Educational Research Review 30 (2020) 100322

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev

Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts

Shurui Bai, Khe Foon Hew*, Biyun Huang

Faculty of Education, The University of Hong Kong, China

A B S T R A C T

Despite the buzz around gamification as an exciting new method to engage students, evidence of its ability to enhance learning is mixed. In fact, gamification has attracted considerable controversy ("*gamification is bullshit*") and some derogatory labels such as "*exploitationware.*" Therefore, in order to make the case for or against gamification in education, it is important to examine the effects (if any) of gamification on student learning achievements. This study is a meta-analysis of 30 independent interventions (3,202 participants) drawn from 24 quantitative studies that have examined the effects of gamification on student academic performance in various educational settings. The results show an overall significant medium effect size in favor of gamification over learning without gamification (Hedges' $g = 0.504$, 95% CI [0.284–0.723], $p < 0.001$). No publication bias is detected. An analysis of 32 qualitative studies reveals four reasons for learners' enjoyment of gamification: (a) gamification can foster enthusiasm; (b) gamification can provide feedback on performance; (c) gamification can fulfill learners' needs for recognition; and (d) gamification can promote goal setting, and two reasons for their dislike of gamification: (a) gamification does not bring additional utility and (b) gamification can cause anxiety or jealousy. We conclude by highlighting two unresolved questions, and suggesting several future research directions concerning gamification in educational contexts.

1. Background

In recent years, there has been a considerable amount of buzz about gamification across various academic disciplines and industries. It has been proposed that gamified practices will become a key element in motivating people, particularly those born in the digital age of online games, to engage in certain tasks such as improving their fitness (Nike+), generating brand loyalty (Starbucks), or answering other people's questions (e.g., Yahoo! Answers). The potential of gamification, however, extends beyond healthy lifestyles and marketing strategies: it may be able to motivate students to learn better (Caponetto, Earp, & Ott, 2014).

The term "gamification" is a neologism derived from the digital media field. It was coined in 2002 (Marczewski, 2013) and documented in 2008, and it has become prevalent in many fields since 2010 (Deterding, Dixon, Khaled, & Nacke, 2011). Gamification is usually portrayed as different from games or serious games. In categorizing games, it is useful to think about the primary function of the game – whether the game was developed initially for entertainment, or for learning (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012).

Games for entertainment such as the Mario Brothers were developed primarily for fun, and recreation, while the main purposes of games for learning (games-based learning and serious games) are knowledge acquisition and behavior change. Serious games are digital games designed to train its players on a particular skill, or learn some academic content (Annetta, 2010). The terms serious games and games-based learning are sometimes used synonymously (Boyle, 2016; Corti, 2006), although serious games have been

* Corresponding author. Director of the Center of Information Technology in Education, Faculty of Education, The University of Hong Kong, China.

E-mail address: kfhew@hku.hk (K.F. Hew).

https://doi.org/10.1016/j.edurev.2020.100322
Received 22 June 2019; Received in revised form 22 December 2019; Accepted 16 February 2020
Available online 02 March 2020
1747-938X/

developed for the broader purposes of training change in healthcare, civic engagement, advertising, as well as in education (Blumberg, Almonte, Anthony, & Hashimoto, 2013).

In contrast, gamification in educational context is not a product in the way a serious game is, but is a process of applying game elements in a non-game context in order to motivate learner behavior (Educause, 2011).

Nevertheless, despite the interest in gamification as an exciting new method to engage individuals, it has attracted controversies and critiques. Several prominent game designers and scholars have argued that current forms of gamification are shallow and superficial (e.g., Bogost, 2011), and others caution about the possible negative effects of gamification (e.g., Toda, Valle, & Isotani, 2017). Bogost's (2011) derogatory description of gamification reflects many people's attitudes:

> "Gamification is marketing bullshit, invented by consultants as a means to capture the wild, coveted beast that is videogames and to domesticate it for use in the grey, hopeless wasteland of big business, where bullshit already reigns anyway."

Therefore, in order to make the case for gamification (or against gamification) in education, it is important that efforts be made to quantify the effects (if any) of gamification on student learning achievements. Despite the popularity of gamification, there is little consensus about whether it contributes to improved academic performance. Previous studies have reported mixed findings, with some reporting positive effects with varying effect sizes (e.g., Marín, Frez, Cruz-Lemus, & Genero, 2018; Zainuddin, 2018), and others reporting no effects (e.g., Rachels & Rockinson-Szapkiw, 2018) or adverse effects on students' exam scores (e.g., de-Marcos, Domínguez, Saenz-de-Navarrete, & Pagés, 2014). The conflicting results reported in the literature make the decision to support (or to dismiss) the use of gamification in education difficult.

This study offers two contributions to this debate. First, it presents a meta-analysis of quantitative intervention studies of the effects of gamification on student achievements in various educational settings. Intervention studies (i.e., empirical investigations that manipulate an independent variable) can tell us whether an intervention is effective (or not) in improving educational outcomes, which observational and correlational research cannot (Lazowski & Hulleman, 2016). Meta-analyses, which integrate the results of several small intervention studies, can provide a more precise estimate of the effect of an intervention than a single study (Cochrane Handbook for Systematic Reviews of Interventions, 2008). Meta-analyses can also determine the strength of the effect, establish whether the effect is positive or negative, and explore the factors, if any, that moderate the effect size. Unlike the narrative synthesis approach, which is insufficient for synthesizing contradictory results when a large number of studies is involved (Hunter & Schmidt, 2004), the meta-analytic approach can address conflicting findings (Paré, Trudel, Jaana, & Kitsiou, 2015). Second, this study examines what learners specifically like and dislike about gamification. Understanding learners' likes and dislikes can help us to gain deeper insights into how gamification may affect learners' academic performance.

The rest of the paper is organized as follows. First, we define the term *gamification* and describe some of its game elements. This is followed by a brief description of several relevant theories and how these theories might underpin the various game elements. Third, we describe some of the previous published reviews on gamification and their limitations. Following this, we describe our data sources, methods, and findings. In the Discussion section, we provide several reasons that might have contributed to the favorable effect of gamification, highlight two unresolved questions concerning the use of gamification, and discuss limitations of the current review. We conclude by highlighting several directions for future research.

1.1. Game elements

Gamification may be defined as the application of digital game elements in non-gaming contexts to motivate user behavior (Educause, 2011). Although game elements are the basic building components of gamification (Deterding et al., 2011; Sailer, Hense, Mayr, & Mandl, 2017; Werbach, 2012), there is no commonly agreed on classification of game elements. For example, Deterding et al. (2011) proposed a classification scheme of game elements based on five levels; Zichermann and Cunningham (2011) categorized game elements into mechanics, dynamics, and aesthetics; and Dicheva, Dichev, Agre, and Angelova (2015) proposed a two-level classification scheme. Sailer et al. (2017) described game elements as a set of elements that are clearly visible to users, which can be easily activated and deactivated in an experimental procedure. Given this lack of standardization in the literature, we do not use any of the aforementioned classification schemes. In this review, we map only the game elements that have been *explicitly* mentioned in the primary studies. Some of the common game elements are badges, challenges, leaderboard/rank, levels/unlock, storyline, points, progress bar, and teams (Giannetto, Chao, & Fontana, 2013; Huang & Hew, 2018; Richter, Raban, & Rafaeli, 2015; Sailer et al., 2017).

1.2. Common theories underpinning gamification

In this section, we describe briefly several theories (goal-setting, self-efficacy, self-determination, social comparison, flow theory, and operant conditioning theory) that have been frequently applied in the gamification literature (Huang & Hew, 2018; Ritcher et al., 2015; Sailer et al., 2017), and suggest how these theories may relate to the various game elements.

Goal-setting theory: A goal gives an individual a purpose, focus, and measurable outcomes that can be used to define what needs to be accomplished (Kapp, 2014). According to goal setting theory, goals that are immediate, specific, and moderately challenging are more motivating than goals that are long-term, vague, and too easy or too difficult (Locke, Shaw, Saari, & Latham, 1981). In addition, when immediate feedback is provided to individuals, as in the case of many gamified practices (e.g., using badges), participants can measure their progress in relation to their goals. In this way, individuals know if they need to adjust their directions or strategies to better pursue their goals (Locke, 1996; Locke & Latham, 2002).

Self-efficacy theory: Self-efficacy is a person's belief in how well he or she can deal with prospective situations (Bandura, 1982). It also determines the effort and the persistence that one exerts to overcome obstacles (Bandura, 1982). Self-efficacy can be increased when an individual experiences success completing a series of tasks that gradually increase in difficulty. Gamified practices that start with smaller, less difficult tasks can thus help build users' self-efficacy. The use of point systems, badges, and progress bars can also stimulate self-efficacy by measuring progression and providing users with direct feedback on their performance (Gnauk, Dannecker, & Hahmann, 2012).

Self-determination theory: The self-determination theory posits that humans possess three innate psychological needs that can motivate them to engage or not to engage in an activity: autonomy, relatedness, and competence (Ryan & Deci, 2000). Gamified practices that offer individuals the autonomy to choose which activities they prefer to complete (e.g., by offering varying levels of difficult tasks) can address this need. Having a sense of autonomy can increase participants' behavioral and emotional engagement (E. Skinner, Furrer, Marchand, & Kindermann, 2008). Relatedness refers to individuals' need to connect or interact with other people (Ryan & Deci, 2000). Gamified practices that allow participants to compete or collaborate with other people serve this need. A heightened sense of relatedness helps promote feelings of enjoyment; it can also motivate participants to further participate in the activity (Skinner et al., 2008). Competence refers to the desire to master one's pursuits or learning. Gamified practices that provide indicators of participants' progression (e.g., progress bars, levels) and immediate feedback (e.g., points, badges) can help foster users' sense of competency (Sailer et al., 2017).

Social comparison theory: According to social comparison theory, human beings evaluate their opinions and abilities by comparing them with those of others (Festinger, 1954). Two types of social comparison may be found: upward-identification, which occurs when individuals compare themselves with more competent people and believe that they can be as good as them, and downward-identification, which occurs when individuals compare themselves with people who are worse off (Molleman, Nauta, & Buunk, 2007; Suls, Martin, & Wheeler, 2002). Social comparison theory can help explain the motivational aspect of badges, points, levels, and leaderboards. For example, the total number of points collected or the ranking an individual achieves in a leaderboard can help drive upward-identification comparisons. Upward-identification can positively influence students to become more engaged in learning (Chen & Chen, 2015).

Flow theory: Flow is often used to describe the experience of being fully focused on an activity (Csikszentmihalyi, 1990; Nakamura & Csikszentmihalyi, 2009, pp. 195–206). Some of the conditions that can promote flow include clear and proximal goals, immediate feedback on performance and progress, and an appropriate level of challenges (Nakamura & Csikszentmihalyi, 2009, pp. 195–206; Shernoff, Csikszentmihalyi, Schneider, & Shernoff, 2003). Gamified practices such as using badges to provide feedback can help promote flow (Hamari & Sjöblom, 2017). The use of different levels, which allow users to choose the appropriate level of challenge, can also help promote flow.

Operant conditioning theory: Behavior can be motivated or impeded by the consequences of the behavior (B. F. Skinner, 1950). One of the most frequently used types of reinforcement is positive reinforcement. Positive reinforcement occurs when a new stimulus, presented as a consequence of a behavior, strengthens the original behavior (B. F. Skinner, 1953; Woolfolk, 1998). Certain behaviors can be reinforced or maintained using a schedule of reinforcement (B. F. Skinner, 1953, 1989). In the first stage, when people are learning a new skill or behavior, a continuous reinforcement schedule can be adopted (Skinner, 1953, 1989; Woolfolk, 1998). Gamified practices that give a reward (e.g., points or badges) for every correct response or for the completion of every task are examples of a continuous reinforcement schedule. Although badges cannot be used to buy tangible materials such as food, these virtual rewards reinforce desirable behavior within the gamified practice (Landers, Bauer, Callan, & Armstrong, 2015). However, after people have mastered a new skill, a random reinforcement schedule can be used to maintain the particular skill or behavior (Skinner, 1953, 1989; Woolfolk, 1998). Gamified practices that provide badges intermittently when a certain number of points have been accumulated can help retain an individual's interest in the activity. Table 1 summarizes the above discussion.

2. Prior reviews

This section briefly reviews some of the relevant literature on gamification (Borges, Durelli, Reis, & Isotani, 2014; Dichev & Dicheva, 2017; Falkner & Falkner, 2014; Hamari, Koivisto, & Sarsa, 2014; Looyestyn et al., 2017) and discusses its limitations. None of the previous reviews have examined the effects of gamification on learner academic performance. Several have reviewed studies that examined a mixture of gamification and games-based learning implementations or studies conducted in a variety of disciplines. Hamari et al. (2014), for example, presented a multi-disciplinary literature review of 24 empirical studies covering a wide variety of contexts including commerce, health, intra-organizational systems, work, and education. However, within the education context,

Table 1
Theories and related game elements (Huang & Hew, 2018; Ritcher et al., 2015; Sailer et al., 2017).

Theory	Game element (example)	Supporting role
Self-efficacy	Progress bar, levels/unlock, points	Feedback, achievements
Self-determination (autonomy)	Levels	Choice
Self-determination (relatedness)	Team	Competition, cooperation
Self-determination (competence)	Progress bar, levels/unlock, badges	Feedback, challenges
Social comparison, goal-setting	Leaderboard, points	Status and reputation, achievements, competition, challenges
Flow theory, goal-setting	Points, badges, progress bar	Goals and objectives, challenges
Operant conditioning theory	Points, badges	Goals and objectives, rewards, status, achievements

only nine papers were discussed, and student learning outcomes were not discussed. Falkner and Falkner (2014) examined the use of badges in 10 studies, but their paper is limited because it only considered the use of badges in computer science education.

In another study, Borges et al. (2014) reviewed 26 studies on gamification. The authors found that most studies focused on higher education, with the most common focus being learner engagement. However, as their review included non-empirical studies that reported the authors' views but contained no supporting empirical evidence, the review's conclusions were limited.

Dicheva et al. (2015) conducted a thematic analysis of published studies of the application of gamification to education. The authors identified which game elements have been explored, where gamification has been applied (e.g., blended learning courses), what educational levels have used gamification, and the evaluation results. However, the analysis did not include a discussion of the overall effects of gamification on student achievement.

Looyestyn et al., 2017 examined 15 studies on the effectiveness of gamification on adult engagement, including studies conducted in non-educational contexts (e.g., marketing, health medication use, survey design, online trading). The individual effect sizes of only three studies that focused specifically on student learning performance were reported, limiting the conclusion.

More recently, Subhash and Cudney (2018) conducted a systematic review of gamification and game-based learning studies. They reported the results mainly in a "descriptive" manner: the studies were presented in chronological order and the authors described how gamification and games were implemented in each study. No discussion of the quantitative effects of gamification on student achievement was given.

Finally, Baptista and Oliveira (2019) conducted a meta-analysis to examine the various factors (e.g., ease of use, hedonic value, enjoyment, usefulness) that can predict user intention to use gamification. A mixture of studies of gamification and games (e.g., serious games, augmented reality games) were analyzed together, making it impossible to isolate the findings related to gamification only. Moreover, the study did not examine the variables specifically related to student academic performance.

2.1. Contributions of this study

This review differs from other reviews of gamification research in four areas. First, it specifically examines *comparison-based* intervention studies (i.e., studies that compared gamified to non-gamified practices) including studies with randomly selected groups (gamified and non-gamified) and non-randomly selected groups (gamified and non-gamified). The main purpose of the comparison is to identify possible causal effects of gamified practices on student academic performance. Comparing learning with gamification with learning without gamification practices makes it possible to make more informed decisions about the use of gamification.

Second, we ascertain the factors (or moderators) under which gamification most effectively affects students' academic performance (Faiella & Ricciardi, 2015). Previous studies have not fully explored the potential factors that may moderate the effect sizes. For example, we do not know how the number and type of game elements used affect student performance.

Third, we focus specifically on gamification studies, excluding all game-based learning studies (e.g., studies of serious games).

Fourth, we include all levels of educational settings – not limiting our search to any particular educational context. The following research questions guide this study.

a) RQ1: How does the effect of gamification on student academic performance compare to the effect of non-gamified learning in K-12 and university contexts?
b) RQ2: What do students like or dislike about gamification?

3. Methods

3.1. Databases searched

The process for selecting relevant studies was guided by the preferred-reporting of items for systematic reviews and meta-analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, & Altman, 2009). Leading academic databases were searched, as they are more likely to yield relevant and high quality papers (Webster & Watson, 2002). The recommended databases are ACM Digital Library, EBSCO host research databases (including Academic Search Premier, British Education Index, ERIC, TOC Premier), Emerald Insight, Science Direct, Scopus, and Web of Science (Baptista & Oliveira, 2019; Chung, Hwang, & Lai, 2019; Connolly et al., 2012). We did not impose any constraints on the language of instruction or the location of the studies, but the studies had to be reported in English and published in peer-reviewed journals. Peer-reviewed journals are a useful criterion for selecting studies of sufficient quality (Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016).

3.2. Search strings

To acquire the broadest scope of qualified papers, we used the Boolean operators (AND, OR). An asterisk was applied to capture a range of common expressions used in gamified learning. For the purpose of this review, two sets of search terms were used. One set consisted of terms for gamification; we used gamif* to cover all morphologic variations of "gamification," "gamified learning," "gamify," and "gameful." The other set of search terms contained terms related to learning, performance, course, outcomes, education, class, impact, effect, or influence. Therefore, the search strings were as follows:gamif* AND (education OR class OR course OR learning OR performance OR behavior OR outcomes OR evaluation OR impacts OR effects OR influence).

The above search string found 828 articles. An additional three articles were identified through other sources (e.g., searching through the references of relevant articles). After removing duplicate articles (due to duplication across the databases) and reading each article's title and abstract to determine whether they fit the research topic, 72 full-text articles were selected and downloaded for further analysis. The other articles were dropped because they did not meet the criteria for this study (e.g., articles using games such as hangman, articles focusing on business practices, or articles not written in English).

3.3. Inclusion criteria: Meta-analysis and qualitative evidence synthesis

To examine the possible effects of gamification on student academic performance, a meta-analysis of eligible articles was conducted. The eligibility criteria were as follows:

3.3.1. Inclusion

a) The study reported on an empirical examination of gamified practices that used at least one game element. Studies that contained mere descriptions or discussions of gamification without any empirical data were excluded. As there were few studies of gamification published before 2010 (Deterding et al., 2011), the period of our search was from January 2010 to December , 2018.
b) The study was conducted in a K-12 or higher education setting.
c) The study contained at least one comparison of learning outcomes in a gamified class versus in a non-gamified class (i.e., a between-subjects research design). We excluded within-subject designs due to the possible carry-over intervention effects in such studies (Charness, Gneezy, & Kuhn, 2012).
d) The study objectively measured student academic performance in the same course topic for both the gamified and non-gamified classes using assessments such as final exams or post-tests. These forms of assessments have been frequently used by researchers assessing the learning outcomes of an intervention (Chen, Wang, Kirschner, & Tsai, 2018; Cheung & Slavin, 2013; Li & Ma, 2010; Rakes, Valentine, McGatha, & Ronau, 2010; Zheng, Warschauer, Lin, & Chang, 2016).
e) The study reported sufficient data such as means, standard deviations, sample sizes, the t value of the t-test, the F-value, or the z score of the Mann–Whitney U test. The corresponding authors of the selected studies were contacted to provide these data (if necessary).

3.3.2. Exclusion

a) Studies were excluded if their datasets or results lacked sufficient information (e.g., no sample size data) to calculate effect sizes.
b) Studies were excluded if they merely contained student self-reported data about their learning achievement.
c) Studies were excluded if they did not explicitly describe the specific game elements used.

Two authors independently read the 72 empirical papers and identified 24 that met the criteria and were therefore included in the meta-analysis. Percent agreement is calculated by adding up the number of cases that received the same rating by two coders and dividing that number by the total number of cases rated by the two coders (Stemler, 2004). The percent agreement was 90%. Disagreement about the selections was resolved by discussions between the two authors.

To be included in the synthesis of qualitative evidence, a study had to meet the following criteria:

a) The study reported empirical gamified practices that used at least one clearly described game element.
b) The study was conducted in a K-12 or higher education setting.
c) The study involved data collected through interviews and/or open-ended survey questions and the analysis of qualitative data concerning students' perceptions (likes and dislikes) of gamification.
d) The study clearly presented the qualitative findings.

Two authors independently read the 72 papers and selected 32 studies for the qualitative synthesis. The percent agreement between the two authors was 100%. Fig. 1 outlines the process of article selection for both the meta-analysis and the qualitative synthesis.

4. Data analysis

4.1. Data extraction

The article selection process identified 24 articles (Aşıksoy, 2018; Cadavid and Corcho, 2018, de-Marcos et al., 2014, 2017; Gafni et al., 2018; Haruna et al., 2018; Hew et al., 2016; Homer et al., 2016; Huang & Hew, 2018; Huang et al., 2018; Jong et al., 2018; Ketyi, 2016; Lam et al., 2018; Marín et al., 2018; Ozdener, 2018; Rachels et al., 2018; Su & Cheng, 2015; Sun-Lin & Chiou, 2017; Tan & Hew, 2016; Tsay et al., 2018; Turan et al., 2016; Van Nuland et al., 2015; Yildirim, 2017; Zainuddin, 2018) discussing 30 independent interventions that met the inclusion criterion for the meta-analysis. The following key information was extracted from each article: (a) geographic origin; (b) intervention duration; (c) flipped or non-flipped class; (d) sample size; (e) game elements (type and number) (f) subject disciplines; (g) whether the students were given tangible rewards as a result of obtaining the game elements; (h) controls used in the research design (student equivalence and instructor equivalence); and (i) statistical results of student

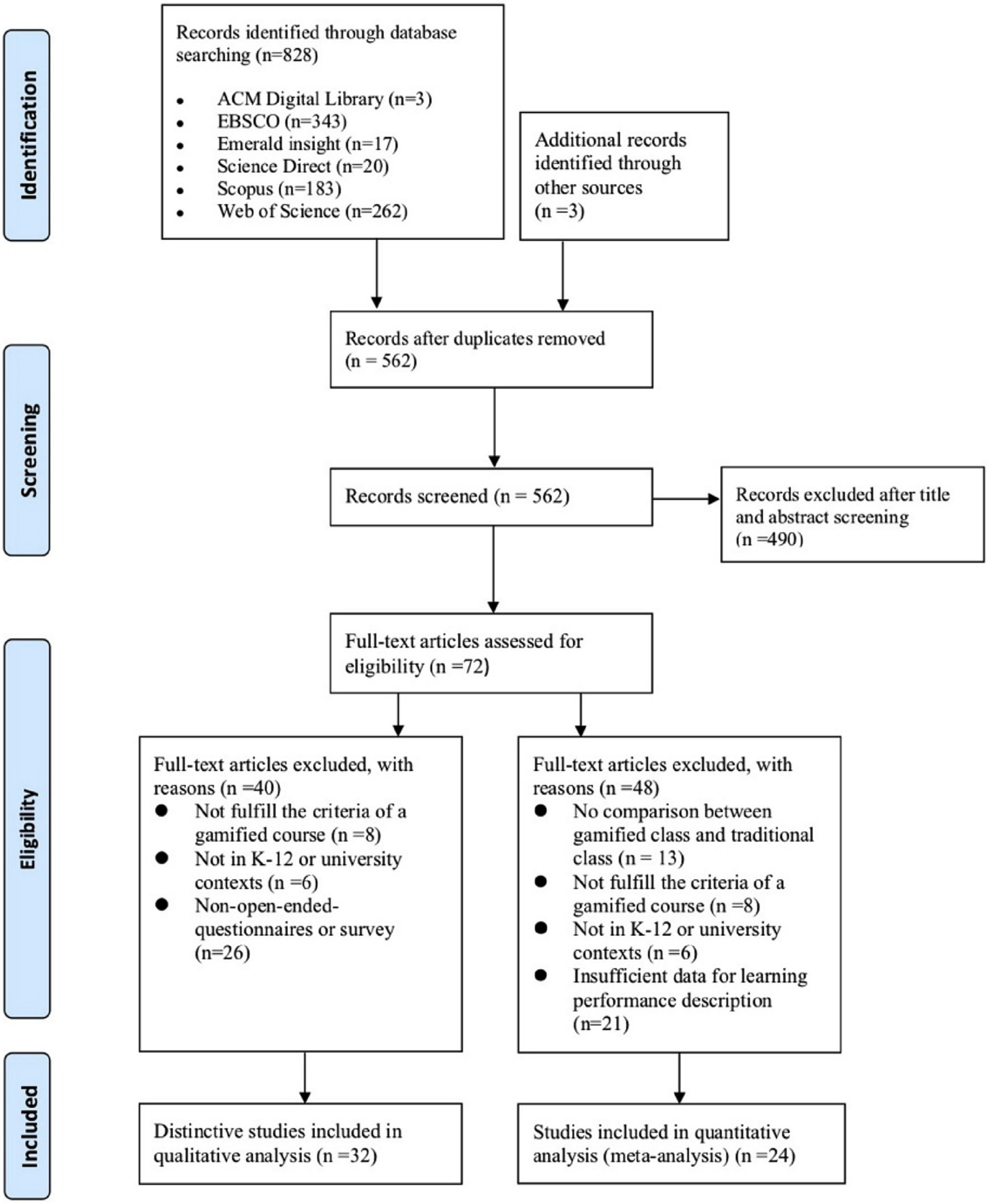

Fig. 1. PRISMA flowchart of article screening process.

achievements (e.g., means, standard deviations). Two coders independently extracted the information from all 24 articles. Intercoder reliability was 89%. The discrepancies were mutually resolved.

4.2. Determining the quality of the quantitative experimental studies

We also evaluated the methodological quality of the quantitative studies included in our meta-analysis, as valid results depend on the quality of the research design. Three instruments have been commonly used to evaluate the quality of quantitative experimental designs: The Medical Education Research Study Quality Instrument (MERSQI); the modified Newcastle-Ottawa Scale (m-NOS); and the Best Evidence in Medical Education (BEME) global rating (Cook, Levinson, & Garside, 2011). The MERSQI and m-NOS have been found to be superior to the global BEME in judging the quality of quantitative research (Cook et al., 2011). However, the MERSQI is more useful than the m-NOS for researchers who wish to examine the methodological rigor of the work, because it has lower rater subjectivity than the m-NOS scale (Cook et al., 2011). Therefore, we used the MERSQI to assess the quality of the quantitative studies in this review. The MERSQI is a 10-item instrument that assesses six aspects of study quality: study design, sampling, data type, validity of evidence, data analysis, and outcomes (for more details, see Reed et al., 2008). The possible MERSQI scores range from 5 to 18.

4.3. Computing effect sizes

We computed the effect sizes using the Comprehensive Meta-Analysis software (Biostat, Inc., Englewood Cliffs, NJ). All of the reported *p*-values were two-tailed unless otherwise reported. The I^2 test was used to test heterogeneity. The random-effects model was used to compute the effect sizes, as the mean scores of the samples were estimated in different population (Gurevitch & Hedges, 1999). The random-effects model was used to account for variation across different studies (Raudenbush, 2009, pp. 295–315).

Hedges' g is the adjusted standardized mean difference between two groups based on the pooled standard deviations, and it is particularly useful for the meta-analysis of studies with varying sample sizes (Korpershoek et al., 2016). If the means and standard deviations were not reported in the reviewed studies, the standardized mean difference was estimated using a variety of sources, including *t*-tests, *z*-scores, and *F*-values (Borenstein, Hedges, Higgins, & Rothstein, 2009; R.; Rosenthal & DiMatteo, 2001). If standard errors were used in the selected empirical studies but not standard deviations, we used the following formula to calculate the standard deviations (Altman & Bland, 2005):

$$SE = \frac{SD}{\sqrt{sample\ size}}$$

To meet the assumption of the independence of the effect sizes of the independent samples of students, one effect size was calculated for each eligible study (Scammacca, Roberts, & Stuebing, 2014). However, if a study reported effect sizes for *separate* sets of students (e.g., grade 1, grade 2, and grade 3 students), where the participants did not overlap, an effect size for each group was used in the meta-analysis, because each group (grade) was an independent sample (Lipsey & Wilson, 2001). Note that the two coders had to agree that the study participants were completely independent for multiple effect sizes to be reported for a single article. If a study reported multiple assessments of a single course subject, we selected the assessment that was most summative, as suggested by Freeman et al. (2014). For example, we chose final examinations over other types of assessments (e.g., mid-term examinations, weekly quizzes).

If a study reported multiple assessments (e.g., test 1, test 2, …) without a summative assessment or overall result from the *same* set of participants, we computed a single combined mean effect size using the Comprehensive Meta-Analysis software. The defining feature here was that the same participants provided data on different outcomes. Treating different outcomes from the same participants as though they were independent is not recommended, because it can lead to incorrect estimates of the overall effect (Halme et al., 2010). Following Freeman et al. (2014), we assumed that the correlation between the different outcomes within a comparison was 1 (i.e., a perfect positive correlation), because the same participants were sampled for each outcome. This was a conservative measure; the actual correlation between outcomes was probably lower than 1 (Freeman et al., 2014).

An important matter related to study coding is the issue of missing data. In every meta-analysis, it is not always possible to code certain items for some of the primary studies because these studies did not report them (Lipsey & Wilson, 2001). Although it is possible for the analyst to completely discard these studies, this procedure is not recommended because drawing conclusions only from studies that do report the items may be misleading (Lipsey & Wilson, 2001). Therefore, following the recommendation by Lipsey and Wilson (2001), we provided a "not reported" option to items in our coding protocol. Later moderator analysis can reveal the differences (if any) between studies with and without missing data.

4.4. Analyses of heterogeneity

The presence of heterogeneity (i.e., a degree of inconsistency in the studies' results) was detected using the I^2 test, where 0%–40% indicated that heterogeneity might not be important; 30%–60% indicated moderate heterogeneity; 50%–90% indicated considerable heterogeneity; and 75%–100% indicated substantial heterogeneity (Shamseer et al., 2015). To identify the possible sources of variance in effect sizes, we conducted moderator analyses by applying the random effects model to several categorical variables.

These variables were grouped into four main categories: (a) the types and number of game elements used; (b) levels of control (student equivalence, instructor equivalence); (c) course characteristics (sample size, subject disciplines, intervention duration, flipped or non-flipped class, instructional activities where game elements were integrated, provision of tangible rewards); and (d) participant characteristics (student level, geographical origin). These moderator variables were selected based on previous other research that examined possible moderators which could cause variance in effect size (e.g., Chen et al., 2018; Freeman et al., 2014; Fu et al., 2011; Landers et al., 2015; Zheng et al., 2016).

The first set of moderator analyses focused on the types of game elements and the number of game elements used in the studies. We wanted to determine whether the type of game element caused any difference in the effect sizes. We also wished to establish whether the use of more (or fewer) game elements moderated the effect sizes. Combinations of different game elements might affect learner outcomes more strongly than any element in isolation (Landers et al., 2015).

The second set of moderator analyses focused on the levels of control between the control (non-gamified) and treatment (gamified) groups. This is related to the methodological rigor of the included studies. Different levels of control may contribute to a variation in effect sizes (Freeman et al., 2014). We found that some of the reviewed studies were well-controlled, for example, they used the same instructor for both the treatment and control groups, but some studies were not. Therefore, following Freeman et al. (2014), we compared the overall effect sizes of the well-controlled and the less well-controlled studies. Two levels of control were examined in our review: student initial equivalence and instructor equivalence.

To evaluate student *initial equivalence*, we had three initial groups: (a) *No significant difference* group – i.e., students were initially equivalent based on the statistical results of some initial assessment such as pre-test. (b) *Significant difference* group – i.e., students were not initially equivalent. (c) *No data reported* group – i.e., the authors did not *explicitly* provide data on student initial equivalence, or they claimed that students were equivalent but did not support the claims with statistical data. The opposite of "no data reported" would be either "no significant difference", or "significant difference". Upon completing the coding, we found 24 interventions reporting no statistical difference between students in the non-gamified (control) and gamified (treatment) conditions. We found 6 interventions in which the authors did not explicitly provide data on student initial equivalence, or they claimed that students were equivalent but did not give any statistical evidence. We did not find any intervention that explicitly reported significant difference between the control and treatment groups.

To evaluate instructor equivalence, two coders checked each individual intervention and assigned it to one of three groups. In the *identical instructor* group, the same instructor was responsible for both the treatment and control groups. In the *different instructors* group, one instructor taught the treatment group and a different instructor taught the control group. In the *no data provided* group, the authors provided no data about the instructor.

The third set of moderator analyses focused on differences in the courses taught, specifically the subject disciplines (e.g., language, computer and information science, business, mathematics), the duration of the individual intervention, the sample size (the number of participants involved), the implementation of flipped classrooms (if any), and the provision of tangible rewards (if any).

Different subject disciplines, sample sizes, and intervention durations may contribute to the heterogeneity of effect sizes (Chen et al., 2018; Zheng et al., 2016). As one academic term is assumed to be 16 weeks, we coded the intervention duration as belonging to one of the following groups: (a) *< 1 week*, (b) *1 week*-1 month, (c) *1 month-3 months*, (d) *3 months-1 semester*, (e) *≥ 1 semester*, and (f) *no data provided*. With regard to sample size, we used a classification scheme adapted from Chen et al. (2018): (a) *< 50,* (b) *50–100,* (c) *100–150,* and (d) *≥150*.

We also looked for heterogeneity in effect sizes based on whether the gamified interventions were conducted in a flipped classroom setting. In a flipped classroom, students typically watch pre-recorded video lectures and answer quizzes before class. They then perform some type of active learning activities (e.g., discussions, projects) in class. Some of the studies (e.g., Zainuddin, 2018) used gamification to motivate learners to complete the pre-class activities (e.g., answering quiz questions). For example, learners who completed the pre-class activities would earn badges. We examined whether there was any significant heterogeneity between studies that used gamification in flipped classroom settings and those that used gamification in non-flipped settings. In addition, we evaluated whether studies that provided tangible rewards to students who obtained the game elements had different overall effect sizes. Learners might be more motivated by tangible rewards (e.g., marks, food) than by mere virtual rewards (e.g., points, badges) that they receive in a gamified system (Landers et al., 2015).

We also attempted to identify (if any) the types of instructional activities *where game elements are integrated*. For the purpose of this review, we grouped these activities into individual activities (e.g., individual quizzes), small-group activities (e.g., group-based questionnaire design activities), or both activities. We conducted an additional moderator analysis to determine whether these activities may cause any difference in the effect sizes. Please note, however, that we could not conduct moderator analyses on a more fine grained level because it was difficult to meaningfully group the various finer grained tasks. More importantly, details of the tasks were not always provided in the primary studies.

The fourth set of moderator analyses focused on differences in participants, specifically student levels and geographical regions. Variability in population-level characteristics may contribute to heterogeneity in effect sizes (Fu et al., 2011). For student level, we use school level categories, but not specific grades (e.g., elementary school, high school, undergraduate, and postgraduate). For geographical region, we assigned each independent intervention to a geographical region based on the United Nations Statistics Division (Lin et al., 2014). The United Nations Statistics Division is better known than other schemes used by United Nations agencies due to its finer categories, which can accurately capture the many different geographical regions.

As recommended by Tondello, Mora, and Nacke (2017), subgroups that had only one intervention were eliminated from the moderator analyses because the number of intervention is considered too few to derive meaningful findings (Fu et al., 2011).

4.5. Analyses of publication bias

Publication bias occurs when researchers publish only favorable results (Borenstein et al., 2009). To determine whether our review suffers from publication bias, we performed three types of analyses: the classic fail-safe *N* test, the funnel plot, and the Begg and Mazumdar rank correlation.

We first calculated the classic fail-safe *N* test to determine how many missing studies with an effect size of 0 would have to be published to reduce the overall effect size of 0.504 (this was the overall effect size found in this study) of the student learning outcomes (i.e., increase the p value associated with the mean effect above $\alpha = 0.05$). Carson, Schriesheim, and Kinicki (1990) stated that "If the fail-safe N (X) is relatively small in comparison to the number of studies in the meta-analysis (k), then only tenuous conclusion should be drawn" (p. 239). The researchers suggested using Rosenthal 's (1979) guideline in which X should reach $5k + 10$ to ensure that X is large relative to k (the number of independent effect sizes). Using this formula, the X in our meta-analysis should be larger than 160 (i.e., 5(30) + 10).

In addition to the classic fail-safe N test, we created a funnel plot of the standard error using Hedges' g (random effects). Funnel plots compare the effect sizes to standard errors. They provide a useful graphic tool for checking whether treatment effects have been distorted by publication bias (Matthias Egger, Schneider, & Smith, 1998). The results from smaller studies are scattered at the bottom, and the spread narrows among larger studies (Sterne & Egger, 2001). If there is no publication bias, the plot resembles a symmetrical inverted funnel (M. Egger, Smith, Schneider, & Minder, 1997). Conversely, in the event of publication bias, the funnel plot is often asymmetrical and skewed. To further examine the presence of symmetry, we computed the Begg and Mazumdar rank correlation (Kendall's Tau with continuity correction, which can "quantify the amount of bias captured by the funnel plot" (Borenstein, 2005, p. 195).

4.6. Analyzing students' perceptions of gamification

This study also identified what learners specifically like and dislike about gamification. Understanding learners' likes and dislikes enables us to gain insights into how gamification affects learners' academic performance. We synthesize the evidence in qualitative studies using the thematic analysis approach, one of several approaches recommended by the Cochrane Qualitative Review Methods Group (Noyes & Lewin, 2011). We follow the six main stages espoused by Braun and Clarke (2006): familiarizing oneself with data; generating initial codes; searching for themes; reviewing themes; defining and naming themes; and producing the report.

5. Results

5.1. General characteristics of the interventions

Nine of the interventions were conducted with elementary school students (e.g., Rachels & Rockinson-Szapkiw, 2018), and five interventions were conducted at the high school level (e.g., Jong, Chan, Hue, & Tam, 2018). One involved a mixture of high school, undergraduate, and postgraduate students (Gafni, Achituv, Eidelman, & Chatsky, 2018). Ten were carried out with undergraduate students (e.g., Marín et al., 2018). Three studies were conducted with postgraduate students (e.g., Tan & Hew, 2016), and two studies provided no information on student school level (e.g., Zainuddin, 2018).

Various subject disciplines were involved in the reviewed studies, including language learning (Homer, Hew, & Tan, 2016), arts (Jong et al., 2018), computer and information science (Huang, Hew, & Lo, 2018), and so on. A majority of the interventions (n = 10) had durations ranging from 4 to 12 weeks. Six interventions had very short durations of less than 1 week, whereas five interventions had durations that were equal to or more than 16 weeks (one semester). Four interventions lasted between 12 and 16 weeks and three had durations ranging from 1 to 4 weeks long. Two studies provided no information on intervention duration.

The types of game elements used in the interventions were also analyzed (Fig. 2). Most of the interventions used multiple game elements. A majority of the interventions used the combination of badges, leaderboard, and points (n = 8). The next most common combination was badges, leaderboard, levels, and points, which was used by six interventions. Four interventions used the

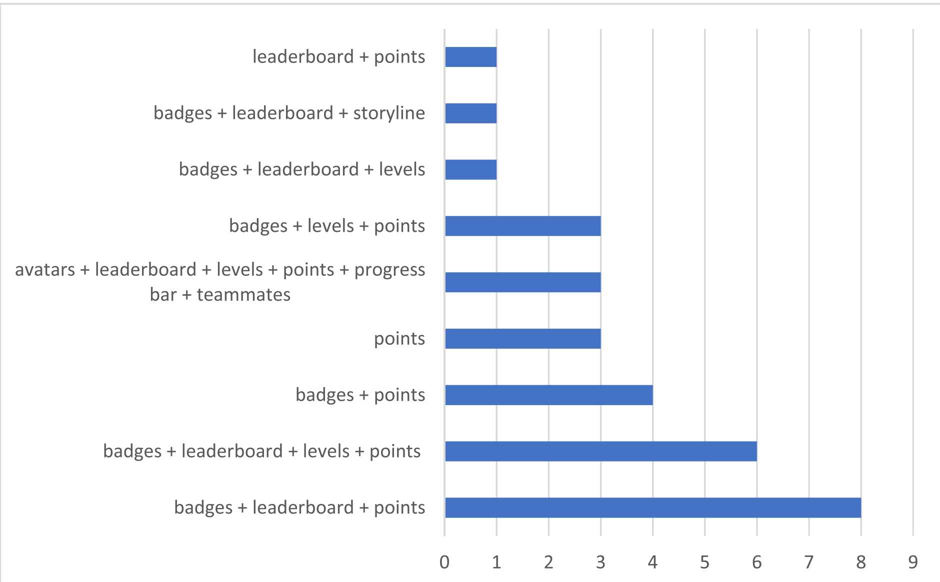

Fig. 2. Game elements in meta-analysis.

Table 2
MERSQI scores for 30 independent interventions.

Domain (max score)	MERSQI items presented (score)	No. present (%)
Study design (3)	Nonrandomized, 2 groups (2)	20 (67)
	Randomized controlled trial (3)	10 (33)
Sampling (3)	Institutions studies: 1 (0.5)	24 (80)
	Institutions studies: 2 (1)	1 (3)
	Institutions studies: 3 (1.5)	5(17)
	Response rate: < 50% or not reported (0.5)	19 (64)
	Response rate: 50–74% (1)	1 (3)
	Response rate: ≥75 (1.5)	10 (33)
Type of data (3)	Objective measurement (3)	30 (100)
Validity evidence (3)	Content (1)	30 (100)
Data analysis (3)	Appropriate for study design, type of data (1)	30 (100)
	Complexity of analysis: beyond descriptive analysis (2)	30 (100)
Outcomes (3)	Knowledge, skills (1.5)	30 (100)

Note: Sampling response rate means completion rate in the gamified class.

combination of badges and points. Only three interventions used a single type of game element (i.e., points).

Methodological quality was tested using the MERSQI. The MERSQI items presented in the eligible studies are listed in Table 2; the mean score was 12.37 on an 18-point scale. The mean score (M = 12.37) of the quantitative studies falls within the published range of other quantitative studies for other specialties (Smith, 2017).

5.2. RQ1: Comparison of learning performance between gamified and non-gamified groups

5.2.1. Overall effect size

Thirty independent interventions, involving 3,202 participants, were examined in this meta-analysis. An overall significant positive effect size (Hedges' g = 0.504, 95% CI [0.284–0.723], $p < 0.001$) in favor of gamification across various subject disciplines was found (Fig. 3). In other words, learners using gamification techniques had significantly better learning outcomes than those who did not use gamification in their courses.

5.2.2. Heterogeneity analyses

A significant Q statistic value ($Q = 246.773$, $I^2 = 88.248\%$, $p < 0.001$) suggested the presence of heterogeneity. We conducted several moderator analyses using the random-effects model to explore the possible causes of this heterogeneity. As stated above, four main categories of variables were examined: (a) the types and number of game elements used; (b) levels of control (student equivalence and instructor equivalence); (c) course characteristics (sample size, subject disciplines, intervention duration, flipped or non-flipped class implementation, provision or non-provision of tangible rewards); and (d) participant characteristics (student level, geographical origin). In this approach, each effect size is treated as an observation, and the effect size variance across studies is analyzed using the Comprehensive Meta-Analysis software (analogous to one-way random effects ANOVA model) (Lazowski & Hulleman, 2016; Lipsey & Wilson, 2001).

When we analyzed the data based on the game element used, we found no evidence of significant differences in heterogeneity for the types of game element ($Q_B = 2.525$, df = 5, $p = 0.773$) or for the number of game elements used ($Q = 0.179$, df = 4, $p = 0.996$) (see Table 3).

Heterogeneity analyses also indicated no significant variation between interventions that reported initial student equivalence and those that did not provide any such data ($Q_B = 0.494$, df = 1, $p = 0.482$) (see Table 4). It is noted that significant difference among students' prior knowledge were not found in the interventions. There was also no evidence of heterogeneity among interventions that reported the use of identical instructors in both the gamified and non-gamified groups, interventions that used different instructors, and interventions that did not provide any such data ($Q_B = 0.659$, df = 2, $p = 0.719$), see Table 6.

5.2.3. Publication bias

A funnel plot resembles a symmetrical inverted funnel if there is no publication bias (Egger et al., 1997). Conversely, in the event of publication bias, a funnel plot will often be asymmetrical and skewed. A visual analysis of Fig. 4 indicated that our funnel plot was symmetrical, suggesting there was no publication bias.

To further examine the symmetry, we computed the Begg and Mazumdar rank correlation (Kendall's Tau with a continuity correction). The result was 0.147 ($p = 0.254$), indicating that publication bias was not significant. In the fail-safe N test, we found 890 missing studies were needed to make the overall effect size statistically insignificant. This figure of 890 missing studies is more than five times larger than the 160 missing studies suggested by Rosenthal's (1979) formula: $5k + 10$ (i.e., 5(30) + 10). There would therefore have to be an unreasonably large number of undetected studies with zero effect to bring the reported effect sizes to values that may be statistically insignificant. Based on the funnel plot and statistical tests, we concluded that the effect sizes were not inflated by publication bias.

Study name	Outcome	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
		Statistics for each study						
Asiksoy (2018)	final score	2.236	0.324	0.105	1.601	2.871	6.903	0.000
Cadavid & Corcho (2018)	Combined	0.618	0.301	0.091	0.028	1.208	2.054	0.040
de-Marcos et al. (2014)	final	-1.254	0.166	0.028	-1.579	-0.929	-7.555	0.000
de-Marcos et al. (2017)	final	-0.233	0.114	0.013	-0.456	-0.010	-2.048	0.041
Gafni et al. (2018)	final score	0.264	0.177	0.031	-0.082	0.611	1.498	0.134
Haruna et al. (2018)	final score	1.835	0.265	0.070	1.316	2.354	6.931	0.000
Hew et al. (2016)	final score	-0.438	0.303	0.092	-1.032	0.156	-1.445	0.148
Homer et al. 1 (2016)	reading G1	-0.046	0.347	0.121	-0.727	0.635	-0.132	0.895
Homer et al. 2 (2016)	reading G2	-0.039	0.345	0.119	-0.714	0.637	-0.112	0.911
Homer et al. 3 (2016)	speaking G3	0.966	0.384	0.148	0.213	1.720	2.513	0.012
Homer et al. 4 (2016)	speaking G4	0.795	0.389	0.151	0.033	1.557	2.044	0.041
Huang & Hew (2018)	Combined	1.050	0.342	0.117	0.380	1.720	3.073	0.002
Huang et al. (2018)	final score	0.408	0.244	0.060	-0.070	0.886	1.673	0.094
Jong et al. 1 (2018)	high-achieving students	0.274	0.147	0.022	-0.014	0.563	1.865	0.062
Jong et al. 2 (2018)	low-achieving students	0.803	0.157	0.025	0.496	1.110	5.131	0.000
Jong et al. 3 (2018)	moderate-achieving students	0.690	0.152	0.023	0.392	0.988	4.538	0.000
Ketyi (2016)	final score	0.573	0.210	0.044	0.162	0.984	2.735	0.006
Lam et al. (2018)	final score	0.589	0.310	0.096	-0.018	1.196	1.901	0.057
Marin et al. (2018)	Combined	0.226	0.104	0.011	0.023	0.430	2.179	0.029
Nuland et al. (2014)	final score	0.654	0.299	0.089	0.069	1.240	2.190	0.029
Ozdener (2018)	Combined	0.643	0.316	0.100	0.022	1.263	2.031	0.042
Rachels & Rockinson-Szapkiw (2018)	final score	-0.052	0.154	0.024	-0.355	0.250	-0.339	0.734
Su & Cheng (2014)	final score	0.857	0.251	0.063	0.365	1.348	3.416	0.001
Sun-Lin & Chiou 1 (2016)	final score	0.883	0.295	0.087	0.305	1.462	2.994	0.003
Sun-Lin & Chiou 2 (2016)	final score	0.288	0.285	0.081	-0.272	0.847	1.009	0.313
Tan & Hew (2016)	final score	0.642	0.421	0.178	-0.184	1.468	1.523	0.128
Tsay et al. (2018)	final score	0.374	0.110	0.012	0.158	0.589	3.393	0.001
Turan et al. (2016)	final	0.916	0.215	0.046	0.495	1.338	4.257	0.000
Yildirim (2017)	final score	0.465	0.204	0.042	0.064	0.865	2.275	0.023
Zainuddin (2018)	Combined	0.817	0.279	0.078	0.270	1.364	2.928	0.003
		0.504	0.112	0.013	0.284	0.723	4.501	0.000

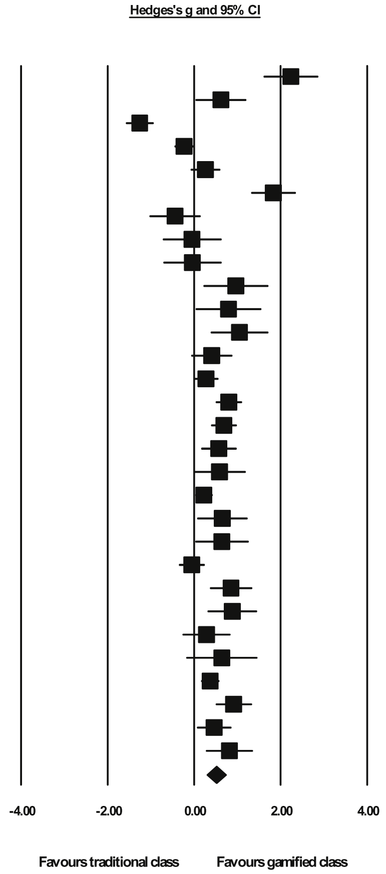

Fig. 3. Forest plot of effect size (Hedges'g) in random-effects model.

Table 3
Effect of game elements on effect size.

Moderator variables	N	Hedges' g	SE	95% CI LL	95% CI UL	Q_B
Type of game elements						2.525
Points	3	0.607	0.331	−0.042	1.255	
Badges + points	4	0.401	0.307	−0.2	1.003	
Badges + levels/unlock + points	3	0.254	0.303	−0.34	0.848	
Badges + leaderboard/rank + points	8	0.773	0.201	0.379	1.166	
Badges + leaderboard/rank + levels/unlock + points	6	0.494	0.221	0.061	0.927	
Avatar + leaderboard/rank + levels/unlock + points + progress bar + team (collaboration, competition)	3	0.588	0.298	0.005	1.172	
Number of game elements						0.179
1	3	0.607	0.381	−0.139	1.353	
2	5	0.444	0.308	−0.16	1.048	
3	13	0.485	0.178	0.136	0.834	
4	6	0.502	0.258	−0.005	1.008	
6	3	0.588	0.352	−0.102	1.279	

Note: N is the number of independent interventions, SE is the standard error, 95% CI is a 95% confidence interval, LL is the lower limit, UL is the upper limit, and Q_B is between-group variance in effect size.

5.3. RQ2 students' likes and dislikes about gamification

The meta-analysis reported above provided us with quantitative insights into the overall effect of gamification on learner academic performance. In this section, we present our synthesis of 32 qualitative studies (Aşıksoy, 2018; Borras-Gene et al., 2016; Buckley et al., 2017; Butler, 2013; Cadavid & Corcho, 2018; Çakıroğlu et al., 2017; Cavalcanti et al., 2018; Cheong et al., 2014; de-

Table 4
Effect of controls on research design on effect size.

Moderator variables	N	Hedges' g	SE	95% CI		
				LL	UL	*QB*
Student equivalence						0.494
No statistical difference	24	0.463	0.13	0.208	0.719	
No data reported	6	0.664	0.254	0.166	1.162	
Instructor equivalence						0.659
Different instructor	3	0.588	0.342	−0.081	1.258	
Identical instructor	18	0.565	0.152	0.266	0.864	
No data reported	9	0.37	0.202	−0.026	0.766	

Table 5 shows the variation in effect sizes due to course characteristics. There was no significant variation between different subject disciplines ($Q_B = 10.914$, df = 6, p = 0.091), between flipped and non-flipped classroom settings ($Q_B = 0.752$, df = 1, p = 0.386), between instructional activities (individual, and individual with small group activities) ($Q_B = 0.447$, df = 2, p = 0.800), or between interventions that used tangible rewards (e.g., marks, gifts) and those that did not ($Q_B < 0.001$, df = 1, p = 0.985). However, the effect sizes appeared to vary significantly with sample size ($Q_B = 10.597$, df = 2, p = 0.005) and intervention duration ($Q_B = 16.987$, df = 5, p = 0.005).

Table 5
Effects of course characteristics on effect size.

Moderator variables	N	Hedges' g	SE	95% CI		
				LL	UL	Q_B
Sample size						10.597**
1 (< 50)	13	0.501	0.167	0.174	0.827	
2 (50–100)	8	0.984	0.198	0.597	1.371	
4 (≥150)	8	0.106	0.184	−0.253	0.466	
Subject discipline						10.914
Arts	3	0.588	0.319	−0.037	1.213	
Computer & information	9	0.249	0.192	−0.128	0.627	
Health education	2	1.26	0.425	0.427	2.094	
Language	7	0.377	0.232	−0.077	0.832	
Mathematics	2	0.583	0.428	−0.255	1.422	
Research methodology	3	0.209	0.355	−0.487	0.905	
Science	3	1.275	0.348	0.593	1.957	
Intervention duration						16.987**
1 (< 1 week)	6	0.533	0.203	0.135	0.93	
2 (1week-1month)	3	0.488	0.31	−0.12	1.096	
3 (1month-3months)	10	0.906	0.166	0.581	1.232	
4 (3months-1 semester)	4	−0.278	0.238	−0.744	0.188	
5 (≥1 semester)	5	0.392	0.245	−0.088	0.871	
No data reported	2	0.421	0.359	−0.282	1.124	
Use of flipped classroom						0.752
Flipped class	8	0.671	0.224	0.233	1.11	
Non-flipped class	22	0.446	0.132	0.187	0.705	
Instructional activities where game elements were integrated						0.447
Individual and small group	6	0.417	0.251	−0.074	0.909	
Individual activity	23	0.512	0.130	0.257	0.768	
Provision of tangible rewards						< 0.001
No	24	0.505	0.125	0.26	0.75	
Yes	6	0.499	0.264	−0.018	1.016	

Note: One academic semester is assumed to be 16 weeks. **$p < 0.01$.

Finally, a heterogeneity test of interventions at different student levels detected no significant difference between levels ($Q_B = 2.173$, df = 4, p = 0.704), although the effect size was noticeably higher in the high school setting than in the elementary, undergraduate or postgraduate settings. In contrast, the geographic location of studies displayed significant evidence of heterogeneity ($Q_B = 16.935$, df = 4, p = 0.002).

Marcos et al., 2017; Fotaris et al., 2016; Goehle, 2013; Hakulinen et al., 2015; Halloluwa et al., 2018; Haruna et al., 2018; Hew et al., 2016; Homer et al., 2016; Huang & Hew, 2018; Hung, 2018; Jagust et al., 2018, Jong et al., 2018; Lam et al., 2018; Leaning, 2015; Lin et al., 2018; Mert and Samur, 2018; Olsson et al., 2015; Papp, 2017; Sahin et al., 2017; Sanchez-Carmona et al., 2017; Tan & Hew, 2016; Turan et al., 2016; Van Nuland et al., 2015; Zainuddin, 2018) that used data gathered from student interviews and open-ended responses. These studies can contribute to our deeper understanding of what students like or dislike about gamification. In this study, we find four main reasons why learners enjoy gamification: (a) gamification can foster enthusiasm; (b) gamification can provide feedback on performance; (c) gamification can fulfill learners' needs for recognition; and (d) gamification can promote goal setting. We also find two major reasons for learners' dislike of gamification: (a) gamification does not bring additional utility and (b) gamification can cause anxiety or jealousy. We will elaborate on these reasons in the following paragraphs.

Table 6
Effect of participant characteristics on effect size.

Moderator variables	N	Hedges' g	SE	95% CI		
				LL	UL	Q_B
Student level						2.173
Elementary school	9	0.501	0.214	0.083	0.92	
High school	5	0.822	0.271	0.291	1.352	
Undergraduate	10	0.363	0.193	−0.016	0.741	
Postgraduate	3	0.391	0.386	−0.366	1.148	
No data reported	2	0.69	0.437	−0.166	1.546	
Geographic location						16.935**
East Asia	15	0.514	0.135	0.25	0.779	
Northern America	2	0.254	0.347	−0.427	0.934	
South America	2	0.39	0.341	−0.279	1.059	
Southern Europe	2	−0.725	0.323	−1.359	−0.092	
Western Asia	5	0.848	0.224	0.409	1.286	

Note: Division standard based on United Nations geoscheme, **$p < 0.01$.

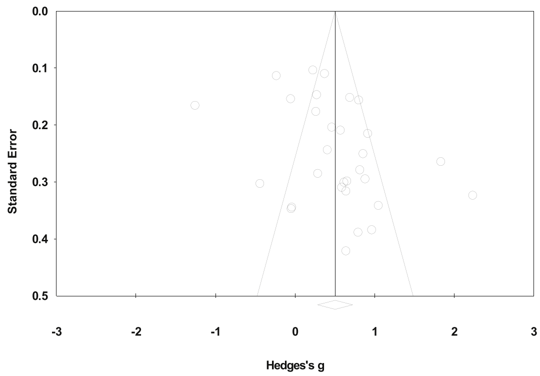

Fig. 4. Funnel plot assessing publication bias.

The most commonly reported "like" (reported in 22 studies) was that gamification fostered enthusiasm. For example, in a qualitative analysis of semi-structured interviews of 28 sixth-grade students in Turkey, 25 participants commented that "*I like this method,*" 24 participants reported that "*I want this method to be used in other courses,*" and 17 participants stated that "*This method has made the course more enjoyable*" (Turan, Avinc, Kara, & Goktas, 2016, p. 67). In a study of postgraduate master students, participants reported that "*The badges and points changed my attitude and made me more motivated to learn the topics presented in this course*" (Tan & Hew, 2016, p. 29).

The second most commonly reported "like," reported in 17 studies, was that gamification can provide feedback on performance. The participants in these studies felt that the use of game elements (e.g., leaderboards) helped them to track their own performance and their peers' performance. For example, Hakulinen, Auvinen, and Korhonen (2015) reported that several participants stated they enjoyed comparing badges with other people: "*Nice to see how others have done and nice to collect [badges]*" (p. 25). In another study, Cheong, Filippou, and Cheong (2014) reported students' comments about the use of several game elements. For example, students considered the point and leaderboard system as a useful feedback mechanism: "*It is always good to see how well you do,*" "*[Helps] keep track of performance,*" and "*… see how good you are compared with others*" (p. 240).

Several studies (n = 11) reported that gamification can also fulfill learners' needs for recognition. Cavalcanti, Filatro, and Presada (2018) reported that students perceived badges as a form of recognition for their achievements: "*At the beginning, I wasn't paying attention, but then I began to desire all [badges]. I believe that we are used to always having a 'prize' for our achievements; is already in our subconscious. I felt like in a video game, looking for the* 'award *for each phase*'" (p. 898). In another study (Goehle, 2013), students reported that gamification heightened their sense of accomplishment upon completing an activity: "*Something that I liked about the [gamification] system was that you were rewarded for doing good on problems, especially the harder problems. After working hard on a problem it was a nice feeling to get an achievement to show for it*" (p. 243).

Finally, several studies (n = 11) reported that students through gamification promoted goal setting. Huang and Hew (2018) reported that a majority of the participants stated that gamification motivated them to set higher goals and complete more tasks. For instance, one participant stated, "*With gamification, I would set up goals each week as to the type of badges that I wanted to earn and the total number of badges I wanted. As such, I would complete my work early*" (p. 268). Another participant remarked, "*If given three questions, and I am able to answer only two, and one question is beyond my ability to answer, I will spend time to search more literature and resources to find answer to that one question. But if without the badge system, I may just answer the two questions that I know. I would not* *attempt to answer the one question that I find difficult*" (p. 268).

Despite the overall positive sentiment about the use of gamification, the studies reported two major "dislikes." First, several studies (n = 7) reported that students disliked that fact that gamification did not bring additional utility. Several participants commented that gamification would have a stronger effect if there was a concrete reward attached to it: "*Extra points or a concrete reward would be good instead [of just a badge]*" and "*The badges alone do not encourage me much to do the exercises, because this is not primary school*" (Hakulinen et al., 2015). This was corroborated by findings reported in other studies (Hew, Huang, Chu, & Chiu, 2016; Huang & Hew, 2018).

Second, several studies (n = 14) reported that gamification can cause anxiety or jealousy. For example, one participant in the study by Turan et al. (2016) remarked, "*If I am to summarize this method [gamification] with a single word, I would say redundant. It cased jealousy between students*" (p. 67). Çakıroğlu, Başıbüyük, Güler, Atabay, and Yılmaz Memiş (2017) elaborated that participants who expressed negative opinions about gamification (e.g., leaderboards) were usually those who failed to get a top position on the board: "*In this sense, sometimes I thought it would have been better if we did not practice this application*" (p. 103). Çakıroğlu et al. (2017) found that all of the students who expressed this opinion were those who did not achieve a top-five place on the leaderboard on any occasion. However, the majority of participants (20 out of 37) were excited by the chance to be on the leaderboard and stated that achieving a place on the leaderboard was a key motivating factor for their participation in course activities.

6. Discussion

Many pundits have promoted the potential of gamification for motivating students to perform desired learning tasks. Yet despite the popularity of gamification, there is little consensus on whether it contributes to improved academic performance. Previous studies have reported mixed findings. This review provides a useful overview of quantitative studies on the use of gamification in educational settings. More specifically, we conduct a meta-analysis of studies of gamification and its impact on student learning performance as measured by objective instruments such as student exam scores.

Our comprehensive and careful selection process yields 24 relevant articles that discuss 30 independent interventions. We find that the majority of the interventions examined the impact of *multiple* game elements instead of single game elements; the most frequently used combination consisted of badges + leaderboards + points. Our appraisal of the quality of these quantitative studies using the MERSQI reveals a mean score of 12.37, which is consistent with the quality of other published quantitative experiment studies (Smith, 2017). This suggests that the quality of the quantitative gamification research studies reviewed in our meta-analysis is generally comparable with those published in other disciplines.

In response to our first research question about the effects of gamification on academic performance, we find an overall significant positive effect of $g = 0.504$, suggesting that gamification can increase student learning performance. The overall effect size of 0.504 may be considered a medium effect (Cohen, 1988) and is above the typical value of 0.4 for achievement that is reported elsewhere (Hattie, 2009). In the education field, Coe (2002) has argued that a small and inexpensive change that could increase achievement by an effect size as little as 0.1 would be a very significant improvement.

What reasons might have contributed to the favorable effect of gamification on student academic performance? Reasons for this may be inferred from the qualitative responses reported in previous studies. In our analysis of the qualitative studies, we identify three specific reasons for the positive effect of gamification on learning outcomes. First, gamification can promote goal setting among the learners. Goal setting can direct an individual's attention to goal-relevant activities and increase their persistence (Locke & Latham, 2002). Therefore, from a goal-setting theoretical perspective, learners who are given a clear goal (e.g., learners who complete the learning tasks before class will earn an "early bird" badge or learners who comment on other students' forum posts will get an "active participation" badge) will be more likely to complete the course-related tasks than learners who are merely told "to do their best" or who are not given an explicit goal (Jung, Schneider, & Valacich, 2010). Learners who complete course-related tasks tend to perform better than learners who put off completing them (Michinov, Brunot, Le Bohec, Juhel, & Delaval, 2011).

Second, gamification can fulfil a learner's needs for recognition. Motivational theorists (e.g., Deci & Ryan, 1985; Malone & Lepper, 1987) generally agree that learners like having their efforts recognized by others. Recognition can serve as a source of pride, and increased pride can lead to continued participation in course tasks and better learning (Landers et al., 2015). However, recognition requires the efforts or results of an individual learner's activities to be made visible to other people (Malone & Lepper, 1987). In a gamified course, this can be easily done through the use of badges. Badges, which are icons or logos that signify the accomplishments of a particular learner, can fulfil a person's need for acknowledgment and work as virtual status symbols (Sailer, Hense, Mandl, & Klevers, 2014). From an operant conditioning theoretical perspective, satisfying an individual's need for recognition can be a form of positive reinforcement. Therefore, when a learner completes a course-related task (i.e., a desired target response) and is given positive reinforcement through recognition of his effort (e.g., a badge), the likelihood of the desired target response occurring again increases (Landers et al., 2015).

Third, gamification can provide feedback on the performance of the individual learner and his or her peers' performance. From the self-determination theoretical perspective, gamified practices that provide indicators of the participants' performance (e.g., points, badges) can help to satisfy learners' need for competence (Sailer et al., 2017). Competence satisfaction can lead to improved task performance and enjoyment (Landers et al., 2015). In addition, from the perspective of the social comparison theory, people have a tendency to compare themselves with others to evaluate their own performances (Baldwin & Mussweiler, 2018). Therefore, when some learners know that their peers have won certain game elements (e.g., badges) or have achieved a high position on the leaderboard, they may strive to do better in the course activities (Huang & Hew, 2018).

Contrary to Landers et al. (2015), the use of more game elements or combinations of game elements does not appear to be associated with larger (or smaller) effect sizes. Although we are not completely sure of the reason for this result, we provide the following rationale based on our own research into gamification. The use of more (or different combinations of) game elements does not necessarily improve learning performance because one game element can fulfil multiple motivational roles. For example, gamification points can stimulate self-efficacy by providing feedback on performance and also function as a positive reinforcement by rewarding desired target behavior. Badges can fulfil similar motivational roles. This overlap in the functions of different game elements may explain why using combinations of elements might not have a stronger effect on learner outcome than using a single element.

Not every study used the same instructor for both the gamified (treatment) and non-gamified (control) courses, supporting claims that methodological rigor needs to be enhanced in research on gamification in educational contexts. Nevertheless, the results of our moderator analysis indicate that the findings are not unduly biased by the use of different instructors. The overall effect size derived from studies that use different instructors is very similar to that obtained from studies that use identical instructors. The results also appear unaffected by variations in the methodological rigor of the analyzed studies, as measured by the reporting of data on student initial equivalence.

We also find that effect sizes are not affected by student level (e.g., elementary school students, high school students, university students), subject disciplines (e.g., computer and information science, math, science), or by whether concrete or tangible rewards are provided. (For a further discussion on the use of tangible rewards, see the *Unresolved questions in existing gamification empirical studies* section.).

However, interventions with fewer subjects had higher effect sizes than interventions with more subjects. This is consistent with other research that shows that small samples generally return larger effect sizes than large samples (Slavin & Smith, 2009). One explanation might be that studies with smaller samples must have larger effect sizes to produce statistically significant results (Slavin & Smith, 2009).

Our results indicate that gamification tends to work better in Asian (e.g., East Asia, Western Asia) contexts, although this finding should be interpreted with caution due to the small number of non-Asian (e.g., European, North American) experimental studies available for comparison. At this juncture, we cannot establish the reasons for this variation. Most extant quantitative studies were conducted in Asia and in post-secondary (e.g., higher education) contexts. More research is required to examine whether gamification has different effects in different countries or educational levels and to establish the reasons for any such differences.

Finally, our results indicate that shorter gamified interventions have greater average effect sizes than longer interventions. More specifically, interventions between 1 and 3 months long have the largest average effect size ($g = 0.906$), whereas longer interventions of more than 1 semester have an average effect size of 0.392. The fact that short-term interventions tend to have larger average effect size raises the issue of the "novelty effect." Research has shown that the results of short-term interventions can be driven by the novelty effect, where participants try the intervention because it is something new and exciting, but do not maintain the engagement (Clark, 1983). It is plausible that in short-term interventions, learners are excited by the use of gamification, which leads to high participation in the course activities and better learning outcomes in the short term. Learners' tendency to pay more attention to new approaches could be a confounding variable in learning gains (Clark, 1983). Once the novelty of a new approach wears off, learners become bored with the same approach and this might diminish their desire to use it.

6.1. *Unresolved questions in existing gamification empirical studies*

We wish to highlight two unresolved questions from our meta-analysis: (a) Do tangible rewards matter to users? and (b) What is the best way to use leaderboards in educational contexts? If the answer to the first unresolved question is yes, then the next logical question is what types of tangible rewards should be given? In many gamified practices, learners are awarded game elements such as badges that do *not* carry any tangible rewards. Tangible rewards refer to those having concrete, visible, and easily measurable characteristics (Yoon, Sung, Choi, Lee, & Kim, 2015) such as course marks. It is possible that in a gamified learning environment, attaining these elements is considered fun (Landers et al., 2015), and obtaining tangible rewards is seen as less important.

Nevertheless, in several studies, learners expressed a lack of interest in non-utilitarian game elements (Hakulinen et al., 2015; Jia, Xu, Karanam, & Voida, 2016). Learners' lack of interest could eventually cause gamification to fail. Some learners expressed their desire to convert badges into tangible marks that count toward actual course grades (Huang & Hew, 2018). In view of this, we suggest that gamification points could be used to exchange for marks that count toward the overall grade. An example of this approach in action could be found in Lee Sheldon's (a professor at Indiana University) class, who determined his students' final letter grades (e.g., A, A-, B+, etc.) by the amount of gamification points the students had earned at the end of the course, in other words, by how much they have accomplished (Suzanne, 2018). Students were given a list of optional and meaningful tasks, along with the corresponding gamification points such as 25 points for presenting an individual project proposal to the class, or 1 gamification point for contributing an entry in the course online Glossary. All the tasks and gamification points earned helped measure learning against the intended course learning outcomes. Sheldon found the "gamification points-for-course grades" approach motivated students to perform better (with class average of a B) than students in classes without this approach (class average of a C) (Laster, 2010).

The gamification points earned in a course may also be exchanged for other types of tangible rewards, besides course marks. Zichermann and Cunningham (2011) divide rewards into four distinct categories - status, access, power, and stuff (SAPS). Status is the relative position in relation to others – for example, learners who obtained a certain number of gamification points will receive the title of *Expert Learner*. Access gives learners early access to new features – for example, learners who receive a certain number of gamification points could be allowed to unlock access to more useful course information (e.g., exam tips). Power can be given by promoting learners to moderators on forums (e.g., viewing deleted posts, deleting questions). Stuff can be a freebie (e.g., free e-book) to reward learners for completing a certain number of learning tasks.

The second unresolved question relates to the use of leaderboards in educational contexts. The leaderboard is one of the most commonly used gamification elements – so much so, that it is considered the workhorse of the gamification literature (Leung, 2019). According to Pedersen, Rasmussen, Sherson, and Basaiawmoit (2017, pp. 531–537), leaderboards can be roughly grouped into two categories: arcade style top N leaderboards that show only the top players and relative leaderboards that rank all of the participants. Most gamified practices in the education literature use the latter version (i.e., relative leaderboards).

Research on relative leaderboards in the context of game play suggests that individuals like to compare their performance with other people's. For example, Butler (2013) randomly assigned 132 individuals to one of three leaderboard versions, i.e., listing only the user's own score (blank version), listing the names and scores of users with higher scores than the player's (winning version), and listing the names and scores of users with lower scores than the player's (losing version), and found that users in the comparison conditions (i.e., the winning or losing versions) replayed the game approximately 50% more often than those in the blank condition. Studies of gamified education practices have also reported that many participants like to compare themselves with other people on the leaderboards, but many (especially those ranked near the bottom of a leaderboard) were unhappy about their positions.

How then should we best use leaderboards? There are no established best practices for the deployment of leaderboards, but several interesting findings have emerged from recent studies. Leung (2019) reported that people were *more* motivated to participate when their rank in the leaderboard was lowered from 1st to 20th but *less* motivated if their rank was further lowered. Leung (2019) therefore proposes the use of small groups of 10–20 individuals when implementing a leaderboard. In another study, Jia, Liu, Yu, and Voida (2017) propose that designers avoid showing the lowest-ranking participants on leaderboards in order to spare them from being negatively perceived by their peers. Further research should be conducted to test the effectiveness of Leung's (2019), and Jia et al.'s (2017) proposals.

6.2. Limitations

The search string used in this study was broad, allowing us to capture as many empirical studies as possible. In addition, the search was conducted in a large number of academic databases. Nevertheless, we had to exclude many studies because they did not use our specified experimental or quasi-experimental research designs to examine the causal effects of gamification on student learning performance. In addition, important basic information (e.g., insufficient statistical data or unclear reporting of students' qualitative data) was missing from many studies. Missing or unclear reporting of findings is a common problem for meta-analyses or research synthesis studies (e.g., Karabulut-Ilgu, Jaramillo Cherrez, & Jahren, 2018).

Second, in some of the moderator analyses, there were only a few interventions per category. For example, there were only two interventions in the health education category and in the mathematics education category. The low number of interventions per category limits the generalizability of our results.

7. Conclusion

Despite the aforementioned limitations, this study contributes to the literature in three ways. First, it augments our understanding of gamification in educational contexts. Put another way, it goes beyond the unqualified assertion that gamification is beneficial by quantifying the effect of gamification on learners' academic performance. The second contribution is the identification of factors that may moderate the effect sizes of gamification. For example, this study provides empirical evidence that increasing the number of game elements does not necessarily lead to better learning outcomes. Third, this study identifies what learners specifically like and dislike about gamification. Understanding learners' likes and dislikes enables us to gain insights into how gamification affects learners' academic performance.

We conclude by highlighting five additional directions for future research on gamification. First, studies with longer durations are urgently needed to examine the long-term effects of gamification. The majority of existing interventions lasted between 1 and 3 months. Longitudinal studies, in which learners are tracked over a long period of time (e.g., more than one year) should be conducted to examine whether and how learners' motivation and their behavioral engagement change over time.

Second, future research could examine the effects of game elements on student learning in non-conventional learning environments such as massive open online courses (MOOCs). Conventional learning environments employ some form of teacher exposition such as lectures or presentations. These lectures may or may not be accompanied with some kinds of active learning approaches such as small group work and discussions (Deslauriers, McCarthy, Miller, Callaghan, & Kestin, 2019; Gillette et al., 2018; Haruna et al., 2018). Conventional classroom may take place in face-to-face, online, or a mixture of online and face-to-face setting (Haruna et al., 2018; Kim, 2018; Vu, 2017). In a conventional classrooms, learners are supervised closely by a professor, and they are familiar with their classmates (Chiu & Hew, 2018). MOOC learners, however, are not familiar with their peers, and they are not closely supervised by a professor (Chiu & Hew, 2018). Almost all conventional college learners plan to finish the course they enrol in, but this is not the case for MOOC students (Reich & Ho, 2014). The typical MOOC completion rates range from 3.5% to 7.3% (Aydin, 2018). Given these unique characteristics of MOOCs, it will be interesting for future research to examine learners' perceptions towards using gamification in MOOCs, and to investigate whether gamification can enhance MOOC completion rate.

Third, future research might want to take a closer look at the influence of user types or traits on their interest in gamification. In this meta-analysis, we could not investigate user types or traits as a moderator variable, as none of the selected studies had measured this specific variable. Nevertheless, several studies not included in this meta-analysis have examined how users' types or traits can be used to develop personalized gamified systems for individual users. Arkün Kocadere and Çağlar Özhan (2018) examined how different player types (killer, achiever, explorer, and socializer) may influence different game elements. Jia et al. (2016) reported that extraversion and neuroticism are the two personality traits that most significantly affect the design of gamified systems. Other studies (e.g., Tondella et al., 2016) proposed other lists of gamification user types, including philanthropists, socializers, free spirits, achievers, players, and disruptors. Further research should first sort out this smorgasbord of a list to identify the most common traits or types found in the boarder user population. Following this, further research could be conducted to examine how different user types or traits can be best supported in a gamified system. We anticipate that this research agenda will help instructors design more personalized gamified systems that can better cater to different users' preferences.

Fourth, our coding and analysis could only be based on what had been explicitly reported in the reviewed articles. The absence of certain elements in the coding implies that the original authors did not explicitly report these elements in their articles. There may be discrepancies between what was documented and the actual research. We urge further work on the question of moderation of effect sizes based on the instructional design arrangements, as well as the types of instructional activities. Future research should not merely examined whether gamification can improve student learning, but *under which condition* gamification might work best. In order to achieve the latter, it is important for the research community to explicitly describe the instructional design arrangements, as well as types of instructional activities used in their studies.

Fifth, future work should examine the teachers' or instructors' attitude toward gamification. Since teachers play a key role in introducing technology-related innovations in the classroom, it will be useful to examine what teachers feel about gamification. Attitude towards a target behavior such as using gamification is an important predictor of an individual's intention toward performing the target behavior (Fishbein & Ajzen, 1975). To date, most gamification studies focus on students' use of gamification. Teachers' attitudes toward gamification remains a neglected research area (Marti-Parreno, Segui-Mas, & Segui-Mas, 2016).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.edurev.2020.100322.

References

References marked with an asterisk * indicate studies included in the meta-analysis.
References marked with two asterisks ** indicate studies included in the qualitative evidence synthesis.
References marked with two asterisks *** indicate studies included both in the meta-analysis and qualitative evidence synthesis.

Altman, D. G., & Bland, J. M. (2005). Statistics notes: Standard deviations and standard errors. *BMJ British Medical Journal, 331*(7521), 903 903.

Annetta, L. A. (2010). The "I's" have it: A framework for serious educational game design. *Review of General Psychology, 14*, 105–112.

Arkün Kocadere, S., & Çağlar Özhan, Ş. (2018). *Gamification from player type perspective: A case study. 21*, Educational Technology & Society12–22.

***Aşıksoy, G. (2018). The effects of the gamified flipped classroom environment (GFCE) on students' motivation, learning achievements and perception in a physics course. *Quality and Quantity, 52*(1), 129–145.

Aydin, C. H. (2018). MOOCs as change agents. In D. Jansen, & L. Konings (Eds.). *The 2018 OpenupEd trend report on MOOCs* (pp. 18–21). Maastricht, NL: EADTU. Retrieved from https://www.openuped.eu/images/Publications/The_2018_OpenupEd_trend_report_on_MOOCs.pdf.

Baldwin, M., & Mussweiler, T. (2018). The culture of social comparison. *Proceedings of the National Academy of Sciences of the United States of America, 115*(39), E9067–E9074. https://doi.org/10.1073/pnas.1721555115.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122.

Baptista, G., & Oliveira, T. (2019). Gamification and serious games: A literature meta-analysis and integrative model. *Computers in Human Behavior, 92*, 306–315. https://doi.org/10.1016/j.chb.2018.11.030.

Blumberg, F. C., Almonte, D. E., Anthony, J. S., & Hashimoto, N. (2013). Serious games: What are they? What do they do? Why should we play them? In K. E. Dill (Ed.). *The oxford Handbook of media psychology* (pp. 334–351). Oxford, United Kingdom: Oxford University Press.

Bogost, I. (2011). *Gamification is bullshit.* The Atlantic. Retrieved from https://www.theatlantic.com/technology/archive/2011/08/gamification-is-bullshit/243338/.

Borenstein, M. (2005). Software for publication bias. In A. J. S. H. R. Rothstein, & M. Borenstein (Eds.). *Publication bias in meta-analysis-prevention, assessment and adjustments* (pp. 193–220). Biostat, Inc., USA: John Wiley & Sons, Ltd.

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

Borges, S.d. S., Durelli, V. H. S., Reis, H. M., & Isotani, S. (2014). A systematic mapping on gamification applied to education. *Gyeongju, Republic of Korea Paper presented at the Proceedings of the 29th annual ACM symposium on applied computing*.

**Borras-Gene, O., Martinez-NuÑEz, M., & Fidalgo-Blanco, Á. (2016). New challenges for the motivation and learning in engineering education using gamification in MOOC. *International Journal of Engineering Education, 32*(2B), 501–512.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., et al. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education, 94*, 178–192.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

**Buckley, P., Doyle, E., & Doyle, S. (2017). Game on! Students' perceptions of gamified learning. *Educational Technology & Society, 20*(3), 1–10.

**Butler, C. (2013). The effect of leaderboard ranking on players' perception of gaming fun. *Paper presented at the Proceedings of the 5th international conference on online communities and social computing*. NV: Las Vegas.

***Cadavid, J.n. M., & Corcho, A.s. F. P. (2018). Competitive programming and gamification as strategy to engage students in computer science courses. *Revista ESPACIOS, 39*(35), 11.

**Çakıroğlu, Ü., Başıbüyük, B., Güler, M., Atabay, M., & Yılmaz Memiş, B. (2017). Gamifying an ICT course: Influences on engagement and academic performance. *Computers in Human Behavior, 69*, 98–107. https://doi.org/10.1016/j.chb.2016.12.018.

Caponetto, I., Earp, J., & Ott, M. (2014). Gamification and education: A literature review. In C. Busch (Ed.). *Proceedings of the 8th European conference on games based learning* (pp. 50–57). Nr Reading: Acad Conferences Ltd.

Carson, K. P., Schriesheim, C. A., & Kinicki, A. J. (1990). The usefulness of the "fail-safe" statistic in meta-analysis. *Educational and Psychological Measurement, 50*(2), 233–243. https://doi.org/10.1177/0013164490502001.

**Cavalcanti, C. C., Filatro, A., & Presada, W. A. (2018). Gamification design for tutor education in an online course. *ETD: Educação Temática Digital, 20*(4), 887–904. https://doi.org/10.20396/etd.v20i4.8648223.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization, 81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009.

Chen, Y.-H., & Chen, P.-J. (2015). MOOC study group: Facilitation strategies, influential factors, and student perceived gains. *Computers & Education, 86*, 55–70. https://doi.org/10.1016/j.compedu.2015.03.008.

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *88*(6), 799–843. https://doi.org/10.3102/0034654318791584.

**Cheong, C., Filippou, J., & Cheong, F. (2014). Towards the gamification of learning: Investigating student perceptions of game elements. *Journal of Information Systems Education, 25*(3), 233–244.

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88–113. https://doi.org/10.1016/j.edurev.2013.01.001.

Chiu, K. F., & Hew, K. F. (2018). Factors influencing peer learning and performance in MOOC asynchronous online discussion forum. *Australasian Journal of Educational Technology, 34*(4), 16–28.

Chung, C.-J., Hwang, G.-J., & Lai, C.-L. (2019). A review of experimental mobile learning research in 2010–2016 based on the activity theory framework. *Computers & Education, 129*, 1–13. https://doi.org/10.1016/j.compedu.2018.10.010.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459. https://doi.org/10.3102/00346543053004445.

Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. *Paper presented at the British educational research association annual conference, Exeter*.

Cohen, J. (1988). *Statistical power ANALYSIS for the Behavioral sciences*, *2nd*. Hillsdale: NJ: Lawrence Erlbaum Associates.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004.

Cook, D., Levinson, A., & Garside, S. (2011). Method and reporting quality in health professions education research: A systematic review. *Medical Education, 45*, 227–238.

Corti, K. (2006). *Games-based learning: A serious business application.* PIXELearning Limited. Retrieved from https://www.cs.auckland.ac.nz/courses/compsci777s2c/lectures/Ian/serious%20games%20business%20applications.pdf.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* Harper & Row.

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

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

Deci, E., & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*, *3*. New York: Plenum.

Deslauriers, L., McCarthy, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences of the United States of America, 116*(39), 19251–19257.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". *Paper presented at the Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. Tampere: Finland.

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

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

Educause (2011). *7 things you should know about gamification.* Retrieved from http://www.educause.edu/library/resources/7-things-you-should-know-about-gamification.

Egger, M., Schneider, M., & Smith, G. D. (1998). Meta-Analysis: Spurious precision? Meta-analysis of observational studies. *BMJ British Medical Journal, 316*(7125), 140–144.

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.

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.

Festinger, L. (1954). *A Theory of Social Comparison Processes, 7*(2), 117–140. https://doi.org/10.1177/001872675400700202.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

**Fotaris, P., Mastoras, T., Leinfellner, R., & Rosunally, Y. (2016). Climbing up the leaderboard: An empirical study of applying gamification techniques to a computer programming class. *Electronic Journal of E-Learning, 14*(2), 94–110.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., et al. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America, 111*(23), 8410–8415.

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., ... Trikalinos, T. A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the effective health care program. *Journal of Clinical Epidemiology, 64*(11), 1187–1197. https://doi.org/10.1016/j.jclinepi.2010.08.010.

*Gafni, R., Achituv, D. B., Eidelman, S., & Chatsky, T. (2018). The effects of gamification elements in e-learning platforms. *Online Journal of Applied Knowledge Management, 6*(2), 37–53.

Giannetto, D., Chao, J., & Fontana, A. (2013). Gamification in a social learning environment. *Issues in Informing Science and Information Technology, 10*, 195–207. https://doi.org/10.28945/1806.

Gnauk, B., Dannecker, L., & Hahmann, M. (2012). Leveraging gamification in demand dispatch systems. *Paper presented at the proceedings of the 2012 joint EDBT/ICDT Workshops, Berlin, Germany*.

**Goehle, G. (2013). Gamification and web-based homework. *Primus, 23*(3), 234–246. https://doi.org/10.1080/10511970.2012.736451.

Gurevitch, J., & Hedges, L. V. (1999). Statistical issues IN ecological meta-analyses. *Ecology, 80*(4), 1142–1149. https://doi.org/10.1890/0012-9658(1999)080[1142 SIIEMA]2.0.CO.

**Hakulinen, L., Auvinen, T., & Korhonen, A. (2015). The effect of achievement badges on students' behavior: An empirical study in a university-level computer science course. *International Journal of Emerging Technologies in Learning, 10*(1), 18–29. https://doi.org/10.3991/ijet.v10i1.4221.

**Halloluwa, T., Vyas, D., Usoof, H., Hewagamage, K. P. J. P., & Computing, U. (2018). Gamification for development: A case of collaborative learning in. *Sri Lankan primary schools, 22*(2), 391–407. https://doi.org/10.1007/s00779-017-1073-6.

Halme, P., Toivanen, T., Honkanen, M., Kotiaho, J. S., Mönkkönen, M., & Timonen, J. (2010). Flawed meta-analysis of biodiversity effects of forest management. *Conservation Biology, 24*(4), 1154–1156. https://doi.org/10.1111/j.1523-1739.2010.01542.x.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? — a literature review of empirical studies on gamification. *Paper Presented At The 47th Hawaii International Conference On System Science, Hawaii, USA*https://people.uta.fi/~kljuham/2014-hamari_et_al-does_gamification_work.pdf.

Hamari, J., & Sjöblom, M. (2017). What is eSports and why do people watch it? *Internet Research, 27*(2), 211–232. https://doi.org/10.1108/IntR-04-2016-0085.

***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), https://doi.org/10.3390/ijerph15092027.

Hattie, J. (2009). *Visible learning : A synthesis of over 800 meta-analyses relating to achievement.* London. New York: 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.

***Homer, R., Hew, K. F., & Tan, C. Y. (2016). 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.

***Huang, B., & Hew, K. F. (2018). Implementing a theory-driven gamification model in higher education flipped courses: Effects on out-of-class activity completion and quality of artifacts. *Computers & Education, 125*, 254–272. https://doi.org/10.1016/j.compedu.2018.06.018.

*Huang, B., Hew, K. F., & Lo, C. K. (2018). Investigating the effects of gamification-enhanced flipped learning on undergraduate students' behavioral and cognitive engagement. *Interactive Learning Environments,* 1–21. https://doi.org/10.1080/10494820.2018.1495653.

**Hung, H.-T. (2018). Gamifying the flipped classroom using game-based learning materials. *ELT Journal: English Language Teaching Journal, 72*(3), 296–308. https://doi.org/10.1093/elt/ccx055.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis : Correcting error and bias in research findings.* Thousand Oaks: SAGE Publications Inc. (US Thousand

Oaks.
**Jagušt, T., Botički, I., & So, H.-J. (2018). Examining competitive, collaborative and adaptive gamification in young learners' math learning. *Computers & Education, 125*, 444–457.
Jia, Y., Liu, Y., Yu, X., & Voida, S. (2017). Designing leaderboards for gamification: Perceived differences based on user ranking, application domain, and personality traits. *Paper presented at the proceedings of the 2017 CHI conference on human factors in computing systems. USA: Denver, Colorado*.
Jia, Y., Xu, B., Karanam, Y., & Voida, S. (2016). Personality-targeted gamification: A survey study on personality traits and motivational affordances. *Paper presented at the Proceedings of the 2016 CHI conference on human factors in computing systems*. USA: San Jose, California.
***Jong, M. S.-Y., Chan, T., Hue, M.-T., & Tam, V. W. L. (2018). Gamifying and mobilising social enquiry-based learning in authentic outdoor environments. *Educational Technology & Society, 21*(4), 277–292.
Jung, J. H., Schneider, C., & Valacich, J. (2010). Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management Science, 56*(4), 724–742.
Kapp, K. M. (2014). *The gamification of learning and instruction fieldbook : Ideas into practice.* San Francisco, CA: Wiley.
Karabulut-Ilgu, A., Jaramillo Cherrez, N., & Jahren, C. T. (2018). A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology, 49*(3), 398–411. https://doi.org/10.1111/bjet.12548.
*Ketyi, A. (2016). From mobile Language Learning to gamification: An overlook of research results with business management students over a five-year period. *Porta Linguarum,* 45–60.
Kim, J. (2018). *Confounding traditional and open online courses. Inside Higher.* Retrieved from https://www.insidehighered.com/blogs/technology-and-learning/confounding-traditional-and-open-online-courses.
Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*(3), 643–680. https://doi.org/10.3102/0034654315626799.
**Lam, Y. W., Hew, K. F., & Chiu, K. F. (2018). Improving argumentative writing: Effects of a blended learning approach and gamification. *Language, Learning and Technology, 22*(1), 97–118.
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 International Publishing.
Laster, J. (2010). *At Indiana U., a class on game design has students playing to win.* The Chronicle of Higher Education Retrieved from.
Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education:A meta-analytic review. *Review of Educational Research, 86*(2), 602–640. https://doi.org/10.3102/0034654315617832.
**Leaning, M. (2015). A study of the use of games and gamification to enhance student engagement, experience and achievement on a theory-based course of an undergraduate media degree. *Journal of Media Practice, 16*(2), 155–170. https://doi.org/10.1080/14682753.2015.1041807.
Leung, W. (2019). *How do one's Peers on a leaderboard affect oneself? Paper presented at the Proceedings of the 2019 CHI conference on human factors in computing systems, Glasgow, Scotland Uk*.
Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22*(3), 215–243. https://doi.org/10.1007/s10648-010-9125-8.
**Lin, D. T. A., Ganapathy, M., & Kaur, M. (2018). Kahoot! It: Gamification in higher education. *Pertanika Journal of Social Science and Humanities, 26*(1), 565–582.
Lin, J.-W., Lai, Y.-C., Szu, Y.-C., Lai, C.-N., Chuang, Y.-S., & Chen, Y.-H. (2014). Development and evaluation of across-unit diagnostic feedback mechanism for online learning. *Educational Technology & Society, 17*(3), 138–153.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA, US: Sage Publications, Inc.
Locke, E. A. (1996). Motivation through conscious goal setting. *Applied and Preventive Psychology, 5*(2), 117–124. https://doi.org/10.1016/S0962-1849(96)80005-9.
Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717. https://doi.org/10.1037/0003-066X.57.9.705.
Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin, 90*(1), 125–152. https://doi.org/10.1037/0033-2909.90.1.125.
Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does gamification increase engagement with online programs? A systematic review. *PloS One, 12*(3), https://doi.org/10.1371/journal.pone.0173403 e0173403.
Malone, T., & Lepper, M. (1987). *Making learning fun: A taxonomy of intrinsic motivations for learning. Cognitive and affective process analysis)*, *3*. Hillsdale: NJ: Erlbaum.
Marczewski, A. (2013). *Gamification: A simple introduction and a bit more* (2nd ed.). (Kindle edition self-published on Amazon Digital Services).
*Marín, B., Frez, J., Cruz-Lemus, J., & Genero, M. (2018). An empirical investigation on the benefits of gamification in programming courses. *ACM Transactions on Computing Education, 19*(1), https://doi.org/10.1145/3231709.
Marti-Parreno, J., Segui-Mas, D., & Segui-Mas, E. (2016). Teachers' attitude towards and actual use of gamification. *Procedia – Social and Behavioral Sciences, 228*, 682–688.
**Mert, Y., & Samur, Y. (2018). Students' opinions toward game elements used in gamification application. *Turkish Online Journal of Qualitative Inquiry, 10*(2), 70–101. https://doi.org/10.17569/tojqi.364130.
Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education, 56*(1), 243–252. https://doi.org/10.1016/j.compedu.2010.07.025.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology, 62*(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005.
Molleman, E., Nauta, A., & Buunk, B. P. (2007). Social comparison-based thoughts in groups: Their associations with interpersonal trust and learning outcomes. *Journal of Applied Social Psychology, 37*(6), 1163–1180. https://doi.org/10.1111/j.1559-1816.2007.00207.x.
Nakamura, J., & Csikszentmihalyi, M. (2009). *Flow theory and researchOxford handbook of positive psychology* (2nd ed.). New York, NY, US: Oxford University Press.
Noyes, J., & Lewin, S. (2011). Supplemental guidance on selecting a method of qualitative evidence synthesis, and integrating qualitative evidence with Cochrane intervention reviews. In B. A. Noyes J, K. Hannes, A. Harden, J. Harries, S. Lewin, & C. Lockwood (Eds.). *Supplementary Guidance for Inclusion of qualitative Research in Cochrane systematic Reviews of interventions: Cochrane collaboration qualitative methods group*.
**Olsson, M., Mozelius, P., & Collin, J. (2015). Visualisation and gamification of e-learning and programming education. *Electronic Journal of E-Learning, 13*(6), 441–454.
*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.
**Papp, T. A. (2017). Gamification effects on motivation and learning: Application to primary and college students. *International Journal for Cross-Disciplinary Subjects in Education, 8*(3), 3199–3201.
Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management, 52*(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008.
Pedersen, M. K., Rasmussen, N. R., Sherson, J., & Basaiawmoit, R. V. (2017). Leaderboard effects on player performance in a citizen science game. *Proceedings of the 11th European conference on games based learning, Ecgbl* 2017.
*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), 72–89.
Rakes, C. R., Valentine, J. C., McGatha, M. B., & Ronau, R. N. (2010). Methods of instructional improvement in algebra: A systematic review and meta-analysis. *Review of Educational Research, 80*(3), 372–400. https://doi.org/10.3102/0034654310374880.
Raudenbush, S. W. (2009). *Analyzing effect sizes: Random-effects modelsThe handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY, US: Russell Sage Foundation.

Reed, D., Beckman, T., Wright, S., Levine, R., Kern, D., & Cook, D. (2008). Predictive validity evidence for medical education research study quality instrument scores: Quality of submissions to JGIM 's medical education special issue. *Journal of General Internal Medicine, 23*(7), 903–907. https://doi.org/10.1007/s11606-008-0664-3.

Reich, J., & Ho, A. (2014). *The tricky task of figuring out what makes a MOOC successful. The atlantic*. Retrieved from https://www.theatlantic.com/education/archive/2014/01/thetricky-task-of-figuring-out-what-makes-a-mooc-successful/283274/.

Richter, G., Raban, D. R., & Rafaeli, S. (2015). Studying gamification: The effect of rewards and incentives on motivation. In T. Reiners, & L. C. Wood (Eds.). *Gamification in education and business* (pp. 21–46). Cham: Springer International Publishing.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638.

Rosenthal, R., & DiMatteo, M. R. (2001). META-ANALYSIS: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*(1), 59–82. https://doi.org/10.1146/annurev.psych.52.1.59.

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. https://doi.org/10.1037/0003-066X.55.1.68.

**Şahin, Y. L., Karadağ, N., Bozkurt, A., Doğan, E., Kılınç, H., Uğur, S., ... Güler, C. (2017). The use of gamification in distance education: A web-based gamified quiz application. *Turkish Online Journal of Qualitative Inquiry, 8*(4), 372–395.

Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2014). Psychological perspectives on motivation through gamification. *Interaction Design and Architecture(s), 19*(1), 28–37.

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. https://doi.org/10.1016/j.chb.2016.12.033.

**Sánchez-Carmona, A., Robles, S., & Pons, J. (2017). A gamification experience to improve engineering students' performance through motivation. *Journal of Technology and Science Education, 7*(2), 150–161.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research, 84*(3), 328–364. https://doi.org/10.3102/0034654313500826.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... Page, M. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ Clinical Research, 349*, g7647. https://doi.org/10.1136/bmj.g7647.

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly, 18*(2), 158–176. https://doi.org/10.1521/scpq.18.2.158.21860.

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review, 57*(4), 193–216. https://doi.org/10.1037/h0054367.

Skinner, B. F. (1953). *Science and human behavior.* Oxford, England: Macmillan.

Skinner, B. F. (1989). The origins of cognitive thought. *American Psychologist, 44*(1), 13–18. https://doi.org/10.1037/0003-066X.44.1.13.

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology, 100*(4), 765–781. https://doi.org/10.1037/a0012840.

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis, 31*(4), 500–506. https://doi.org/10.3102/0162373709352369.

Smith, T. (2017). Gamified modules for an introductory statistics course and their impact on attitudes and learning. *Simulation & Gaming, 48*(6), 832–854. https://doi.org/10.1177/1046878117731888.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research and Evaluation, 9*(4), Retrieved from https://pareonline.net/getvn.asp?v=9&n=4.

Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54*(10), 1046–1055. https://doi.org/10.1016/S0895-4356(01)00377-8.

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.

*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.

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science, 11*(5), 159–163. https://doi.org/10.1111/1467-8721.00191.

*Sun-Lin, H.-Z., & Chiou, G.-F. (2017). Effects of gamified comparison on sixth graders' algebra word problem solving and learning attitude. *Educational Technology & Society, 22*(1), 120–130.

Suzanne, H. (2018). *Gamification in education: 4 ways to bring games to your classroom.* Retrieved from https://tophat.com/blog/gamification-education-class/.

***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.

Toda, A. M., Valle, P. H. D., & Isotani, S. (2017). *The dark side of gamification: An overview of negative effects of gamification in education. Paper presented at the higher education for all.* Maceió, Brazil: From Challenges To Novel Technology-Enhanced Solutions.

Tondello, G. F., Mora, A., & Nacke, L. E. (2017). Elements of gameful design emerging from user preferences. *Paper presented at the Proceedings of the annual symposium on computer-human interaction in play, Amsterdam, The Netherlands*.

*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.

***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, 11*(7), 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.

Vu, L. (2017). A case study of peer assessment in a composition MOOC: Students' perceptions and peer-grading scores versus instructor-grading scores. In P. Vu, S. Fredrickson, & C. Moore (Eds.). *Handbook of research on innovative pedagogies and technologies for online learning in higher education* (pp. 178–217). IGI Global.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii–xxiii.

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

Woolfolk, A. E. (1998). *Educational psychology.* Boston: Allyn & Bacon.

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

Yoon, H. J., Sung, S. Y., Choi, J. N., Lee, K., & Lim, S. (2015). Tangible and intangible rewards and employee creativity: The mediating role of situational extrinsic motivation. *Creativity Research Journal, 27*(4), 383–393.

***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.

Zheng, B., Warschauer, M., Lin, C.-H., & Chang, C. (2016). Learning in one-to-one laptop environments:A meta-analysis and research synthesis. *Review of Educational Research, 86*(4), 1052–1084. https://doi.org/10.3102/0034654316628645.

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in Web and mobile apps.* O'Reilly Media, Inc.

Falkner, N. J. G., & Falkner, K. E. (2014). “Whither, badges?” or “wither, badges!”: a metastudy of badges in computer science education to clarify effects, significance and influence. *Paper presented at the proceedings of the 14th koli calling international conference on computing education research.* Koli: Finland.