Gaming, 51*(1), 33–54.

Faiella, F., & Ricciardi, M. (2015). Gamification and learning: A review of issues and research. *Journal of e-Learning and Knowledge Society, 11*(3), 13–21.

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

Higgins, J. P., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Vol. 4). John Wiley & Sons.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557–560.

Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., et al. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development, 68*(4), 1875–1901.

Hung, H. C., Young, S. S. C., & Lin, C. P. (2015a). No student left behind: A collaborative and competitive game-based learning environment to reduce the achievement gap of EFL students in Taiwan. *Technology, Pedagogy and Education, 24*(1), 35–49.

Hwang, G. H., Chen, B., Chen, R. S., Wu, T. T., & Lai, Y. L. (2019). Differences between students' learning behaviors and performances of adopting a competitive game-based item bank practice approach for learning procedural and declarative knowledge. *Interactive Learning Environments, 27*(5–6), 740–753.

Immorlica, N., Stoddard, G., & Syrgkanis, V. (2015, May). Social status and badge design. In *proceedings of the 24th international conference on world wide web* (pp. 473-483). International world wide web conferences steering committee.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89*(1), 47–62.

Karakostas, A., & Demetriadis, S. (2011). Enhancing collaborative learning through dynamic forms of support: The impact of an adaptive domain-specific support strategy. *Journal of Computer Assisted Learning, 27*(3), 243–258.

Katz, B., Jaeggi, S., Buschkuehl, M., Stegman, A., & Shah, P. (2014). Differential effect of motivational features on training improvements in school-based cognitive training. *Frontiers in Human Neuroscience, 8*(1), 242–251.

Kim, E., Rothrock, L., & Freivalds, A. (2018). An empirical study on the impact of lab gamification on engineering students' satisfaction and learning. *International Journal of Engineering Education, 34*(1), 201–216.

Lam, S. F., Yim, P. S., Law, J. S., & Cheung, R. W. (2004). The effects of competition on achievement motivation in Chinese classrooms. *British Journal of Educational Psychology, 74*(2), 281–296.

Landers, R. N., Bauer, K. N., Callan, R. C., & Armstrong, M. B. (2015). Psychological theory and the gamification of learning. In T. Reiners & L. Wood (Eds.), *Gamification in education and business*. Springer.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*(3), 449–521.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G. J., & Koedinger, K. R. (2013). Studying the effect of a competitive game show in a learning by teaching environment. *International Journal of Artificial Intelligence in Education, 23*(1–4), 1–21.

Michaels, J. W. (1977). Classroom reward structures and academic performance. *Review of Educational Research, 47*(1), 87–98.

Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer-Supported Collaborative Learning, 6*(3), 421–443.

Pareto, L., Haake, M., Lindström, P., Sjödén, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: Evaluating math comprehension and motivation. *Educational Technology Research and Development, 60*(5), 723–751.

Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of Educational Psychology, 105*(4), 1050–1066.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223–231.

Reid, A. J., Paster, D., & Abramovich, S. (2015). Digital badges in undergraduate composition courses: Effects on intrinsic motivation. *Journal of Computers in Education, 2*(4), 377–398.

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

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. Publication bias in meta-analysis: Prevention, assessment and adjustments, 1–7.

Rubio-Aparicio, M., Sánchez-Meca, J., López-López, J. A., Botella, J., & Marín-Martínez, F. (2017). Analysis of categorical moderators in mixed-effects meta-analysis: Consequences of using pooled versus separate estimates of the residual between-studies variances. *British Journal of Mathematical and Statistical Psychology, 70*(3), 439–456.

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. (2020). The gamification of learning: A Meta-analysis. *Educational Psychology Review, 32*, 77–112.

Scott, S., & Palincsar, A. (2013). Sociocultural theory. Education. http://dr-hatfield.com/theorists/resources/sociocultural_theory.pdf

Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research, 50*(2), 315–342.

Skon, L., Johnson, D. W., & Johnson, R. T. (1981). Cooperative peer interaction versus individual competition and individualistic efforts: Effects on the acquisition of cognitive reasoning strategies. *Journal of Educational Psychology, 73*(1), 83–92.

Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin, 94*(3), 429–445.

Slavin, R. E., Madden, N. A., & Leavey, M. (1984). Effects of team assisted individualization on the mathematics achievement of academically handicapped and nonhandicapped students. *Journal of Educational Psychology, 76*(5), 813–819.

Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics, 31*, 275–288.

Thiebes, S., Lins, S., & Basten, D. (2014). Gamifying information systems-a synthesis of gamification mechanics and dynamics. In *proceedings of the 20th European conference on information system,* Tel Aviv, Israel.

Verenikina, I., Harris, P., & Lysaght, P. (2003, July). Child's play: Computer games, theories of play and children's development. In *proceedings of the International Federation for Information Processing Working Group 3.5 open conference on Young children and learning technologies-Vol. 34* (pp. 99-106). Australian computer society, Inc.

Virvou, M., & Katsionis, G. (2008). On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. *Computers & Education, 50*(1), 154–178.

Vygotsky, L. S. (1977). The development of higher psychological functions. *Soviet Psychology, 15*(3), 60–73.

Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al.(2014) The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.

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

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.

Zainuddin, Z. (2018a). Students' learning performance and perceived motivation in gamified flipped-class instruction. *Computers & Education, 126*, 75–88.

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

Studies Included in Meta-analysis

Abad, E., Gil, J., & Suárez, P. (2017). A game-based educational method relying on student-generated questions. *The International Journal of Engineering Education, 33*(6), 1786–1797.

Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014). L2 writing practice: Game enjoyment as a key to engagement. *Language Learning & Technology, 18*(2), 124–150.

Bawa, P. (2019). Using Kahoot to inspire. *Journal of Educational Technology Systems, 47*(3), 373–390.

Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments, 24*(6), 1162–1175.

Daubenfeld, T., & Zenker, D. (2014). A game-based approach to an entire physical chemistry course. *Journal of Chemical Education, 92*(2), 269–277.

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.

De-Marcos, L., Garcia-Cabot, A., & Garcia-Lopez, E. (2017). Towards the social gamification of e-learning: A practical experiment. *International Journal of Engineering Education, 33*(1), 66–73.

Eseryel, D., Law, V., Ifenthaler, D., Ge, X., & Miller, R. (2014b). An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning. *Journal of Educational Technology & Society, 17*(1), 42–53.

Gauthier, A., Corrin, M., & Jenkinson, J. (2015). Exploring the influence of game design on learning and voluntary use in an online vascular anatomy study aid. *Computers & Education, 87*, 24–34.

Göksün, D. O., & Gürsoy, G. (2019). Comparing success and engagement in gamified learning experiences via Kahoot and Quizizz. *Computers & Education, 135*, 15–29.

Hsu, C. C., & Wang, T. I. (2018). Applying game mechanics and student-generated questions to an online puzzle-based game learning system to promote algorithmic thinking skills. *Computers & Education, 121*, 73–88.

Hulse, T., Daigle, M., Manzo, D., Braith, L., Harrison, A., & Ottmar, E. (2019). From here to there! Elementary: A game-based approach to developing number sense and early algebraic understanding. *Educational Technology Research and Development, 67*(2), 423–441.

Hung, H. C., Young, S. S. C., & Lin, C. P. (2015b). No student left behind: A collaborative and competitive game-based learning environment to reduce the achievement gap of EFL students in Taiwan. *Technology, Pedagogy and Education, 24*(1), 35–49.

Jo, J., Jun, H., & Lim, H. (2018). A comparative study on gamification of the flipped classroom in engineering education to enhance the effects of learning. *Computer Applications in Engineering Education, 26*(5), 1626–1640.

Liao, C. C., Chang, W. C., & Chan, T. W. (2018). The effects of participation, performance, and interest in a game-based writing environment. *Journal of Computer Assisted Learning, 34*(3), 211–222.

Lin, K. C., Wei, Y. C., & Hung, J. C. (2012). The effects of online interactive games on high school students' achievement and motivation in history learning. *International Journal of Distance Education Technologies (IJDET), 10*(4), 96–105.

Lu, S. J., & Liu, Y. C. (2015). Integrating augmented reality technology to enhance children's learning in marine education. *Environmental Education Research, 21*(4), 525–541.

Milner, R., Parrish, J., Wright, A., Gnarpe, J., & Keenan, L. (2015). Research and teaching: Exploring the use of an online Quiz game to provide formative feedback in a large-enrollment, introductory biochemistry course. *Journal of College Science Teaching, 45*(2), 86–97.

Neef, N. A., Perrin, C. J., Haberlin, A. T., & Rodrigues, L. C. (2011). Studying as fun and games: effects on college students'quiz performance. *Journal of Applied Behavior Analysis, 44*(4), 897–901.

Ortiz-Rojas, M., Chiluiza, K., & Valcke, M. (2019). Gamification through leaderboards: An empirical study in engineering education. *Computer Applications in Engineering Education, 27*(4), 777–788.

Panoutsopoulos, H., & Sampson, D. (2012). A study on exploiting commercial digital games into school context. *Educational Technology and Society, 15*(1), 15–27.

Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *International Journal of Artificial Intelligence in Education, 24*(3), 251–283.

Tsay, C. H. H., Kofinas, A., & Luo, J. (2018). Enhancing student learning experience with technology-mediated gamification: An empirical study. *Computers & Education, 121*, 1–17.

Tsai, F. H. (2017). An investigation of gender differences in a game-based learning environment with different game modes. *Eurasia Journal of Mathematics Science and Technology Education, 13*(7), 3209–3226.

Turan, Z., & Meral, E. (2018). Game-based versus to non-game-based: The impact of student response systems on Students' achievements, engagements and test anxieties. *Informatics in Education, 17*(1), 105–116.

van den Heuvel-Panhuizen, M., Kolovou, A., & Robitzsch, A. (2013). Primary school students' strategies in early algebra problem solving supported by an online game. *Educational Studies in Mathematics, 84*(3), 281–307.

Yen, L., Chen, C. M., & Huang, H. B. (2016, June). Effects of mobile game-based English vocabulary learning APP on learners' perceptions and learning performance: A case study of Taiwanese EFL learners. In *international conference on e-learning* (p. 255). Academic conferences international limited.

Young, S. S. C., & Wang, Y. H. (2014). The game embedded CALL system to facilitate English vocabulary acquisition and pronunciation. *Journal of Educational Technology & Society, 17*(3), 239–251.

Zainuddin, Z. (2018b). Students' learning performance and perceived motivation in gamified flipped-class instruction. *Computers & Education, 126*, 75–88.