Received: 20 January 2021 | Revised: 5 August 2021 | Accepted: 1 May 2022
DOI: 10.1111/jcal.12696

REVIEW ARTICLE

Journal of Computer Assisted Learning WILEY

Effect of blended learning on student performance in K-12 settings: A meta-analysis

Shuqin Li | **Weihua Wang**

School of Educational Science, Hunan Normal University, Changsha, China

Correspondence
Weihua Wang, School of Educational Science, Hunan Normal University, 36 Lushan Street, Changsha 410081, China.
Email: wangweihua@hunnu.edu.cn

Funding information
The National Office for Education Sciences Planning of China, Grant/Award Number: BAA170020

Abstract

Background: Blended learning programs in Kindergarten through Grade 12 (K-12) classrooms are growing in popularity; however, previous studies assessing their effects have yielded inconsistent results. Further, their effects have not been completely quantitatively synthesized and evaluated.

Objectives: The purpose of this study is to synthesize the overall effects of blended learning on K-12 student performance, distinguish the most effective domains of learning outcomes, and examine the moderators of the overall effects.

Methods: For the purpose, this study conducted a meta-analysis of 84 studies published between 2000 and 2020, and involved 30,377 K-12 students.

Results and Conclusions: Results revealed that blended learning can significantly improve K-12 students' overall performance [$g = 0.65$, $p < 0.001$, 95% CI = (0.54–0.77)], particularly in the cognitive domain [$g = 0.74$, $p < 0.001$, 95% CI = (0.61–0.88)). The testing of moderators indicates that the factors moderating the impact of blended learning on student performance in these studies included group activities, educational level, subject, knowledge type, instructor, sample size, intervention duration and region.

Implications: The results indicate that blended learning is an effective way to improve K-12 students' performance compared to traditional face-to-face (F2F) learning. Additionally, these findings highlight valuable recommendations for future research and practices related to effective blended learning approaches in K-12 settings.

KEYWORDS
blended learning, K-12, meta-analysis, student performance

1 | INTRODUCTION

Blended learning is unique, because it preserves the benefits of face-to-face (F2F) learning, while maximizing the advantages of technology-enhanced online learning environments (Lim et al., 2007; Osguthorpe & Graham, 2003; Rasheed et al., 2019; Xu et al., 2019). According to the 2014 report by the International Association for Kindergarten through Grade 12 (K-12) Online Learning (iNACOL, 2014), there is increasing interest in blended learning practices for K-12 classrooms. This was particularly true in the recent years, when COVID-19 impacted the world, significantly impacting many K-12 schools. Blended learning offered an effective temporary solution to issues caused by school closures and distancing requirements. In a blended learning environment, K-12 students impacted by the pandemic could participate in online learning at home and attend schools on a rotating basis for F2F learning (Arnett, 2020).

However, can blended learning really improve the academic performance of K-12 students? Many empirical studies have compared blended learning with traditional F2F learning and have led to inconsistent conclusions. Some scholars argue that blended learning can significantly improve the K-12 students' performance (e.g. Bhagat et al., 2016; Kirvan et al., 2015; Sergis et al., 2017), while others find no significant improvement (e.g. Gelgoot et al., 2020; Hwang et al., 2019; Lo & Hew, 2020). Some studies even suggest that blended learning can significantly reduce academic performance (e.g. Rockman, 2007; Tse et al., 2017).

These ambiguous findings highlight the need for a more rigorous review of the effects of blended learning in K-12 settings. However, we found no relevant quantitative reviews. Therefore, we conducted a meta-analysis of 84 studies published from 2000 to 2020 to systematically synthesize the overall effect of blended learning on the K-12 students' performance, and to distinguish which domains of student performance-blended learning most effectively impacts. Also, we used moderator analysis to determine the factors that may increase the effectiveness of blended learning in K-12 settings. This study provides valuable insights for research and practice related to effective blended learning approaches in K-12 settings, which teachers, researchers, educational policymakers and those interested in blended learning might find useful.

1.1 | Blended learning in K-12

Blended learning, also known as mixed or hybrid learning, refers to the combination of traditional F2F and online learning (Bonk & Graham, 2005; Garrison & Kanuka, 2004; Osguthorpe & Graham, 2003; Young, 2002). The F2F learning component generally means that students acquire knowledge or receive instructions through F2F lectures, while the online learning component refers to "a formal education program, in which a student learns at least in part through online delivery of content and instruction with some element of student control over time, place, path, and/or pace" (Staker & Horn, 2012, p. 3). We adhere to this definition as well. Under this definition, those instances in which teachers use online materials when teaching in a F2F classroom, or where students use digital textbooks but attend class in a physical environment, are not blended learning (Horn & Staker, 2011); studies that feature such situations are not included in our meta-analysis.

In the recent years, blended learning has become increasingly popular in K-12 contexts (e.g. Alabama State Department of Education, 2018; Barbour & Labonte, 2017; Fazal & Bryant, 2019). Several studies have compared the effectiveness of blended learning with F2F learning in K-12 classrooms, and have identified many advantages of blended learning in these environments. For students, the blended learning environment is more supportive and flexible, because part of the learning is untethered to time and place. Also, it can improve the students' confidence and enable them to actively participate in their learning process (e.g. Aidinopoulou & Sampson, 2017; Flores, 2018; Sergis et al., 2017). Several experimental studies have shown that blended learning can significantly improve the students' academic achievement (Macaruso et al., 2020), interest in learning (Nee, 2014), engagement (Clark, 2015), motivation (Bhagat et al., 2016), attitudes (Lin et al., 2017), skills (Kirvan et al., 2015) and abilities (Wilkes et al., 2020). Blended learning supports whole-class, small-group and independent learning, which is beneficial for teachers interested in increasing student-centred activities and offering differentiated instruction (Freeland, 2015, Morgan, 2002, Powell et al., 2014). Teachers can also check online learning records to adjust instructional strategies based on the students' learning progress (Hilliard, 2015; Powell et al., 2015). For schools, blended learning can alleviate problems related to a lack of classrooms and insufficient teachers, because part of the learning is done online (Lorenzo, 2017).

However, studies have also identified several challenges related to blended learning in K-12 settings. First, blended learning is effective only if students have high motivation, self-regulation and technology skills (Kettle, 2013; Lo, Lie, & Hew, 2017b; Van Laer & Elen, 2017). Second, the digital divide is an issue with blended learning; some students in underdeveloped areas do not own adequate technology and have poor network signals, impacting the effect of blended learning (Graham, 2006; Lorenzo, 2017). Third, the correct blending of 'time, people, place and resources' is also challenging for teachers, and requires them to have sufficient technical literacy. Unfortunately, Arnett (2019) found that many K-12 teachers struggle to manage the many technology platforms they are required to work with, and many teachers feel that they are not adequately trained in the use of computer and internet technology (Fraillon et al., 2014). Finally, educational institutions also face challenges with respect to providing effective training support for teachers (Rasheed et al., 2019), and paying for technology and maintenance costs (Akcayir & Akcayir, Akçayir & Akçayir, 2018).

A recent narrative review provided general views on the effects and potential variables of blended learning in K-12 contexts (Poirier et al., 2019). However, it is difficult to generalize and extend their results, because of the small number of studies considered in this review ($k = 11$). Further, the review did not address inconsistent reports on the effects of blended learning. Therefore, we conducted a meta-analysis to combine the quantitative results of empirical studies. This approach is more objective than a narrative review, and the comprehensive effect estimate has greater statistical power than individual studies (Borenstein et al., 2009; Lipsey & Wilson, 2001). As such, the study supports more general conclusions about the effects of blended learning on the academic performance of K-12 students.

1.2 | Previous meta-analyses and purpose of this study

As shown in Table 1, we found six previous meta-analyses exploring the impact of blended learning on student performance.

To our knowledge, the earliest meta-analysis of the effects of blended learning on student performance can be traced back to a study conducted by Means et al. (2010), which included 23 studies (1996–2008). The results confirmed that blended learning had a significant, but small effect on student performance ($g = 0.35$, $p < 0.001$) compared to F2F learning.

TABLE 1 Meta-analyses of the impact of blended learning on student performance

Studies	Year covered	Educational level	K	Effect size
Means et al., 2010	1996–2008	All	23	$g = 0.35^{***}$
Bernard et al., 2014	1990–2010	High education	117	$g = 0.334^{***}$
Spanjers et al., 2015	2005–201	All	24	$g = 0.34^{**}$
			11	$g = 0.27^{*}$
			30	$g = 0.11^{**}$
			4	$g = -1.04^{**}$
Liu et al., 2016	2004–2014	Health Professions	20	$SMD = 1.40^{***}$
	1991–2014	Health Professions	56	$SMD = 0.81^{***}$
Vo et al., 2017	2001–2015	High education	51	$g = 0.385^{***}$
Li et al., 2019	1980–2015	Nursing students	8	$SMD_1 = 0.70^{***}$
			5	$SMD_2 = 0.72^{*}$
			3	$SMD_3 = 0.58$

Note: k = number of studies.
Abbreviation: *SMD*, Standard Mean Difference.
$^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

In 2014, Bernard et al. (2014) analysed 117 studies published between 1990 and 2010. They concluded that blended learning can significantly improve student performance in higher education ($g = 0.334$, $p < 0.001$).

Using a sample of 47 articles (2005–2013), Spanjers et al. (2015) summarized the effects of blended learning on objective effectiveness ($g = 0.34$), subjective effectiveness ($g = 0.27$), satisfaction ($g = 0.11$), and the investment effect of evaluations ($g = -1.04$). The authors concluded that blended learning was more effective than traditional F2F learning, and the moderated analysis indicated that quizzes positively affected the effectiveness of blended learning and the students' perception of its attractiveness.

Liu et al. (2016) investigated the effectiveness of blended learning in Health Professions programs. They conducted two meta-analyses. The first meta-analysis used 20 studies that compared blended learning with no intervention. It produced a large effect size of 1.40 ($p < 0.001$). The second meta-analysis compared the effects of blended learning with non-blended learning (including F2F and online learning) and the effect size was 0.81 ($p < 0.001$). This result concluded that blended learning had a significant positive impact on health professions students' performance. However, the authors also noted that conclusions should be treated with caution due to the high degree of heterogeneity.

In 2017, Vo et al. (2017) conducted a meta-analysis of 51 studies and identified a small mean effect ($g = 0.385$, $p < 0.001$) of blended learning on student performance at the course level in higher education. They also found that blended learning significantly improved achievement in STEM subjects.

Li et al. (2019) found that compared to traditional instruction, blended learning effectively increased nursing students' knowledge ($SMD = 0.70$, $p < 0.001$) and satisfaction ($SMD = 0.72$, $p < 0.05$), but did not significantly increase skill levels ($SMD = 0.58$, $p = 0.13$).

The meta-analyses in Table 1 indicate that blended learning had a significant positive effect on the students' performance. However, these studies had limitations. For example, although there have been many studies on blended learning in K-12, few meta-analyses have systematically explored the effect of this type of learning on the K-12 students' academic performance. Further, findings in higher education, professional education and adult education environments may not directly apply to K-12 classrooms (Means et al., 2010), because of differences in the students' needs, abilities and limitations (Drysdale et al., 2013). As such, this meta-analysis systematically examines literature since 2000 on the effectiveness of blended learning in K-12 settings.

Additionally, in K-12 classrooms, it is unclear which learning outcomes from blended learning will most improve and under what circumstances it will be most effective. Therefore, our meta-analysis considers as many variables as possible. We examined the effects of blended learning on three domains of student performance: cognitive domain (e.g. exam scores), affective domain (e.g. satisfaction, motivation) and psychomotor domain (e.g. skill, ability) (Bloom & Krathwohl, 1956). We also conducted extensive moderator analyses to investigate if blended learning design, educational context, research methodology, or study characteristics impact the effect of blended learning on the students' academic performance.

We designed the study to explore the following two research questions (RQ):

> **RQ1.** *How does blended learning affect the overall academic performance of K-12 students compared to F2F learning? What kinds of learning outcomes are more suitable for blended learning?*

> **RQ2.** *What factors influence the overall effect of blended learning on K-12 students' performance?*

1.3 | Potential moderator variables considered

This meta-analysis investigated whether different implementations of blended learning produce different effects to reveal the characteristics of effective blended learning. We selected moderator variables based on the previous research. First, we consulted the previous reviews of blended learning to examine which theoretical grounding could explain the variance of effects. We then identified variables that are frequently reported in blended learning intervention studies, by conducting a preliminary literature search. Finally, we grouped the hypothetical moderating variables into four categories: blended learning design, educational context, and research methodology and literature characteristics.

1.3.1 | The characteristics of blended learning design

The following variables related to blended learning design characteristics were featured prominently in the existing research studies: blended learning models, media features in online learning, communication in online learning, and group activities.

First, we examined whether different blended learning models might impact its effectiveness. Until now, there has been no empirical evidence demonstrating which blended learning model is the most effective. Consequently, according to Staker and Horn's (2012) classification, we considered seven models, including Station-Rotation, Lab-Rotation, Flipped-Classroom, Individual-Rotation, Flex model, Self-blend model, and Enriched-Virtual model.

Second, we analysed the media features in online learning as a potential moderator variable. Rossett and Frazee (2003) noted that technological tools are a key component for successful blended learning. In the blended learning environment, many intervention studies refer to the use of a learning management system (LMS) (e.g. Moodle; Edmodo) for online learning (e.g. Jia et al., 2013; Wendt & Rockinson-Szapkiw, 2015), while fewer studies used video only (e.g. Bhagat et al., 2016; Gelgoot et al., 2020) or other media (such as a combination of video and websites, for example, in Sartepeci & Cakir, 2015; Wang et al., 2018).

Third, we investigated whether communication in online learning could explain the variance between studies concerning the effects of blended learning. Garrison (2011) divided online learning into two categories: synchronous and asynchronous. The asynchronous mode is believed to provide learners with more flexibility and to encourage students to spend more time thinking and reflecting about their learning, thereby facilitating the learning performance (Dey & Bandyopadhyay, 2019).

Fourth, group activities were included as a potential moderator variable. Many studies have shown that group activities have a significant positive impact on student performance (Kyndt et al., 2013; Lo et al., 2017).

1.3.2 | The characteristics of the educational context

We selected the following variables as characteristics of the educational context: educational level, discipline, knowledge type and instructor.

We created three categories for the educational-level variable: Kindergarten, Elementary (grades 1–6), and Secondary (grades 7–12). This was done to determine whether the effectiveness of blended learning would vary by context. Blended learning programs usually involve student-centred instruction; as such, students must embrace the role of self-regulated learners (Shea et al., 2010). However, the ability to self-regulate differs among students of different ages (Wigfield et al., 2011). The teacher's absence during learning activities may cause difficulties for learners, especially those with low self-regulation abilities (Dabbagh & Kitsantas, 2005).

Next, we investigated subjects (e.g. computer, reading and science) as a variable. Some studies have found that blended learning can lead to some positive outcomes; however, its effect on student performance may depend on how it is implemented across different content areas (Lo, Hew, & Chen, 2017a; Pane et al., 2017).

Similarly, knowledge type was also tested as a potential moderating variable. Scholars found that the effect of blended learning may vary depending on different knowledge types (Sitzmann et al., 2006). Therefore, we adopt the framework by Kraiger et al. (1993) to classify knowledge types into two categories: declarative knowledge and procedural knowledge.

Lastly, we included the instructor variable in our moderator analysis. Previous research has shown that a teacher's expertise and technological literacy can impact learning outcomes (Darling-Hammond, 2000). However, in many studies, the experimental and control groups did not have the same instructor (e.g. Hwang et al., 2019; Wilkes et al., 2020).

1.3.3 | The characteristics of the research methodology

The research methodology category includes four moderating variables: study design, sample size, intervention duration and region.

We divided the study designs of different studies into quasi-experimental and random experimental approaches. As Cheung and Slavin (2016) note, different study designs may produce significantly different effects depending on the rigour of the study.

It is also possible that the different sample sizes in blended learning studies affect the sizes of their effects. Previous research has shown that sample size can significantly impact the results of a study; for example, a smaller sample size can have a larger effect (Chen et al., 2018; Cheung & Slavin, 2013; Hillmayr et al., 2020). Thus, based on the distribution of the sample size of 84 included studies, we adapted a classification scheme of sample size from Slavin and Smith (2009), and Bai et al. (2020): Small ($N \leq 50$), Medium ($51 \leq N \leq 100$), and Large ($N \geq 101$), with N referring to the sample size.

We created two broad categories for intervention duration to examine its moderating effect: less than one semester and equal to one semester. Chandra and Lloyd (2008) found that blended learning is a new approach to many students, and it may take time for them to adjust to the pace of blended learning, so intervention time is likely to influence the effectiveness of blended learning.

Moreover, the region in which a study took place is also a potential influencing factor. Studies have shown that the people's attitudes towards, and usage of, information technology vary by context (Collis & Williams, 2001; Li & Kirkup, 2007). Therefore, we have divided studies according to the continents in which they were conducted.

1.3.4 | The characteristics of the literature

We conducted a meta-regression analysis using the publication year as a continuous variable. The studies included in this meta-analysis span 21 years. Over time, several emerging technologies create more flexible and diverse forms of applied blended learning and may improve its effectiveness. Examples include augmented reality (Dunleavy et al., 2009), learning management systems (Dias & Diniz, 2013), learning analytics (Lu et al., 2018) and virtual reality (Nortvig et al., 2020). Cheung and Slavin (2012) analysed the effect of publication year on the effectiveness of educational technology, and they found that the effect differed significantly across periods. Therefore, we assume that the reported effect of blended learning in K-12 settings may change in studies from different years, as educational technology advances over time.

2 | METHOD

Our meta-analysis strictly adhered to the PRISMA guidelines (Moher et al., 2009), to ensure that we would scientifically and systematically estimate the effects of blended learning on student performance in K-12 settings. The main processes included the development of literature inclusion and exclusion criteria, the collection of literature, quality assessment, coding, effect size calculation and data analysis.

2.1 | Inclusion and exclusion criteria

The studies included in this meta-analysis meet the following criteria.

1. The study is conducted with students enrolled in regular K-12 educational institutions. We excluded special education and vocational schools, and higher education institutions.
2. The intervention in the study uses blended learning, which was defined as a combination of F2F and online learning.
3. The control group in the study experienced F2F instruction. We excluded studies that did not have a control group or that used a control group that was not F2F.
4. Reported outcomes include the effects of blended learning on learning performance (e.g. academic achievement, skills and attitudes). Outcomes that are not related to learning performance are excluded. The study should also report sufficient statistical information to ensure that effect sizes can be calculated.
5. The study is grounded in one of these experimental designs: true experiment, quasi-experiment, or crossover design. Other types of experiments were excluded.
6. The study is published in English in a peer-reviewed journal and published between 2000 and 2020.

2.2 | Literature search

Our search process for studies that met the above criteria involved several steps. First, to retrieve all of the available research on blended learning, we searched for publications on the following platforms: Web of Science, EBSCOhost, Science Direct, Association for Computing Machinery, Spring Link, Wiley Online Library, and Taylor & Francis Online. We applied the Boolean search parameters in the 'subject', 'title' and 'abstract' search fields: #1 AND #2 AND #3 AND #4 AND #5 (as Table 2 shows). Second, references from the previous review studies on blended learning were manually checked. Third, a snowballing method was used to search for references in retrieved articles. We did not include unpublished studies because an assessment of their quality cannot be guaranteed in the absence of a peer-review process.

The last literature search was conducted on August 15, 2020. As shown in Figure 1, after a multi-method search, 1618 records were obtained; 910 remained after the removal of duplicates. After checking titles and abstracts, 733 studies were excluded, because they did not fit the inclusion criteria that are stated above. Therefore, the full-

TABLE 2 Search parameters

	Search parameters
#1	(blend* OR hybrid* OR integrate* OR computer-assist* OR flip* OR invert*) AND (learn* OR teach* OR class* OR course* OR environment*)
#2	("academic performance" OR "academic achievement" OR "school grades" OR "school achievement" OR "GPA" OR "school performance" OR "school marks" OR "educational outcome" OR "class grades" OR "standardized test" OR "course grade" OR skill* OR attitude* OR satisfaction OR motivation)
#3	(K-12 OR kindergarten OR "primary school" OR "elementary school" OR "middle school" OR "high school" OR "secondary school" OR "pre-college student")
#4	(quasi-experiment OR experiment* OR "random* control*" OR compar* OR trial* OR evaluat* OR assess* OR effect* OR pretest* OR pre-test OR posttest* OR post-test OR pre-interven* OR post-interven)
#5	Language = English; Time = 2000–2020; Document type = Article

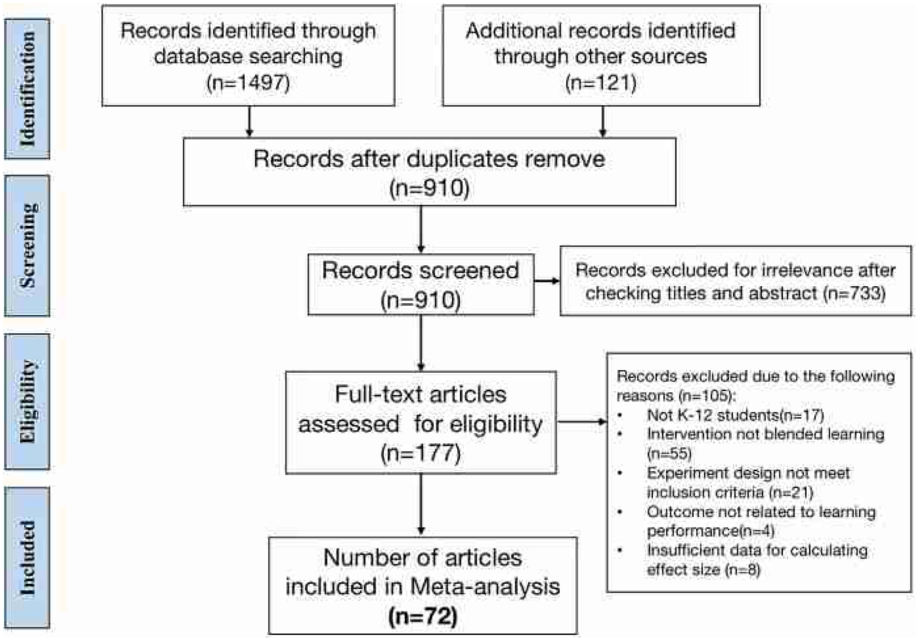

FIGURE 1 Flow chart of the inclusion and exclusion of studies, following the PRISMA statement

text versions of the remaining 177 articles were assessed. We also sent emails to corresponding authors of 11 studies due to the missing data in these studies; we received three responses. The remaining 8 studies for which we did not hear back from the authors were excluded, because the data could not be used to calculate effect sizes. Ultimately, this meta-analysis included 84 studies from 72 articles (see Appendix A).

2.3 | Quality assessment

Two authors independently assessed the quality of the studies included in the meta-analysis according to the Medical Education Research Study Quality Instrument (MERSQI). The instrument has been rigorously evaluated, is neutral in practice subjects, and can also be used to assess the quality of non-medical, education research (Jensen & Konradsen, 2018; Wu et al., 2020). The MERSQI is designed to measure the quality of experimental, quasi-experimental and observational studies, and it consists of six domains: experimental design, sampling, type of data, validity of evaluation instrument, data analysis, and outcomes, with a maximum domain score of 3 and potential total score range of 5 to 18 (Reed et al., 2007). The higher the total score, the higher the quality of the study. The mean MERSQI score of the 84 studies included in this meta-analysis was 12.58 (SD = 1.05), with a range of 9.5 to 15.5 Therefore, the studies we selected are of relatively high quality (see Appendix B).

2.4 | Coding

Two researchers spent two rounds improving coding schemes. They then coded 15 randomized publications before discussing and resolving differences. They continued coding the remaining 57 publications. Cohen's kappa statistic was used to check the reliability of the two authors' coding, and the value range from 0.88 to 0.94 ($p < 0.001$). Disagreements about the codes were discussed until the two authors reached an agreement.

TABLE 3 Coding of studies

Categories	Codes
Dependent variable	1. Student performance: Cognitive domain; Affective domain; Psychomotor domain
blended learning design characteristics	2. Type of blended learning model: Station-rotation model; Lab-rotation model; Flipped-classroom model; Flex model; Others or not reported 3. Media type in online learning: Learning manage system; Video only; Others 4.Communication in online learning: Asynchronous; Synchronous and Asynchronous; Not reported 5. Group activities: Yes; No; Not reported
Educational context characteristics	6. Educational level: Kindergarten; Elementary (grades 1–6); Secondary (grades 7–12); Mixed 7.Discipline: Mathematics; Computer; Biology; Reading; Language; Physics; Science; Others 8. Knowledge type: Declarative; Declarative and procedural; Procedural 9. Instructor: Same; Different; Not reported
Research methodology characteristics	10. Study design: Quasi-experimental; Random experimental 11. Sample size: Small ($N \leq 50$); Medium ($51 \leq N \leq 100$); Large ($N \geq 101$) 12. Intervention duration: ≤ 1 semester; >1 semester; Not reported 13. Region: Africa; Asia; Europe; North America; Oceania
Literature characteristics	14. Publication year (A continuous variable)

Finally, the included studies were systematically analysed using two coding schemes: one at the effect size level and one at the study level (Lipsey & Wilson, 2001). For the first coding scheme, we divided the dependent variable into three domains. For the second coding scheme, a total of 13 variables were coded for each study. The main codes are shown in Table 3. For those studies that did not report the corresponding information, the "not reported" category was used.

2.5 | Effect size calculation

The effect size can reflect the effect of the intervention. A larger effect size indicates a more effective blended learning environment. Hedges' *g* was used as the standardized mean difference effect size metric. It is a corrected version of Cohen's *d* and has been regarded as more unbiased and conservative estimate than Cohen's *d*.

According to Cohen's *d* (Cohen, 1988) and Hedges' *g* equation (Hedges, 1982), the calculation formulas of effect size in our meta-analysis are expressed as:

$$\text{Cohen's}\, d = \frac{M_{E\_post} - M_{C\_post}}{\sqrt{\frac{(N_{E\_post}-1)SD^2_{E\_post} + (N_{C\_post}-1)SD^2_{C\_post}}{(N_{E\_post}+N_{C\_post}-2)}}}$$

$$\text{Hedges'}\, g = \left[1 - \frac{3}{4(N_{E\_post}+N_{C\_post})-9}\right] d$$

Where, M_{E_post} and M_{C_post} are the post-test mean scores of the experimental group and control group, respectively; N_{E_post} and N_{C_post} are the post-test sample sizes in of the experimental group and control group, respectively; SD_{E_post} and SD_{C_post} are the post-test standard deviation in of the experimental group and control group, respectively.

In order to calculate the effect sizes of all studies, there were instances where we needed to make additional calculations, as described here. The basis for these calculations and selections followed the guidelines of Borenstein et al. (2009). First, if data from the same experiments were published in different papers, we only counted the first-reported effect size (e.g. Jia et al., 2012). In other words, the 84 studies included in this meta-analysis were independent of each other. Second, some studies did not report *M* and *SD* for the experimental and control groups. Thus, we calculated effect size based on the available data the study provided, such as *t*-values, *p*-values, etc. (e.g. Lee et al., 2013; Sartepeci & Cakir, 2015). Third, when a study reported two sub-groups (e.g. males vs. females) that were not related to our meta-analysis (e.g. Afrilyasanti et al., 2016; Wang et al., 2018), we combined the sub-group data to generate a single effect size. Fourth, when some studies provided multiple data on the same outcome domain, we conducted an individual, fixed-effect meta-analysis to calculate the average effect size of this domain (e.g. Kirvan et al., 2015; Lin et al., 2017). Fifth, if a study reported results for different periods (e.g. unit tests, final exams, immediate post-test and delayed post-test), we selected only the immediate post-test data (e.g. Jia et al., 2013; Graziano & Hall, 2017).

2.6 | Data analyses

In this meta-analysis, we used the comprehensive meta-analysis (CMA 3.0) software to perform an effect size synthesis, moderator analyses, and a publication bias test. For the overall effect size estimate, we used a random-effects model rather than a fixed-effects model. This choice was made because differences in effect sizes are due to sampling error under the fixed-effects model; whereas, under the random-effects model, systematic differences between studies are also considered (Borenstein et al., 2009). There are many differences in the selected studies, such as their study population and interventions; as such, the random-effects model was determined to be most appropriate (Cooper, 2017; Lipsey & Wilson, 2001). Therefore, the findings are highly practical and can be generalized to different educational contexts (Borenstein et al., 2009).

2.6.1 | Overall effect size

When calculating the overall effect size, a study uses only one effect size as an independent unit of meta-analysis to avoid statistical dependence and bias (Lipsey & Wilson, 2001). Forest plots were used to describe the effect size of each study, and one-study removal analysis was used to check for any outliers that might distort the overall results (Borenstein et al., 2009). Additionally, to further compare the differences in effect size across student performance, separate meta-analyses were conducted for the three dependent variables.

The *Q* statistic and the I^2 statistic (Borenstein et al., 2009) were used to test for statistical heterogeneity between studies. The statistic *Q* obeys a cardinality distribution with *k*-1 degrees of freedom (*df*), which indicates the presence of heterogeneity when the *p*-value is less than 0.05. The I^2 statistic was used to measure the proportion of variance in the individual study effect measure due to non-sampling error. The I^2 value of 25%, 50%, and 75% indicates low, medium, and high levels of heterogeneity, respectively (Higgins et al., 2003). Heterogeneity is unavoidable due to the clinical, methodological and statistical diversity of studies (Higgins et al., 2003). However, when heterogeneity is found between studies, Borenstein et al. (2009) suggest that moderate analyses can be performed to further explore potential factors of heterogeneity, which is more meaningful than simply quantifying heterogeneity (Ioannidis et al., 2008).

2.6.2 | Moderator analyses

We performed a moderator analysis using a mixed-effects model to analyse which factors enhance or inhibit the overall effect of blended learning on student academic performance in K-12 contexts. For categorical variables in the coding, we used the sub-group analysis method of CMA and the meta-regression method for continuous variables. As Lipsey and Wilson (2001) suggest, we also used *Q*-tests to examine heterogeneity between categories. When Q_{between} (Q_B) is significant, it indicates significant heterogeneity between categories and can partially explain the heterogeneity between effects size (Hedges & Pigott, 2004).

2.6.3 | Publication bias

We used Egger's regression test, the funnel plot with the Trim and Fill method, and the Fail-safe *N* to test publication bias, because the results of this meta-analysis may be affected by multiple publication biases (Egger et al., 1997). Egger's regression test reports publication bias as a quantitative result with high statistical power. A funnel plot is a simple distribution of each effect size of studies. This method assumes that in the absence of bias, the plot will resemble a symmetrical, inverted funnel (Egger et al., 1997). If publication bias is present, then the Trim and Fill method can trim the mismatched values and fill in the missing values in the distribution to obtain a more symmetrical funnel plot, and then recalculate the overall effect size (Duval & Tweedie, 2000). Fail-safe *N* is a procedure for assessing whether publication bias (if present) can be safely ignored; if the value of fail-safe *N* is greater than $5n + 10$ (*n* is the number of studies included in the meta-analysis), then it indicates that the estimated effect size of the unpublished studies is unlikely to affect the overall effect size of the meta-analysis (Rosenthal, 1979).

3 | RESULTS

This meta-analysis included 72 publications that reported on 84 independent experiments, with a total of 112 effect sizes assessing the learning outcomes of 30,377 K-12 students. The 112 effect sizes ranged from $g = -2.27$ to $g = 2.73$ (see Appendix C for effect sizes and codes for 84 studies).

3.1 | Overall effect of blended learning on K-12 student

The forest plots (Figure 2) summarized the *g*, stand error (*SE*), 95% confidence interval (95% CI), and *p*-values of each study under the random-effects model. As illustrated in Figures 2, 84 studies had an overall Hedges' *g* of 0.65 ($p < 0.001$) with a 95% CI of 0.54–0.77. According to Cohen's (1992) classification of effect sizes (i.e., $d = 0.2$, 0.5 and 0.8 as small, medium and large effect size, respectively), the effect of blended learning has a significantly medium magnitude for K-12 student performance. The one-study removal analysis did not monitor extreme outliers and the overall effect size remained within the 95% CI, indicating that there were no studies that would distort the overall results.

The overall test for heterogeneity of effect sizes indicated that the study was heterogeneous ($Q_B = 1112.26$, $df[Q] = 83$, $p < 0.001$, $I^2 = 92.54\%$). This also demonstrates that one or more factors other than sampling error may have contributed to the heterogeneity of these effect sizes.

Further, we synthesized the observed effect size between different student performance domains. Table 4 displays the average effect sizes for the cognitive, affective and psychomotor domains. The strongest effect size occurred in the cognitive domain ($g = 0.74$) with a significant effect ($p < 0.001$). The average effect on the affective domain was smaller than the cognitive domain, with a mean effect size of $g = 0.52$ ($p < 0.001$). For the psychomotor domain, there was a small and significant effect size ($g = 0.46$, $p < 0.001$). The results of the Q-tests for three dependent variables were significant, and I^2 was greater than 75%, which further illustrates the presence of high heterogeneity between effect sizes.

3.2 | Moderator analysis

Tables 5–7 show the results of the moderator analyses under the random-effects model. Studies without information about moderator variables were included in the analyses (named ‘not reported’).

3.2.1 | The characteristics of the blended learning design

Table 5 presents the variables related to the blended learning design characteristics of the studies. The first variable is the type of blended learning model. Most studies used flipped classrooms ($k = 48$) for blended learning, which had a large effect ($g = 0.79$, $p < 0.001$). The station-rotation model and the ‘Others or not reported’ category both had an effect size of $g = 0.50$ ($p < 0.001$). The flex model also produced a close to moderately significant positive effect ($g = 0.45$, $p < 0.01$), while the lab-rotation model had the smallest effect size ($g = 0.30$, $p > 0.05$) without being statistically significant. However, there was no significant difference between blended learning models ($Q_B = 7.83$, $p > 0.05$).

The effect is better when using a LMS ($g = 0.71$, $p < 0.001$) and “others” ($g = 0.62$, $p < 0.001$) for online learning, compared to using only video ($g = 0.42$, $p > 0.05$). However, there was no statistically significant between-levels variance ($Q_B = 1.18$, $p > 0.05$).

Blended learning interventions that used a mixture of synchronous and asynchronous learning ($g = 1.01$, $p < 0.01$), or that relied on asynchronous learning alone ($g = 0.71$, $p < 0.001$),

Study name	Outcome	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Hecht & Close, 2002	Combined	0.654	0.129	0.017	0.400	0.907	5.058	0.000
vernadakis et al., 2004	Combined	0.278	0.348	0.121	-0.404	0.959	0.799	0.424
Macaruso et al.,2006	Cognitive dimension	0.065	0.154	0.024	-0.238	0.367	0.419	0.675
Ozmen,2008	Combined	1.906	0.339	0.115	1.242	2.570	5.624	0.000
Vernadakis et al.,2008	Cognitive dimension	0.268	0.280	0.078	-0.280	0.816	0.957	0.338
Korkmaz & Karacus, 2009	Combined	0.811	0.277	0.076	0.269	1.353	2.933	0.003
Securro et al., 2010	Cognitive dimension	0.895	0.226	0.051	0.453	1.338	3.967	0.000
Jia et al.,2012	Cognitive dimension	0.280	0.204	0.041	-0.119	0.679	1.375	0.169
Yapici & Akbayin,2012	Cognitive dimension	1.541	0.220	0.048	1.109	1.973	6.997	0.000
Chandra & Watters,2012	Combined	0.253	0.180	0.033	-0.100	0.606	1.403	0.161
Smith, 2013	Cognitive dimension	0.756	0.381	0.145	0.008	1.503	1.981	0.048
Jia et al.,2013	Combined	0.111	0.092	0.009	-0.070	0.292	1.202	0.229
Lee et al., 2013	Cognitive dimension	0.041	0.316	0.100	-0.579	0.661	0.130	0.896
Pane et al., 2013	Cognitive dimension	-0.045	0.014	0.000	-0.074	-0.017	-3.159	0.002
Schultz et al., 2014	Cognitive dimension	0.679	0.260	0.068	0.169	1.190	2.608	0.009
Kazu & Demirkol, 2014	Cognitive dimension	0.567	0.274	0.075	0.031	1.104	2.072	0.038
Wendt & Rockinson-Szapkiw,2015	Affective dimension	0.311	0.162	0.026	-0.007	0.629	1.920	0.055
Al-Madani,2015	Combined	0.266	0.365	0.133	-0.450	0.982	0.727	0.467
Schechter et al. ,2015	Cognitive dimension	0.261	0.219	0.048	-0.169	0.691	1.191	0.234
Saritepeci & Cakir, 2015	Combined	0.372	0.194	0.038	-0.008	0.752	1.918	0.055
Chao et al., 2015	Combined	0.908	0.219	0.048	0.480	1.337	4.153	0.000
Tsai et al. 2015	Cognitive dimension	0.589	0.205	0.042	0.187	0.990	2.875	0.004
Charles-Ogan & Williams, 2015	Cognitive dimension	1.673	0.232	0.054	1.218	2.127	7.214	0.000
Kirvan et al., 2015	Psychomotor dimension	0.503	0.161	0.026	0.187	0.819	3.124	0.002
Yousefzadeh & Salimi, 2015(1)	Cognitive dimension	1.631	0.323	0.104	0.999	2.264	5.056	0.000
Yousefzadeh & Salimi, 2015(2)	Cognitive dimension	0.515	0.283	0.080	-0.040	1.070	1.818	0.069
Yousefzadeh & Salimi, 2015(3)	Cognitive dimension	1.826	0.333	0.111	1.174	2.479	5.485	0.000
Yousefzadeh & Salimi, 2015(4)	Cognitive dimension	1.575	0.320	0.102	0.948	2.201	4.923	0.000
Yousefzadeh & Salimi, 2015(5)	Cognitive dimension	1.223	0.304	0.092	0.627	1.819	4.021	0.000
Nair & Bindu, 2016	Cognitive dimension	0.270	0.217	0.047	-0.156	0.696	1.243	0.214
Sulisworo et al., 2016	Cognitive dimension	2.219	0.315	0.099	1.601	2.837	7.037	0.000
Bhagat et al., 2016	Combined	0.666	0.177	0.031	0.318	1.013	3.757	0.000
Al-Harbi & Alshumaimeri, 2016	Cognitive dimension	0.326	0.302	0.091	-0.267	0.918	1.077	0.281
Afrilyasanti et al., 2016	Psychomotor dimension	2.735	0.351	0.123	2.046	3.423	7.788	0.000
Casem, 2016	Combined	0.805	0.411	0.169	-0.001	1.610	1.959	0.050
Leo & Puzio, 2016	Cognitive dimension	0.156	0.242	0.058	-0.318	0.629	0.645	0.519
Mohanty & Parida, 2016(1)	Cognitive dimension	1.164	0.226	0.051	0.721	1.608	5.145	0.000
Mohanty & Parida, 2016(2)	Cognitive dimension	0.308	0.210	0.044	-0.104	0.720	1.464	0.143
Atwa et al., 2016	Cognitive dimension	0.500	0.193	0.037	0.121	0.879	2.588	0.010
Graziano & Hall, 2017	Cognitive dimension	0.428	0.227	0.051	-0.016	0.873	1.888	0.059
Akgunduz & Akinoglu, 2017	Combined	0.765	0.292	0.085	0.194	1.337	2.625	0.009
Cimen & Yilmaz, 2017	Cognitive dimension	0.879	0.250	0.062	0.390	1.369	3.520	0.000
Ige & Hlalele, 2017	Cognitive dimension	2.008	0.366	0.134	1.292	2.725	5.494	0.000
Lin et al., 2017	Combined	0.727	0.210	0.044	0.315	1.138	3.459	0.001
Olakanmi, 2017	Combined	1.067	0.263	0.069	0.552	1.582	4.062	0.000
Sezer, 2017	Combined	0.887	0.252	0.063	0.394	1.380	3.524	0.000
Aidinopoulou & Sampson, 2017	Combined	1.063	0.313	0.098	0.449	1.676	3.396	0.001
Kostaris et al.,2017	Combined	1.389	0.245	0.060	0.908	1.870	5.660	0.000
Abdelrahman et al., 2017	Psychomotor dimension	1.110	0.274	0.075	0.572	1.647	4.046	0.000
Tugun et al.,2017	Cognitive dimension	1.847	0.328	0.108	1.203	2.491	5.624	0.000
Sergis et al., 2017(1)	Combined	1.728	0.352	0.124	1.039	2.417	4.916	0.000
Sergis et al., 2017(2)	Combined	1.163	0.338	0.114	0.502	1.825	3.447	0.001
Sergis et al., 2017(3)	Combined	1.725	0.370	0.137	1.000	2.449	4.666	0.000
Lo et al., 2017(1)	Cognitive dimension	0.720	0.275	0.075	0.181	1.258	2.621	0.009
Lo et al., 2017(2)	Cognitive dimension	0.293	0.128	0.016	0.041	0.544	2.281	0.023
Lo et al., 2017(3)	Cognitive dimension	0.989	0.419	0.176	0.168	1.811	2.360	0.018
Lo et al., 2017(4)	Cognitive dimension	0.072	0.410	0.168	-0.733	0.876	0.175	0.861
Lam et al., 2018	Cognitive dimension	0.725	0.293	0.086	0.151	1.300	2.473	0.013
Cetin & Ozdemir, 2018	Combined	0.261	0.126	0.016	0.014	0.508	2.068	0.039
Wong et al., 2018	Combined	0.317	0.165	0.027	-0.006	0.639	1.923	0.054
Utami, 2018	Cognitive dimension	1.403	0.279	0.078	0.857	1.949	5.036	0.000
Wang et al., 2018	Psychomotor dimension	0.218	0.169	0.028	-0.113	0.548	1.291	0.197
Zupanec et al., 2018	Cognitive dimension	2.132	0.236	0.056	1.670	2.594	9.048	0.000
Alsalhi et al., 2019	Cognitive dimension	1.206	0.205	0.042	0.804	1.608	5.885	0.000
CH & Saha, 2019	Cognitive dimension	0.728	0.320	0.103	0.100	1.356	2.272	0.023
Almasseri & AlHojailan, 2019	Cognitive dimension	0.169	0.242	0.059	-0.306	0.643	0.697	0.486
Hwang et al., 2019	Combined	-0.341	0.198	0.039	-0.729	0.047	-1.724	0.085
Fazal & Bryant, 2019	Cognitive dimension	0.298	0.099	0.010	0.104	0.491	3.013	0.003
Tse et al., 2019(1)	Affective dimension	-0.796	0.180	0.032	-1.149	-0.443	-4.422	0.000
Tse et al., 2019(2)	Affective dimension	-1.736	0.266	0.071	-2.257	-1.215	-6.526	0.000
Yang et al., 2019	Cognitive dimension	-0.059	0.213	0.045	-0.476	0.358	-0.276	0.783
Edward et al., 2019	Cognitive dimension	1.068	0.118	0.014	0.836	1.301	9.019	0.000
Inal & Korkmaz, 2019	Combined	-0.809	0.362	0.131	-1.518	-0.101	-2.238	0.025
Seage & Turegun, 2020	Cognitive dimension	0.805	0.182	0.033	0.448	1.162	4.421	0.000
Hijazi & AlNatour, 2020	Combined	0.990	0.211	0.044	0.577	1.403	4.698	0.000
Erbil & Kocabas, 2020	Combined	0.701	0.207	0.043	0.295	1.107	3.383	0.001
Seitan et al., 2020	Cognitive dimension	1.671	0.362	0.131	0.962	2.381	4.618	0.000
Macaruso et al., 2020(1)	Cognitive dimension	0.089	0.078	0.006	-0.064	0.241	1.141	0.254
Macaruso et al., 2020(2)	Cognitive dimension	0.153	0.037	0.001	0.080	0.226	4.129	0.000
Wilkes et al., 2020	Psychomotor dimension	0.085	0.095	0.009	-0.101	0.271	0.899	0.369
Tan et al., 2020	Combined	-0.078	0.285	0.081	-0.636	0.481	-0.273	0.785
Gelgoot et al., 2020	Cognitive dimension	-0.117	0.314	0.099	-0.733	0.499	-0.372	0.710
Lo & Hew, 2020	Affective dimension	-0.337	0.157	0.025	-0.645	-0.029	-2.146	0.032
Nunez et al.,2020	Affective dimension	0.562	0.262	0.069	0.048	1.076	2.144	0.032
		0.653	0.059	0.004	0.537	0.769	11.028	0.000

Statistics for each study

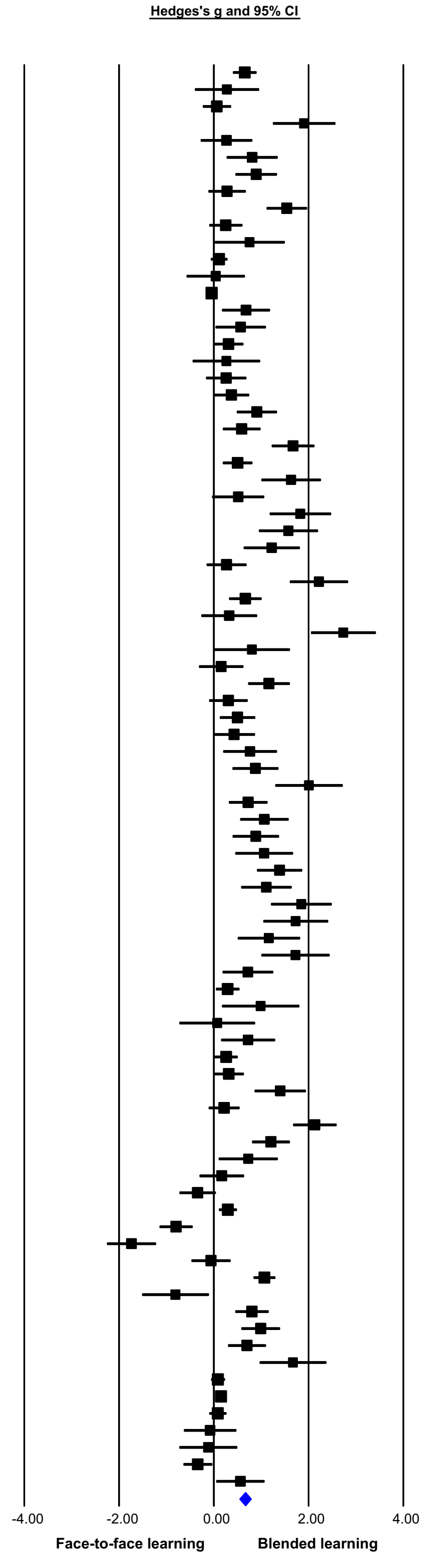

FIGURE 2 Forest plot

yielded larger effect size than in cases where the approach was not reported ($g = 0.60$, $p < 0.001$). However, the difference is not significant.

Interestingly, the variable of group activities produced statistically significant between-level variance ($Q_B = 15.21$, $p < 0.001$). The effect size of blended learning with group activities ($g = 0.94$,

$p < 0.001$) was greater compared to without group activities ($g = 0.18$, $p > 0.05$).

3.2.2 | The characteristics of the educational context

Table 6 provides the results of the moderator variable analysis of the educational context characteristics. The overall effect of blended learning on students at different educational levels reflects significant differences between levels variance ($Q_B = 25.84$, $p < 0.001$). The overall effect of blended learning on elementary school students ($g = 0.70$, $p < 0.001$) was slightly greater than secondary students ($g = 0.67$, $p < 0.001$), whereas the Kindergarten ($g = 0.36$, $p > 0.05$) and Mixed ($g = 0.09$, $p > 0.05$) categories had small and non-significant effects.

The overall effect of blended learning on the K-12 student performance significantly varied across subjects ($Q_B = 48.29$, $p < 0.001$). The effect size was largest when blended learning was used in a computer course ($g = 1.09$, $p < 0.001$), followed by other types of courses ($g = 0.79$, $p < 0.001$), language course ($g = 0.58$, $p < 0.001$), mathematics course ($g = 0.53$, $p < 0.01$) and physics course ($g = 0.50$, $p < 0.01$). There was a weak positive effect in reading courses ($g = 0.15$, $p < 0.01$). All these effect sizes were statistically significant. In contrast, the effect sizes for blended learning were not statistically significant in the biology ($g = 0.65$, $p > 0.05$) and science ($g = 0.61$, $p > 0.05$) classrooms.

TABLE 4 Results of the univariate random-effects meta-analysis

		Random-effects model							
		Effect size		95% CI		Heterogeneity			
Dependent variable	k	g	SE	Lower	Upper	Q_B	df(Q)	p	I^2
Cognitive domain	73	0.74***	0.07	0.61	0.88	1192.64	72	< 0.001	93.96%
Affective domain	25	0.52***	0.14	0.25	0.79	217.68	24	< 0.001	88.97%
Psychomotor domain	14	0.46***	0.13	0.20	0.72	111.55	14	< 0.001	88.34%

Note: k = number of studies.
***$p < 0.001$.

TABLE 5 Results of the moderator analyses of blended learning design characteristics

		Random-effects model					
		Effect size		95% CI		Heterogeneity	
Moderator variables	k	g	SE	lower	upper	Q_B	p
Type of blended learning model						7.83	0.09
Station-rotation model	9	0.50***	0.13	0.24	0.75		
Lab-rotation model	9	0.30	0.16	−0.02	0.62		
Flipped-classroom model	48	0.79***	0.11	0.57	1.00		
Flex model	6	0.45**	0.18	0.01	0.80		
Others or not reported	12	0.50***	0.13	0.24	0.76		
Media type in online learning						1.18	0.55
Learning management system	57	0.711***	0.07	0.57	0.85		
Video only	12	0.42	0.28	−0.13	0.97		
Others	15	0.62***	0.15	0.33	0.91		
Communication in online learning						1.65	0.44
Asynchronous	27	0.71***	0.10	0.52	0.90		
Asynchronous+ Synchronous	3	1.01**	0.41	0.21	1.80		
Not reported	54	0.60***	0.07	0.46	0.75		
Group activities						15.21	0.00
Yes	41	0.94***	0.12	0.72	1.16		
No	4	0.18	0.33	−0.46	0.82		
Not reported	39	0.43***	0.08	0.28	0.58		

Note: k = number of studies.
**$p < 0.01$. ***$p < 0.001$.

TABLE 6 Results of the moderator analysis of educational context characteristics

		Random-effects model					
		Effect size		95% CI		Heterogeneity	
Moderator variables	k	*g*	*SE*	*Lower*	*Upper*	Q_B	*p*
Educational level						25.84	0.00
Kindergarten	2	0.36	0.28	−0.19	0.91		
Elementary (Grades 1–6)	19	0.70***	0.13	0.44	0.96		
Secondary (Grade 7–12)	62	0.67***	0.08	0.51	0.83		
Mixed	1	0.09	0.09	−0.10	0.27		
Discipline						48.29	0.00
Biology	7	0.65	0.33	−0.01	1.30		
Computer	10	1.09***	0.18	0.73	1.44		
Language	18	0.58***	0.13	0.33	0.84		
Mathematics	13	0.53**	0.16	0.23	0.84		
Physics	7	0.50**	0.16	0.19	0.81		
Reading	5	0.15**	0.05	0.05	0.24		
Science	5	0.61	0.32	−0.01	1.23		
Others	19	0.79***	0.16	0.48	1.12		
Knowledge type						11.00	0.00
Procedural	13	0.85***	0.19	0.48	1.22		
Declarative	6	1.29***	0.23	0.84	1.74		
Declarative + procedural	65	0.56***	0.06	0.44	0.68		
Instructor						15.04	0.00
Same	29	0.54***	0.14	0.27	0.82		
Different	19	0.35***	0.07	0.21	0.49		
Not reported	36	0.91***	0.13	0.66	1.16		

Note: k = number of studies.
**$p < 0.01$. ***$p < 0.001$.

Similarly, there was heterogeneity in the effect sizes of knowledge type ($Q_B = 11$, $p < 0.01$). The overall effect size for declarative knowledge ($g = 1.29$, $p < 0.001$) was larger compared to procedural knowledge ($g = 0.85$, $p < 0.001$) and the combination of declarative and procedural knowledge ($g = 0.56$, $p < 0.001$).

Our findings indicate that the instructor variable significantly moderates the overall effect size ($Q_B = 15.04$, $p < 0.001$). The largest effect size ($g = 0.91$) was seen in studies that did not report the instructor variable. When the experimental and control groups used the same instructor ($g = 0.54$, $p < 0.001$), the effect of blended learning was significantly greater than with a different instructor ($g = 0.35$, $p < 0.001$).

3.2.3 | The characteristics of the research methodology

Table 7 shows that the experimental design does not moderate the overall effect of blended learning ($Q_B = 4.16$, $p > 0.05$). In other words, the difference between the effect sizes in quasi-experimental ($g = 0.61$, $p < 0.001$) and random experimental ($g = 0.99$, $p < 0.001$) designs was not statistically significant.

The results indicate that sample size can significantly affect the overall effect ($Q_B = 8.91$, p < 0.05). Small sample studies produced the largest effect size ($g = 0.80$, $p < 0.001$); medium sample studies produced a moderate effect size ($g = 0.72$, $p < 0.001$); and large sample studies produced the smallest effect size ($g = 0.40$, $p < 0.001$).

The overall effect of blended learning was significantly moderated by the reported intervention duration ($Q_B = 16.66$, $p < 0.001$). Those studies that did not report the intervention duration ($g = 0.85$, $p < 0.001$) had a larger positive effect. Interventions that lasted less than one semester ($g = 0.71$, $p < 0.001$) were considered to be more effective than those lasting more than one semester ($g = 0.29$, $p < 0.001$).

Region also became one of moderators of heterogeneity ($Q_B = 44.64$, $p < 0.001$). Three studies from Africa had the largest mean effects ($g = 1.55$, $p < 0.001$), followed by Europe ($g = 1.22$, $p < 0.001$) and Asia ($g = 0.59$, $p < 0.001$). The studies in North America ($g = 0.30$, $p < 0.001$) showed only a small effect size. The effect sizes in these areas were significant. However, the effect size in Oceania, $g = 0.39$ ($p > 0.05$), was not statistically significant.

3.2.4 | Publication year

We used a meta-regression analysis to test the relationship between the effect and the year of publication. Figure 3 shows that more

TABLE 7 Results of the moderator analysis of research methodology characteristics

		Random-effects model					
		Effect size		95% CI		Heterogeneity	
Moderator variables	*k*	*g*	*SE*	*lower*	*upper*	Q_B	*p*
Experimental design						3.75	0.05
Quasi-experimental	74	0.61***	0.06	0.50	0.73		
Random experimental	10	0.99***	0.19	0.63	1.36		
Sample size						8.91	0.01
Small (N ≤ 50)	21	0.80***	0.15	0.50	1.09		
Medium(51 ≤ N ≤ 100)	45	0.72***	0.10	0.52	0.92		
Large (N ≥ 101)	18	0.40***	0.08	0.24	0.56		
Intervention duration						16.66	0.00
≤1 semester	58	0.71***	0.09	0.53	0.89		
>1 semester	14	0.29***	0.07	0.15	0.43		
Not reported	12	0.85***	0.19	0.47	1.22		
Region						44.64	0.00
Africa	3	1.55***	0.26	1.03	2.06		
Asia	49	0.59***	0.09	0.41	0.78		
Europe	14	1.22***	0.17	0.88	1.55		
North America	16	0.30***	0.07	0.17	0.43		
Oceania	2	0.39	0.22	−0.05	0.83		

Note: *k* = number of studies.
***$p < 0.001$.

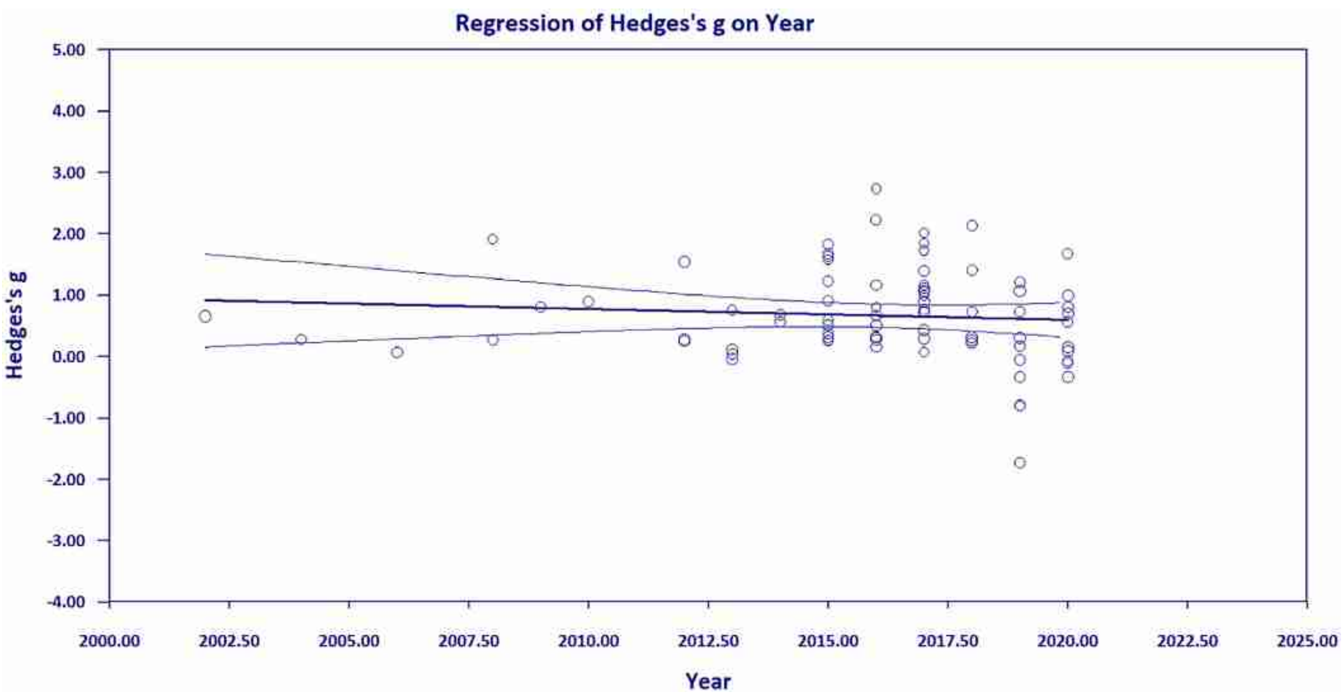

FIGURE 3 Regression of Hedges' *g* on year

studies were published after 2012. There appears to be a slight decrease in effect size from 2000 to 2020, with a correlation coefficient of $r = -0.02$ with publication year, however, there is no statistical significance ($p = 0.40$). Thus, there is no evidence that the reported effect of blended learning in K-12 settings varied based on the year the study was published.

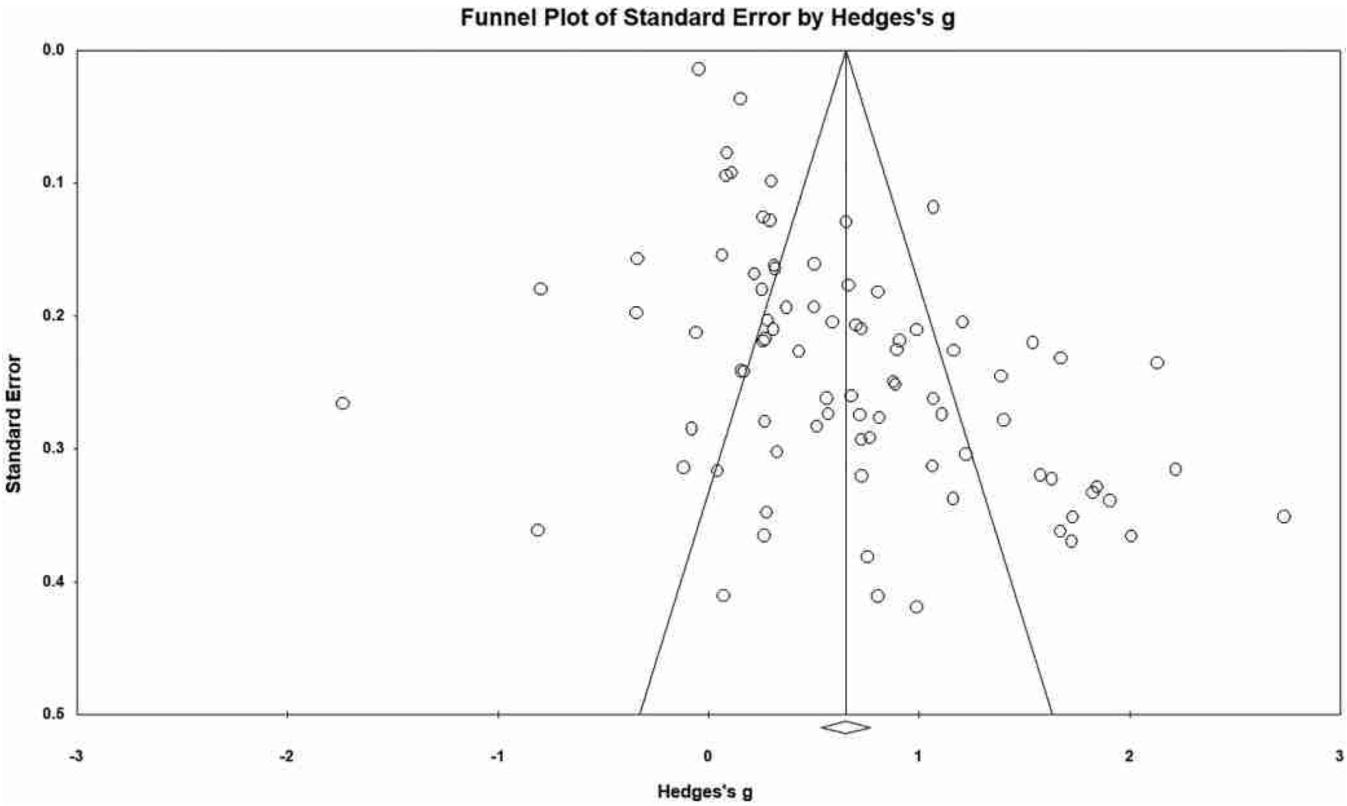

FIGURE 4 Funnel plot of standard error by Hedges' *g*, distribution of all 84 included studies

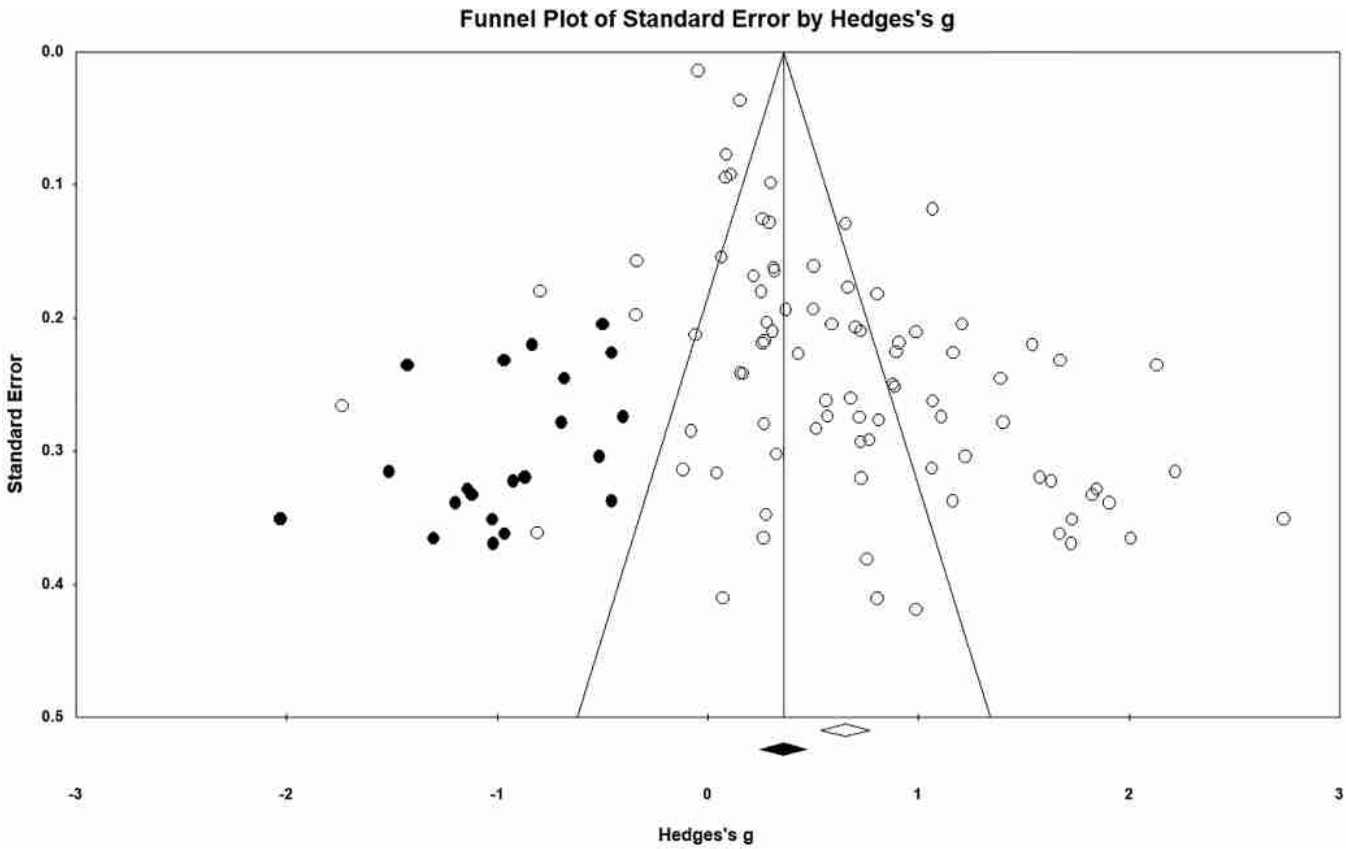

FIGURE 5 Funnel plot of standard error by Hedges' *g*, distribution of 84 observed and 21 imputed studies (black circles)

3.3 | Publication bias

We used four methods to test for publication bias. First, Egger's regression test indicates a potential publication bias in our data ($p < 0.001$). Second, the slight asymmetry of the funnel plot in Figure 4 also indicates a potential risk of publication bias. Therefore, we adjusted the funnel plot using the Trim and Fill method. Figure 5 shows that 21 studies on the left side of the funnel plot were missing. The mean effect of blended learning was recalculated to obtain the adjusted effect size of $g = 0.38$ ($p < 0.001$) after adding the 21 additional values. Finally, the value of the fail-safe *N* shows that only 3147 missing studies with non-significant effects can invalidate the observed overall effect size. The 3147 number exceeds by far the limit of $5n + 10$ studies (430 in this meta-analysis) suggested by Rosenthal (1979). In conclusion, there was a slight publication bias in our meta-analysis, which may be related to unpublished studies with insignificant conclusions or small effects. Although the adjusted effect size ($g = 0.38$) was smaller than the observed effect size ($g = 0.65$), we found no evidence that the positive overall effect size was significantly affected by publication bias. Therefore, we conclude that our finding on the significant, positive effect of blended learning is robust.

4 | DISCUSSION

4.1 | RQ1: How does blended learning affect the overall academic performance of K-12 students compared with F2F learning? What kinds of learning outcomes are more suitable for blended learning?

Our study found a significant medium positive effect of blended learning on K-12 students' performance ($g > 0.5$, Cohen, 1992). This finding is meaningful for educational practice, because the overall effect size in this meta-analysis ($g = 0.65$) is larger than the hinge point ($d = 0.4$) of the average effect of educational interventions found by Hattie (2012). Moreover, according to Hattie (2017), the effect size of 0.65 in blended learning is greater compared to when using gaming/simulations (0.34), intelligent tutoring systems (0.51) and interactive video methods (0.54).

Blended learning appears to have better effects in K-12 settings compared to higher education. The overall effect size is larger than that was found in Bernard et al. (2014) ($g = 0.33$, $p < 0.001$) and Vo et al. (2017) ($g = 0.39$, $p < 0.001$). This may be because the blended learning environment in K-12 classrooms generally has more teacher–student interaction, and teacher control and supervision than in higher education. This means that needs of students for relationships (with the instructor and each other), autonomy, and competence are more likely to be met, generating higher levels of motivation to facilitate learning (Abeysekera & Dawson, 2015).

This overall effect size is also larger than Means et al. (2010). They concluded that the effect of blended learning in K-12 settings is $g = 0.17$ ($p > 0.05$). There may be several reasons for this. One explanation may be the widespread popularity of blended learning in K-12 schools that started after 2012. We included studies from 2000 to 2020, whereas the previous meta-analyses did not include the recent studies. Similarly, technology is not ‘new’ for K-12 students who are part of Generation Z and grew up in the digital, networked era (Prensky, 2001; Fernandez-Cruz & Fernandez-Diaz, Fernández-Cruz & Fernández-Díaz, 2016). Thus, their existing digital literacy skills may make them more comfortable with blended learning projects, leading to clear benefits.

Moreover, our analysis reflects that blended learning positively impacted all student performance dimensions: the effect of the cognitive domain ($g = 0.74$) is larger than the effects for the affective ($g = 0.51$) and psychomotor domains ($g = 0.48$). This finding is consistent with the previous research, which indicates that the effect of blended learning may vary based on different performance domains. For example, previous meta-analyses found that blended learning had a small effect on student satisfaction (Spanjers et al., 2015; Van Alten et al., 2019). Also, Li et al.'s (2019) s work found that blended learning did not significantly improve nursing the students' skill levels. There may be several reasons for these differences. First, students may be more likely to remember, understand, and apply correct information in blended learning environments, thus scoring higher in cognitive tests. However, skill acquisition cannot be achieved quickly. Rather, it is a gradual process, which explains the minimal improvement of the psychomotor domain. Further, students spent more time on blended learning with more tasks compared to traditional F2F learning, which also could easily lead students to have negative attitudes.

4.2 | RQ2: What factors influence the overall effect of blended learning on the K-12 students' performance?

To address the second research question, we analysed the possible moderating roles of blended learning design, educational context, and research methodology and literature characteristics.

First, among the categories of blended learning design characteristics, only group activities were found to significantly moderate the overall effect of blended learning. This finding indicates that designing appropriate group tasks and activities in blended learning environments can enhance its overall impact on the K-12 students' performance. We expected this finding, as many studies have found collaborative learning methods to be effective (Foldnes, 2016; Liu & Beaujean, 2017). Although the differences between several other variables were not statistically significant, the results are important for further exploring beneficial blended learning conditions. For example, our meta-analysis shows that the effect size of a flipped classroom ($g = 0.79$) is larger than other models; the effect size is also higher than the overall effect size of blended learning ($g = 0.65$). In the flipped classroom, students study the material before class (e.g. by watching an online lecture), while engaging in practice and projects in class (Staker & Horn, 2012). The self-paced nature of such pre-classroom activities and personalized, customized instruction can reduce cognitive burdens and thus improve learning outcomes (Abeysekera & Dawson, 2015).

Second, the variables making up educational context characteristics are significant moderators. In terms of educational level, the effect in elementary interventions is slightly larger compared to those in secondary school, whereas the difference is relatively small. The effect sizes for blended learning interventions in kindergarten are not significant; however, it is difficult to draw meaningful conclusions because of the smaller kindergarten sample ($k = 2$). In terms of subjects, the effect size of interventions taking place in computer courses is the most significant ($g = 1.09$), while the effect size in biology ($g = 0.65$, $p > 0.05$) and science ($g = 0.61$, $p > 0.05$) interventions are not significant. The reason for this difference is unclear. It may be that other factors in the study (e.g. the same subject in different grades and different knowledge types in the same subject) may affect the effectiveness of blended learning. Further research is needed to answer this question. Moreover, students effectively acquired declarative knowledge in a blended learning environment more than other types of knowledge. However, this result should be interpreted with caution, because few studies examined declarative knowledge ($k = 3$). Finally, the effect size was greater for interventions that used the same instructor ($g = 0.54$) compared to those that had different ones ($g = 0.35$) in both the experimental and control groups. This may be because the use of different teachers can reduce the effectiveness of blended learning, due to varying levels of professionalism and digital literacy.

Third, our moderator analysis for research methodology characteristics shows that sample size, intervention duration and region can affect the reported overall effect size of blended learning, while the experimental design does not. The overall effect size of the small sample was significantly higher than that of the large sample. Because small sample studies with significantly larger effect sizes are more likely to be published, this may lead to publication bias (Rothstein et al., 2005; Sterne et al., 2000). Also, small sample studies may have lower methodological quality, amplifying the effect size (Slavin & Smith, 2009; Sterne et al., 2000). For example, designers of small sample studies are more likely to use experimental conditions and measures that bias the study towards a large positive effect; these conditions are difficult to replicate in large sample studies. Further, the results were also consistent with previous meta-analyses, in which interventions with duration of less than one term were more effective than those of one term (Hillmayr et al., 2020; Vo et al., 2017). Clark (2015) acknowledges that the novelty effect may be a factor that explains why student performance improves in the short term, when new technologies are adopted. Besides, there is a significant difference in effect size across regions such as Africa ($g = 1.55$) and Europe ($g = 1.22$), where the effect size is higher than the overall effect size. In Oceania, the effect size is small and insignificant. The reason for this discrepancy could be that social, economic and cultural factors influence the students' attitudes towards blended learning. However, because the sample sizes included in Africa ($k = 3$) and Oceania ($k = 2$) were small, we approach this result with caution, and encourage researchers to explore the use of blended learning in these regions in the future studies.

Fourth, there is no evidence that the overall effect of blended learning vary by publication year. Some researchers have posited that recent studies on blended learning feature more advanced technologies and sophisticated instructional designs, which may lead to better results (Cheung & Slavin, 2012). Many researchers have also tried to incorporate emerging technologies into blended learning in K-12 settings (e.g. Lee et al., 2021; Zhang et al., 2020). However, our findings contradict this hypothesis. While this may be due to the influence of other variables, it also may indicate that 'later' does not necessarily mean 'better'. As we focus on the technological improvement of blended learning, it is important to focus on how to use technology to make learning more effective.

4.3 | Limitations and direction for future study

Like all studies, this one had some limitations. First, the current meta-analysis needed to exclude many studies, because they did not meet our inclusion criteria. In the future, researchers could consider combining qualitative analysis to explore the impact of blended learning on the K-12 students' performance.

Second, in developing the classification scheme for each moderator variable, we considered the characteristics of the included studies and the results of relevant studies; however, the classification of each moderator variable may slightly affect the results of moderator analysis. These would benefit from future research.

Third, our moderator analysis mainly focuses on the influence of different factors on the overall effectiveness of blended learning. A systematic review and meta-analysis focus on the effect of blended learning for a single domain (e.g. affective domain and psychomotor domain) of K-12 student development would be valuable, and researchers could further explore potential moderators.

Finally, many studies did not provide very detailed information, so we only considered 13 moderate variables. Also, some dependent variables (e.g. psychomotor domain), educational context characteristics (e.g. Kindergarten, Declarative), and research methodology characteristics (e.g. Africa, Oceania) were only represented in a few studies. This indicates that these results need to be treated with caution. We encourage researchers to initiate additional rigorous and comprehensive studies and explore more impact factors concerning the effect of blended learning.

5 | CONCLUSION

This meta-analysis evaluated 84 studies published from 2000 to 2020; this included synthesizing the effects of blended learning on student performance in K-12 settings. Overall, blended learning had a significant moderate positive effect on improving academic performance compared to traditional F2F learning. Specifically, our finding indicates that blended learning was most effective in improving academic performance in the cognitive domain, followed by the affective domain. Both produced a significant moderate effect size. In the psychomotor domain, blended learning only has a significantly small positive effect. Our study offers the latest supportive evidence that blended learning is a powerful approach for enhancing K-12 students' academic performance.

The moderator analysis results imply that the overall effect may vary due to the specific implementation of blended learning and research methodology. Therefore, we recommend that group activities be added in blended learning programs; we further recommend that educators consider educational context factors, such as educational level, subject, knowledge type and instructor in blended learning practice. In addition, variables related to the included studies themselves, such as sample size, intervention duration and region in which the intervention took place, are found to moderate the overall effect of blended learning.

Given previous inconsistent reports of the effects of blended learning, this meta-analysis fills specific research gaps. These results also enrich our understanding of the effective design of blended learning and provide a basis for improving blended learning practices in K-12 classrooms. We encourage researchers to conduct more high-quality and large-scale research, with a particular focus on the influence of different factors in blended learning design, to develop more effective blended forms of learning in K-12 schools.

AUTHOR CONTRIBUTIONS

All authors contributed equally to this work. All authors wrote, reviewed, and commented on the manuscript. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to thank the editors and the anonymous reviewers for their comments, which have greatly improved our paper.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was funded by the National Office for Education Sciences Planning of China (BAA170020).

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12696.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Shuqin Li https://orcid.org/0000-0002-3161-5548
Weihua Wang https://orcid.org/0000-0001-6652-8332

REFERENCES

Marked with asterisks are studies included in the meta-analysis.

Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: Definition, rationale and a call for research. *Higher Education Research and Development*, *34*(1), 1–14. https://doi.org/10.1080/07294360.2014.934336

*Afrilyasanti, R., Cahyono, B. Y., & Astuti, U. P. (2016). Effect of flipped classroom model on Indonesian EFL students' writing achievement across cognitive styles. *International Journal of English Language and Linguistics Research*, *4*(5), 65–81. https://doi.org/10.21093/di.v19i1.1479

*Aidinopoulou, V., & Sampson, D. G. (2017). An action research study from implementing the flipped classroom model in primary school history teaching and learning. *Journal of Educational Technology & Society*, *20*(1), 237–247.

Akçayir, G., & Akçayir, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, *126*(9), 334–345. https://doi.org/10.1016/j.compedu.2018.07.021

Alabama State Department of Education. (2018). *Alabama high school graduation requirements*. Retrieved from https://www.alsde.edu/sec/sct/Graduation%20Information/AHSG%20Requirements%20May%202018.pdf

Arnett, T. (2019). *One major barrier to high-quality blended learning*. Retrieved from https://www.christenseninstitute.org/blog/one-major-barrier-to-high-quality-blended-learning/?_sft_topics=k-12-education

Arnett T. (2020). The blended learning models that can help schools reopen. Retrieved from https://www.christenseninstitute.org/blog/the-blended-learning-models-that-can-help-schools-reopen/

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, *30*, 1–20. https://doi.org/10.1016/j.edurev.2020.100322

Barbour, M. K., & LaBonte, R. (2017). *State of the nation study: K-12 e-learning in Canada*. Canadian E-Learning Network.

Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, *26*(1), 87–122. https://doi.org/10.1007/s12528-013-9077-3

*Bhagat, K. K., Chang, C. N., & Chang, C. Y. (2016). The impact of the flipped classroom on mathematics concept learning in high school. *Educational Technology and Society*, *19*(3), 134–142.

Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, the classification of educational goals*. McKay.

Bonk, C. J., & Graham, C. R. (Eds.). (2005). *Handbook of blended learning: Global perspectives, local designs*. Pfeiffer.

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

Chandra, V., & Lloyd, M. (2008). The methodological nettle: ICT and student achievement. *British Journal of Educational Technology*, *39*(6), 1087-1098. https://doi. org/10.1111/j.1467-8535.2007.00790.x

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

Cheung, A. C. K., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, *7*(3), 198–215. https://doi.org/10.1016/j.edurev.2012.05.002

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

Cheung, A., & Slavin, R. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283-292. https://doi.org/10.3102/0013189X16656615

Clark, K. (2015). The effects of the flipped model of instruction on student engagement and performance in the secondary mathematics classroom. *The Journal of Educators Online*, *12*(1), 91–115. https://doi.org/10.9743/jeo.2015.1.5

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Collis, B. A., & Williams, R. L. (2001). Cross-cultural comparison of gender differences in adolescents' attitudes toward computers and selected school subjects. *The Journal of Educational Research*, *81*(1), 17–27. https://doi.org/10.1080/00220671.1987.10885792

Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). Sage Publications.

Dabbagh, N., & Kitsantas, K. (2005). Using web-based pedagogical tools as scaffolds for self-regulated learning. *Instructional Science*, *33*(5), 513–540. https://doi.org/10.1007/s11251-005-1278-3

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, *8*(1), 1–44. doi:10.14507/epaa.v8n1.2000

Dey, P., & Bandyopadhyay, S. (2019). Blended learning to improve quality of primary education among underprivileged school children in India. *Education and Information Technologies*, *24*(3), 1995–2016. https://doi.org/10.1007/s10639-018-9832-1

Dias, S. B., & Diniz, J. A. (2013). Fuzzyqoi model: A fuzzy logic-based modelling of users' quality of interaction with a learning management system under blended learning. *Computers & Education*, *69*, 38–59. https://doi.org/10.1016/j.compedu.2013.06.016

Drysdale, J. S., Graham, C. R., Spring, K. J., & Halverson, L. R. (2013). An analysis of research trends in dissertations and theses studying blended learning. *The Internet and Higher Education*, *17*, 90–100. https://doi.org/10.1016/j.iheduc.2012.11.003

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education & Technology*, *18*(1), 7–22. https://doi.org/10.1007/s10956-008-9119-1

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

*Fazal, M., & Bryant, M. (2019). Blended learning in middle school math: The question of effectiveness. *Journal of Online Learning Research*, *5*(1), 49–64.

Fernández-Cruz F. J., & Fernández-Díaz, M. J. (2016). Teachers generation Z and their digital skills. *Comunicar*, *24*(46), 97–105. https://doi.org/10.3916/C46-2016-10

Flores, L. (2018). *Taking stock of 2017: What we learned about personalized learning*. Retrieved from https://www.christenseninstitute. org/blog/taking-stock-2017-learned-personalized-learning/.

Foldnes, N. (2016). The flipped classroom and cooperative learning: Evidence from a random experiment. *Active Learning in Higher Education*, *17*(1), 39–49. https://doi.org/10.1177/1469787415616726

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). Preparing for life in a digital age: The IEA international computer and information literacy study international report. http://link.springer.com/content/pdf/bfm%3A978-3-319-14222-7%2F1.pdf

Freeland, J. (2015). *Three false dichotomies in blended learning*. Clayton Christensen Institute.

Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, *7*(2), 95–105. https://doi.org/10.1016/j.iheduc.2004.02.001

Garrison, D. R. (2011). *E-learning in the 21st century: A framework for research and practice*. Taylor & Francis.

*Gelgoot, E. S., Bulakowski, P. F., & Worrell, F. C. (2020). Flipping a classroom for academically talented students. *Journal of Advanced Academics*, *31*(4), 1–19. https://doi.org/10.1177/1932202X20919357

Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In C. J. Bonk & C. R. Graham (Eds.), *(in press). Handbook of blended learning: Global perspectives, local designs*. Pfeiffer Publishing.

*Graziano, K. J., & Hall, J. D. (2017). Flipped instruction with English language learners at a newcomer high school. *Journal of Online Learning Research*, *3*(2), 175-196. Retrieved from https://files.eric.ed.gov/fulltext/EJ1151092.pdf

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning* (1st ed.). Routledge.

Hattie, J. A. (2017). *Visible learning plus: 250+ influences on student achievement*. Retrieved from https://us.corwin.com/sites/default/files/250_influences_10.1.2018. pdf.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*(2), 490–499. https://doi.org/10.1037/0033-2909.92.2.490

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, *9*(4), 426–445. https://doi.org/10.1037/1082-989X.9.4.426

Higgins, J., Thompson, S., Deeks, J., & Altman, D. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hilliard, A. T. (2015). Global blended learning practices for teaching and learning, leadership and professional development. *Journal of International Education Research*, *11*(3), 179–187. https://doi.org/10.19030/jier.v11i3.9369

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, *153*, 1–25. https://doi.org/10.1016/j.compedu.2020.103897

Horn, M B & Staker, H (2011). *The rise of K-12 blended learning*. Retrieved from http://www.christenseninstitute.org/wp-content/uploads/2013/04/The-rise-of-K-12-blended-learning.pdf.

*Hwang, R. H., Lin, H. T., Sun, J. C. Y., & Wu, J. J. (2019). Improving learning achievement in science education for elementary school students via blended learning. *International Journal of Online Pedagogy and Course Design*, *9*(2), 44–62. https://doi.org/10.4018/IJOPCD.2019040104

iNACOL. (2014). *Transforming K-12 rural education through blended learning: Teacher perspectives*. Retrieved from http://www.inacol.org/resources_August 2015

Ioannidis, J. P. A., Patsopoulos, N. A., & Rothstein, H. R. (2008). Reasons or excuses for avoiding meta-analysis in forest plots. *British Medical Journal*, *336*, 1413–1415. https://doi.org/10.1136/bmj.a117

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*(11), 1515–1529. https://doi.org/10.1007/s10639-017-9676-0

Jia, J., Chen, Y., Ding, Z., Bai, Y., Yang, B., Li, M., & Qi, J. (2013). Effects of an intelligent web-based English instruction system on students' academic performance. *Journal of Computer Assisted Learning*, *29*(6), 556–568. https://doi.org/10.1111/jcal.12016

Jia, J., Chen, Y., Ding, Z., & Ruan, M. (2012). Effects of a vocabulary acquisition and assessment system on students' performance in a blended learning class for English subject. *Computers & Education*, *58*(1), 63–76. https://doi.org/10.1016/j.compedu.2011.08.002

Kettle, M. (2013). Flipped physics. *Physics Education*, *48*(5), 593–596. https://doi.org/10.1088/0031-9120/48/5/593

*Kirvan, R., Rakes, C. R., & Zamora, R. (2015). Flipping an algebra classroom: Analyzing, modeling, and solving systems of linear equations. *Computers in the Schools*, *32*(3–4), 201–223. https://doi.org/10.1080/07380569.2015.1093902

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*(2), 311–328. https://doi.org/10.1037//0021-9010.78.2.311

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review*, *10*, 133–149. https://doi.org/10.1016/j.edurev.2013.02.002

*Lee, C., Cheung, W. K. W., Wong, K. C. K., & Lee, F. S. L. (2013). Immediate web-based essay critiquing system feedback and teacher follow-up feedback on young second language learners' writings: An experimental study in a Hong Kong secondary school. *Computer Assisted Language Learning*, *26*(1), 39–60. https://doi.org/10.1080/09588221.2011.630672

Lee, J. H., Lee, H. K., Jeong, D., Lee, J. E., Kim, T. R., & Lee, J. H. (2021). Developing museum education content: AR blended learning. *International Journal of Art & Design Education*, *40*(3), 473–491. https://doi.org/10.1111/jade.12352

Li, C., He, J., Yuan, C., Chen, B., & Sun, Z. (2019). The effects of blended learning on knowledge, skills, and satisfaction in nursing students: A meta-analysis. *Nurse Education Today*, *82*(8), 51–57. https://doi.org/10.1016/j.nedt.2019.08.004

Li, N., & Kirkup, G. (2007). Gender and cultural differences in internet use: A study of China and the UK. *Computers & Education*, *48*(2), 301–317.

Lim, D. H., Morris, M. L., & Kupritz, V. W. (2007). Online vs. blended learning: Differences in instructional outcomes and learner satisfaction. *Journal of Asynchronous Learning Networks*, *11*(2), 27–42. https://doi.org/10.24059/olj.v11i2.1725

*Lin, Y. W., Tseng, C. L., & Chiang, P. J. (2017). The effect of blended learning in mathematics course. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(3), 741–770. https://doi.org/10.12973/eurasia.2017.00641a

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.

Liu, Q., Peng, W., Zhang, F., Hu, R., & Yan, W. (2016). The effectiveness of blended learning in health professions: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *18*(1), e2. https://doi.org/10.2196/jmir.4807

Liu, S. N. C., & Beaujean, A. (2017). The effectiveness of team-based learning on academic outcomes: A meta-analysis. *Scholarship of Teaching and Learning in Psychology*, *3*, 1–14. https://doi.org/10.1037/stl0000075

*Lo, C. K., & Hew, K. F. (2020). Developing a flipped learning approach to support student engagement: A design-based research of secondary school mathematics teaching. *Journal of Computer Assisted Learning*, *37*(1), 142–157. https://doi.org/10.1111/jcal.12474

Lo, C. K., Hew, K. F., & Chen, G. (2017a). 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

*Lo, C. K., Lie, C. W., & Hew, K. F. (2017b). Applying "first principles of instruction" as a design theory of the flipped classroom: Findings from a collective study of four secondary school subjects. *Computers & Education*, *118*, 150–165. https://doi.org/10.1016/j.compedu.2017.12.003

Lorenzo, A. R. (2017). Comparative study on the performance of bachelor of secondary education students in educational technology using blended learning strategy and traditional face-to-face instruction. *Turkish Online Journal of Educational Technology*, *16*(3), 36–46.

Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Lin, A. J. Q., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society*, *21*(2), 220–232.

*Macaruso, P., Wilkes, S., & Prescott, J. E. (2020). An investigation of blended learning to support reading instruction in elementary schools. *Educational Technology Research and Development*, *68*(6), 2839–2852. https://doi.org/10.1007/s11423-020-09785-2

Means, B., Toyama, Y., Murphy, R., Baki, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Retrieved from https:// www2.ed.gov/rschstat/eval/tech/evidence-based-practices/ finalreport.pdf.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Morgan, K. R. (2002). *Blended learning: A strategic action plan for a new campus*. University of Central Florida.

Nee, C. K. (2014). The effect of educational networking on students' performance in biology. *International Journal on Integrating Technology in Education*, *3*(1), 21–41. https://doi.org/10.5121/ijite.2014.3102

Nortvig, A. M., Petersen, A. K., Helsinghof, H., & Brnder, B. (2020). Digital expansions of physical learning spaces in practice-based subjects-blended learning in art and craft & design in teacher education. *Computers & Education*, *159*(4), 1–11. https://doi.org/10.1016/j.compedu.2020.104020

Osguthorpe, R. E., & Graham, C. R. (2003). Blended learning environments: Definitions and directions. *The Quarterly Review of Distance Education*, *4*(3), 227–233.

Pane, J. F., Steiner, E., Baird, M., Hamilton, L. & Pane, J. D. (2017). *Informing progress: Insights on personalized learning implementation and effects*.: RAND Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RR2042.html

Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, *9*(5), 1–6. https://doi.org/10.1108/10748120110424843

Poirier, M., Law, J. M., & Veispak, A. (2019). A spotlight on lack of evidence supporting the integration of blended learning in K-12 education: A systematic review. *International journal of Mobile and blended learning*, *11*(4), 1–14. https://doi.org/10.4018/IJMBL.2019100101

Powell, A., Rabbitt, B., & Kennedy, K. (2014). *iNACOL blended learning teacher competency framework*. International Association for K-12 Online Learning.

Powell, A., Watson, J., Staley, P., Patrick, S., Horn, M., Fetzer, L., & Verma, S. (2015). *Blended learning: The evolution of online and face-to-face education from 2008–2015*. International Association for K-12 Online Learning.

Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2019). Challenges in the online component of blended learning: A systematic review. *Computers & Education*, *144*, 103701. https://doi.org/10.1016/j.compedu.2019.103701

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *The Journal of the American Medical Association*, *298*(9), 1002–1009. https://doi.org/10.1001/jama.298.9.1002

Rockman, S. (2007). *ED PACE final report*.: Rockman. http://www.rockman.com/projects/146.ies.edpace/finalreport

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

Rossett, A. & Frazee, R. F. (2003). Strategies for building blended learning. Learning circuit. Retrieved from http://www.learningcircuits.org/2003/ju12003/rossett.htm.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. John Wiley.

*Sartepeci, M., & Cakir, H. (2015). The effect of blended learning environments on student motivation and student engagement: A study on social studies course. *Education and Science*, *40*(177), 203–216. https://doi.org/10.15390/EB.2015.2592

*Sergis, S., Sampson, D. G., & Pelliccione, L. (2017). Investigating the impact of flipped classroom on students' learning experiences: A self-determination theory approach. *Computers in*

Human Behavior, *78*(8), 368–378. https://doi.org/10.1016/j.chb.2017.08.011

Shea, P., Vickers, J., & Hayes, S. (2010). Online instructional effort measured through the lens of teaching presence in the community of inquiry framework: A re-examination of measures and approach. *The International Review of Research in Open and Distributed Learning*, *11*(3), 127–154. https://doi.org/10.19173/irrodl.v11i3.915

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology*, *59*(3), 623–664. https://doi.org/10.1111/j.1744-6570.2006.00049.x

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

Spanjers, I. A. E., Könings, K. D., Leppink, J., Verstegen, D. M. L., de Jong, N., Czabanowska, K., & van Merriënboer, J. J. G. (2015). The promised land of blended learning: Quizzes as a moderator. *Educational Research Review*, *15*(3), 59–74. https://doi.org/10.1016/j.edurev.2015.05.001

Staker, H., & Horn, M. B. (2012). *Classifying K-12 blended learning*. Clayton Christenson Institute.

Sterne, J., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in literature. *Journal of Clinical Epidemiology*, *53*(11), 1119–1129.

*Tse, W. S., Choi, L. Y. A., & Tang, W. S. (2017). Effects of video-based flipped class instruction on subject reading motivation. *British Journal of Educational Technology*, *50*(1), 385–398. https://doi.org/10.1111/bjet.12569

Van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, *28*(7), 1–18. https://doi.org/10.1016/j.edurev.2019.05.003

Van Laer, S., & Elen, J. (2017). In search of attributes that support self-regulation in blended learning environments. *Education and Information Technologies*, *22*(4), 1395–1454. https://doi.org/10.1007/s10639-016-9505-x.

Vo, H. M., Zhu, C., & Diep, N. A. (2017). The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation*, *53*, 17–28. https://doi.org/10.1016/j.stueduc.2017.01.002

*Wang, J., Jou, M., Lv, Y., & Huang, C. C. (2018). An investigation on teaching performances of model-based flipping classroom for physics supported by modern teaching technologies. *Computers in Human Behavior*, *84*, 36–48. https://doi.org/10.1016/j.chb.2018.02.018

*Wendt, J. L., & Rockinson-Szapkiw, A. J. (2015). The effect of online collaboration on adolescent sense of community in eighth-grade physical science. *Journal of Science Education and Technology*, *24*(5), 671–683. https://doi.org/10.1007/s10956-015-9556-6

Wigfield, A., Klauda, S. L., & Cambria, J. (2011). Influences on the development of academic self-regulatory processes. In D. H. S. Barry & J. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance*. Routledge. https://doi.org/10.4324/9780203839010.ch3

*Wilkes, S., Kazakoff, E. R., Prescott, J. E., Bundschuh, K., Hook, P. E., Wolf, R., … Macaruso, P. (2020). Measuring the impact of a blended learning model on early literacy growth. *Journal of Computer Assisted Learning*, *36*(2), 595–609. https://doi.org/10.1111/jcal.12429

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, *51*(6), 1991–2005. https://doi.org/10.1111/bjet.13023

Xu, D., Glick, D., Rodriguez, F., Cung, B., Li, Q., & Warschauer, M. (2019). Does blended instruction enhance English language learning in developing countries? Evidence from Mexico. *British Journal of Educational Technology*, *51*(1), 211–227. https://doi.org/10.1111/bjet.12797

Young, J. R. (2002). Hybrid teaching seeks to end the divide between traditional and online instruction. *The Chronicles of Higher Education*, *48*(28), 33–34.

Zhang, H., Yu, L., Ji, M., Cui, Y., Liu, D., Li, Y., …, & Wang, Y. (2020). Investigating high school students' perceptions and presences under VR learning environment. *Interactive Learning Environments*, *28*(5), 635–655. https://doi.org/10.1080/10494820.2019.1709211

How to cite this article: Li, S., & Wang, W. (2022). Effect of blended learning on student performance in K-12 settings: A meta-analysis. *Journal of Computer Assisted Learning*, 1–19. https://doi.org/10.1111/jcal.12696