Computers & Education 164 (2021) 104123

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

Effect sizes and research directions of peer assessments: From an integrated perspective of meta-analysis and co-citation network

Ching-Yi Chang [a], De-Chih Lee [b], Kai-Yu Tang [c,*], Gwo-Jen Hwang [d]

[a] School of Nursing, College of Nursing, Taipei Medical University, 250 Wuxing Street, Taipei, 11031, Taiwan
[b] Department of Information Management, Da-Yeh University, No.168, University Rd., Dacun, Changhua, 51591, Taiwan
[c] Department of International Business, Ming Chuan University, 250 Zhong Shan N. Rd., Sec. 5, Shilin Dist., Taipei, 111, Taiwan
[d] Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, 43, Sec.4, Keelung Rd., Taipei, 106, Taiwan

ARTICLE INFO

Keywords:
Peer assessment
Cooperative/collaborative learning
Interactive learning environments
meta-Analysis
Co-citation network analysis

ABSTRACT

Peer assessments (PA) have been widely explored by researchers over the past few decades. However, their impact on various learning environments remains unclear. Based on a meta-analysis, this study aimed to analyze PA-based research performed across multiple learning environments. Unlike existing meta-analyses, however, this study integrated co-citation network analysis with the meta-analysis to examine the bibliometric relationships between the analyzed studies. This method provides citation-based evidence to secure the association of research data for conducting the meta-analysis. In addition, the visualized research network offers a fresh micro-to-macro understanding of the PA literature. A total of 20 PA-related core articles consisting of their follow-up 340 citing studies were collected based on systematic keyword searches. All analyzed articles were published between 1998 and 2020, and were sourced from two renowned databases, the Web of Science and Scopus, in order to maintain the quality of the data used for this present research. The main findings are twofold. First, the meta-analysis reported that the standard error at the 95% confidence interval for the average difference between the experimental and control groups is [0.288, 0.424], indicating that each effect value in the meta-analysis is heterogeneous. We also found that technology combined with PA teaching methods has certain learning effects in the contexts of web-based problem-solving activities. In addition, following Fu and Hwang's (2018) constructivist theoretical framework, we propose some criteria for conducting peer assessment research. Second, the co-citation network analysis presented a visualized networking structure of 20 PA articles, of which 11 were bibliographically connected. This result echoed the previous research by reviewing the relationships between the studies in the meta-analysis. Moreover, the co-citation network of PA research reveals some heavy co-citation links from the 11 core articles. The results further identified the main research foci include combining PA with technology-enhanced learning environments (e.g., game- and web-based learning) and PA activities for collaborative learning in the field. In conclusion, this study presents an innovative approach to examining the effects of empirical PA studies and, accordingly, offers discussions and directions for future research.

* Corresponding author.
E-mail addresses: frinng.cyc@gmail.com (C.-Y. Chang), dclee@mail.dyu.edu.tw (D.-C. Lee), ky.nctu@gmail.com (K.-Y. Tang), gjhwang.academic@gmail.com (G.-J. Hwang).

https://doi.org/10.1016/j.compedu.2020.104123
Received 27 May 2020; Received in revised form 4 December 2020; Accepted 29 December 2020
Available online 6 January 2021

1. Introduction

Educators have explored various ways of effectively improving students' metacognitive and cognitive skills, learning motivation, creativity, and critical thinking. A series of innovative learning methods and courses using educational technologies and theories have also been used to enhance the effectiveness of student learning. Among them, peer assessment (PA) has been considered an effective strategy to help students reflect on and experience deep thinking during learning activities (Ng, 2016).

Over the past decade, scholars have applied PA strategies in various teaching environments, such as high school science courses (Demir, 2018; Falchikov & Goldfinch, 2000; Harland, Wald, & Randhawa, 2017), higher education (Luaces, Díez, & Bahamonde, 2018; Rico-Juan, Gallego, & Calvo-Zaragoza, 2019), English courses (Samaie, Mansouri Nejad, & Qaracholloo, 2018), and have achieved promising results (Carey, Chick, Kent, & Latour, 2018). PA promotes students' higher-order thinking, analysis, evaluation, and organizational skills (Zheng, Chen, Li, & Huang, 2016). Relevant research has verified the benefits of integrating PA in education (Fu & Hwang, 2018; Kuo, Chen, Chu, Yang, & Chen, 2017). Recently, some researchers have suggested PA as a novel pedagogical strategy involving peer evaluations with scoring methods and feedback to improve learning performance (Chien, Hwang, & Jong, 2019; Lin, Hsia, Sung, & Hwang, 2019). More specifically, the use of PA records helps teachers and educators conduct in-depth analysis and develop a comprehensive understanding of students' learning behaviors, thus supporting the notion that PA helps students think and learn effectively. Hsia and Sung (2020) emphasized the need for appropriate technology support systems when designing PA to facilitate the integration of learning content and to aid students with mutual assessments, integrated thinking, and reflections.

In the meantime, the benefits of conducting review studies have been highlighted by researchers. For example, Lai and Bower (2019) indicated that a quality review could assist researchers, instructors, or educational policymakers to gain a whole picture of the important trends and research streams of a particular field, which is crucial for conducting innovative research, improving teaching quality, and making correct decisions in the future. Zheng, Chen, Cui, and Zhang (2019) conducted a systematic review to analyze some characteristics of the technology-supported peer assessment research, such as evaluation criteria, learning objectives, and tools. Similarly, Tenório, Bittencourt, Isotani, and Silva (2016) used a scoping review to outline the approaches of peer assessment adopted in online learning environments for the understanding of the field.

For a single review study, however, it is hard to systematically evaluate some potential biases of published research results, such as the heterogeneity test of research designs (Yu et al., 2019). Researchers have suggested that a meta-analysis is a quantitative approach for systematically combining the results of previous research to arrive at conclusions about the body of research (Kates, Wu, & Coryn, 2018). For example, based on 48 empirical studies, Falchikov and Goldfinch (2000) used a meta-analysis to examine the effect sizes of student peer assessment in higher education. They found that the overall correlation between a peer and teachers' assessments was 0.69. Similarly, Li and colleagues analyzed a total of 70 eligible studies via meta-analysis to compare the effects of peers' ratings in peer assessment and teachers' views (Li et al., 2016). The results showed that peer and teacher ratings had a moderate correlation of 0.63.

The abovementioned studies conducted some meta-analyses to estimate the overall effect sizes of peer assessments adopted in diversified learning environments. According to Begg's (1994) suggestions, it is critical to identify eligible studies for conducting a meta-analysis. Recently, a methodological paper by Zupic and Čater (2015) have suggested that using bibliometric citations as a measure of objectivity is helpful for evaluating the potential impact of research papers included in a meta-analysis. They also concluded that citation-based sources provide evidence which complements the research review using meta-analysis and some qualitative approaches. Echoing the validity issues of research data, Janssens and Gwinn (2015) have conducted a series of methodological surveys from a medical research perspective to highlight the importance of using a citation-based approach (e.g., co-citation links) to provide direct relationships between selected studies for a meta-analysis. They found that about half of the studies included in the previous meta-analyses had not been accredited by other publications, indicating that the association of such studies may not have been sufficient for inclusion in the meta-analyses. Janssens and his colleagues suggested that the degree of co-citation evidence (e.g., a co-citation relationship between two focal papers) that triangulated from large follow-up citers' view provides a way to increase both the quality and the relevance of core papers used in a meta-analysis (Janssens & Gwinn, 2015; Janssens, Gwinn, Brockman, Powell, & Goodman, 2020).

The previous studies have suggested a useful solution to the problem, that is, using co-citations as an indicator to enhance the validity of the research data used in a meta-analysis. Researchers have also used various approaches to identify the sub-networking structure for the understanding of research streams in the field as well as the follow-up directions for future research. For example, some researchers have used the bibliometric meta-analytic approach to identify the research agenda and the features of meta-knowledge in various areas, including the consumer-brand relationship research (bibliometric citation meta-analysis; Fetscherin & Heinrich, 2015), innovation and entrepreneurial ecosystems (co-citation analysis and network meta-analysis; Zhang & Guan, 2015), and brand personality (bibliometric citation meta-analysis; Radler, 2018).

However, few studies have performed co-citation network analysis and meta-analysis to co-review the literature, because these two forms of analysis both have specific requirements regarding the research data. For example, the correlation coefficient of analyzed constructs is required for meta-analysis, while citations and co-citation counts are necessary for co-citation network analysis. Following the methodological suggestions from previous studies, this study provides the very first attempt at integrating meta-analysis and citation-based network analysis to review the literature. In this study, a meta-analysis was used to specify the effect sizes of the selected PA studies. Some research characteristics and publication biases were also investigated. On the other hand, co-citation network analysis was performed to further examine the representativeness of the research data and to identify the most influential research streams in the field. Through the visualized research network, some potential relationships between the meta-analyzed studies (e.g., extended research directions of the follow-up studies) were also identified. In sum, this study aimed to investigate the

overall effect sizes of peer assessment in various learning environments and to provide research directions based on large citers' perspectives. Detailed research questions are as follows:

(1) What are the characteristics of the major articles in the context of PA?
(2) What is the risk of bias within the articles in the context of PA?
(3) What is the overall effect of PA under each condition?
(4) From citers' perspectives, what are the main streams of the PA research network?

2. Literature review

Researchers have suggested peer assessment (PA) as an effective pedagogical strategy involving peer evaluations with scoring methods and feedback to improve learning performance (Rico-Juan, Gallego, Valero-Mas, & Calvo-Zaragoza, 2018). Studies conducted over the past decade have examined the effects of PA using different methods (Li, 2016). Among them, Zheng, Chen, Cui, and Zhang's (2019) study is a noteworthy exception which integrated a meta-analysis and co-citation network analysis. The authors conducted a comprehensive quantitative analysis of multiple studies that had the same objective but were independent of each other. Further, they increased the accuracy of the statistical testing power and testing results, perfected the hypothesis testing, and applied systematic and clear comparison programs to reduce the subjectivity of their research (Harder & Siedler, 2019).

Several scholars have conducted meta-analysis research by extracting and integrating empirical research data to obtain the overall EFs between different quantitative works and to contrast the average effect of research groups with different characteristics (Borenstein, Hedges, Higgins, & Rothstein, 2009; Harder & Siedler, 2019). According to Hedges and Vevea (1998), a meta-analysis is used to measure effects on the basis of the size of distinct reasoning models. A fixed-effects model is suitable when changes in studies are caused by a sampling error. On the other hand, a random-effects model is needed when the aim is to generalize inferred results beyond a given population (Hedges & Vevea, 1998). Given its objectives, the present study adopted a statistical model with random effects. To obtain the OES, descriptive statistics and independent result data—including distinct results derived from a single analysis or independent group evaluations—were combined to determine the EF. This explains the greater EF compared with those in previous research. We used Cohen's *d* to measure the EFs and convert the value to Hedges' *g* to solve the problem of small sample deviations using the corrected standardized average effects and sample weights (Lipsey & Wilson, 2001).

Even though meta-analysis could obtain the overall EFs between different quantitative works, it is unable to identify the core topics investigated in these articles. Education scholars have suggested that the use of bibliometric co-citation analysis could help field researchers single out the most influential research streams from a scientometrics perspective with longitudinal citation-based evidence (Aryadoust, 2020; Hallinger, 2020; Pan & An, 2020). As such, co-citation network analysis was performed in this study. By combining meta-analysis and a co-citation network, researchers can examine the associations between analyzed articles. It is also expected that the most influential research topics and trends can be identified to provide suggestions for future studies.

2.1. Review of the peer assessment research

In education research, peer assessment (PA) refers to the activities which engage students in reviewing and providing comments and/or ratings on peers' work (Topping, 1998). In the contexts of higher education, several scholars have recognized the benefits of PA for improving students' learning performances and facilitating their in-depth thinking based on the rubrics provided by the teacher (Luaces et al., 2018; Rico-Juan et al., 2019). Moreover, some researchers have reported that, via reviewing peers' work and receiving feedback from peers, students are guided to make reflections and reconstruct knowledge by linking the new and old concepts (Chien et al., 2019); therefore, participating in PA activities enables students to reflect on the problems to be coped with, and hence their learning performances, motivation, and their critical thinking could be promoted (Sanaeifar & Mirshojaee, 2020).

In the past decade, PA has been applied to various courses. For example, Demir (2018) reported that conducting PA activities using social media could promote students' learning engagement with instructional technology and material design course; Chien et al. (2019) adopted the PA strategy in a virtual reality-based English practice context to enhance students' English-speaking performance and self-confidence, and to reduce their language learning anxiety. Li (2019) used AP in a game-based training activity and found that it not only enhanced students' learning achievement but also promoted their learning motivation. In addition, Lin et al. (2019) employed PA in a college flipped dance classroom using mobile technologies to improve the students' dance skills and self-efficacy. From these empirical studies, it can be seen that PA has been widely adopted in various domains with diverse technologies and research directions. In addition, researchers exploring cognitive functions are aware of the core role of PA in digital learning (Li et al., 2016; Panadero & Alqassab, 2019). The result also shows that researchers and practitioners in the field of humanities and education have been extensively studying PA in the design of teaching activities, and continuously developing or comparing various factors. The extent of their research efforts suggests that PA could improve learning efficiency and technology. Further, technology-enhanced PA also supports ubiquitous learning through various teaching strategies, cultivates learners' self-regulation and self-reflection, and promotes cross-cultural interactions. Combining PA teaching with technology has good potential to encourage students' participation in interactive learning environments and peer learning, thus promoting social interactivity and seamless social interactions (Hwang, Hung, & Chen, 2014; Wang, Hwang, Liang, & Wang, 2017). From a perspective of constructivism, Fu and Hwang (2018) show that learners could develop evaluative knowledge by interacting with peers and experts with different viewpoints and by constructing, reflecting on, and transforming the information into personal knowledge. Thus, it is worth exploring technology-based PA learning under different conditions and from various perspectives to strengthen learners' knowledge construction. During technology-based PA

learning, it is also important to consider learners' personal factors and thus, future research should explore learning and cognitive styles. Furthermore, investigating students' long-term PA learning performance and behaviors could offer valuable insights. When adopting structured technology-based PA learning strategies such as shared decision-making in medicine, other strategies should be used to help students establish common learning goals and responsibilities. It is also necessary to encourage and support learner interactions and self-reflections.

In sum, some PA review studies have been conducted (e.g., Li et al., 2006; Fu, Lin, & Hwang, 2019; Panadero & Alqassab, 2019; Zheng et al., 2019). They are systematic reviews that report the research trends and foci of PA. Scholars have pointed out that in order to further confirm the effectiveness of a learning approach, it is necessary to analyze the outcomes of the published articles using statistical methods (Gijbels, Dochy, Van den Bossche, & Segers, 2005). To address this issue, meta-analysis, which is an effective way to evaluate the effectiveness of a learning strategy via statistical analysis (Vo, Zhu, & Diep, 2017), was adopted in the present study.

2.2. Meta-analysis

According to Hedges and Vevea (1998), a meta-analysis is used to measure effects on the basis of the size of distinct reasoning models. A fixed-effects model is suitable when changes in studies are caused by a sampling error. On the other hand, a random-effects model is needed when the aim is to generalize inferred results beyond a given population. Given its objectives, the present study adopts a statistical model with random effects. To obtain the overall effect size (OES), descriptive statistics and independent result data—including distinct results derived from a single analysis or independent group evaluations—are combined to determine EFs. This explains the greater EFs compared with those in previous research. We used Cohen's *d* to measure the EFs and converted the value to Hedges' *g* to solve the problem of small sample deviations using the corrected standardized average effects and sample weights (Lipsey & Wilson, 2001). The criteria for Cohen's *d* are as follows: the effects less than 0.2, between 0.2 and 0.8, and greater than 0.8 are considered small, medium, and large, respectively. Finally, we employed Cochran's *Q* to examine the heterogeneity of the effect in each case. A *p*-value less than 0.05 indicates a significant dispersion of the effect, and thus, a real difference between the groups, whereas a *p*-value greater than 0.05 suggests that the difference between the effects is affected by a sampling error (Hedges, 1992). The Comprehensive Meta-Analysis 3.0 is the software used for the calculation.

In this study, the correlation coefficients were treated as the effect quantity to integrate the relationship between the two coders. During the coding process, certain articles did not directly report the correlation coefficient between the various indicators, although the statistical *F*-value, *t*-value, and $\eta 2$ value were estimated. Four steps were carried out: 1) search for and identify target research articles; 2) compile a coding table according to the research conditions and adjustment variables; 3) calculate the EFs for each condition; and 4) test the EFs of the potential adjustment variables.

2.3. Co-citation network analysis

In line with the previous meta-analyses (Janssens & Gwinn, 2015; Janssens et al., 2020), the main purpose of co-citation network analysis used in this study was to provide the associations between the analyzed data, and to further identify the most frequently referenced research streams in the field. The co-citation network analysis was introduced by Small in 1973 to scientific map out the bibliographic relationships among 11 physics studies (Small, 1973). By definition, the co-citation counts occurred if more than two core papers were cited together by a follow-up citing paper (Small, 1973). As such, the analysis is also used to profile the relationships among selected core papers to reveal highly joint-referenced research topics in the field. In this study, all the selected empirical papers of peer assessment ($n = 20$) were included for the co-citation network analysis. First, all citation data of 20 selected papers were manually collected from the re-updated databases. All citing papers were searched both from the Web of Science and Scopus. As a result, a total of 340 citing references were obtained for the analysis. This practice was to maximize all the possible citation links, but also to maintain the quality of the research. The collection process was carried out over a period of one week and was completed on April 25, 2020. Next, after the co-cited frequency of each pair was obtained, the result of the co-citation matrix was then visualized as a network diagram to reveal the underlying structure of all selected PA articles.

In the education domain in recent years, researchers have adopted the co-citation approach to profile various important topics. For example, Yilmaz, Topu, and TakkaçTulgar (2019) examined the relationships among 596 papers of language learning in the pre-school education context. Based on the results, the authors identified the most referenced journals and the top-three productive countries, as well as some of the most cross-referenced topics (e.g., ICT in foreign language learning) in the field. Similarly, Aryadoust (2020) explored the mainstream of comprehension subskills in the applied linguistics area based on a total of 192 empirical studies. The author found that most of the research focused on the issues of processing abilities in reading comprehension of first language learning, while the rest emphasized the comprehension subskills in second/foreign language learning. For general educational leadership and management research, Hallinger (2020) used the bibliographic technique to map out the relationships of international research collaboration across three emerging regions. The networking diagram showed that the research focus of the literature has become more diversified and the collaborations were more globally oriented than in earlier decades. More recently, two forthcoming articles focused on the development of higher education research (Pan & An, 2020). Using large-scale publication and citation data as evidence, both of these studies' findings indicated some evolutionary characteristics in the research contexts, central themes, and international research collaborations. Following previous review studies, we used the co-citation network analysis to profile and extract the main research foci of 20 selected PA articles based on the joint perspectives of 340 follow-up references.

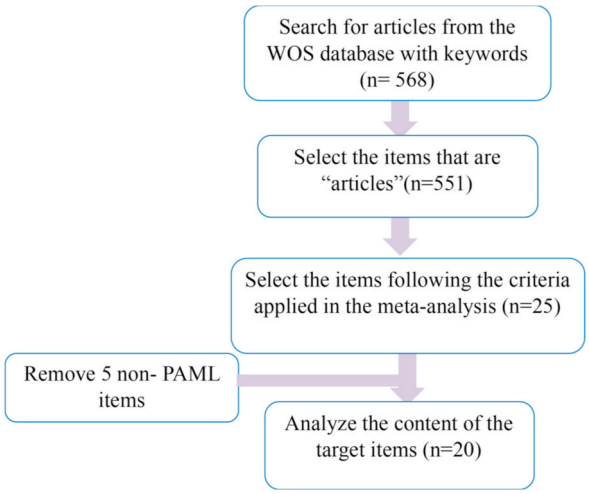

Fig. 1. A step-by-step flowchart of data refining for selecting the PA articles.

3. Methodology

3.1. Data collection

Data selection is critical to conducting a meta-analysis, and thus, this study employed a systematic process (see Fig. 1). First, we performed a comprehensive search of the social science citation index (SSCI) from the Web of Science database. Using keywords such as "peer assessment" and "education," we identified 568 articles published between 1998 and 2019. Second, the following criteria were applied for the meta-analysis: the article must (1) empirically analyze the experimental or survey data, and thus, all theoretical and literature review articles were excluded; (2) have at least one correlation or indicators that can be converted into an effect index; (3) have independent samples and specify the sample size; (4) include information supporting PA activities; (5) be experimental or quasi-experimental; and (6) report data for the experimental and quasi-experimental research groups. To avoid unpredictable deviations caused by publication bias, we adopted a widely accepted method to calculate the fail-safe numbers (Kaminski, Valle, Filene, & Boyle, 2008). A total of 20 articles met the abovementioned criteria.

3.2. Theoretical framework: coding schemes

Following Fu and Hwang (2018), this study adopted a constructivist theoretical framework to review the PA studies (see Fig. 2). The framework recommends considering basic items such as research purpose, learning domain, and sample group to systematically analyze the trends of personal adaptive technologies and learning. From a constructivist perspective, learning is a positive and personal

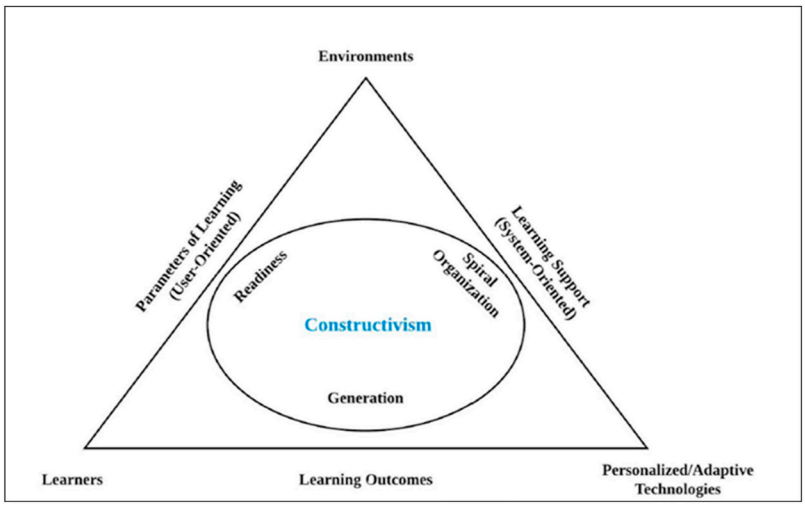

Fig. 2. The coding scheme framework based on constructivism.

process involving learning support, system parameters, and learning results that are based on learner interactions. Highlighting technology-enhanced learning, Hwang (2018) demonstrated an interactive relationship among learners, environments, and technologies. Constructivism theorizes that personal knowledge can be developed in an external environment through experiences, beliefs, and learner interactions. Constructivism has been adopted not only for in-class activities but also to establish an online learning environment supported by mobile devices and other technologies (Tsai, Tsai, & Hwang, 2012). Accordingly, this study adopted the coding schemes based on constructivism (see Table 1) to code PA studies published over the past 21 years, including technologies, learning environments, application areas, PA methods, and research questions (Fu & Hwang, 2018; Zheng et al., 2019).

3.3. Data analysis

This study extracts and integrates empirical research data for the meta-analysis to obtain the overall effect sizes (EFs) between different quantitative works and to contrast the average effect of research groups with different characteristics. The main characteristics of 20 collected articles were analyzed to identify the potential moderators in peer assessment research. The Comprehensive Meta Analysis (CMA) with version 3.307 was adopted to conduct the meta-analysis in the present study, inducing the estimation of the potential publication bias (Research Question 2), and the overall effect size (Research Question 3). To further examine the associations among 20 articles used for the meta-analysis, a network diagram was used to visualize the bibliometric co-citation relationships of 20 articles. The social network software 'UCINET' with 'NETDRAW' version 6.992 (Borgatti, Everette, & Freeman, 2002) was utilized to manage all the citation relationships of research data (i.e., 20 articles with the 340 follow-up citations) and to visualize the co-citation structure of PA research. In this study, the integration of the two analyses (i.e., meta-analysis and citation-based network analysis) aimed to provide researchers with a micro-to-macro understanding of the peer assessment literature. More specifically, based on the joint perspectives of 340 follow-up references, the co-citation network analysis was used to identify the most frequently referenced peer assessment research from 20 selected articles, shedding light on the directions of research streams in the field (Research Question 4).

4. Results

4.1. Article selection and characteristics for meta-analysis (research question 1)

The meta-analysis offers deeper insights into the EF of the selected PA studies, the results of which are presented in Table 2. The 20 studies were published between 2010 and 2019. All of the studies evaluated their participants immediately after the interventions. We found that 13 were web-based peer assessment (WBPA) studies and two were general learning management system (GLMS) studies. Further, 12 were quantitative studies, while eight were both quantitative and qualitative.

4.2. Risk of bias within articles (research question 2)

Rothstein, Sutton, and Borenstein (2005) highlighted the issue of publication bias in meta-analyses. A publication bias indicates that the published articles cannot methodically and comprehensively represent the overall articles in the field. The three main methods used for analyzing potential publication bias are a funnel plot, Rosenthal's classic fail-safe *N* test, and Egger's test (Megías, Gómez-Leal, Gutiérrez-Cobo, Cabello, & Fernández-Berrocal, 2018). Following previous research, this study employed a funnel plot to test for potential bias in the selected studies. Fig. 3 presents the results and shows that the integrated effect is robust (two-tailed test: $\alpha = 0.05$; the number of classic fail-safe $N = 1035$). The funnel plot in this study is symmetrical, indicating no serious bias in this present analysis.

4.3. Meta-analysis: the overall effect of peer assessments (research question 3)

Existing meta-analyses mainly employed a fixed- or random-effects model. A fixed-effects model assumes that all studies have a true effect size (EF) and that the results apply only to the studied population and cannot be generalized to other population groups. The random-effects model reports true effect quantity, and its results vary by study group and research tool. The present meta-analysis employs a random-effects model. We performed a heterogeneity test to verify the rationality of the model selection. More

Table 1
The codebook for the moderators and their subgroup definitions.

Category	Definition
Technologies tools	Dedicated Web-based peer assessment system, general learning management system, social media, and mobile application.
Learning environments	Primary school, junior and senior high school, and higher education
Application domains	Natural science (science, mathematics, physics, biology, geography, medicine, and so on); Social science (politics, history, education, psychology, linguistics, art, and so on); Engineering and technological science (engineering, computer science, educational technology, and so on).
PA method	Assessment method quantitative only, qualitative only, and both quantitative and qualitative.
Research issues	Learning outcomes (learner's competency, learning achievements, motivations, problem-solving skills, performance, metacognitive skills, meta-cognitive awareness, cognitive load, attitude, self-efficacy, learning anxiety, satisfaction, and so on).

Table 2
Characteristics of studies (n = 20).

Article	Technologies tools	Learning environments	Application domains	PA method	Research issues	No. of subjects	EE mean (SD)	DP mean (SD)
Learning and Instruction (2010)	n/a	Junior and senior high school: Dutch secondary-vocational education	Engineering	Quantitative	Psychological safety, Value diversity, Interdependence, Trust, and Conceptions of peer assessment	E = 45C = 17	7.42 (1.56)	n/a
Education and Science (2012)	n/a	Primary School	Mathematics	Quantitative	GWS and CLL checklists	E = 33C = 37	−0.62	−0.15
Journal of The Medical Library Association (2013)	Web-based peer assessment system: basic types of PubMed searches	Higher education: medicine	Natural science	Quantitative	EBM PubMed searching competency	E = 47C = 24	45.7 (4.0)	43.5 (9.5)
Educational Technology Research and Development (2014)	Web-based peer assessment system: peer assessment-based game	Primary school: elementary school	Natural science	Quantitative and qualitative: open-ended questionnaire, Interview	Learning achievements, Motivations, and Problem-solving skills	E = 82C = 85	40.00 (6.47)	38.36 (5.67)
Computers & Education (2015)	Mobile application: Handheld devices	Primary school: fifth graders of an elementary school	Social science	Quantitative	Students' learning achievement, Learning motivation, Meta-cognitive awareness, and Cognitive load	E = 51C = 52	83.06 (5.31)	70.31 (5.33)
Saudi Pharmaceutical Journal (2015)	n/a	High school: pharmacy students	Natural science	Quantitative and qualitative: open questionnaire	academic performance and metacognitive skills	E = 22C = 31	73.5 (3.1)	69.7 (2.8)
Assessment & Evaluation in Higher Education (2016)	Web-based peer assessment system:online peer assessment	Higher education: public university	Engineering	Quantitative	Student learning levels, Project performance	E = 63C = 67	36.06 (1.98)	34.48 (1.49)
Eurasia Journal of Mathematics Science and Technology Education (2016)	Web-based peer assessment system: Facebook-based Online Peer Assessment	Higher education: university students	Social science	Quantitative and qualitative:open questionnaire	Attitude toward peer assessment, Perceived learning	E = 16C = 15	4.08	4.78
British Journal of Educational Technology (2016)	Web-based peer assessment system: web-based peer-assessment approach	Junior and senior high school	Arts	Quantitative	Students' performance, Self-efficacy and motivation, Satisfaction	E = 14C = 15	2.41 (0.50)	1.66 (0.32)
International Journal of Engineering Education (2016)	Web-based peer assessment system: TeachSys e-learning 2.0 platform	Higher education: university	Engineering	Quantitative and qualitative	Learning satisfaction, Academic achievement, Interviews	E = 51C = 60	81.92 (12.33)	75.44 (7.87)

(*continued on next page*)

Table 2 (*continued*)

Article	Technologies tools	Learning environments	Application domains	PA method	Research issues	No. of subjects	EE mean (SD)	DP mean (SD)
Educational Technology & Society (2017)	Web-based peer assessment system: online peer assessment attempt	Junior and senior high school: ninth graders of a junior high school	Computer science	Quantitative	Computer programming performances, Students' computer programming performances, Critical thinking	E = 80C = 86	74.1 (10.5)	62.9 (14.8)
Journal of Educational Research (2018).	General learning management system instruction and coaching on formative assessment	Primary school: fifth-grade students	Arts	Quantitative	Learning outcome	E = 298C = 222	10	9.36
European Journal of Psychology of Education (2018)	General learning management system: informative brochure	Primary school: sixth grade pupils	Natural science	Quantitative and qualitative: measured with a questionnaire andfeedback	Written feedback	E = 469C = 271	5.15 (1.86)	4.18 (1.75)
Assessment & Evaluation in Higher Education (2018)	Web-based peer assessment system: web-based peer assessment	High education	Social science	Quantitative and qualitative:30-min interview.	Writing performance, Feedback quality, Metacognitive awareness, and Self-efficacy	E = 32C = 32	17.41 (1.57)	16.63 (1.06)
Behaviour & Information Technology (2018)	Web-based peer assessment system: computer-supported collaborative learning (CSCL) environment	High education	Computer Science	Quantitative	Management Information System & SRL questionnaire	E = 43C = 41	30.53 (5.10)	26.02 (4.22)
Instructional Science (2018)	Web-based peer assessment system: through online rubrics	High education	Natural science	Quantitative and qualitative	Argumentation competence	E = 57C = 41	66.7	47.2
European Journal of Dental Education (2018)	Web-based peer assessment system: The digital training system	High education	Natural science	Quantitative	RDTES and OPRS questionnaire	E = 33C = 33	86.70 (4.17)	83.29 (4.81)
BMC Medical Education (2019)	Balanced, consensus-based peer assessment system	High education	Natural science	Quantitative	CATME questionnaire	E = 94C = 98	3.92 (0.81)	4.01 (0.86)
Journal of International Medical Research (2019)	gold standard video	High education	Natural science	Quantitative	Physical examination	E = 126C = 126	n/a	n/a
Computers & Education (2020)	SVVR environment	High education	Social science	Quantitative and qualitative	English-speaking performance, Motivation, Critical thinking, Interview outline, English learning anxiety	E = 33C = 36	19.85 (2.22)	18.13 (2.57)

Note: SRL: self-regulated learning; RDTES: real-time dental training and evaluation system; OPRS: online peer-review system; CATME: comprehensive assessment team member effectiveness; SVVR: spherical video-based virtual reality.

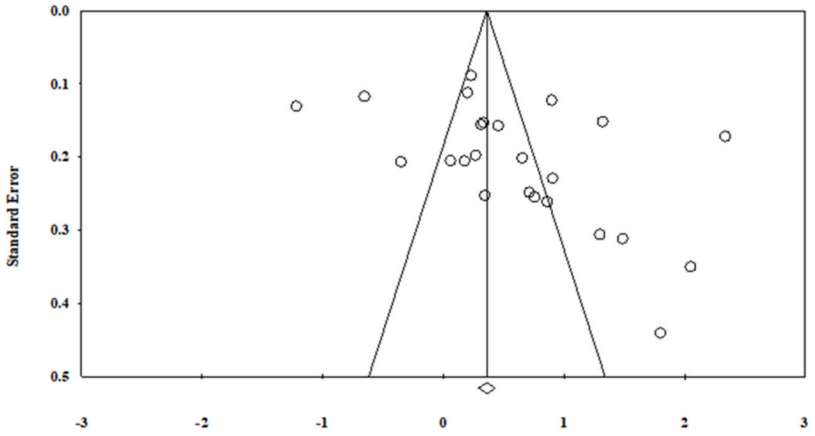

Fig. 3. funnel plot of standard error.

specifically, we tested whether the EFs measured among the studies were heterogeneous. The heterogeneity test mainly includes a Dixon's Q test, an I-square test, and Kruskal-Wallis' H test. The Q test examines total variation, assuming that the EFs follow a chi-squared distribution (Hoaglin, 2016). A *p*-value of less than 0.05 denotes obvious heterogeneity. The I-square test mainly reflects the proportion of true variation in the EFs of the total variation. According to existing research, I-square values of 25%, 50%, and 75% represent the low, medium, and high heterogeneity limits, respectively (Borenstein et al., 2009). Research designs are generally prone to heterogeneity depending on the incidence rate in the unexposed group, the length of the follow-up studies, adjustment variables or content, statistical method or model, source of bias, and research quality.

Fig. 4 illustrates the forest charts compiled in this study. Of the selected 18 articles, Hsia, Huang, and Hwang (2016) and Yurdabakan (2012) showed that certain subjects in the experimental group are not significantly better than those in the control group. In

Study name	Standard error	Lower limit	Upper limit	p-Value
Hwang, Hung, & Chen 2014	0.156	0.008	0.618	0.044
Lai & Hwang 2015	0.198	-0.119	0.657	0.173
Eldredge, Bear, Wayne, & Perea 2013	0.253	-0.150	0.840	0.172
Hsia, Huang, & Hwang 2016	0.440	0.937	2.664	0.000
Yusuff 2015(a)	0.306	0.698	1.898	0.000
Yusuff 2015(b)	0.312	0.877	2.099	0.000
Wang, Hwang, Liang, & Wang 2017	0.157	0.148	0.764	0.004
Chen & Andrade 2018	0.089	0.056	0.405	0.010
Leenknecht & Prins 2018(c)	0.205	-0.342	0.463	0.767
Leenknecht & Prins 2018(d)	0.207	-0.755	0.056	0.091
Leenknecht & Prins 2018(e)	0.206	-0.225	0.581	0.388
Zheng, Cui,Li, & Huang 2018	0.261	0.351	1.376	0.001
Lin 2016	0.229	0.457	1.356	0.000
Biesma, et al. 2019(f)	0.172	2.002	2.676	0.000
Biesma, et al. 2019(g)	0.118	-0.883	-0.422	0.000
Biesma, et al. 2019(h)	0.123	0.659	1.139	0.000
Biesma, et al. 2019(i)	0.112	-0.019	0.421	0.074
Biesma, et al. 2019(j)	0.131	-1.474	-0.961	0.000
Chen, et al, 2019	0.152	1.024	1.619	0.000
Liu, et al. 2018	0.255	0.258	1.257	0.003
Yurdabakan 2012	0.350	1.367	2.739	0.000
Chien, Hwang, & Jong 2019	0.249	0.227	1.201	0.004
Van Gennip, Segers, & Tillema 2010	0.153	0.034	0.634	0.029
Li & Gao 2016	0.202	0.259	1.050	0.001
Cebrian-Robles, et al. 2018(k)	11.860	-138.820	-92.331	0.000
Cebrian-Robles, et al. 2018(l)	3.504	-40.962	-27.227	0.000
	0.034	0.288	0.424	0.000

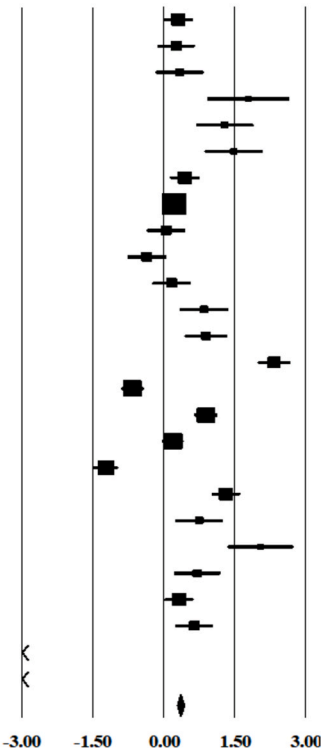

Fig. 4. The effect of PA strategies.

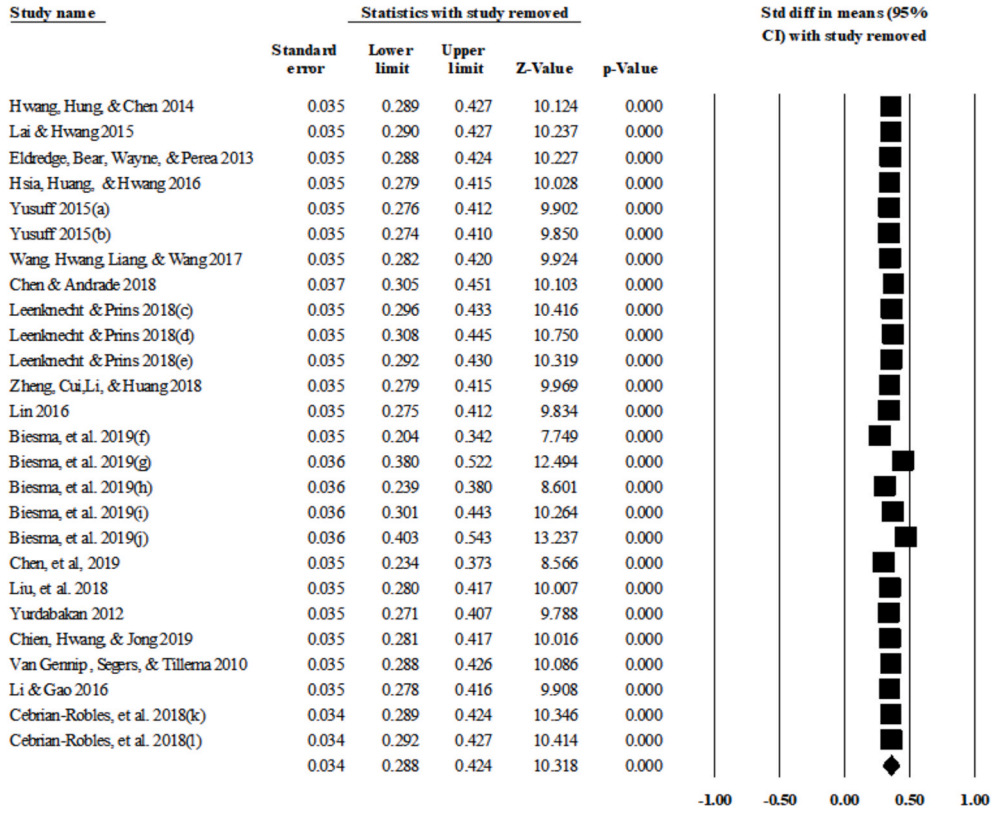

Fig. 5. Sensitivity analysis.

addition, Biesma (2019) demonstrated that the experimental group reported worse results than those for the control group. Nevertheless, the standard error at the 95% confidence interval for the average difference between the experimental and control groups is [0.288, 0.424] with $p < .05$, indicating that each effect value in the meta-analysis is heterogeneous. In other words, the experimental group reports significantly better results than those for the control group. The results of the heterogeneity test are $Q = 503.117$, $p < .05$, $I_2 = 95.428$, suggesting serious heterogeneity between the studies. Similarly, the random-effects model shows strict heterogeneity ($Tau^2 = 0.613$ (>0.01)). This can be attributed to the use of different activity designs across the studies.

In addition, subjects and the types of subjects can contribute to heterogeneity. Cebrián-Robles, Franco-Mariscal, and Blanco-López (2018), for example, employed a small sample that did not obey normal distribution. Their Mann-Whitney *U* test showed that the experimental group did not significantly outperform in the tree growth activity and the purity of bottled water activity. Simply put, the results for the experimental group were worse than those for the control group, suggesting that the researchers were expecting the former to perform significantly better than the latter. The reasons underpinning their findings warrant further verification through more evidence.

The heterogeneity test showed that the effect values between the studies were highly heterogeneous. This study used the removal method to conduct sensitivity analysis. Fig. 5 presents the results of the sensitivity analysis. After excluding certain articles, we found a difference in the mean of the overall effect at the 95% confidence interval. In particular, we omitted Biestma et al. (2019) and observed a relatively large change in the upper and lower limits of the confidence interval and the range of the Z-value.

4.4. A co-citation network of PA research: the main streams in the field (research question 4)

Based on the co-citation analysis, the whole network of PA research was visualized. As shown in Fig. 6, all research articles were denoted as 20 nodes in the network. While a total of 11 articles were linked (network density = 11%) and presented as the main component in the center of the map, the remaining nine articles which have not been co-cited were located as isolated components on the left side. Most of the isolated papers were newly published after 2018, leaving relatively limited time to be cited. Researchers may update the development of these selected papers in the future. Therefore, the following analysis is based on the main component.

To make a clear presentation of the network and an index for the linking of two analyses, the label of each node is presented as the full last name of the first author, followed by the co-authors' initials and the publication year. For example, the highest cited article that was co-authored by three authors (Van Gennip, Segers, & Tillema, 2010) and was published in 2010 is presented as VangennipST2010 in the diagram. Two thicker links (in red) are used to highlight the two highest co-cited pairs in this study. While the research of Lai and Hwang (2015) and Hsia et al. (2016) was co-cited 10 times, another pair of studies (Lai & Hwang, 2015 and Hwang et al., 2014) was

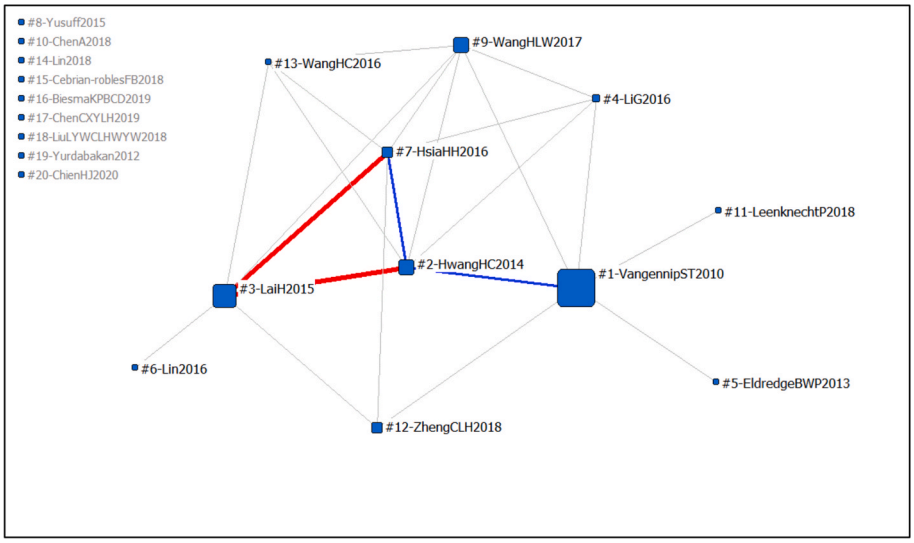

Fig. 6. The co-citation network of PA.

co-cited eight times. As seen in the map, three highly cross-referenced papers are grouped as a triangle, indicating the importance and bibliometric similarity among them. Note that the main focus within the triangular network is the use of technology-enhanced peer-assessment (e.g., game-based learning, web-based learning) to improve the learning outcomes. It is also noteworthy that the research of Van Gennip et al. (2010) is not only the highest cited paper in the field, but it acts as a critical bridging paper between Hwang et al. (2014) (which is the only node to link the other two highly co-cited papers in the triangle) and the two furthest nodes (Leenknecht and Prins, 2018; Eldredge, Bear, Wayne, & Perea, 2013).

5. Discussion

With a focus on the research questions, this section discusses the findings of the analyses using the constructivist theoretical framework.

5.1. The technological applications for PA design

Technology and learning environments are critical in PA learning activities. This study re-confirms that improvements in the PA learning effect should be attributed to customized computer equipment (Eldredge et al., 2013). Teachers could utilize learning equipment, such as the widely used mobile devices, smartphones, and tablets, in a learning environment to develop interactive teaching approaches using PA. Eldredge et al. (2013) used a combination of PA and computer equipment to promote learners' ability of searching. Technology influences learner outcomes and interactive behaviors in PA learning activities. Learners could use their own mobile learning devices to improve the effectiveness of their learning through PA (Lai & Hwang, 2015). However, there is a lack of retrospective research on the use of wearable devices to conduct PA activities. The planning and design of PA activities using wearable devices will promote autonomous learning and gradual integration into daily life. Further, exploring ways for researchers and teachers to support these new technologies could improve students' self-learning, cooperative learning, and lifelong learning concepts and effectiveness, knowledge building, and higher-order thinking skills.

5.2. The major PA-based learning environments

Over the past decade, PA-based education using learning environments supported by digital technology is being increasingly valued. For example, researchers highlight the value of mobile devices in education using the constructivist theoretical framework to integrate computer technology in learning environments (Fu & Hwang, 2018). Eldredge et al. (2013) explored the development of PA learning activities to promote student learning in technology-based learning environments. Hwang et al. (2014) combined PA with computer games to develop problem-solving skills among students in elementary school. Education researchers are, thus, adopting the constructivist theoretical framework to connect learners with their learning environments through related research and design. Their studies focused on learners' needs and integrating information technologies in teaching strategies and learning environments to enhance learning outcomes and to attempt to understand various learning behaviors by surveying learning attitudes and opinions.

This study further reveals that teachers are designing PA activities that foster more meaningful student participation and real learning environments and that provide students with better opportunities to associate learning content with real-life experiences (Chen et al., 2019). Scholars are investigating ways to integrate PA into learning environments, which is highly valued by teachers. For

example, Chien et al. (2019) adapted PA to a spherical video-based virtual reality environment to improve EFL students' English-speaking performance. While researchers state that technology creates numerous education opportunities and changes teacher and learner attitudes toward technology, there is a lack of research design that uses relevant technologies for workplace learners (e.g., training in various fields).

5.3. The major PA-based research domains

Few PA-based studies have been conducted on health and nursing, particularly from a research participant perspective. PA teaching research still has significant educational value, and researchers are largely adopting the constructivist theoretical framework to integrate computer technologies with teaching themes. A majority of the current research focuses on multiple disciplines such as social science and natural science, whereas few studies are based on business and management, health and nursing, and architecture. For example, Chen and Andrade (2018) combined PA teaching with applied science and technology to conduct a formative assessment of fifth-grade students' theater arts achievements and, accordingly, offered suggestions to effectively improve their academic performance. Thus, there is scope to explore PA in various domains, to integrate challenging technologies and PA teaching approaches in various domains, and to examine how such methods could benefit learners from various perspectives such as learning attitudes, interests, and views; academic performance; and emotional factors. Similarly, there is a need to study the impact of PA on higher-order thinking, cognitive skills, and self-efficacy.

5.4. The major methods and research issues related to PA

The study results show that most researchers examining the development of students' knowledge and skills have tended to adopt PA strategies that could facilitate student learning and knowledge building more flexibly and interactively. From a social constructivism perspective, learning through PA supported by action technology creates more opportunities for peer and expert interactions. Further, thinking about and reflecting on various information and opinions is conducive to meaningful knowledge construction. Lai and Hwang (2015) examined peer groups comprising two to three students and showed that the proper handling of social interactions offers students the opportunity to think about specific issues, to construct various aspects of knowledge and theory, and to develop a belief in the nature of knowledge and learning. While Chien et al. (2019) proposed the emerging phenomenon of combining technology with PA teaching and suggested considering not only quantitative methods but also effect size. Hwang et al. (2014) quantitatively examined PA experimental design and reported a positive impact on students' academic performance and motivation. Scholars have also shown that students who study through PA report significantly better learning performance than those who learn using traditional web-based problem-solving approaches (Li & Gao, 2016). As shown in Table 2, Wang et al.'s (2016) experimental results suggested that students who learn through technology-integrated PA teaching show a marked increase in acceptance and learning satisfaction. Combining technology with PA teaching helps students organize the searched and prior knowledge during learning activities, and clarifies possible misunderstandings. Meanwhile, Wang et al. (2017) stated that combining PA teaching methods with technology helps students deal with complex web-based learning tasks. This finding is in line with the view of the present retrospective study: Technology not only demonstrates learning significance but also emphasizes peer learning, mutual assessments, new knowledge, and humanism and care when integrated with PA teaching activities.

5.5. The overall effect of PA under each condition

This study conducts a meta-analysis to examine peer review methods. In general, PA strategies have a significantly negative correlation with the learning effectiveness of a control group. The relationship between PA strategies and learning effectiveness is affected by measurement tools, the type of method (e.g., the items of each scale), the subject group, and the cultural background and its dimensions. A sensitivity analysis of the 18 studies after excluding Biestma et al. (2019) revealed a concentrated mean at the 95% confidence level (Fig. 5). The literature contains experimental design research that is worth exploring and referencing for further analysis and discussion. Moreover, the 18 analyzed studies could improve students' learning effectiveness, all of which experimentally verified significant improvements in learners' competency, learning achievements, motivation, problem-solving skills, performance, meta-cognitive skills and awareness, cognitive load, attitude, self-efficacy, learning anxiety, and satisfaction. However, the STEM-based research on teaching issues, learning style or behaviors, and path performance is yet to be explored. This may be because developing STEM skills requires more time and there is a need for appropriate tools and methods to examine learning styles and behaviors. Moreover, learners who are located in different environments may experience increasing learning anxiety when faced with challenging tasks. Thus, combining technology with PA teaching for STEM training and other cross-disciplines is worthwhile. To summarize, in addition to emerging PA learning models that use a constructivist theoretical framework, integrated analysis, and network analysis, it is important to develop and improve learning strategies that guide more effective learning among students. Notably, not all studies could be included in the analysis or subgroup of adjustment variables, and the outcome index could not be converted into a model test to obtain the overall effect. There is also a need for research on emerging mobile PA learning environments and the impact of learning models on student performance, including social negotiation and higher-order thinking.

5.6. The main streams in the PA research network

According to the co-citation network of PA research (Fig. 6), the four studies (i.e., Van Gennip et al., 2010; Hwang et al., 2014; Lai &

Hwang, 2015; Hsia et al., 2016) that are closely connected and related to all other articles were identified as the core papers in this study. This result strengthened the bibliometric associations between the core papers. In addition, the empirical findings of the four core papers have shed light on some important research streams of peer assessment literature. For example, Van Gennip et al.'s (2010) study confirmed that interpersonal peer assessment is an effective collaborative learning activity for vocational education. Three highly cross-referenced studies, on the other hand, highlighted the effects of technology-enhanced peer assessment approaches for various learning contexts (Hsia et al., 2016; Hwang et al., 2014; Lai & Hwang, 2015). Hwang et al. (2014) highlighted the effects of peer assessment with a digital game-based learning environment to improve users' motivations and learning achievement. Similarly, Lai and Hwang (2015) designed interactive peer-assessment criteria for the use of handheld devices to enhance the art design performance of elementary school students, while Hsia et al. (2016) provided a web-based peer assessment approach to increase young adults' learning performance and motivation in arts courses. From the citers' perspectives, the empirical findings of the four core papers have become an important foundation in the field. Future researchers may extend the research directions from the main streams.

6. Conclusions and research limitations

Peer assessment (PA) has become a key learning strategy for educational technology research. In this review, we address some important research questions regarding the overall effect sizes, conditions for related technologies, learning environments, and the most frequently referenced research streams from 1998 to 2019. Taken altogether, this study makes two key contributions. First, this study provides the very first attempt by integrating meta-analysis and co-citation network analysis to examine 20 empirical research of peer assessment, including the effect sizes, publication biases, research characteristics, and the association of research data for conducting the meta-analysis. The findings provide the overall effect sizes of peer assessment in various learning environments and research directions based on large citers' perspectives. The visualized the networking relationships of 20 PA articles also provide a fresh micro-to-macro understanding of the PA literature. Second, echoing Fu and Hwang's (2018) constructivist theoretical framework through our integration of the analysis and verification processes, we propose a series of criteria for PA research as a basic condition for future researchers. This research will serve as a good reference for researchers adopting social constructivist theory and combining technology with PA teaching and learning systems to improve learning effectiveness.

This study is subject to the following limitations. First, it does not account for the moderating effects of characteristics such as gender and regional differences. Second, the literature presents numerous measurement tools; although we examined students' learning effectiveness, the variation may affect the reliability of the adjustment effect of the measurement tools. Last, the analysis is limited to articles published in international journals. The uneven distribution of integrated analyses also affects the results. It is necessary to pay attention to the regulatory role of research subgroups and more comprehensively and accurately explain the relationship of effect sizes. Future research should consider more comprehensive measurement tools with higher frequency to improve their representativeness.

Credit author statement

Ching-Yi Chang, De-Chih Lee, Kai-Yu Tang: Conceptualization, Methodology, Formal analysis, Validation, Investigation, Data curation, Writing – original draft. **Ching-Yi Chang, De-Chih Lee, Kai-Yu Tang, and Gwo-Jen Hwang**: Writing – review & editing, Supervision, Resources, and Project administration

Acknowledgment

This paper has benefited from the comments and suggestions of the two anonymous reviewers. Partly financial supports from the Ministry of Science and Technology, Taiwan, under grant number MOST 109-2635-H-227- 001, 109-2511-H-130-002, and 109-2511-H-011-002-MY3 as well as by the Ministry of Education under contract number PMN1090292, are also acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2020.104123.

References

Aryadoust, V. (2020). *A review of comprehension subskills: A scientometrics perspective*. System, 102180.
Begg, C. B. (1994). Publication bias (pp. 339- 422). In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation.
Biesma, R., Kennedy, M. C., Pawlikowska, T., Brugha, R., Conroy, R., & Doyle, F. (2019). Peer assessment to improve medical student's contributions to team-based projects: Randomised controlled trial and qualitative follow-up. *BMC Medical Education, 19*(1), 371–378.
Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). Fixed-effect versus random-effects models. In *Introduction to meta-analysis* (pp. 77–86). Chichester: Wiley.
Borgatti, S. P., Everette, M. G., & Freeman, L. C. (2002). *UCINET for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Carey, M. C., Chick, A., Kent, B., & Latour, J. M. (2018). An exploration of peer-assisted learning in undergraduate nursing students in paediatric clinical settings: an ethnographic study. *Nurse Education Today, 65*, 212–217.
Cebrián-Robles, D., Franco-Mariscal, A. J., & Blanco-López, Á. (2018). Preservice elementary science teachers' argumentation competence: Impact of a training programme. *Instructional Science, 46*(5), 789–817.
Chen, F., & Andrade, H. (2018). The impact of criteria-referenced formative assessment on fifth-grade students' theater arts achievement. *The Journal of Educational Research, 111*(3), 310–319.
Chen, L., Chen, H., Xu, D., Yang, Y., Li, H., & Hua, D. (2019). Peer assessment platform of clinical skills in undergraduate medical education. *Journal of International Medical Research, 47*(11), 5526–5535.
Chien, S. Y., Hwang, G. J., & Jong, M. S. Y. (2019). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-speaking performance and learning perceptions. *Computers & Education, 146*, 103751.
Demir, M. (2018). Using online peer assessment in an instructional technology and material design course through social media. *Higher Education, 75*(3), 399–414.
Eldredge, J. D., Bear, D. G., Wayne, S. J., & Perea, P. P. (2013). Student peer assessment in evidence-based medicine (EBM) searching skills training: An experiment. *Journal of the Medical Library Association, 101*(4), 244–251.
Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322.
Fetscherin, M., & Heinrich, D. (2015). Consumer brand relationships research: A bibliometric citation meta-analysis. *Journal of Business Research, 68*(2), 380–390.
Fu, Q. K., & Hwang, G. J. (2018). Trends in mobile technology-supported collaborative learning: A systematic review of journal publications from 2007 to 2016. *Computers & Education, 119*, 129–143.
Fu, Q. K., Lin, C. J., & Hwang, G. J. (2019). Research trends and applications of technology-supported peer assessment: A review of selected journal publications from 2007 to 2016. *Journal of Computers in Education, 6*(2), 191–213.
Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*(1), 27–61.
Hallinger, P. (2020). Science mapping the knowledge base on educational leadership and management from the emerging regions of Asia, Africa and Latin America, 1965–2018. *Educational Management Administration & Leadership, 48*(2), 209–230.
Harder, T., & Siedler, A. (2019). Systematic review and meta-analysis of chickenpox vaccination and risk of herpes zoster: A quantitative view on the 'exogenous boosting hypothesis'. *Clinical Infectious Diseases, 69*(8), 1329–1338.
Harland, T., Wald, N., & Randhawa, H. (2017). Student peer review: Enhancing formative feedback with a rebuttal. *Assessment & Evaluation in Higher Education, 42*(5), 801–811.
Hedges, L. V. (1992). Meta-analysis. *Journal of Educational Statistics, 17*(4), 279–296.
Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486–504.
Hoaglin, D. C. (2016). Misunderstandings about Q and 'Cochran's Q test' in meta-analysis. *Statistics in Medicine, 35*(4), 485–495.
Hsia, L. H., Huang, I., & Hwang, G. J. (2016). A web-based peer-assessment approach to improving junior high school students' performance, self-efficacy and motivation in performing arts courses. *British Journal of Educational Technology, 47*(4), 618–632.
Hsia, L. H., & Sung, H. Y. (2020). Effects of a mobile technology-supported peer assessment approach on students' learning motivation and perceptions in a college flipped dance class. *International Journal of Mobile Learning and Organisation, 14*(1), 99–113.
Hwang, G. J., Hung, C. M., & Chen, N. S. (2014). Improving learning achievements, motivations and problem-solving skills through a peer assessment-based game development approach. *Educational Technology Research & Development, 62*(2), 129–145.
Janssens, A. C. J., & Gwinn, M. (2015). Novel citation-based search method for scientific literature: Application to meta-analyses. *BMC Medical Research Methodology, 15*(1), 84. https://doi.org/10.1186/s12874-015-0077-z
Janssens, A. C. J., Gwinn, M., Brockman, J. E., Powell, K., & Goodman, M. (2020). Novel citation-based search method for scientific literature: A validation study. *BMC Medical Research Methodology, 20*(1), 25. https://doi.org/10.1186/s12874-020-0907-5
Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology, 36*(4), 567–589.
Kates, A. W., Wu, H., & Coryn, C. L. (2018). The effects of mobile phone use on academic performance: A meta-analysis. *Computers & Education, 127*, 107–112.
Kuo, F. C., Chen, J. M., Chu, H. C., Yang, K. H., & Chen, Y. H. (2017). A Peer-assessment mobile kung fu education approach to improving students' affective performances. *International Journal of Distance Education Technologies, 15*(1), 1–14.
Lai, J. W., & Bower, M. (2019). How is the use of technology in education evaluated? A systematic review. *Computers & Education, 133*, 27–42.
Lai, C. L., & Hwang, G. J. (2015). An interactive peer-assessment criteria development approach to improving students' art design performance using handheld devices. *Computers & Education, 85*, 149–159.
Leenknecht, M. J., & Prins, F. J. (2018). Formative peer assessment in primary school: The effects of involving pupils in setting assessment criteria on their appraisal and feedback style. *European Journal of Psychology of Education, 33*(1), 101–116.
Li, L. (2019). Using game-based training to improve students' assessment skills and intrinsic motivation in peer assessment. *Innovations in Education & Teaching International, 56*(4), 423–433.
Li, L., & Gao, F. (2016). The effect of peer assessment on project performance of students at different learning levels. *Assessment & Evaluation in Higher Education, 41*(6), 885–900.
Lin, Y. N., Hsia, L. H., Sung, M. Y., & Hwang, G. H. (2019). Effects of integrating mobile technology-assisted peer assessment into flipped learning on students' dance skills and self-efficacy. *Interactive Learning Environments, 27*(8), 995–1010.
Lipsey, M. W., & Wilson, D. B. (2001). The way in which intervention studies have "personality" and why it is important to meta-analysis. *Evaluation & the Health Professions, 24*(3), 236–254.
Li, H., Xiong, Y., Zang, X., Kornhaber, L. M., Lyu, Y., Chung, K. S., et al. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education, 41*(2), 245–264.
Luaces, O., Díez, J., & Bahamonde, A. (2018). A peer assessment method to provide feedback, consistent grading and reduce students' burden in massive teaching settings. *Computers & Education, 126*, 283–295.
Megías, A., Gómez-Leal, R., Gutiérrez-Cobo, M. J., Cabello, R., & Fernández-Berrocal, P. (2018). The relationship between trait psychopathy and emotional intelligence: A meta-analytic review. *Neuroscience & Biobehavioral Reviews, 84*, 198–203.
Ng, E. M. (2016). Fostering pre-service teachers' self-regulated learning through self-and peer assessment of wiki projects. *Computers & Education, 98*, 180–191.
Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education, 44*(8), 1253–1278.
Pan, L., & An, T. (2020). *The evolutionary characteristics of higher education studies worldwide: Central themes and regions*. Studies in Higher Education. https://doi.org/10.1080/03075079.2020.1735331
Radler, V. M. (2018). 20 Years of brand personality: A bibliometric review and research agenda. *Journal of Brand Management, 25*(4), 370–383.
Rico-Juan, J. R., Gallego, A. J., & Calvo-Zaragoza, J. (2019). Automatic detection of inconsistencies between numerical scores and textual feedback in peer-assessment processes with machine learning. *Computers & Education, 140*, 103609.
Rico-Juan, J. R., Gallego, A. J., Valero-Mas, J. J., & Calvo-Zaragoza, J. (2018). Statistical semi-supervised system for grading multiple peer-reviewed open-ended works. *Computers & Education, 126*, 264–282.
Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. West Sussex: John Wiley & Sons.
Samaie, M., Mansouri Nejad, A., & Qaracholloo, M. (2018). An inquiry into the efficiency of WhatsApp for self-and peer-assessments of oral language proficiency. *British Journal of Educational Technology, 49*(1), 111–126.

Sanaeifar, S. H., & Mirshojaee, S. B. (2020). Optimizing students' engagement through implementing peer-assessment practice in Iranian public high school: An action research. *Theory and Practice in Language Studies, 10*(8), 940–946.
Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265–269.
Tenório, T., Bittencourt, I. I., Isotani, S., & Silva, A. P. (2016). Does peer assessment in on-line learning environments work? A systematic review of the literature. *Computers in Human Behavior, 64*, 94–107.
Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.
Van Gennip, N. A., Segers, M. S., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*(4), 280–290.
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.
Wang, H. C., Hsieh, Y. J., & Chen, W. F. (2016). The effect of online peer assessment in engineering education: A quasi-experimental study. *International Journal of Engineering Education, 32*(1), 199–208.
Wang, X. M., Hwang, G. J., Liang, Z. Y., & Wang, H. Y. (2017). Enhancing students' computer programming performances, critical thinking awareness and attitudes towards programming: An online peer-assessment attempt. *Educational Technology & Society, 20*(4), 58–68.
Yilmaz, R. M., Topu, F. B., & Takkaç Tulgar, A. (2019). *An examination of the studies on foreign language teaching in pre-school education: A bibliometric mapping analysis*. Computer Assisted Language Learning. https://doi.org/10.1080/09588221.2019.1681465
Yu, H., He, Y., Mao, Z., Dong, W., Fu, X., & Lei, X. (2019). Hypertensive disorders during pregnancy and elevated blood pressure in the offspring: A systematical review and meta-analysis protocol. *Medicine, 98*(20), 1–5.
Yurdabakan, I. (2012). The effect of co and peer assessment training on self-assessment. *Egitim Ve Bilim (Education and Science), 37*(163), 190–202.
Zheng, L., Chen, N. S., Cui, P., & Zhang, X. (2019). A systematic review of technology-supported peer assessment research: An activity theory approach. *International Review of Research in Open and Distributed Learning, 20*(5), 168–191.
Zheng, L., Chen, N. S., Li, X., & Huang, R. (2016). The impact of a two-round, mobile peer assessment on learning achievements, critical thinking skills, and meta-cognitive awareness. *International Journal of Mobile Learning and Organisation, 10*(4), 292–306.
Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472.