Received: 26 April 2024 | Revised: 7 August 2024 | Accepted: 15 September 2024

DOI: 10.1111/jcal.13077

REVIEW ARTICLE

Journal of Computer Assisted Learning WILEY

The use of leaderboards in education: A systematic review of empirical evidence in higher education

Chunqi Li | Lishi Liang | Luke K. Fryer | Alex Shum

Teaching and Learning Innovation Centre, The University of Hong Kong, Hong Kong, China

Correspondence
Chunqi Li, Teaching and Learning Innovation Centre, The University of Hong Kong, Hong Kong 999077, China.
Email: chunqili@connect.hku.hk and llchunqili@163.com

Abstract

Background: Leaderboards are among the most popular gamification elements in education. Some studies have implemented leaderboards and reported their individual effects on students' learning. Despite the emergence of relevant empirical studies, most of the existing reviews have only investigated the holistic impact of gamification. No previous systematic reviews were identified examining the individual use of leaderboards.

Objective: To address this gap, this review aims to systematically synthesise the existing empirical evidence concerning leaderboard use in education, examine their designs and effectiveness, and propose leaderboard design recommendations in gamified educational settings.

Method: This systematic review drew upon 20 articles (22 studies; 29 interventions) published from 2014 to 2023.

Results: The results found that using leaderboards can have a beneficial influence on students' learning motivation, engagement, and performance, but their effectiveness largely depends on their designs. Thus, this review examined the effectiveness of specific leaderboard design practices on students' learning and proposed four corresponding leaderboard design recommendations based on well-established educational and motivational theories as well as pertinent empirical studies.

Discussions and Conclusions: Notably, this review found that all included studies were undertaken in higher education and around half of them had short durations (less than or equal to 1 h). More longitudinal studies in other educational levels (e.g., primary and secondary schools) are thus called for to examine the validity and generalisability of the recommendations proposed.

KEYWORDS
empirical evidence, gamification, leaderboard, students' leaning, systematic review

1 | INTRODUCTION

Gamification has gained burgeoning popularity in education in recent years because it has been demonstrated to enhance students' learning outcomes (e.g., motivation, engagement, and achievement) effectively when appropriately implemented (Bai et al., 2020; Sailer and Homner, 2020; Zainuddin et al., 2020). Gamification is defined as using game design elements in non-game contexts (Deterding

et al., 2011). Most previous studies have deemed gamification as a unitary construct (Schürmann and Quaiser-Pohl, 2022). However, gamification can combine various gamification elements in different ways and thus it is inappropriate to study gamification as a generic construct (Sailer et al., 2017). In line with this argumentation, an emerging research trend emphasises examining one gamification element at a time to identify its individual effects (Cao et al., 2022; Sanchez et al., 2020). These studies can help provide a more granular understanding of how a particular gamification element functions (Mekler et al., 2017). These studies can also inform educators on how a particular gamification element should or should not be implemented (Bai et al., 2021).

Leaderboards are among the most frequently used gamification elements in education (Bai et al., 2020). Leaderboards measure how an individual performs compared to others (Zichermann and Cunningham, 2011). A few studies indicated that using leaderboards could positively influence students' learning such as their motivation, engagement, and academic achievement (e.g., Landers and Landers, 2014; Sailer and Sailer, 2021). Nevertheless, some research has identified nonsignificant or small negative effects of leaderboards on students' learning (Philpott and Son, 2022; Zahedi et al., 2021). The inconsistent findings can be attributed to the different designs of leaderboards used in these studies (Cao et al., 2022). Despite the inconclusive findings of leaderboards, no single review was identified examining the current empirical studies about leaderboard use and many existing reviews still inspect gamification as a holistic concept (e.g., Zainuddin et al., 2020). To address these research gaps, this review aims to systematically synthesise the existing empirical evidence concerning leaderboard use in education, examine their designs and effectiveness, propose design recommendations for leaderboard use in gamified educational settings, and suggest future research agendas. Notably, all empirical studies identified by the present review were conducted in higher education. More research in other educational levels is called for to examine the validity and generalisability of the findings and design recommendations in this review.

1.1 | Leaderboards in education

Leaderboards show how students perform versus others by providing comparative feedback (Philpott and Son, 2022), which normally involves points and rank (Philpott and Son, 2022; Pi et al., 2023). Leaderboards rank students based on a series of criteria that aim to promote desired behaviours by showing a positive representation of an individual's points and rank and decrease undesirable behaviours by showing a negative representation of points and rank (Chou, 2015; Kapp, 2012). However, some research has cautioned that too much feedback (e.g., showing ranking and points concurrently) on leaderboards could occupy learners' cognitive resources and impede their learning (Lam et al., 2011; Pi et al., 2023).

The existing studies found that leaderboards had mixed effects on students' learning (e.g., Balci et al., 2022; Ortiz-Rojas et al., 2019). The inconsistent findings can be attributed to leaderboards with different designs being implemented (Bai et al., 2021; Cao et al., 2022). In terms of whether all competitors are displayed, leaderboards can normally be divided into two types: absolute leaderboards and relative leaderboards (Zichermann and Cunningham, 2011). Absolute leaderboards show the rankings of all participants (Bai et al., 2021). In contrast, relative leaderboards do not display every user's ranking and users can only see their own and their neighbours' rankings (Ortiz-Rojas et al., 2019). Some relative leaderboards only show users' rankings relative to their neighbours without displaying their overall ranking (Bai et al., 2021), while others display users' and their neighbours' overall ranking (Ortiz-Rojas et al., 2019). *Top N* leaderboards are considered as a special form of relative leaderboards in many studies (e.g., Zichermann and Cunningham, 2011). However, unlike common relative leaderboards that display rankings relative to neighbours, *Top N* leaderboards are not customised for each user, as all players can only see the top N players and cannot see their own rankings unless they are the top N players (Ortiz-Rojas et al., 2019). Thus, this review perceives *Top N* leaderboards as a third type of leaderboard distinct from absolute and relative leaderboards because this review intends to inspect different leaderboard designs and their effectiveness.

In terms of competitive mechanics, leaderboards can be classified as individual and team-based leaderboards (Höllig et al., 2020). A meta-analysis suggested that combining cooperation and competition could be an effective gamification strategy (Sailer and Homner, 2020). Compared to individual leaderboards, the competition provided by team leaderboards could be more constructive as they encourage cooperation, support, and skill development among members (Rigby and Ryan, 2011; Sailer and Sailer, 2021).

1.2 | Common theories underpinning leaderboard use

In this section, the common theories underpinning leaderboard use in education are briefly discussed.

1.2.1 | Goal-setting theory

Goal-setting theory (GST) suggests that goals can help boost performance by guiding attention, enhancing motivation and persistence, and promoting strategy use (Locke and Latham, 2002). Moreover, GST posits that goals, which are immediate, specific, and moderately challenging, motivate participants more effectively than long-term, ill-defined, and too easy or too tough goals (Locke et al., 1981). Leaderboards can provide multiple goal options (Landers et al., 2017). For example, students can set their goal as ranking 1st or 20th in the next examination based on their current ranking. Leaderboards can also provide feedback for participants to measure their progress in

comparison to their goals and adjust their strategies accordingly (Bai et al., 2020; Locke and Latham, 2002).

1.2.2 | Social comparison theory

Social comparison theory (SCT) postulates that people continually evaluate their characteristics and abilities by comparing themselves with others, which results in discrepancy assessment (Festinger, 1954). There are two types of social comparisons: downward comparison (a positive discrepancy), which occurs when one compares themselves with worse-off individuals for maintaining or enhancing self-esteem, and upward comparison (a negative discrepancy), which occurs when individuals compare themselves with those who are more competent to improve their abilities (Diel et al., 2021; Garcia et al., 2013). A prior meta-analysis suggested that people had a strong preference to perform upward comparison rather than downward comparison (Gerber et al., 2018). Leaderboards could provide participants with opportunities for upward comparison (Bai et al., 2020). Upward comparison with attainable negative discrepancy could motivate an individual to reduce the discrepancy between himself/herself and the target, which could subsequently enhance their learning engagement and performance (Balci et al., 2022; Chen and Chen, 2015; Michinov and Primois, 2005). However, if the gap between the individual's standing and the target is too large (extreme upward comparison) and the standard of the target seems out of reach, the individual might give up pursuing the "out-of-reach" standard, resulting in a decrease in motivation and effort (Diel et al., 2021). For example, if a student finds the discrepancy between himself/herself and a top player on leaderboards is attainable, he/she might be motivated to work hard to rank high. However, if he/she finds the discrepancy between himself/herself and a top player on leaderboards is too large, he/she might give up directly and cease to pay more effort.

1.2.3 | Self-determination theory

Self-determination theory (SDT) posits that humans have three innate psychological needs: autonomy, relatedness, and competence, which when satisfied, can enhance an individual's motivation (Ryan and Deci, 2000). Leaderboards allow participants to self-assess their performance and set their goals voluntarily, which could help satisfy their need for autonomy (Aldemir et al., 2018; Burgers et al., 2015). Leaderboards can provide competition or collaboration (team leaderboard) that enable participants to interact with others to serve relatedness needs (Bai et al., 2020). Additionally, leaderboards that provide positive comparative feedback can help fulfil students' competence needs (Cao et al., 2022; Ryan and Rigby, 2019). For example, when team leaderboards are used, students can collaborate with team members to support each other and compete against other teams, which can help satisfy their need for relatedness. In this learning scenario, one team can compare their performance with other teams and voluntarily pick the target(s) they want to exceed in the next round of ranking rather than being provided with a specific standard by the instructor, which can thus satisfy their autonomy need. When students feel they outperform other teams, their competence needs can be satisfied.

1.2.4 | Theory of gamified learning

Leaderboards can represent three (conflict/challenge, rules/goals, and assessment) of the game attributes which are deemed valuable in Landers and Landers's (2014) theory of gamified learning (TGL). In terms of conflict, Landers's TGL suggests that conflicts yield maximum benefits when participants have relatively equal opportunities to rank high on leaderboards through exerting similar efforts (Landers and Landers, 2014). This indicates that leaderboards emphasising efforts (e.g., number of works read) rather than results (e.g., understanding of materials) might be more effective (Landers and Landers, 2014). Regarding rules, Landers's theory postulates that rules/goals should be clearly defined (Landers, 2014). This theory endorses that GST can be used to understand the game attribute of rules (Landers et al., 2015). For instance, goals that are specific, measurable, achievable, and time-bound can encourage players most effectively (Locke and Latham, 1990). The last game attribute is assessment, which refers to recognising the accomplishment of rules/goals (Landers, 2014). Leaderboards can present high-ranking players or points to realise assessment (Landers and Landers, 2014).

2 | PREVIOUS REVIEWS OF GAMIFICATION AND LEADERBOARDS IN EDUCATION

The majority of existing reviews and meta-analyses deemed gamification as a holistic construct and examined the overall effects of gamification in education (Bai et al., 2020; Sailer and Homner, 2020; Zainuddin et al., 2020). For example, the systematic review conducted by Zainuddin et al. (2020) reviewed 46 articles to evaluate the existing gamification literature in education and its impact on students' learning and instruction. This review found gamification positively influenced learners' learning in primarily three aspects: learning achievement, motivation and engagement, and interaction and social connection. Sailer and Homner (2020) performed a meta-analysis including 38 publications to systematically summarise research findings on the impact of gamification. They found that gamification had significantly small and positive effects on students' cognitive, motivational and behavioural learning outcomes. While these research syntheses provided a comprehensive overview regarding gamification use and its effects on students' learning, their research findings cannot be completely applied to specific gamification elements as they examined gamification incorporating various elements as a unitary construct.

Although the previous reviews suggested that leaderboards were among the most popular gamification elements in education (Bai et al., 2020; Zainuddin et al., 2020), only one review was identified as

focusing on leaderboards. Park and Kim (2021) performed a literature review and proposed three leaderboard design principles: (1) macro and micro leaderboards should be used concurrently; (2) each measurable element in the gamified learning environment should be integrated into micro leaderboards (3) leaderboards should incorporate both learning and non-learning activities. They further analysed how leaderboard cases (e.g., leaderboards used in applications/platforms) were related to these design principles. Their research enriched the literature by proposing possible leaderboard design principles. However, the design principles they proposed were mostly from theoretical rationales and lacked guidance from existing empirical evidence. Moreover, they did not examine how leaderboards influenced students' learning. Overall, empirical evidence regarding leaderboard use has not been specifically, nor comprehensively reviewed, whereas most extant reviews have examined gamification as a whole.

3 | RESEARCH RATIONALE

The current systematic review differs from previous reviews in three aspects. First, this review specifically examines the use of one popular gamification element, leaderboards, in education rather than focusing on the holistic concept of gamification. Second, this review comprehensively examines existing empirical research regarding leaderboard use in education rather than leaderboard cases. Third, this review investigates both leaderboard designs and effectiveness rather than focusing on only leaderboard design as the previous review (Park and Kim, 2021). As the inconsistent findings of leaderboards on students' learning can be attributed to the different designs of leaderboards used (Bai et al., 2021; Cao et al., 2022), this review does not aim to conclude the holistic effects of leaderboards with different designs and instead focuses on the effects of specific leaderboard design practices on students' learning.

4 | RESEARCH AIMS, SIGNIFICANCE AND QUESTIONS OF THE CURRENT REVIEW

Specifically, this systematic review synthesised the existing empirical studies using leaderboards in education to examine leaderboard designs and their effects on students' learning, to propose leaderboard design recommendations, and to direct future investigation. This review is the first research synthesis that examines leaderboard use in education by inspecting empirical evidence. The following research questions were posed to guide this review:

1. What are the general characteristics (e.g., educational settings and disciplines) of studies that have been conducted?
2. What are the common leaderboard design practices in the included studies?
3. What learning outcomes have been reported when leaderboards were implemented?
4. What are the effects of specific leaderboard design practices on students' learning outcomes?

5 | METHOD

Before initiating this systematic review, an a priori protocol was registered in the Open Science Framework to ensure lucid and impartial reporting (Li and Fryer, 2022). This review consulted Petticrew and Roberts (2006)'s systematic review guideline for social sciences, as well as the preferred items for systematic reviews and meta-analysis (PRISMA) statement (Page et al., 2021). The steps taken for the present systematic review are outlined below and shown in Figure 1.

5.1 | Inclusion criteria

Articles included for the review need to satisfy the following criteria:

1. Articles should have at least one intervention that focused on leaderboard use in education and did not mix leaderboards with other gamification elements except points. Studies that used points along with leaderboards were included as points are fundamental gamification elements and are frequently used along with leaderboards (Zichermann and Cunningham, 2011).
2. Participants needed to be attending preschool, primary, secondary, or higher education.
3. Articles were required to report empirical data using quantitative, qualitative, or mixed-methods design.
4. Articles had to be published in English and in peer-reviewed journals.

5.2 | Study search

A systematic search was conducted in November 2023 using five educational (Education Research Complete and ERIC), psychological (PsycINFO), or multi-disciplinary databases (Web of Science and Scopus). A search string was carefully designed to capture relevant articles: ("gamified" OR "gamification" OR "gameful" OR "game-based learning" OR "serious game*" OR "educational game*" OR "educational videogame*") AND "leaderboard*" AND ("education" OR "learning" OR "teaching" OR "student*" OR "teacher*" OR "class*" OR "course*").

The search yielded 759 articles. After removing duplicates, titles and abstracts of these articles were examined based on the above inclusion criteria. The full texts of the remaining articles were further screened to examine whether they met the inclusion criteria. During full-text reviews, no additional articles were identified through snowballing. Twenty-one articles were obtained and further evaluated for their quality.

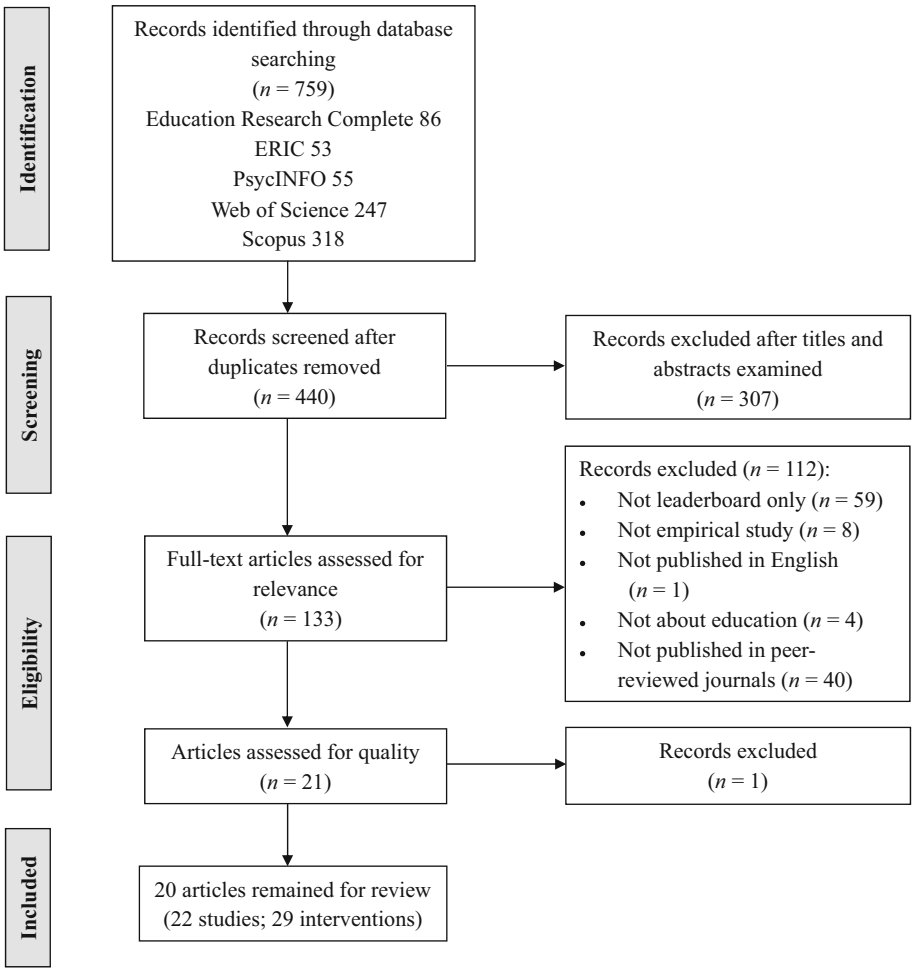

FIGURE 1 Preferred items for systematic reviews and meta-analysis (PRISMA)-SR flow diagram for article selection.

5.3 | Quality appraisal

This review adopted Hong, Fàbregues, et al. (2018)'s Mixed Methods Appraisal Tool (MMAT) to appraise the quality of obtained papers. MMAT was selected for its validated psychometric properties (Pace et al., 2012) and flexibility to assess qualitative, quantitative, and mixed methods studies (Hong, Gonzalez-Reyes, and Pluye, 2018). MMAT comprises two fundamental screening questions (e.g., are there clear research questions?) and five sections with five criteria under each section. The five sections involve five types of study design: qualitative, quantitative randomised controlled trials, quantitative non-randomised, quantitative descriptive, and mixed methods. Included articles had to fulfil the two screening questions and at least two of the five criteria under each section (Chan and Chen, 2022). The scores were calculated as high (four or five yes out of five criteria), medium (three yes), or low quality (two yes) based on the checklist (Chan and Chen, 2022; Evangelio et al., 2022; Hong, Fàbregues, et al., 2018). Articles that did not fulfil the two screening questions or satisfy none (0 Yes) or only one (one Yes) of the five criteria under each section were excluded.

After the quality assessment, one article (Flinton et al., 2023) was excluded because it was found to have only one Yes after being examined under the mixed-methods design section. Twenty articles (22 studies; 29 interventions) published from 2014 to 2023 were included in this review. Two articles had two studies each (Bai et al., 2021; Balci et al., 2022). The other included articles had one study each. Most included articles were rated as having high quality ($n = 18$; see Table A1). The detailed information for quality assessment of the twenty included articles and the one excluded article can be found in Tables A3A–A3E.

5.4 | Data extraction and data analysis

The key data extracted included (1) authors, year, and title; (2) research aims; (3) method; (4) location; (5) students' educational level; (6) subjects; (7) leaderboard design practices; (8) reported research outcomes; (9) instructional design (macro versus micro level). Leaderboards on macro-level instructional design refer to using leaderboards when a whole class is gamified (Philpott and Son, 2022). For example, the teacher awards points to students in multiple activities (e.g., completing homework/quizzes, reading books) and displays a class leaderboard (Philpott and Son, 2022). Leaderboards on a micro-level instructional design refer to using leaderboards for a specific learning activity (e.g., ranking students after they complete homework/quizzes) rather than gamifying different activities within a class

(Philpott and Son, 2022). For leaderboard design practices, leaderboard types were first coded. Specifically, this review classified leaderboards according to three standards: (1) real leaderboards versus artificial leaderboards (this classification standard was first raised by this review after examining the features of used leaderboards). The difference between them is that real leaderboards present comparisons between actual scores of real participants, while artificial leaderboards manipulate comparisons using fictitious or selected competitors for certain research purposes (see Figure 2 for an example); (2) individual versus team leaderboards based on competition mechanics; (3) absolute leaderboards, relative leaderboards, *Top N* leaderboards or other special leaderboards in terms of whether all participants (or teams) were presented. Afterwards, specific leaderboard design practices were generated inductively during the coding process.

Following common practices from previous reviews (e.g., Jia and Hew, 2021; Lo et al., 2017), 20% of articles were randomly selected and coded by two authors. The interrater reliability of Cohen's kappa ($\kappa = 0.89$; see Table A4) can be considered almost perfect (Landis and Koch, 1977). The discrepancies were noted and discussed until a consensus was reached. The remaining articles were coded independently by the two authors and all uncertainties were resolved through discussion during the coding process. This review utilised a combination of content and thematic analysis. This methodology enables all quantitative and qualitative data to be converted into meaningful codes, themes, and categories (Zainuddin et al., 2020). The data obtained via content and thematic analysis were descriptively summarised and critically appraised in terms of frequencies and identified themes.

6 | RESULTS

6.1 | General characteristics of included studies

6.1.1 | Educational settings and disciplines

All interventions were conducted in higher education, and none were conducted in primary or secondary education. Leaderboard interventions were implemented in various locations. The most frequently mentioned location was Germany ($n = 6$), followed by the US ($n = 5$), and China ($n = 5$). As for instructional design, 24 interventions used micro-level leaderboards. Among these interventions, 15 used leaderboards for learning activities which lasted for less than or up to 60 min, and nine used leaderboards for specific learning activities within a course without gamifying the whole course. Only five used leaderboards designed on a macro level. The included interventions were implemented for diverse subjects covering STEM, social science, and humanities. The most frequently addressed disciplines included English ($n = 7$), physics ($n = 4$), computer science ($n = 3$), math, arts, education and electrical engineering ($n = 2$ for each).

6.1.2 | Methodological features

Fourteen studies (21 interventions) used a quantitative research design, and eight studies (8 interventions) used a mixed-methods research design. Fifteen interventions had a short duration of less than or equal to 60 min. Eleven interventions had a much longer duration of one semester (around 10 weeks or above) to 1 year. Moreover, 15 interventions utilised self-developed platforms/applications to implement leaderboards, whereas 12 interventions used existing platforms such as Blackboard Learning Management System or Moodle (both $n = 2$). The general characteristics of the included studies discussed in Section 6.1 can be found in Table A1. The leaderboard design practices and reported outcomes of the included studies discussed in the following Sections 6.2 and 6.3 can be found in Table A2.

6.2 | Common leaderboard design practices

6.2.1 | Leaderboard types

Across all interventions, 18 used real leaderboards, and 11 used artificial leaderboards (Figure 3). Among interventions using real leaderboards, nine used absolute leaderboards (1, 2, 3, 10, 12, 17, 19). One intervention used relative leaderboards (1). One used a top five leaderboard (20). Three interventions (18) utilised leaderboards which presented only the user and indicated his/her rank (e.g., 70/180) and/or points (e.g., 60%). Four interventions used different types of leaderboards concurrently (6, 7, 8, 16). For instance, Ortiz-Rojas et al. (2019) used a leaderboard that displayed the first quartile of high-ranking students and relative leaderboards. Two interventions used special leaderboards for low-performing students (6, 7). For instance, Cigdem et al. (2023) presented both the top three and the bottom three students in quizzes.

Among interventions using artificial leaderboards, five used relative leaderboards (13, 14, 15). Two interventions used top 10 leaderboards with fabricated players (5). Four studies (seven interventions) examined how leaderboards' difficulty level (i.e., how difficult it is to rank high on leaderboards) influenced students' learning (4, 9, 13, 14). Two studies suggested that high difficulty could improve students' retention performance (13) or brainstorming task performance (9). In contrast, two studies indicated that difficulty level did not influence students' retention knowledge (14) or English vocabulary learning performance (4).

In terms of competitive mechanisms, most interventions ($n = 25$) used individual leaderboards (Figure 4). One used individual and team-based leaderboards simultaneously (3). One used individual leaderboards and a special type of team leaderboards which did not compare the performance of teams within class but compared the performance of different classes (6). Two interventions used team-based leaderboards (10, 19). For instance, in Lee et al. (2022), students were assigned to a four-member team and participated in management-related simulation games. The simulation activities

LEADERBOARD (real)		
Rank	**Username**	**Score**
1	Jenny	100
2	Mike	96
3	Kate	90
4	Tony	85
5	**Mary**	**80**
6	Laura	70
7	Rachel	65
8	Carol	62
9	Tony	55
10	Bob	50

Real leaderboards allow for comparisons between the actual scores of real participants. Please imagine the participant is Mary who is ranked 5th with a total score of 80, the leaderboard above is a real leaderboard if other students on the leaderboard are real participants with actual scores.

LEADERBOARD (artificial with high difficulty)		
Rank	**Username**	**Score**
1	Jack	100
2	Cindy	99
3	Dennis	97
4	Karl	95
5	John	90
6	Linda	88
7	Bill	86
8	Vivian	85
9	Robert	83
10	**Kitty**	**80**

LEADERBOARD (artificial with low difficulty)		
Rank	**Username**	**Score**
1	**Helen**	**80**
2	Kitty	75
3	Jane	72
4	David	70
5	Mia	65
6	Julie	60
7	Grace	55
8	Rita	50
9	Sunny	48
10	Gary	45

Artificial leaderboards manipulate comparisons using fictitious participants or selected competitors for certain research purposes. For example, a study can use artificial leaderboards to examine how leaderboards with different difficulty levels (high difficulty versus low difficulty) influence students' learning. Please imagine one participant, Kitty, got a total score of 80. For the leaderboard with high difficulty, the researchers can fabricate other nine participants with fake scores over 80 or intentionally select nine participants who got scores over 80. In this case, Kitty is only ranked 10th. Please imagine another participant, Helen, who also got a total score of 80. For the leaderboard with low difficulty, the researchers can fabricate other nine participants with fake scores below 80 or intentionally select nine participants who got scores below 80. In this case, Helen is ranked 1st though she got the same score as Kitty.

FIGURE 2 An example to illustrate the difference between real and artificial leaderboards.

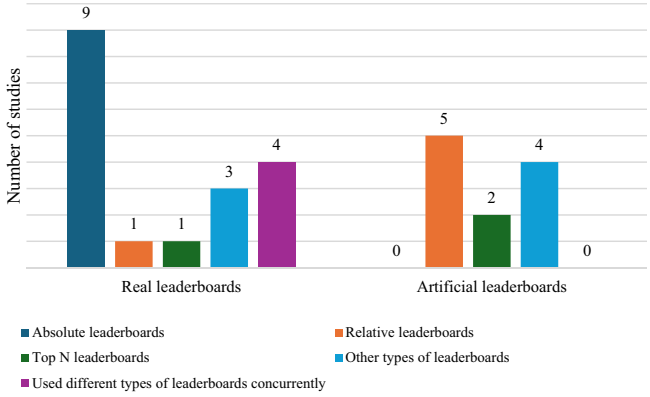

FIGURE 3 Leaderboard types (Real vs. Artificial) in the included studies ($N = 29$).

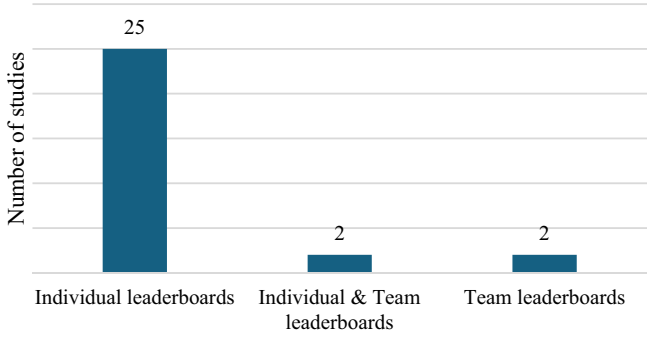

FIGURE 4 Leaderboard types (Individual vs. Team) in the included studies ($N = 29$).

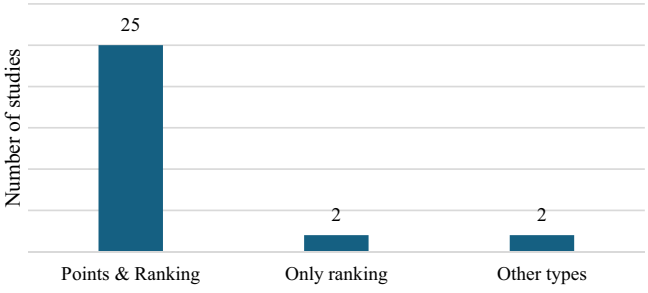

FIGURE 5 Whether leaderboards in the Included Studies presented points and/or ranking ($N = 29$). For other types, one intervention presented only points. Another intervention used one leaderboard with only ranking as well as one leaderboard with both ranking and points.

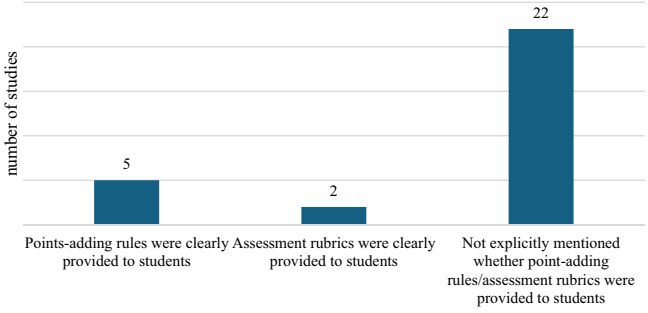

FIGURE 6 Whether points-adding rules/assessment rubrics were clearly provided to participants in the Included Studies ($N = 29$).

involved team discussions, making decisions, writing reflection reports and so on. Teams were ranked based on their cash balance in simulation activities.

6.2.2 | Points and ranking

Twenty-five interventions presented both participants' points and rankings (Figure 5). Lee et al. (2022) used leaderboards that showed students' ranking and their total cash balance, which resembled points in essence. Two interventions did not show points and only indicated ranking (16, 18). For instance, one intervention in Pi et al. (2023) showed only the participant's ranking (e.g., 70/180) without points. Pi et al. (2023) performed three leaderboard interventions (point/rank/point and rank condition) and examined specifically whether presenting points and/or ranking could influence students' learning. Their research suggested using leaderboards regardless of designs improved students' attention compared to the non-gamified control group, while only presenting ranking alone enhanced students' learning performance. They explained that leaderboards displaying much information (points and ranking simultaneously) could distract students' cognitive sources from learning tasks and subsequently impair their learning performance.

However, Mekler et al. (2017) found that presenting both points and ranking improved students' image annotation task quantity compared to showing only points, though this positive influence was not found in task quality. Additionally, two leaderboard interventions showed detailed records of students' performance/engagement besides their points and ranking (7, 17). For example, He and Loewen (2022) displayed students' study time and difficult words along with their points and ranking.

6.2.3 | Transparency of points-adding rules and assessment rubrics

Only five interventions explicitly mentioned that participants were informed of points-adding rules on leaderboards (1, 2, 17; see Figure 6). Bai et al. (2021) mentioned that clearly announcing the rules to participants before the intervention could support their perceived fairness. Balci et al. (2022) mentioned that explaining rules to participants could enable them to set goals and increase goal-related behaviours (e.g., ranking high on leaderboards) according to previous research (Sailer et al., 2017). Zahedi et al. (2021) investigated students' experiences of using leaderboards and reported that students could become confused by the scoring system and subsequently ignore it if it was not clear. Moreover, only two interventions mentioned that assessment rubrics for assigning points were provided to

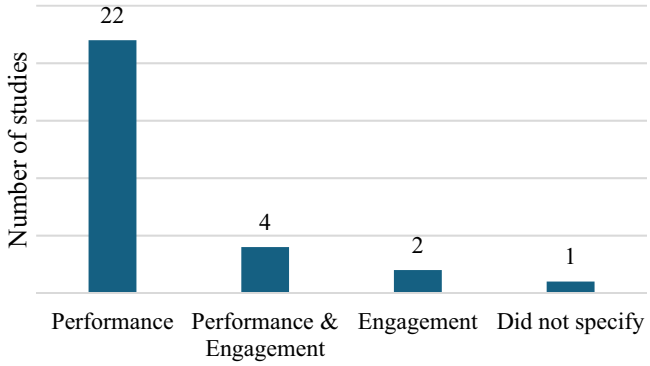

FIGURE 7 How participants were ranked in the included studies ($N = 29$).

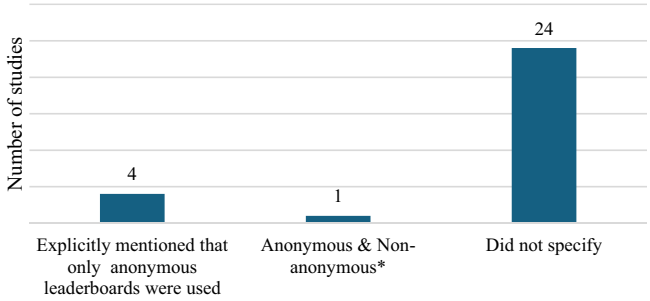

FIGURE 8 Anonymity of leaderboards in the included studies ($N = 29$). One intervention used the top 1/4 leaderboard that displayed students' real names and relative leaderboards which blurred the names of neighbouring students.

graders (3, 8). For instance, Bovermann and Bastiaens (2019) mentioned that assessment recommendations were available to students who conducted peer review.

6.2.4 | Performance and engagement

Leaderboard interventions normally ranked students based on their performance (learning results, e.g., how well students perform in a quiz/task) and/or their engagement (learning efforts, e.g., whether they complete a quiz/task).

Twenty-two interventions ranked students based on their performance only (Figure 7). Four interventions ranked students according to both their performance and engagement (7, 8, 17, 20). For example, Philpott and Son (2022) awarded points for both passing quizzes (performance) and reading books (engagement). Two interventions ranked students based on only engagement such as viewing forums and submitting assignments (1). For the interventions rewarding students' engagement, the primary rationale mentioned is that small, specific, and achievable activities that focused on participants' engagement could help build their self-efficacy and motivate students more effectively (1, 7, 8), which is supported by GST (Locke and Latham, 1990). TGL (Landers, 2014) also posits that leaderboards should be designed in a way that engagement/efforts (instead of ability) are the primary drivers of success, otherwise, challenging could be demotivating (8).

6.2.5 | Anonymity

Four interventions explicitly mentioned that they only used anonymous leaderboards (2, 17, 19; Figure 8). These interventions anonymised students by letting them self-create pseudonyms/codes or assigning pseudonyms for students. Among these interventions, only Balci et al. (2022) explained that anonymity was utilised because it could minimise the negative effects of leaderboards. Bai et al. (2021) investigated students' perceptions towards the anonymity of absolute and relative leaderboards qualitatively. Students who preferred public leaderboards thought that presenting real names could let them target competitors and increase the sense of competition. Absolute leaderboards displaying real names could identify top-ranked peers and let students consult them for help. Moreover, students ranked in the bottom third preferred anonymous comparison on an absolute leaderboard to save face but favoured public comparison on a relative leaderboard.

6.2.6 | Maximum points

Three interventions mentioned explicitly that leaderboards had a maximum point limit (1, 17). Bai et al. (2021) mentioned that the maximum points of leaderboards were 8000 but students were not informed of it. Philpott and Son (2022) used leaderboards that had a weekly maximum of 100 points to avoid an overwhelming discrepancy between high- and low-performing students. However, this study found that leaderboards encouraged performance up to the reward threshold and performance ceased once the maximum had been obtained.

6.2.7 | Repeatability

Seven interventions mentioned explicitly that students were allowed to repeat some tasks or quizzes (2, 6, 13, 14). Nebel et al. (2016) examined the influence of choice (i.e., whether students could repeat a level) and found that choice did not significantly influence students' learning performance but improved their motivation and emotions.

6.2.8 | Grading

Five interventions mentioned explicitly that student performance on leaderboards was not counted in students' final grades (1, 2, 8). Only Lee et al. (2022) emphasised that students' final ranking on leaderboards was counted in their course grades. Balci et al. (2022) suggested that feelings of indifference towards leaderboard performance could be because it was not counted in the final grading.

6.3 | Reported learning outcomes

After careful examination, this review found that three major learning outcomes (performance, motivation, and engagement) were reported when leaderboards were implemented. Students' perceptions towards leaderboard use are also reported in this section.

6.3.1 | Performance

Nineteen studies (25 interventions) discussed the impact of leaderboards on students' academic performance. Six interventions were reported to improve students' performance compared to non-gamification (8, 10, 11, 16, 18, 19). Across these interventions, two used absolute leaderboards (10, 19). One used an artificial five-place leaderboard (11), which seems to resemble an absolute leaderboard as the participant did not know the other four participants were fictional and only knew he/she was compared to other available participants. This is different from *Top N* leaderboards on which participants know they only see several best players or relative leaderboards on which participants know they are compared to their neighbours. Four interventions used individual leaderboards (8, 11, 16, 18), and two used team leaderboards (10, 19). Four ranked students based on their performance only (10, 11, 18, 19), and one ranked students based on their performance and engagement (8). Half of the interventions presented students' points and rankings (10, 11, 19).

Six interventions did not report statistically significant differences on students' performances compared to non-gamification (2, 6, 11, 18). Among these interventions, two used absolute leaderboards (2), and one used an artificial five-place leaderboard similar to an absolute leaderboard as previously discussed (11). Most interventions used individual leaderboards ($n = 5$). Moreover, all these interventions ranked students based on their performance only and most interventions used leaderboards showing both points and ranking ($n = 5$). No intervention was reported to negatively influence students' performance compared to non-gamification. The interventions which reported positive or nonsignificant effects on students' performance were found to have similar designs. Specifically, half of the interventions used absolute leaderboards or leaderboards resembling an absolute leaderboard. Most of the interventions used individual leaderboards, ordering students based on their performance only and presented both points and ranking.

Additionally, two studies examined moderators of leaderboard effects on students' performance (4, 9). Landers et al. (2017) found that goal commitment (attachment/determination to reach a goal; Locke and Latham, 1990) positively moderated leaderboard effects on students' task performance. Cao et al. (2022) examined whether goal orientation (performance orientation: adopting more social comparison standards; learning orientation: typically using absolute/self-reference standards) moderated leaderboard effects on students' performance. However, no statistically significant moderating effects were found on performance.

6.3.2 | Motivation

Fifteen studies (20 interventions) reported how leaderboards influenced students' motivation. Seven interventions were found to positively affect students' motivation compared to non-gamification (2, 3, 12, 17, 19). Among these interventions, two were reported to positively affect intrinsic motivation specifically (3, 19), one improved extrinsic motivation specifically (17), and the remaining four did not distinguish motivation types (2, 12). Interestingly, all interventions were found to use absolute leaderboards, among which four explicitly mentioned using anonymous leaderboards (2, 17, 19) and four explicitly mentioned that points-adding rules (2, 17) or assessment rubrics (3) were provided to students. Across the interventions, most used individual leaderboards ($n = 5$; 2, 12, 17). Six interventions ranked students based on their performance only (2, 3, 12, 19), and one ranked students based on both performance and engagement (17). All interventions displayed both points and ranking.

Five interventions were found to have no statistically significant effects on motivation compared to non-gamification (11, 16, 18). Among these interventions, two were reported not to influence intrinsic motivation specifically (11, 16), and three did not specify motivation types (18). Three interventions used leaderboards that presented only one participant whose ranking (e.g., 70/180) and/or points (60%) were presented (18). Only one intervention mentioned anonymous leaderboard was used (16) and no interventions mentioned explicitly that points-adding rules/assessment rubrics were provided. All interventions used individual leaderboards, and most interventions ranked students based on their performance only ($n = 4$; 11, 18). The interventions which reported positive or nonsignificant effects on motivation had some similar designs. Most interventions used individual leaderboards and ranked students based on only their performance. More importantly, some different design patterns were also identified. Specifically, interventions which reported positive effects on motivation were found to use absolute leaderboards, anonymous leaderboards, and provide points-adding rules/assessment rubrics more frequently than interventions reporting nonsignificant effects.

Additionally, Philpott and Son (2022) reported that they found leaderboards improved students' extrinsic motivation, but impaired their intrinsic motivation compared to non-gamification. Their findings contradict most studies which found that leaderboards could enhance (3, 19) or at least had nonsignificant effects on students' intrinsic motivation (11, 16). Philpott and Son (2022) suggested that according to SDT, if the rewards provided by leaderboards were perceived as controlling rather than informational, leaderboards could hinder students' intrinsic motivation. This is supported by the empirical findings reported by Mekler et al. (2017). Their study found that goal causality orientation moderated leaderboard effects on students' intrinsic motivation. Autonomy-oriented (tending to interpret external events as informational) participants had higher intrinsic motivation than control-oriented (more likely to perceive external events as pressuring) participants.

6.3.3 | Engagement

Nine studies (10 interventions) discussed the impact of leaderboards on students' engagement. Two interventions were reported to enhance engagement compared to non-gamification (8, 12). One intervention used absolute leaderboards (12), and one used the top three and one-versus-one leaderboards (8). Two interventions used individual leaderboards. One intervention ranked students based on their performance only (12), while one ranked students based on their performance and engagement (8).

Three interventions were found not to have statistically significant effects on engagement in contrast to non-gamification (6, 16, 17). One intervention used absolute leaderboards (17), and two used *Top N* leaderboards along with the bottom three leaderboards (6) or relative leaderboards (16). Two interventions used individual leaderboards (16, 17), and one used individual leaderboards and team (class) leaderboards (6). Additionally, the interventions ranked students based on their performance only (6), or performance and engagement (17) or did not specify (16). The interventions which reported positive or nonsignificant effects on engagement were also found to have similar designs. Specifically, interventions typically used absolute leaderboard/top N leaderboard along with other leaderboard types, individual leaderboards, and ranked students based on only their performance/performance and engagement. No study was found reporting a detrimental impact of leaderboards on students' engagement compared to non-gamification.

Additionally, one study examined how goal setting with feedback influenced students' learning (7). They implemented one treatment group that used leaderboards only and another treatment group which utilised leaderboards with goal-setting-and-checking activities (goal setting with feedback) for 8 weeks. Goal-setting-and-checking activities required students to decide their study goals (e.g., learning how many words per day). Students needed to check whether they achieved their prior goal and if not, they needed to explain the reasons. Their research found that using leaderboards with goal-setting-and-checking activities improved students' engagement compared to the leaderboard-only group, though this positive influence was not found in performance.

6.3.4 | Perceptions

Eight studies reported students' perceptions of using leaderboards qualitatively (1, 2, 3, 7, 17, 20). Of these, students generally had positive perceptions towards leaderboards (seven studies; 1, 2, 3, 7, 17), reporting that leaderboards supported increased motivation to work hard (seven studies; 1, 2, 3, 7, 17), setting goals and/or monitoring their learning process by providing feedback (six studies; 1, 2, 7, 17). Four studies reported that students engaged in upward comparisons more frequently than downward comparisons (1, 2). However, five studies also reported some negative opinions towards leaderboards (2, 3, 7, 17). One study reported that students felt discouraged when comparing ranking with other students (2), and one study found that some students felt negative towards not being able to improve their rank (17). Two studies reported that students did not care about leaderboards (3, 17).

Additionally, Bai et al. (2021) examined students' perceptions towards absolute and relative leaderboards. Their research suggested that absolute leaderboards seemed to promote a sense of competition more intensively than relative leaderboards. Students on absolute leaderboards generally reported they made continuous efforts to rank higher after viewing their positions. In contrast, students on relative leaderboards reported that they made little effort to improve their ranks and tended to be satisfied with their current ranking. Bovermann and Bastiaens (2019) reported students' perceptions towards team leaderboards. Students generally reported team leaderboards could promote their team spirit and collaborative learning, whereas some students reported it was sometimes exhausting to achieve a group result for peer review.

6.4 | Effects of specific leaderboard design practices on students' learning outcomes

The identified interventions were generally reported to have positive or nonsignificant effects on students' learning performance, motivation and engagement compared to non-gamification. The inconsistent findings could be ascribed to the leaderboards with different designs being implemented. These empirical studies indicate that leaderboards can positively influence students' learning, but their effectiveness largely hinges on their designs. Hence, this review did not highlight the overall leaderboard effects on students' learning outcomes regardless of their designs and instead focused on the effects of specific leaderboard design practices on student's learning outcomes.

6.4.1 | Leaderboard types

All the interventions reporting positive effects on learning motivation were found to use absolute leaderboards. As for performance and engagement, only half of the interventions used absolute leaderboards. Hence, this review suggests that using absolute leaderboards might be more effective for promoting motivation than relative or *Top N* leaderboards, but this finding might not generalise to engagement and performance due to the limited empirical evidence.

6.4.2 | Anonymity

This review found that leaderboard interventions reporting positive effects on learning motivation tended to use anonymous absolute leaderboards than interventions reporting nonsignificant effects. This indicates that using absolute leaderboards, especially when participants were anonymised, may positively affect their learning motivation.

6.4.3 | Performance and engagement

This review found that most interventions reporting a positive/nonsignificant influence on students' learning performance, motivation, and/or engagement ranked students based on performance only. None of the interventions ranking students based on performance only reported negative effects on students' learning. Thus, this review suggests that using leaderboards ranking students based on performance exclusively, might be the "safest" way to improve learning. Two interventions ranked students merely based on engagement (1), but they did not examine leaderboard effects on learning compared to non-gamification and the findings were thus not discussed here. Four interventions ranked students based on both performance and engagement, which generally reported positive effects on learning performance, motivation, engagement or positive perceptions towards leaderboards. Notably, this review does not suggest that ranking students based exclusively on performance is the most effective way to improve learning. This review argues that ranking students based on both performance and engagement instead of only performance might improve learning most effectively based on the aforementioned empirical evidence and pertinent theories (see Section 7.2.3).

6.4.4 | Transparency of points-adding rules and assessment rubrics

Five interventions which mentioned explicitly that points-adding rules/ assessment rubrics were provided to students examined leaderboard effects compared to non-gamification and all of them positively affected students' learning motivation/performance. Moreover, Zahedi et al. (2021) reported that students became confused or disappointed and thus ignored the leaderboards when they felt the scoring system was unclear. Given the empirical evidence, this review suggests that providing students with points-adding rules/ assessment rubrics may positively influence their motivation and performance.

7 | DISCUSSION

7.1 | General characteristics of the included studies

This review found that all included studies were implemented in higher education. This concurs with previous review findings that most gamification studies were conducted in tertiary education (e.g., Zainuddin et al., 2020), which is likely the result of convenience sampling (Ritzhaupt et al., 2021). Moreover, prior meta-analyses found educational levels influenced the effectiveness of gamification yet reported inconsistent findings (e.g. Huang et al., 2020; Sailer and Homner, 2020). For example, Sailer and Homner (2020) found that gamification was more effective for school-aged learners than students in tertiary education, whereas Huang et al. (2020) found that gamification was more effective for undergraduates than for school-aged students. The difference could be attributed to different developmental characteristics and educational expectations across educational contexts (Johnson et al., 2023; Knowles, 1984; Li et al., 2024). Future research is thus highly recommended to focus on other educational levels (e.g., primary/secondary schools) to examine the generalisability of this review's research findings.

Moreover, this review found that the included studies were conducted in various geographic locations. A few included studies cautioned that culture could influence the effectiveness of leaderboards (1, 2, 14, 15). Specifically, individuals from East Asian cultures are less likely to take a positive view of themselves than those from Western cultures and thus might have more negative feelings when ranking low on leaderboards (Bai et al., 2021). In line with this, this review found that out of six interventions reporting positive effects on performance, four were conducted in Western culture (e.g., Germany, USA), only one was performed in East Asian culture (China) and one did not specify the location. Future research might consider comparing the effectiveness of leaderboards with the same designs in East Asian and Western cultures directly and propose culture-oriented leaderboard design recommendations (Bai et al., 2021; Hofstede, 2001).

As for instructional design, this review found that out of six interventions reporting a positive influence on performance, five used micro-level designs and one used macro-level design (16). Moreover, out of seven interventions reporting a positive influence on motivation, six used micro-level designs, while one used macro-level design (17). The two interventions reporting a positive impact on students' engagement all used micro-level designs. This seemingly suggested that leaderboards designed on a micro-level could be effective for students' learning outcomes (i.e., motivation, engagement and performance). However, whether macro-level leaderboards are effective for students' learning remains unclear due to the very small number of macro-level leaderboard interventions identified. Moreover, whether micro-level or macro-level leaderboards are more effective is also unclear. Future research should consider using macro-level leaderboards and compare the effectiveness of leaderboards on a micro and macro level directly.

Furthermore, prior reviews suggested that gamification was used in various disciplines such as social science, science, health, business, arts, humanities (Bai et al., 2020; Huang et al., 2020) and was most effective for social science and engineering (e.g., Huang et al., 2020; Ritzhaupt et al., 2021). Similarly, this review found that leaderboards were also used for various disciplines. However, this review only found that leaderboards improved learning across subjects (e.g., English, management, computer science, arts, engineering) and could not identify leaderboards worked most effectively for what disciplines due to the limited number of studies. Future research should conduct more leaderboard research in different subjects. With enough empirical evidence, meta-analyses could examine the effectiveness of leaderboards across subjects, comparing results to holistic gamification.

Apart from the contextual factors (educational level, culture, instructional design, and subjects), three studies have examined how

individual differences (goal commitment, goal orientation, and causality orientation) could moderate leaderboard effects (4, 9, 11). Given the limited empirical evidence, future research is suggested to further validate the reported influence of the three moderators on students' learning. Moreover, the included studies cautioned other learner characteristics (e.g., prior knowledge, gender, level of competition) may also influence leaderboard effects (1, 18). Future research could also examine the effects of these moderators to gain a more in-depth understanding of how leaderboards affect the learning of different individuals.

As for duration, although this review found that around half of the interventions lasted less than or equal to 1 h, most interventions reporting positive effects on students' learning lasted for longer durations (generally 10 weeks or above). Specifically, six out of seven interventions reporting positive effects on motivation lasted for at least one semester, and both interventions reporting positive effects on engagement lasted for 10 weeks/one semester. Half of the interventions reporting positive effects on performance lasted for 2, 4, and 10 weeks respectively. This indicates relatively reliable evidence that leaderboards with appropriate designs could improve students' motivation, engagement, and performance on the scale of weeks/months. However, only one study was identified to implement leaderboards for as long as 1 year (3). This study used micro-level leaderboards (e.g., providing individual leaderboards after students complete quizzes on Kahoot, a gamified learning platform) for different seminars across 1 year. Moreover, no studies were identified to directly examine whether the novelty effect might be present when using leaderboards: a pattern of heightened engagement when a novel phenomenon (e.g., gamification) is introduced followed by a drop in engagement once the novelty wears off (Tsay et al., 2019). Future longitudinal studies with longer durations (e.g., 1 year or above) are thus highly recommended to further examine the sustainability of learning outcomes and potential novelty effects in leaderboard use (e.g., comparing the effectiveness of the same leaderboards for the same students after hours/weeks/months/years).

7.2 | Effects of specific leaderboard design practices on students' learning outcomes and corresponding design recommendation

In this section, relevant theories are discussed alongside the empirical findings in Section 6.4 to propose corresponding leaderboard design recommendations. The design recommendations are proposed based on empirical evidence in higher education, so their generalisability to other educational levels needs to be further examined.

7.2.1 | Leaderboard types

The finding that using absolute leaderboards might be more effective for promoting motivation than relative or *Top N* leaderboards could be explained by SCT (Balci et al., 2022; Festinger, 1954) and GST (Landers et al., 2017; Locke et al., 1981). Compared to relative leaderboards, both absolute and *Top N* leaderboards can provide students with opportunities for upward comparison and higher goals, which could better motivate students according to the two theories (Festinger, 1954; Locke et al., 1981). However, the upward comparisons provided by *Top N* leaderboards could bring too large negative discrepancies and the displayed goals could be too difficult for some students as *Top N* leaderboards only present a few best-ranking participants. Too extreme negative discrepancy or too difficult goals could not motivate students effectively according to SCT (Diel et al., 2021) and GST (Locke et al., 1981). In contrast, absolute leaderboards display the ranking of all participants and provide participants with opportunities to set goals and perform upward comparisons with others whose achievement would require moderate difficulty to attain, which could positively affect students' learning (Festinger, 1954; Locke et al., 1981).

Considering the empirical evidence and related theories, this review proposes *Design Recommendation 1. Absolute leaderboard may come first*: *Using absolute leaderboards to improve learning motivation*. This finding that absolute leaderboards might be more beneficial to learning motivation was obtained based on the induction of the empirical studies available. No studies were identified comparing the influence of absolute, relative, and *Top N* leaderboards directly. To validate this finding, future research is thus recommended to implement the three types of leaderboards separately and compare their effects on motivation directly. Moreover, improved motivation is considered a precursor to enhanced engagement and performance (Zainuddin et al., 2020). Hence, future research is also encouraged to examine whether absolute leaderboards could better improve learning engagement and performance besides motivation.

7.2.2 | Anonymity

This review indicates that using absolute leaderboards, especially when participants were anonymised, may positively affect their learning motivation. In line with this finding, Bai et al. (2021) reported that low-ranking students preferred anonymity on absolute leaderboards to preserve their self-esteem. This finding can be supported by SDT (Ryan and Deci, 2000). Presenting low-performing students publicly might have more detrimental effects on their competence need satisfaction compared to anonymising these students, which can subsequently harm their motivations. Therefore, this review proposes *Design Recommendation 2. Anonymise low-ranking students*: *Absolute leaderboards should allow students, especially low-ranking students, to use pseudonyms to improve their learning motivation*. Moreover, as previously discussed, motivation could positively influence learning engagement and performance (Zainuddin et al., 2020). Future research is also suggested to examine whether absolute leaderboards which allow low-ranking students to be anonymous could better support students' learning engagement and performance in addition to motivation than public absolute leaderboards.

7.2.3 | Performance and engagement

This review argues that ranking students based on both performance and engagement might improve learning most effectively. This argument can be supported by relevant theories. GST postulates that goals with moderate challenges can motivate students better than too easy or too difficult goals (Locke et al., 1981). In line with the theory, a few empirical studies found that tasks deemed both accomplishable and challenging could better motivate participants (e.g., Abuhamdeh and Csikszentmihalyi, 2012). Ranking students only based on their performance might bring overwhelming challenges for low-performing students and undermine their motivations (Landers and Landers, 2014). Landers and Landers's (2014) TGL posits that leaderboards should emphasise students' engagement rather than their performance so leaderboards could be motivating as participants have equal chances of getting on leaderboards if they pay a similar level of effort. However, primarily focusing on engagement (e.g., number of times a user report is checked), as suggested by TGL, could provide little challenge for students with higher abilities, which might not support their motivation effectively according to GST (Locke et al., 1981) and previous empirical studies (Abuhamdeh and Csikszentmihalyi, 2012).

Given the pertinent theories and empirical studies, this review proposes *Design Recommendation 3*. *Performance and engagement leaderboards*: *Leaderboards that focus on both engagement and performance might provide a more optimal level of challenge to support learning*. This recommendation is consistent with Park and Kim (2021)'s design principle 3: leaderboards should incorporate "non-learning" (engagement) and "learning" (performance) activities. However, it is noteworthy that despite sufficient theoretical guidance, the empirical evidence that supports using performance and engagement leaderboards is relatively scarce. No studies were identified comparing students' perceived challenge level of performance/engagement/performance and engagement leaderboards and their effects. Future research is suggested to implement more performance and engagement leaderboards and compare how these three leaderboard types affect students' perceived challenge levels and learning.

7.2.4 | Transparency of points-adding rules and assessment rubrics

This review suggests that providing students with points-adding rules/assessment rubrics may positively influence their motivation and performance. This finding can also be supported by relevant theories. GST proposes that specific goals could motivate participants more effectively than vague goals (Locke et al., 1981). Landers and Landers's (2014) TGL also suggests that gamification rules/goals should be clearly defined. Many empirical studies have demonstrated the benefits of setting clear goals compared to ill-defined goals (Höpfner and Keith, 2021; Locke and Latham, 2006).

Moreover, He and Loewen (2022) found that using leaderboards with goal-setting-and-checking activities (whether students achieved goals set last week and if not, students needed to explain the reasons) promoted students' engagement more effectively than only using leaderboards after 8 weeks. Their research was based on Mercer and Dörnyei's (2020)'s proposal of goal-setting with feedback, which further developed classic goal-setting strategies (Dörnyei, 2001; Locke et al., 1981). Many previous studies have also demonstrated that setting specific goals with goal-related feedback (feedback that highlights the discrepancy between current and goal performances, and supports students on reducing that discrepancy) fostered learning motivation, engagement, and performance (Hattie and Timperley, 2007; Locke and Latham, 2015). Based on He and Loewen (2022)'s empirical study, relevant theoretical rationales and substantial previous research, this review thus suggests that providing leaderboards with goal-related feedback could foster students' learning. In conclusion, this review proposes *Design Recommendation 4*. *Clear rules with goal-related feedback*: *Leaderboards should provide clear scoring rules with goal-related feedback to foster learning*.

7.3 | Theoretical implications

This review makes theoretical contributions by examining the empirical studies regarding leaderboard use through the lens of key educational and motivation theories (i.e., GST, SCT, SDT, TGL). Section 6.4 discusses the empirical findings concerning the effects of specific leaderboard design practices on students' learning outcomes. Section 7.2 discusses how the empirical findings are aligned with the theories and proposes the corresponding design recommendations. The theories provide theoretical support for the empirical findings and design recommendations. Conversely, the empirical findings provide empirical support for principles within the theories.

Apart from the leaderboard design practices discussed in Section 6.4, this review identified a few other leaderboard design practices whose effects on students' learning remained unclear due to limited empirical evidence or theoretical guidance. Specifically, the unclear design practices included whether individual/team leaderboards are more effective, what information should be presented on leaderboards (e.g., points, ranking, points and ranking, points and ranking and other detailed information), whether leaderboards should have maximum points, whether leaderboards should allow students to repeat some tasks, and whether students' leaderboard performance should be counted into final grading. Future research is recommended to conduct more empirical research to identify effective leaderboard design practices and use relevant theories to explain the findings. Future research may also consider conducting theory-driven research to examine the possible effective design practices indicated by theories. One special caution is raised when conducting relevant empirical research. As motivation is a psychological construct of breadth and complexity (Pi et al., 2023), research is suggested to specify motivation types (e.g., intrinsic/extrinsic motivation). For example, according to SDT, team leaderboards may better satisfy students' need for relatedness than individual leaderboards because team members may feel more related to each other because they have shared goals (Ryan and

Deci, 2000; Sailer et al., 2017), which may better support students' intrinsic motivation. Future research could compare whether team leaderboards are more conducive to students' intrinsic motivation than individual leaderboards as suggested by SDT.

This review only focused on the most commonly used theories (i.e., GST, SCT, SDT, and TGL) in the existing leaderboard research. The included studies also referred to other theories/models to underpin their investigation, which included self-identity theory, achievement goal theory, control-value theory, cognitive load theory, feedback intervention theory, theory of relative performance feedback, elaboration retrieval theory, and ARCS model of motivation. Future research is suggested to further examine leaderboard use from the perspectives of different theories identified rather than being confined to the four theories highlighted in this review. Future research is also suggested to explore other theories which can be used to address leaderboard use in education.

8 | LIMITATIONS AND FUTURE DIRECTIONS

First, despite careful searching, this review only obtained 20 articles from 2014 to 2023. This indicated that examining the individual use of leaderboards is a nascent research field and more studies are greatly needed to gain a more in-depth understanding of leaderboard use in education. Second, all included studies were implemented in tertiary education. Future research is highly recommended to implement leaderboards in other educational levels to examine the generalisability of the current findings. Third, around half of the interventions lasted for less than or equal to 60 min, only one intervention lasted for as long as 1 year, and research has yet to extensively examine whether novelty effects might exist in leaderboard use. Future research should conduct more longitudinal investigations and compare leaderboard effects after hours/weeks/months/years to further examine potential novelty effects and validate long-term effects of leaderboards. Notably, despite around half of the interventions had very short durations, most interventions reporting positive effects on students' learning lasted for longer durations (normally 10 weeks or above). Thus, this review still provides relatively reliable evidence that leaderboards designed appropriately could foster students' learning in weeks/months. Lastly, although the four design recommendations are raised based on robust theories and related empirical evidence with generally high quality, more empirical studies are called for to further validate these design recommendations in different contexts. For recommendations one and three in particular, future researchers are advised to directly compare the effectiveness of different leaderboard types (absolute/relative/*Top N*; performance/engagement/performance and engagement).

9 | CONCLUSIONS

This review is the first research synthesis that inspected existing empirical evidence regarding leaderboard use in education. This review contributes to the literature by providing an overview concerning leaderboard research in education, identifying the effects of leaderboard design practices on students' learning, proposing a set of leaderboard design recommendations, and ascertaining future research avenues. This review indicates that using leaderboards can have a beneficial influence on students' learning motivation, engagement, and performance. However, leaderboards are not a panacea for learning and their effectiveness largely depends on their designs. Based on well-established theories and relevant empirical evidence, this review proposes a set of leaderboard design recommendations. These design recommendations suggest insights for instructors and researchers who want to implement leaderboards to foster students' learning. Moreover, this review found that all identified studies were conducted in higher education and around half of them had short durations (i.e., less or equal to 1 h). More longitudinal research at other educational levels is therefore called for to examine the validity and generalisability of the design recommendations proposed. Additionally, future research is suggested to explore other effective leaderboard design recommendations and further develop the four rudimentary design recommendations presented here into a fully-fledged leaderboard design framework alongside other recommendations identified in the future.

AUTHOR CONTRIBUTIONS

Chunqi Li: Conceptualization; methodology; data curation; writing – original draft; formal analysis; validation. **Lishi Liang:** Writing – review and editing; formal analysis; visualization; validation. **Luke K. Fryer:** Supervision; writing – review and editing; conceptualization; methodology; validation. **Alex Shum:** Validation; supervision; writing – review and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Chunqi Li https://orcid.org/0000-0003-4402-1631
Lishi Liang https://orcid.org/0009-0000-3513-9051
Luke K. Fryer https://orcid.org/0000-0001-6250-5950
Alex Shum https://orcid.org/0000-0001-6223-787X

REFERENCES

Abuhamdeh, S., & Csikszentmihalyi, M. (2012). The importance of challenge for the enjoyment of intrinsically motivated, goal-directed activities. *Personality and Social Psychology Bulletin*, *38*(3), 317–330. https://doi.org/10.1177/0146167211427147

Aldemir, T., Celik, B., & Kaplan, G. (2018). A qualitative investigation of student perceptions of game elements in a gamified course. *Computers in Human Behavior*, *78*, 235–254. https://doi.org/10.1016/j.chb.2017.10.001

Bai, S., Hew, K., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of

qualitative data in educational contexts. *Educational Research Review*, *30*, 100322. https://doi.org/10.1016/j.edurev.2020.100322

Bai, S., Hew, K. F., Sailer, M., & Jia, C. (2021). From top to bottom: How positions on. Different types of leaderboard may affect fully online student learning performance, intrinsic motivation, and course engagement. *Computers and Education*, *173*, 104297. https://doi.org/10.1016/j.compedu.2021.104297

Balci, S., Secaur, J. M., & Morris, B. J. (2022). Comparing the effectiveness of badges and leaderboards on academic performance and motivation of students in fully versus partially gamified online physics classes. *Education and Information Technologies*, *27*(6), 8669–8704. https://doi.org/10.1007/s10639-022-10983-z

Bovermann, K., & Bastiaens, T. (2019). How gamification can foster motivation and collaboration in blended learning: A mixed methods case study. *Journal of Interactive Learning Research*, *30*(3), 275–300.

Burgers, C., Eden, A., van Engelenburg, M. D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior*, *48*, 94–103. https://doi.org/10.1016/j.chb.2015.01.038

Cao, Y., Gong, S., Wang, Z., Cheng, Y., & Wang, Y. (2022). More challenging or more achievable? The impacts of difficulty and dominant goal orientation in leaderboards within educational gamification. *Journal of Computer Assisted Learning*, *38*(3), 845–860. https://doi.org/10.1111/jcal.12652

Chan, C. K. Y., & Chen, S. W. (2022). Students' perceptions on the recognition of holistic competency achievement: A systematic mixed studies review. *Educational Research Review*, *35*, 100431. https://doi.org/10.1016/j.edurev.2021.100431

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

Chou, Y. (2015). *Actionable gamification: Beyond points, badges, and leaderboards*. Octalysis Media.

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

Cigdem, H., Korkusuz, M. E., & Karaçaltı, C. (2023). Gamified learning: Assessing the influence of leaderboards on online formative quizzes in a computer programming course. *Computer Applications in Engineering Education*, *32*, e22697. https://doi.org/10.1002/cae.22697

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

Diel, K., Grelle, S., & Hofmann, W. (2021). A motivational framework of social comparison. *Journal of Personality and Social Psychology*, *120*(6), 1415–1430. https://doi.org/10.1037/pspa0000204

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge University Press.

Evangelio, C., Rodríguez-González, P., Fernández-Río, J., & Gonzalez-Villora, S. (2022). Cyberbullying in elementary and middle school students: A systematic review. *Computers and Education*, *176*, 104356. https://doi.org/10.1016/j.compedu.2021.104356

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Flinton, D., Khine, R., Mannion, L., O'Sullivan, C., & Cherry, P. (2023). Gamification in radiotherapy education: Adopting competitive task elements in simulation using the virtual environment of a radiotherapy treatment room (VERT) system. *Journal of Radiotherapy in Practice*, *22*, 1–6. https://doi.org/10.1017/S1460396923000262

Garcia, S. M., Tor, A., & Schiff, T. M. (2013). The psychology of competition: A social comparison perspective. *Perspectives on Psychological Science*, *8*(6), 634–650. https://doi.org/10.1177/1745691613504114

Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, *144*(2), 177–197. https://doi.org/10.1037/bul0000127

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

He, X., & Loewen, S. (2022). Stimulating learner engagement in app-based L2 vocabulary self-study: Goals and feedback for effective L2 pedagogy. *System*, *105*, 102719. https://doi.org/10.1016/j.system.2021.102719

Hofstede, G. (2001). *Culture's consequences : Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Sage Publications.

Höllig, C. E., Tumasjan, A., & Welpe, I. M. (2020). Individualizing gamified systems: The role of trait competitiveness and leaderboard design. *Journal of Business Research*, *106*, 288–303. https://doi.org/10.1016/j.jbusres.2018.10.046

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Hong, Q. N., Gonzalez-Reyes, A., & Pluye, P. (2018). Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the mixed methods appraisal tool (MMAT). *Journal of Evaluation in Clinical Practice*, *24*(3), 459–467. https://doi.org/10.1111/jep.1288

Höpfner, J., & Keith, N. (2021). Goal missed, self hit: Goal-setting, goal-failure, and their affective, motivational, and behavioral consequences. *Frontiers in Psychology*, *12*, 704790. https://doi.org/10.3389/fpsyg.2021.704790

Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development*, *68*(4), 1875–1901. https://doi.org/10.1007/s11423-020-09807-z

Jia, C., & Hew, K. F. (2021). Toward a set of design principles for decoding training: A. Systematic review of studies of English as a foreign/second language listening education. *Educational Research Review*, *33*, 100392. https://doi.org/10.1016/j.edurev.2021.100392

Johnson, C. C., Walton, J. B., Strickler, L., & Elliott, J. B. (2023). Online teaching in K-12 education in the United States: A systematic review. *Review of Educational Research*, *93*(3), 353–411. https://doi.org/10.3102/00346543221105550

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. Pfeiffer.

Knowles, M. S. (1984). *Andragogy in action. Applying modern principles of adult education*. Jossey Bass.

Lam, C. F., DeRue, D. S., Karam, E. P., & Hollenbeck, J. R. (2011). The impact of feedback frequency on learning and task performance: Challenging the "more is better" assumption. *Organizational Behavior and Human Decision Processes*, *116*, 217–228. https://doi.org/10.1016/j.obhdp.2011.05.002

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

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

Landers, R. N., Bauer, K. N., Callan, R. C., & Armstrong, M. B. (2015). Psychological theory and the gamification of learning. In T. Reiners & L. Wood (Eds.), *Gamification in education and business* (pp. 165–186). Springer.

Landers, R. N., & Landers, A. K. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation and Gaming*, *45*(6), 769–785. https://doi.org/10.1177/1046878114563662

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Lee, B. W., Lee, M. K., & Fredendall, L. (2022). Response to relative performance feedback in simulation games. *The International Journal of Management Education*, *20*(3), 100698. https://doi.org/10.1016/j.ijme.2022.100698

Li, C. and Fryer, L. K. (2022). The use of points, badges, and leaderboards individually or concurrently in gamified educational settings: A systematic review protocol. https://doi.org/10.17605/OSF.IO/BXJ7Z

Li, C., Fryer, L. K., & Chu, S. K. W. (2024). Utilising gamified formative assessment to support English language learning in schools: A scoping review. *Innovation in Language Learning and Teaching*, *1–17*, 1–17. https://doi.org/10.1080/17501229.2024.2352789

Lo, C. K., Hew, K. F., & Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: A synthesis of research in mathematics education. *Educational Research Review*, *22*, 50–73. https://doi.org/10.1016/j.edurev.2017.08.002

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Prentice Hall.

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*, 705–717. https://doi.org/10.1037/0003-066X.57.9.705

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science*, *15*(5), 265–268. https://doi.org/10.1111/j.1467-8721.2006.00449.x

Locke, E. A., & Latham, G. P. (2015). Breaking the rules: A historical overview of goal-setting theory. *Advances in Motivation Science*, *2*, 99–126. https://doi.org/10.1016/bs.adms.2015.05.001

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

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

Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms*. Cambridge University Press.

Michinov, N., & Primois, C. (2005). Improving productivity and creativity in online groups through social comparison process: New evidence for asynchronous electronic brainstorming. *Computers in Human Behavior*, *21*(1), 11–28. https://doi.org/10.1016/j.chb.2004.02.004

Nadeem, M., Oroszlanyova, M., & Farag, W. (2023). Effect of digital game-based learning on student engagement and motivation. *Computers*, *12*(9), 177. https://doi.org/10.3390/computers12090177

Nebel, S., Beege, M., Schneider, S., & Rey, G. D. (2016). The higher the score, the higher the learning outcome? Heterogeneous impacts of leaderboards and choice within educational videogames. *Computers in Human Behavior*, *65*, 391–401. https://doi.org/10.1016/j.chb.2016.08.042

Nebel, S., Beege, M., Schneider, S., & Rey, G. D. (2020). Competitive agents and adaptive difficulty within educational video games. *Frontiers in Education*, *5*, 1–14. https://doi.org/10.3389/feduc.2020.00129

Nebel, S., Schneider, S., Beege, M., & Rey, G. D. (2017). Leaderboards within educational videogames: The impact of difficulty, effort and gameplay. *Computers and Education*, *113*, 28–41. https://doi.org/10.1016/j.compedu.2017.05.011

Ortiz-Rojas, M., Chiluiza, K., & Valcke, M. (2019). Gamification through leaderboards: An empirical study in engineering education. *Computer Applications in Engineering Education*, *27*(4), 777–788. https://doi.org/10.1002/cae.12116

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J., & Seller, R. (2012). Testing the reliability and efficiency of the pilot mixed methods appraisal tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, *49*(1), 47–53. https://doi.org/10.1016/j.ijnurstu.2011.07.002

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 1–11. https://doi.org/10.1186/s13643-021-01626-4

Park, S., & Kim, S. (2021). Leaderboard design principles to enhance learning and motivation in a gamified educational environment: Development study. *JMIR Serious Games*, *9*(2), e14746. https://doi.org/10.2196/14746

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing.

Philpott, A., & Son, J. B. (2022). Leaderboards in an EFL course: Student performance and motivation. *Computers and Education*, *190*, 104605. https://doi.org/10.1016/j.compedu.2022.104605

Pi, Z., Yu, Q., Zhang, Y., Li, Y., Chen, H., & Yang, J. (2023). Presenting points or rank: The impacts of leaderboard elements on English vocabulary learning through video lectures. *Journal of Computer Assisted Learning*, *40*(1), 104–117. https://doi.org/10.1111/jcal.12871

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

Ritzhaupt, A. D., Huang, R., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., & Li, J. (2021). A meta-analysis on the influence of gamification in formal educational settings on affective and behavioral outcomes. *Educational Technology Research and Development*, *69*(5), 2493–2522. https://doi.org/10.1007/s11423-021-10036-1

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

Ryan, R. M., & Rigby, C. S. (2019). Motivational foundations of game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 153–176). MIT Press.

Sailer, M., 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

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, *32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Sailer, M., & Sailer, M. (2021). Gamification of in-class activities in flipped classroom. Lectures. *British Journal of Educational Technology*, *52*(1), 75–90. https://doi.org/10.1111/bjet.12948

Sanchez, D. R., Langer, M., & Kaur, R. (2020). Gamification in the classroom: Examining the impact of gamified quizzes on student learning. *Computers and Education*, *144*, 1–16. https://doi.org/10.1016/j.compedu.2019.103666

Schürmann, L., & Quaiser-Pohl, C. (2022). Digital badges affect need satisfaction but not frustration in males in higher education. *Computers and Education*, *182*, 104484. https://doi.org/10.1016/j.compedu.2022.104484

Tsay, C. H., Kofinas, A. K., Trivedi, S. K., & Yang, Y. (2019). Overcoming the novelty. Effect in online gamified learning systems: An empirical evaluation of student engagement and performance. *Journal of Computer Assisted Learning*, *36*(2), 128–146. https://doi.org/10.1111/jcal.12385

Zahedi, L., Batten, J., Ross, M., Potvin, G., Damas, S., Clarke, P., & Davis, D. (2021). Gamification in education: A mixed-methods study of gender on computer science students' academic performance and identity

development. *Journal of Computing in Higher Education*, *33*(2), 441–474. https://doi.org/10.1007/s12528-021-09271-5

Zainuddin, Z., Chu, S., Shujahat, M., & Perera, C. (2020). The impact of gamification. On learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, *30*, 100326. https://doi.org/10.1016/j.edurev.2020.100326

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

How to cite this article: Li, C., Liang, L., Fryer, L. K., & Shum, A. (2024). The use of leaderboards in education: A systematic review of empirical evidence in higher education. *Journal of Computer Assisted Learning*, *40*(6), 3406–3442. https://doi.org/10.1111/jcal.13077

APPENDIX A

TABLE A1 General characteristics of included studies.

No.	Author(s) and year	Location	Instructional design	Subject	Research paradigm	Duration	Platform/ application	Quality assessment
1	Bai et al. (2021)[a]	East Asia	Macro	Courses about technology-enhanced learning	Mixed-methods	One semester (10 weeks)	Model with a plugin "level-up"	High
2	Balci et al. (2022)[a]	USA	Micro[b]	Physics	Mixed-methods	One semester	Blackboard Learn System	High
3	Bovermann and Bastiaens (2019)	NA	Micro[b]	NA (Six seminars)	Mixed-methods	1 year	Kahoot	Medium
4	Cao et al. (2022)[a]	China	Micro[c]	English	Quantitative	30 min	Self-developed platform/app	High
5	Christy and Fox (2014)[a]	Midwestern university	Micro[c]	Math	Quantitative	15 min	Self-developed platform/app	High
6	Cigdem et al. (2023)	Turkey	Micro[b]	Computer science	Quantitative	One semester	The university's LMS	High
7	He and Loewen (2022)	Japan	Micro[b]	English	Mixed-methods	8 weeks	Memrise	High
8	Landers and Landers (2014)	USA	Micro[b]	Psychology	Quantitative	One semester	Web-based wiki	High
9	Landers et al. (2017)	USA	Micro[c]	Brainstorming task	Quantitative	12 min	NA	High
10	Lee et al. (2022)	USA	Micro[b]	Management	Quantitative	2 weeks	Littlefield Technologies	High
11	Mekler et al. (2017)	NA	Micro[c]	Image annotation task	Quantitative	22 min	Web-based wiki	High
12	Nadeem et al. (2023)[a]	NA	Micro[b]	Electrical engineering	Quantitative	One semester	Self-developed platform/app	Low
13	Nebel et al. (2016)[a]	Germany	Micro[c]	Learn facts about three allegorical paintings (art)	Quantitative	Between 32 and 61 min	Self-developed platform/app	High
14	Nebel et al. (2017)[a]	Germany	Micro[c]	Physics	Quantitative	40.32 min on average	Self-developed platform/app	High
15	Nebel et al. (2020)	Germany	Micro[c]	Simple factual knowledge about animals (biology)	Quantitative	20.87 min	NA	High
16	Ortiz-Rojas et al. (2019)	Ecuador	Macro	Computer science	Quantitative	4 weeks	The university's LMS	High
17	Philpott and Son (2022)	Japan	Macro	English	Mixed-methods	14 weeks	Self-developed platform/app	High
18	Pi et al. (2023)[d]	China	Micro[c]	English	Quantitative	60 min	Self-developed platform/app	High
19	Sailer and Sailer (2021)	Germany	Micro[c]	Science	Quantitative	45 min	Quizalize	High
20	Zahedi et al. (2021)	NA	Macro	Computer science	Mixed-methods	Three semesters	Self-developed platform/app	High

Note: Leaderboards on a macro-level instructional design refer to using leaderboards when a whole class is gamified, while leaderboards on a micro-level instructional design refer to using leaderboards for a specific learning activity rather than gamifying different activities within a class. Moreover, the two articles Bai et al. (2021) and Balci et al. (2022) had two studies each and the other articles had one study each.

[a]An article that includes two interventions.

[b]Leaderboards designed on a micro-level for specific learning activities within a class without gamifying the whole class.

[c]Leaderboards designed on a micro-level for independent learning activities, which are not within a class.

[d]An article that includes three interventions.

TABLE A2 Study aims, leaderboard designs and study findings of included studies.

No.	Author(s) and year	Study aim	Sample size and condition	Leaderboard type	Design principle	Outcomes	Key findings
1	Bai et al. (2021)[a]	To examine how positions on different types of leaderboards may affect students' learning performance, intrinsic motivation, and course engagement	*N* total = 50; *Absolute leaderboard intervention* (*study 1*) *Relative leaderboard intervention* (*study 2*)	Absolute leaderboard intervention: Real, absolute, individual leaderboard Relative leaderboard intervention: Real, relative, individual leaderboard	Points and ranking Students were ranked for engagement. Participants were informed of points-adding rules. Performance on leaderboards was not counted into students' final grades. Leaderboards had maximum points.	Performance Engagement Perceptions	Absolute leaderboard: students at different positions showed similar levels of learning performance and course engagement, but a higher position was associated with higher intrinsic motivation. Relative leaderboard: students ranked in the top third tended to display better learning performance than their peers in the lower two-thirds did. Students who ranked in different positions showed similar levels of course engagement and intrinsic motivation for learning. Students ranked in the bottom third preferred anonymous comparison on an absolute leaderboard but favoured public comparison on a relative leaderboard.
2	Balci et al. (2022)[a]	To compare the effectiveness of two frequently used gamification tools, badges and leaderboards, on students' academic performance and motivation in fully or partially gamified online physics classes	*N* total = 190; Gamify quizzes (study 1: *N* = 102): *Leaderboard only intervention* Badge only condition Badge and leaderboard condition Non-gamified control condition; Gamify quizzes and assignments (study 2: *N* = 88): *Leaderboard only intervention* Badge only condition	Real, absolute, individual leaderboard	Points and ranking Students were ranked for quiz/task performance. Participants were informed of points-adding rules. Anonymous leaderboards were used. Students were allowed to repeat some quizzes/tasks. Performance on leaderboards was counted into students' final grades.	Performance Motivation Perceptions	Badges and leaderboards did not affect participants' academic performance; however, most students approached them positively as motivational tools and wanted to see them in future online classes. This nonsignificant result did not change when gamification tools were implemented partially or fully in the grading system. Individually or together, badges and leaderboards yielded similar results in terms of effectiveness on academic performance and motivation. Leaderboard-prompted social comparison did not affect student performance negatively.

(Continues)

TABLE A2 (Continued)

No.	Author(s) and year	Study aim	Sample size and condition	Leaderboard type	Design principle	Outcomes	Key findings
			Badge and leaderboard condition Non-gamified control condition				Most students perceived badges and leaderboards positively.
3	Bovermann and Bastiaens (2019)	To research the influence of points and leaderboards on students' intrinsic motivation and collaborative learning in class	*N* total = 107; *Leaderboard intervention* Non-gamified control condition	Real, absolute, individual and team leaderboard	Points and ranking Students were ranked for quiz/task performance. Assessment rubrics were provided to graders.	Intrinsic motivation Perceptions Collaborative learning	There is a significant difference for both "Intrinsic Motivation" and 'Collaborative Learning' between both groups. Students appreciated an active engagement through the gamification concept, as it helped towards purposeful learning behaviour, fostered team spirit in the working groups, and supported increased collegial and professional interaction.
4	Cao et al. (2022)[a]	To explore which difficulty is more conducive to learning performance in leaderboard context, and when and how it plays a role	*N* total = 156; *Leaderboard intervention with high difficulty* *Leaderboard intervention with low difficulty*	Leaderboard intervention with high difficulty: Artificial, presenting the target participant and four best participants, individual Leaderboard intervention with low difficulty: Artificial, presenting the target participant and four worst participants, individual	Points and ranking Students were ranked for quiz/task performance.	Performance Motivation Dominant goal orientation Emotions	Participants in the low difficulty group experienced more positive emotions, less negative emotions, and higher learning motivation, but the effect of difficulty on performance was not significant. Dominant learning-oriented and performance-oriented learners were equally affected by difficulty. Negative emotions and learning motivation mediated the relationship between difficulty and learning performance.
5	Christy and Fox (2014)[a]	To compare the impact of male-dominated and female-dominated leaderboards on women's consequent performance on a math quiz and their academic self-concept, both in general and for math specifically	*N* total = 57; *Male-dominated leaderboard intervention* *Female-dominated leaderboard intervention* Non-gamified	Male-dominated leaderboard intervention: Artificial, male majority top ten leaderboard, individual Female-dominated leaderboard intervention: Artificial, female majority top ten	Points and ranking Students were ranked for quiz/task performance.	Performance Academic identification Math identification	Participants in the female majority leaderboard condition performed more poorly on the math test than those in the male leaderboard condition, yet demonstrated a higher level of academic identification than those in the male and control conditions

TABLE A2 (Continued)

No.	Author(s) and year	Study aim	Sample size and condition	Leaderboard type	Design principle	Outcomes	Key findings
			control condition	leaderboard, individual			Results show that leaderboard condition had no significant effect on post-test math identification.
6	Cigdem et al. (2023)	To investigates the impact of gamification on learning outcomes and course engagement in a computer programming course	*N* total = 96; *Leaderboard intervention* Non-gamified control condition	Real, top three and bottom three, individual and team (class) leaderboard	Points and ranking Students were ranked for quiz/task performance. Students were allowed to repeat some quizzes/tasks.	Performance Course engagement Number of completed quizzes Time spent on task	No significant differences were found between the EG (experimental group) and CG (control group) in terms of the theoretical exam. The results of the t tests conducted to examine the differences in the results of PAES (Perceived Academic Engagement Scale) and its subdimensions between the EG and CG showed no significant differences. The number of students who completed the weekly online quizzes showed a decrease in both the CG and EG, particularly after the third quiz. No significant differences were found in completion rates and time spent on all other quizzes, including the overall number of completions.
7	He and Loewen (2022)	To explore app-based L2 vocabulary self-study as a course assignment by drawing on Mercer and Dörnyei's (2020) latest proposal of goal-setting with feedback for stimulating learner engagement and motivation	*N* total = 63; *Leaderboard intervention* Leaderboard with goal-setting-and-checking activities condition	Real, absolute and unfinished list, individual leaderboard	Points and ranking Students were ranked for quiz/task performance and engagement.	Performance Engagement Motivation Perceptions	Quantitative results showed the treatment group studied more words than the control group. All pedagogical interventions received an average rating higher than 3 ("Neither Agree or Disagree"), indicating that they were regarded by learners as helpful to some extent. Leaderboards were often commented on positively for boosting motivation, stimulating a sense of competition, and providing information about self and others'

(Continues)

TABLE A2 (Continued)

No.	Author(s) and year	Study aim	Sample size and condition	Leaderboard type	Design principle	Outcomes	Key findings
							performance. Nonetheless, some learners did not care about the leaderboards, and others disliked them.
8	Landers and Landers (2014)	To link specific game elements common to leaderboards (conflict/challenge, rules/goals, and assessment) with a focal learner behaviour, time on-task, by exploring educational research on competition and psychological research on goal-setting theory	*N* total = 86; *Leaderboard intervention* Non-gamified control condition	Real, top three and one versus one (A VS B), individual leaderboard	One leaderboard presented ranking without points, one leaderboard presented ranking and points. Students were ranked for quiz/task performance and engagement. Assessment rubrics were provided to graders. Performance on leaderboards was not counted into students' final grades.	Performance Engagement	Random assignment to leaderboards supported a causal effect. Students with leaderboards interacted with their project 29.61 more times, on average, than those in a control condition. Bootstrapping was used to support the mediation of the effect of gamification on academic achievement by this amount of time.
9	Landers et al. (2017)	To examine goal-setting theory, one of the most well-established motivational theories in psychology, as a framework to understand how leaderboards actually affected participants' behaviour	*N* total = 339; *Leaderboard intervention* Do-your-best condition Easy goal condition Difficult goal condition Impossible goal condition	Artificial, presenting target participant and fictitious participants corresponding to different difficulty levels of goals, individual leaderboard	Points and ranking Students were ranked for quiz/task performance.	Performance Goal commitment	The presence of a leaderboard was successful in motivating participants to performance levels similar to that of difficult and impossible goal-setting, suggesting participants implicitly set goals at or near the top of the leaderboard without any prompting to do so. Goal commitment, a common individual difference moderator in goal-setting theory, was also assessed and behaved similarly in the presence of the leaderboard as when traditional goals were provided.
10	Lee et al. (2022)	To empirically examine how relative performance feedback (RPF) in simulation games influences students' learning and outcome performance	*N* total = 219; *Leaderboard intervention* Non-gamified	Real, absolute, team leaderboard	Points and ranking Students were ranked for quiz/task performance.	Performance Feedback response Perceived outcome performance	The study findings suggest that providing RPF (relative performance feedback) via leaderboard improves FR (feedback response), which in turn

TABLE A2 (Continued)

No.	Author(s) and year	Study aim	Sample size and condition	Leaderboard type	Design principle	Outcomes	Key findings
			control condition		Performance on leaderboards was counted into students' final grades.	Rank outcome performance	leads to both higher learning and outcome performance (perceived outcome performance (POP) and a rank outcome performance (ROP).
11	Mekler et al. (2017)	To investigate the effects of points, levels and leaderboards on participants' performance and motivation in an image annotation task	*N* total = 273; *Leaderboard intervention* Points condition Level condition Non-gamified control condition	Artificial, presenting target participant and fictitious participants with appropriate difficulty level, individual leaderboard	Points and ranking Students were ranked for quiz/task performance.	Performance (tag quantity and tag quality) Intrinsic motivation Satisfaction of autonomy needs Satisfaction of competence needs Causality orientation	Compared to a control condition, game elements did not significantly affect competence or intrinsic motivation, irrespective of participants' causality orientation. In this particular study context, points, levels and leaderboards functioned as extrinsic incentives, effective only for promoting performance quantity. While points, levels and leaderboards increased tag quantity, the lack of effects on intrinsic motivation, need satisfaction or tag quality suggest that they may have actually functioned as extrinsic incentives.
12	Nadeem et al. (2023)[a]	To examine the effect of digital game-based learning on student engagement and motivation levels and the gender differences in online learning settings	*N* total = 208; *Game-based quiz activities (GBQAs) with leaderboards intervention (TCG2)* *Moodle-based quiz activities (MBQAs) with leaderboards intervention (TCG5)* GBQAs without leaderboards condition (TCG3) MBQAS without leaderboards condition (TCG4) GBQAs/MBQAs (TCG1)	Real, absolute, individual leaderboard	Points and ranking Students were ranked for quiz/task performance.	Motivation Engagement	It can be observed that the group with the leaderboard (TCG2) outperformed the groups without a leaderboard, namely, TCG1 and TCG3, in nearly all activities, suggesting that the presence of a leaderboard may have positively impacted performance in these activities. Students in groups TCG2 and TCG5 had more fun playing and were more motivated than the groups not using the leaderboard.

(Continues)

TABLE A2 (Continued)

No.	Author(s) and year	Study aim	Sample size and condition	Leaderboard type	Design principle	Outcomes	Key findings
13	Nebel et al. (2016)[a]	To investigate the effects of different amounts of standard discrepancy and the choice to repeat levels on students' cognitive, motivational and learning outcomes	*N* total = 85; *Leaderboard (high standard discrepancy) intervention* **Leaderboard (low standard discrepancy) intervention**	Artificial, relative leaderboard with seemingly realistic points of fictitious participants, individual leaderboard	Points and ranking Students were ranked for quiz/task performance. Students were allowed/not allowed to repeat some quizzes/tasks.	Performance Motivation Perceived competition Emotion Choice	Results revealed an effect of standard discrepancy on retention performance with higher scores for the high standard discrepancy condition. Choice did not influence learning outcomes, but improved motivational and emotional measures. Findings underpin the new role of leaderboards as feedback mechanisms.
14	Nebel et al. (2017)[a]	To examine how leaderboards and competitive gameplay has to be designed to facilitate learning by manipulating core concepts of leaderboards which include competitive effort, perceived difficulty and connected gameplay	*N* total = 103; *Leaderboard intervention with high difficulty Leaderboard intervention with low difficulty*	Artificial, relative leaderboard with seemingly realistic points of fictitious participants, individual leaderboard	Points and ranking Students were ranked for quiz/task performance. Students were allowed to repeat some quizzes/tasks	Performance (retention knowledge) Detail knowledge Process efficiency Outcome efficiency Goal orientation	Results show that players with penalties scored higher on retention tests than players without penalties. Whereas detail knowledge is enhanced by a low difficulty in contrast to a high difficulty. Students with penalties learned and recalled the presented knowledge more efficient An explorative analysis of goal orientations revealed influences of individual dispositions on affective responses and positive correlations with performance goals.
15	Nebel et al. (2020)	To experimentally explore the impact of different social entities and the potential of adaptive elements within artificial entities on students' learning process	*N* total = 102; *Leaderboard (adaptive artificial competition with weak social experiences) intervention* Agent (adaptive artificial competition with strong social experiences) condition	Artificial, relative leaderboard with seemingly realistic points of fictitious participants, individual leaderboard	Points and ranking Students were ranked for quiz/task performance.	Performance (retention knowledge) Learning efficiency Emotion mental starins	The results of the study revealed a beneficial effect of adaptive mechanisms on learning performance and efficiency. The participants reported a lowered feeling of shame, increased empathy, and behavioural engagement when facing competitive agents. No significant impact on mental strains by potentially demanding social competitors.

TABLE A2 (Continued)

No.	Author(s) and year	Study aim	Sample size and condition	Leaderboard type	Design principle	Outcomes	Key findings
			Human (nonadaptive competition with strong social experiences) condition				
16	Ortiz-Rojas et al. (2019)	To investigate how gamification built on leaderboards affect students learning performance, and to examine mediating variables, such as intrinsic motivation, self-efficacy, engagement, and background variables, such as sex, previous gaming experience, and undergraduate major	*N* total = 89; *Leaderboard intervention* Non-gamified control condition	Real, top 1/4 participants and relative leaderboard, individual leaderboard	Ranking Students were ranked based on one in or out-of-class activity which was randomly selected each week. Anonymous and non-anonymous leaderboards were used	Performance Intrinsic motivation Engagement Self-efficacy	We observed a significant improvement in the LP (learning performance) of students in the gamified condition. Studying in a gamified condition did not affect either students' intrinsic motivation, self-efficacy, or engagement. No interaction effect was detected, due to mediating and background variables.
17	Philpott and Son (2022)	To explore how leaderboards affect student performance (i.e., amount of work completed) and foreign language (FL) motivation by using self-determination theory	*N* total = 44; *Leaderboard intervention* Non-gamified control condition	Real, absolute, individual leaderboard	Points and ranking Students were ranked for quiz/task performance and engagement Participants were informed of points-adding rules. Anonymous leaderboards were used. Leaderboards had maximum points.	Extrinsic motivation Intrinsic motivation Engagement Perceptions	The results of the study suggest that the participants' focus on the extrinsic rewards used by the leaderboard encouraged performance up to the reward threshold but once the threshold had been achieved, performance ceased. The leaderboard's use of points, rank, and forced social comparison to control behaviour resulted in the participants' internally leaning extrinsic motivation shifting to externally grounded extrinsic motivation, undermining intrinsic FL(foreign language) motivation more than supporting it. Overall, the participants more frequently talked positively about the leaderboard encouraging performance than they talked about the negative aspects of the leaderboard.

(Continues)

TABLE A2 (Continued)

No.	Author(s) and year	Study aim	Sample size and condition	Leaderboard type	Design principle	Outcomes	Key findings
18	Pi et al. (2023)[b]	To examine how different leaderboard elements (i.e., points and rank) may affect students' English vocabulary learning through video lectures	*N* total = 34; *Leaderboard with only points intervention* *Leaderboards with only ranking intervention* *Leaderboards with points and ranking intervention* Non-gamified control condition	Real, only presenting the target participant, individual leaderboard	*Leaderboard with only points intervention* Points Students were ranked for quiz/task performance. *Leaderboards with only ranking intervention* Ranking Students were ranked for quiz/task performance. *Leaderboards with points and ranking intervention* Points and ranking Students were ranked for quiz/task performance.	Performance Motivation Attention	Participants had low prior knowledge about the IELTS words. No statistically significant difference in students' motivations was found across the four conditions: We found significant differences in the percentage dwell times (attention) across the four conditions. We found significant differences in the saccade counts (attention) between the question and choices areas across the four conditions. We observed significant differences in learning performance between the four conditions.
19	Sailer and Sailer (2021)	To investigate the effects of a gamified flipped classroom interven-tion on learning and motivation	*N* total = 205; *Leaderboard intervention* Non-gamfied control condition	Real, absolute, team leaderboard	Points and ranking Students were ranked for quiz/task performance. Anonymous leaderboards were used.	Performance (application-oriented knowledge) Intrinsic motivation Satisfaction of needs for social relatedness Satisfaction of competence needs	The results show a positive indirect effect of gamification on application-oriented knowledge that is mediated by learning process performance. Based on SDT (self-determination theory), positive effects of gamified in-class activities on intrinsic motivation and social relatedness, but no significant effect on competence need satisfaction.
20	Zahedi et al. (2021)	To explore the effects of virtual points and leaderboards on students' CS (computer science) identity development, self-efficacy, and performance	*N* total = 181; *Leaderboard intervention*	Real, top five, individual leaderboard	Points and ranking Leaderboards had two versions (anonymous or not), but the authors did not mention which version was used.	Performance Motivation Engagement Perceptions CS identity Self-efficacy Interest	The results show that virtual points and the leaderboard contributed to improved performance for students of all genders. Regardless of improved performance, most women did not actively enjoy or were motivated by the virtual points or leaderboard in Software Engineering and Programming

TABLE A2 (Continued)

No.	Author(s) and year	Study aim	Sample size and condition	Leaderboard type	Design principle	Outcomes	Key findings
							Cyberlearning Environment (SEP-CyLE). Gamification had no significant impact on CS (computer science) identity development or self-efficacy constructs and had little to no impact on women's interest and engagement in the field of computing.

Note: The two articles Bai et al. (2021) and Balci et al. (2022) had two studies each and the other included articles had one study each.
[a]An article that includes two interventions.
[b]An article that includes three interventions.

TABLE A3A Quality assessment results of the included articles: quality assessment results of the included articles and the excluded article—qualitative study.

No.	Author(s) and year	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	1.1. Is the qualitative approach appropriate to answer the research question?	1.2. Are the qualitative data collection methods adequate to address the research question?	1.3. Are the findings adequately derived from the data?	1.4. Is the interpretation of results sufficiently substantiated by data?	1.5. Is there coherence between qualitative data sources, collection, analysis and interpretation?	Number of yes (except S1 and S2)
1	Bai et al. (2021)	Y	Y	Y	Y	Y	Y	Y	5
2	Balci et al. (2022)	Y	Y	Y	N (The authors only used one open-ended question to collect qualitative data.)	Y	Y	Y	4
3	Bovermann and Bastiaens (2019)	Y	Y	Y	Y	N (The number of students who reported their opinions was not recorded in the text).	Y	Y	4
7	He and Loewen (2022)	Y	Y	Y	Y	Y	Y	Y	5
17	Philpott and Son (2022)	Y	Y	Y	Y	N (The authors did not present qualitative data adequately.)	Y	Y	4
20	Zahedi et al. (2021)	Y	Y	Y	Y	Y	Y	Y	5
-	Flinton et al. (2023)[a]	Y	Y	Y	N (the authors used open-ended questions to collect qualitative data. They should have provided what specific questions were used. Moreover, interviews might provide more in-depth information about students' perceptions towards simulation and gamification/competition elements than open-ended questions).	N (For some statements, the number of students who reported their opinions was not recorded in the text).	Y	CT (The authors did not provide specific questions used to collect qualitative data, so it is not that clear whether the data analysed and interpreted were coherent with the data sources collected).	2

Note: The two articles Bai et al. (2021) and Balci et al. (2022) had two studies each and the other included articles had one study each.
Abbreviation: CT, Cannot tell.
[a]The excluded article after the quality assessment.

TABLE A3B Quality assessment results of the included articles—quantitative (randomised) study.

No.	Author(s) and year	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	2.1. Is randomization appropriately performed?	2.2. Are the groups comparable at baseline?	2.3. Are there complete outcome data?	2.4. Are outcome assessors blinded to the intervention provided?	2.5 did the participants adhere to the assigned intervention?	Number of yes (except S1 and S2)
2	Balci et al. (2022)	Y	Y	CT	Y	Y	Y	Y	4
3	Bovermann and Bastiaens (2019)	Y	Y	CT	Y	N (Only a relatively small proportion of students who enrolled in this course agreed to participate in this study.)	Y	Y	3
4	Cao et al. (2022)	Y	Y	CT	Y	Y	Y	Y	4
5	Christy and Fox (2014)	Y	Y	CT	Y	Y	Y	Y	4
8	Landers and Landers (2014)	Y	Y	CT	Y	Y	Y	Y	4
9	Landers et al. (2017)	Y	Y	Y	Y	Y	Y	Y	5
11	Mekler et al. (2017)	Y	Y	Y	Y	Y	Y	Y	5
13	Nebel et al. (2016)	Y	Y	Y	Y	Y	Y	Y	5

Note: The two articles Bai et al. (2021) and Balci et al. (2022) had two studies each and the other included articles had one study each.
Abbreviation: CT, cannot tell.

TABLE A3C Quality assessment results of the included articles—quantitative (non-randomised) study.

No.	Author (s) and year	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	3.1. Are the participants representative of the target population?	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	3.3. Are there complete outcome data?	3.4. Are the confounders accounted for in the design and analysis?	3.5 during the study period, is the intervention administered (or exposure occurred) as intended?	Number of yes (except S1 and S2)
1	Bai et al. (2021)	Y	Y	Y	Y	Y	Y	Y	5
6	Cigdem et al. (2023)	Y	Y	Y	Y	Y	N (The authors did not mention how confounders were addressed in this study.)	Y	4
7	He and Loewen (2022)	Y	Y	Y	N (The learner perception survey used was unvalidated.)	Y	Y	Y	4
10	Lee et al. (2022)	Y	Y	Y	Y	Y	Y	Y	5
16	Ortiz-Rojas et al. (2019)	Y	Y	Y	Y	Y	Y	Y	5
20	Zahedi et al. (2021)	Y	Y	Y	Y	Y	Y	Y	5

TABLE A3D Quality assessment results of the included articles and the excluded article—quantitative (descriptive) study.

No.	Author (s) and year	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	4.1. Is the sampling strategy relevant to address the research question?	4.2. Is the sample representative of the target population?	4.3. Are the measurements appropriate?	4.4. Is the risk of nonresponse bias low?	4.5. Is the statistical analysis appropriate to answer the research question?	Number of yes (except S1 and S2)
12	Nadeem et al. (2023)	Y	Y	Y	Y	N (The five-point scale investigated several different constructs, but the authors only reported one Cronbach's alpha).	N (The nonresponse rate in the first phase was relatively high.)	N (The authors could have compared the results across different groups using ANOVAs rather than describing the data.)	2
-	Flinton et al. (2023)[a]	Y	Y	Y	N (The proportion of male participants was low.)	N (The authors did not provide the reliability and validity of the questionnaires used although they mentioned that the questionnaires had been piloted to have face validity. Moreover, they did not provide the specific open-ended questions used.)	Y	N (The authors reported medians instead of means of the items. The authors mentioned that a comparative analysis indicated that there was no significant difference among the three cohorts, but did not specify what comparative analysis was used).	2

Note: The two articles Bai et al. (2021) and Balci et al. (2022) had two studies each and the other included articles had one study each.
[a]The excluded article after the quality assessment.

TABLE A3E Quality assessment results of the included articles and the excluded article—mixed-method study.

No.	Author(s) and year	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	5.1. Is there an adequate rationale for using a mixed methods design to address the research question?	5.2. Are the different components of the study effectively integrated to answer the research question?	5.3. Are the outputs of the integration of qualitative and quantitative components adequately interpreted?	5.4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?	5.5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?	Number of yes (except S1 and S2)
1	Bai et al. (2021)	Y	Y	N ((The authors did not provide rationales for using a mixed methods design.)	Y	Y	Y	Y	4
2	Balci et al. (2022)	Y	Y	N (The authors did not provide rationales for using a mixed methods design.)	Y	Y	Y	Y	4
3	Bovermann and Bastiaens (2019)	Y	Y	Y	N (The quantitative and qualitative results seemed to be separated.)	N (The discussion about the integration of quantitative and qualitative results is relatively scare).	Y	Y	3
7	He and Loewen (2022)	Y	Y	N (The authors did not provide rationales for using a mixed methods design.)	Y	Y	Y	Y	4
17	Philpott and Son (2022)	Y	Y	Y	Y	N (The authors did not present qualitative data adequately.)	Y	Y	4
20	Zahedi et al. (2021)	Y	Y	Y	Y	Y	Y	Y	5
-	Flinton et al. (2023)[a]	Y	Y	N (The authors did not provide rationales for using a mixed methods design.)	N (The qualitative data mentioned students' stress, but the quantitative data did not measure students' stress. Thus, the findings regarding stress were only	N (The authors provided detailed explanations about the integration of qualitative and quantitative data regarding feedback on gamification/ competition	N (The qualitative data mentioned a lot about students' perceptions of stress/pressure, but the quantitative data did not measure students' stress.)	Y	1

TABLE A3E (Continued)

No.	Author(s) and year	S1. Are there clear research questions?	S2. Do the collected data allow to address the research questions?	5.1. Is there an adequate rationale for using a mixed methods design to address the research question?	5.2. Are the different components of the study effectively integrated to answer the research question?	5.3. Are the outputs of the integration of qualitative and quantitative components adequately interpreted?	5.4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?	5.5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?	Number of yes (except S1 and S2)
					discussed from the qualitative perspective.)	elements. Meanwhile, they should have provided more details about the integration of qualitative and quantitative data regarding feedback on the simulation. Moreover, the authors seemingly made assertive statements about the differences in males and females found in both qualitative and quantitative data using reference 18 published a decade ago, which might be less appropriate).			

Note: The two articles Bai et al. (2021) and Balci et al. (2022) had two studies each and the other included articles had one study each.
[a]The excluded article after the quality assessment.

TABLE A4 Inter-rater reliability full table.

Case processing summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	*N*	Percent	*N*	Percent
Researcher 1 * Researcher 2 * Categories	125	100.0%	0	0.0%	125	100.0%

Symmetric measures						
Categories			Value	Asymptotic standard Error[a]	Approximate T[b]	Approximate significance
	Measure of Agreement	Kappa	.[c]			
	N of valid cases		25			
Applications/platforms	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Author	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Disciplines	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Duration	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Educational contexts	Measure of Agreement	Kappa	0.692	0.225	3.133	0.002
	N of valid cases		4			
Future research	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Location	Measure of Agreement	Kappa	0.692	0.225	3.133	0.002
	N of valid cases		4			
Implications[d] (practical/theoretical implications)	Measure of Agreement	Kappa	0.200	0.139	2.000	0.046
	N of valid cases		4			
Instructional design	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Instruments (e.g., tests/questionnaires)	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001

TABLE A4 (Continued)

Symmetric measures						
Categories			**Value**	**Asymptotic standard Error[a]**	**Approximate *T*[b]**	**Approximate significance**
	N of valid cases		4			
Intervention number	Measure of Agreement	Kappa	0.692	0.225	3.133	0.002
	N of valid cases		4			
Journal	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Leaderboard types	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Leaderboard design framework	Measure of Agreement	Kappa	0.692	0.225	3.133	0.002
	N of valid cases		4			
Limitations[d]	Measure of Agreement	Kappa	0.429	0.204	2.683	0.007
	N of valid cases		4			
Outcome types[d] (summarising learning outcomes reported as well as the intervention(s) and comparison group(s) involved)	Measure of Agreement	Kappa	0.200	0.139	2.000	0.046
	N of valid cases		4			
Participants (sample size)	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Design practices	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Research aims	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Research outcomes (original sentences reporting the research outcomes in the article)	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Research paradigms	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Theoretical frameworks	Measure of Agreement	Kappa	0.692	0.225	3.133	0.002

(Continues)

TABLE A4 (Continued)

Symmetric measures						
Categories			**Value**	**Asymptotic standard Error[a]**	**Approximate *T*[b]**	**Approximate significance**
	N of valid cases		4			
Title	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Types of research (leaderboard presence/absence or leaderboard designs)	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Year	Measure of Agreement	Kappa	1.000	0.000	3.464	<0.001
	N of valid cases		4			
Total	Measure of Agreement	Kappa	0.891	0.028	53.525	<0.001
	N of valid cases		125			

Note: The discrepancies between the two coders were noted and discussed until a consensus was reached.
[a]Not assuming the null hypothesis.
[b]Using the asymptotic standard error assuming the null hypothesis.
[c]No statistics are computed because Researcher 1 and Researcher 2 are constants.
[d]The category had a relatively low value. These categories were paid special attention when the two coders discussed the differences to reach a consensus.