A Systematic Review of Technology-Supported Peer Assessment Research: An Activity Theory Approach

Lanqin Zheng [1], Nian-Shing Chen[2*], Panpan Cui[1], and Xuan Zhang[1]

[1]*Beijing Normal University, Beijing,* [2]*National Yunlin University of Science and Technology, Taiwan,* * *Corresponding author*

Abstract

With the advancement of information and communication technologies, technology-supported peer assessment has been increasingly adopted in education recently. This study systematically reviewed 134 technology-supported peer assessment studies published between 2006 and 2017 using a developed analysis framework based on activity theory. The results found that most peer assessment activities were implemented in social science and higher education in the past 12 years. Acting assignments such as performance, oral presentations, or speaking were the least common type of assignments assessed across the studies reviewed. In addition, most studies conducted peer assessment anonymously and assessors and assessees were randomly assigned. However, most studies implemented only one round of peer assessment and did not provide rewards for assessors. Across studies, it was more often the case that students received unstructured feedback from their peers than structured feedback. Noticeably, collaborative peer assessment did not receive enough attention in the past 12 years. Regarding the peer assessment tools, there were more studies that adopted general learning management systems for peer assessment than studies that used dedicated peer assessment tools. However, most tools used within these studies only provide basic functionalities without scaffolding. Furthermore, the results of cross analysis reveal that there are significant relationships between learning domains and anonymity as well as learning domains and assessment durations. Significant relationships also exist between assignment types and learning domains as well as assignment types and assessment durations.

*Keywords***:** peer assessment, systematic review, activity theory, collaborative learning

Introduction

Peer assessment is a process by which learners can evaluate peers' products based on assessment criteria (Sadler & Good, 2006). Applying peer assessment can engage learners in providing constructive comments for peers and improving their own works, making peer assessment a meaningful assessment model (Topping, 2017). Moreover, there are theoretical and empirical evidences of the positive effects of peer assessment on higher-order thinking skills (Topping, 2017), social skills (Ching & Hsu, 2016), learning motivations (Hsia, Huang, & Hwang, 2016), and learning outcomes (Zheng, Chen, Li, & Huang, 2016). Due to rapid technological advancement, the implementation of technology-supported peer assessment is becoming more and more effective (Yu & Wu, 2011). More specifically, technology-supported peer assessment can facilitate online submission of works, random assignments, reciprocal peer reviews, and structured feedback (Hsu, 2016). Compared to the traditional peer assessment, the benefits of technology-supported peer assessment include: anonymity, speed and efficiency, random distribution of essays, automatic calculation of marks, and feedback availability (Mostert & Snowball, 2013). In addition, online peer assessment systems can automatically record emotional responses through the self-assessment manikin measurement (Cheng, Hou, & Wu, 2014). Despite of all the advantages described above, many instructors are struggling with how to design and improve technology-supported peer assessment in real practices. Literature also suggests that optimizing peer assessment design is crucial for improving assessment practices (Bearman et al., 2016). A systematic review of technology-supported peer assessment literature can provide better insights for instructors to design and implement peer assessment.

A design feature refers to a particular consideration for making a decision during the process of peer assessment design, which ensures the success of peer assessment to a large extent (Adachi, Tai, & Dawson, 2018). However, the design features of peer assessments are often neglected by instructors due to focusing on peers' works or final scores (Adachi et al., 2018). More specifically, 'front-line' educators and practitioners often find it very challenging to implement and improve peer assessment in practice (Bearman et al., 2016). They often design peer assessment activities based on their assumptions and experiences which leads to problems in selecting appropriate peer assessment tasks, learning domains, and criteria development (Adachi et al., 2018). In addition, there is a lack of a systematic review of technology-supported peer assessment studies in previous literature. These research gaps and problems underlying peer assessment drive us to conduct a comprehensive review of technology-supported peer assessment. A systematic review of technology-supported peer assessment can shed light on how peer assessment works as well as provide useful references for implementing peer assessment. The findings can also contribute to the design of peer assessment and inform educators on how technology can be effectively applied in peer assessment.

The purpose of this study is twofold. One is to investigate the research status of the technology-supported peer assessment studies in the past 12 years. Another is to conduct a correlation analysis among assignment types, learning domains, anonymity, and assessment duration so as to provide insights into the design of peer assessment activities. Based on the research purpose described above, the following eight research questions (RQ) are addressed in this study:

RQ1: What school levels participated in the technology-supported peer assessment research?

RQ2: What kinds of rules were adopted in the technology-supported peer assessment research?

RQ3: What kinds of evaluation criteria were adopted in the technology-supported peer assessment research?

RQ4: How were labors divided in the technology-supported peer assessment research?

RQ5: What were the learning objectives in the technology-supported peer assessment research?

RQ6: What kinds of tools were used in the technology-supported peer assessment research?

RQ7: Are there any significant relationships among anonymity, learning domains, and assessment durations in the technology-supported peer assessment research?

RQ8: Are there any significant relationships among assignment types, learning domains, and assessment durations in the technology-supported peer assessment research?

Literature Review

Technology-Supported Peer Assessment

Peer assessment was conceptualized as an instructional method that requires learners to evaluate the amount, quality, value, and success of the products or learning outcomes of peers (Topping, 1998). Typically, two kinds of learning activities were involved in peer assessment. One was evaluation of peers' works and the other was revision of self-work (Cheng, Liang, & Tsai, 2015). Currently there are many technologies that can support and facilitate peer assessment, including Wiki environments (Gielen & De Wever, 2015), massive open online courses (MOOCs) (Wulf, Blohm, Leimeister, & Brenner, 2014), and mobile technologies (Hwang & Wu, 2014). Furthermore, in research by Tsai (2009), a Web-based peer assessment system was used to automatically record student participation and interactions within peer assessment activities. In addition, Shih (2011) used Facebook to conduct peer assessment and found that the emoticons stimulated learners' motivations for English writing and enhanced interpersonal relationships. Xiao and Lucking (2008) conducted online peer assessment in a Wiki environment and found that students' writing performance and satisfactions were improved through the provision of both quantitative and qualitative feedback. To sum up, technologies can facilitate the efficiency and effectiveness of peer assessment. However, previous studies did not systemically analyze how to use technologies to facilitate peer assessment. The present review aims to identify how technology-supported peer assessment is being designed and implemented.

Deficiency of the Previous Peer Assessment Reviews

The initial literature review done by Topping (1998) revealed that peer assessment had positive effects on learners' attitudes and achievements. Recently, some reviews were conducted to investigate the status of peer assessment. These reviews of peer assessment mainly addressed students' perceptions toward peer assessment (Chang, 2016), peer assessment diversity (Gielen, Dochy, & Onghena, 2011), the effectiveness of peer assessment (Topping, 2017), as well as reliability and validity of peer assessment (Speyer, Pilz, Van Der Kruis, & Brunings, 2011). However, none of the previous reviews systemically analyzed how technology-supported peer assessment activities were designed and

implemented. Furthermore, these review studies were not carried out based on a well-recognized analysis framework such as activity theory. Challenges like how to choose anonymity and assessment durations based on learning domains as well as how to choose learning domains and assessment durations based on assignment types remain lacking. A systematic analysis of 12 years of studies on technology-supported peer assessment may provide better understanding and insights about the current research status and future trends for researchers, educators, and practitioners. Such an analysis may also be helpful to teachers, to provide guidelines on the design and implementation of technology-supported peer assessment.

Activity Theory

Activity theory was initially proposed by Vygotsky (1978) and extended by Engeström (1999) who proposed six elements to be included in this theory, namely: subject, object, tools, community, rules, and division of labor. In the literature, Engeström (2001) claimed that activity theory can effectively represent how learning activities occur as well as highlight the dynamics of learning activities. Furthermore, activity theory has been used to analyze and evaluate various kinds of learning activities (Chung, Hwang, & Lai, 2019; Park & Jo, 2017). Therefore, activity theory is adopted as a framework in this study for analyzing technology-supported peer assessment studies published in the past 12 years.

Methodology

Data Sources and Paper Selection

Papers related to peer assessment and published from 2006 to 2017 were selected from the Web of Science databases, including: the database of science citation index expanded, social sciences citation index, arts and humanities citation index, and emerging sources citation index. These databases were selected because they are well received by academia. There were two stages included in the paper selection process (Zheng, Huang, & Yu, 2014). In the first stage, specific keywords that are closely related to peer assessment were chosen to search papers in the aforementioned databases. These keywords included: "Peer assessment" OR "Peer feedback" OR "Peer review" OR "Peer evaluation" OR "Peer rating" OR "Peer scoring" OR "Peer grading;" "Online peer assessment" OR "Online peer feedback" OR "Web-based peer assessment." In the second stage, the full text of each paper was screened based on the following criteria:

1. Only journal research papers were included in the present study. Book reviews, conference papers, book chapters, abstracts, news, editorials, and reviews were excluded.

2. The papers should be closely related to peer assessment.

3. The papers should address using information and communication technologies (ICT) to support peer assessment.

4. The papers should report the subjects, objects, tools, rules, criteria, and division of labor for peer assessment. Conceptual papers that did not describe the details of peer assessment were excluded.

5. The papers should be published from 2006 to 2017.

6. The papers should be reported in English.

At this stage, three coders were given training regarding the inclusion criteria, to ensure that they had a common understanding of this criteria. Then, 15 papers were chosen from the search results of Stage 1 and read the full text to decide whether the three coders achieved a common understanding of the criteria. Finally, the rest of papers were coded independently by the three coders according to the criteria. Discrepancies were discussed and resolved face-to-face.

Search Results

Figure 1 shows the search results. Initially, 1628 papers were located using the aforementioned keywords. Among 1628 papers, 184 were not research articles, 8 papers were not written in English, 726 papers were not closely related to peer assessment, 354 papers did not adopt technology to support peer assessment, and 222 papers did not report six elements of peer assessment activities. Finally, 134 papers were selected for further analysis.
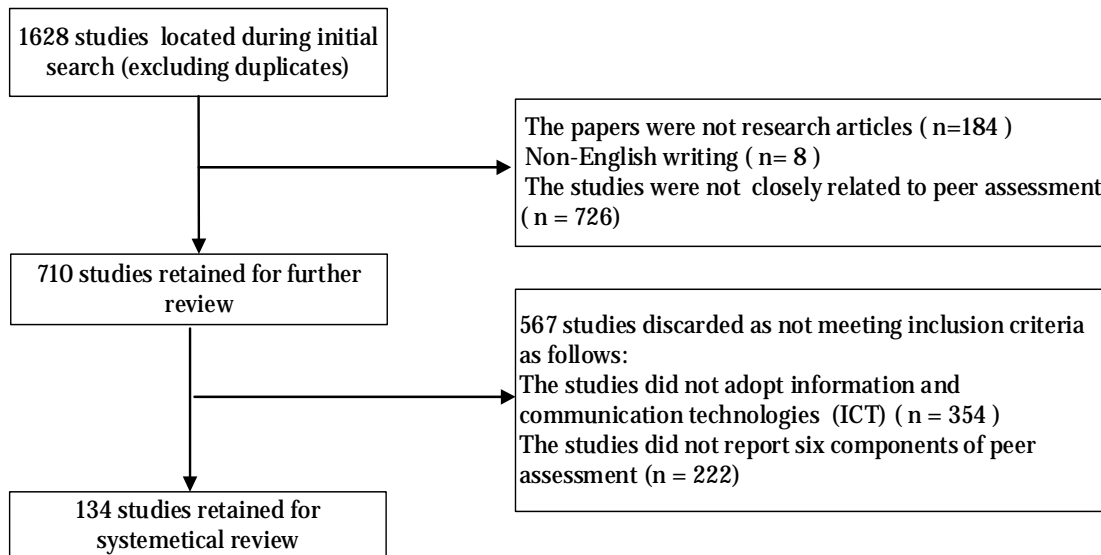

Figure 1. Papers selection process.

Data Analysis

The content analysis method was adopted to analyze the collected papers. Specifically, an analysis framework as shown in Figure 2 was developed based on activity theory. This analysis framework includes six components, namely: subjects, objects, tools, rules, criteria, and division of labor (Engeström, 1999). Table 1 shows the coding scheme based on this analysis framework. The coding scheme and development of subcategories were based on the research questions and purpose. The subcategories and associated coding values for peer assessment features were identified according to the 134 technology-supported peer assessment research articles published from 2006 to 2017. However, the subcategories of peer assessment criteria and tools were developed by the authors. All collected papers were coded independently by the three well-trained coders majoring in educational technology. Furthermore, the adjusted residual value (AR) was adopted to investigate the relationships among assignment types, learning domains, and assessment duration as well as the

relationships among learning domains, anonymity, and assessment duration. If the absolute value of AR is larger than 1.96 then the correlation between the two attributes is significant.
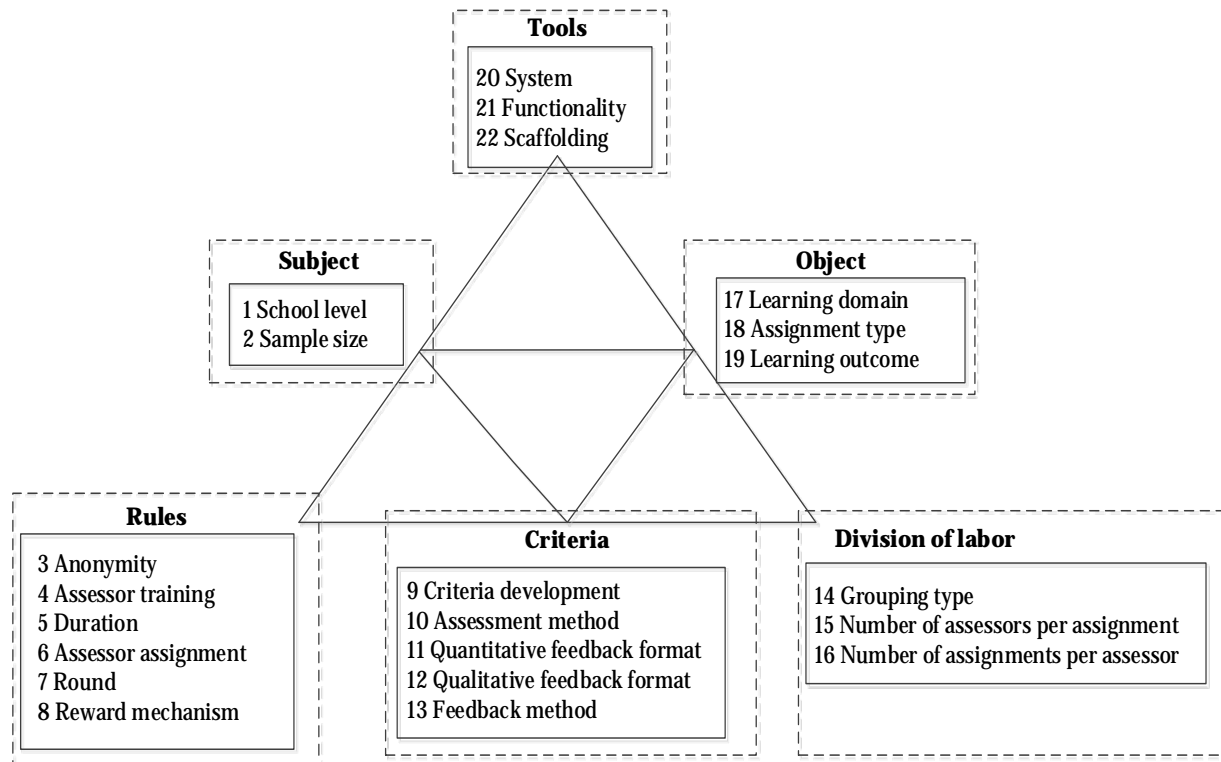

Figure 2. The analysis framework for peer assessment research based on activity theory. Adapted from "Activity theory and individual and social transformation," by Y. Engeström, in Y. Engeström, R. Miettinen, & R.-L. Punamäki (Eds.), *Perspectives on activity theory* (pp. 19–38), 1999, New York, NY: Cambridge University Press. Copyright 1999 by Cambridge University Press. Adapted with permission.

Table 1

The Coding Scheme

Component	Category	Subcategory
Subject	School level	1. Primary school.
		2. Junior and senior high school.
		3. Higher education.
	Sample size	1. 1-50.
		2. 51-100.
		3. More than 100.
Rules	Anonymity	1. Anonymous.
		2. Non-anonymous.
	Assessor training	1. Received training.

		2. No training.
	Assessment duration	1. Less than one week.
		2. 2-5 weeks.
		3. 6-10 weeks.
		4. More than 10 weeks.
	Assessor assignment	1. By system.
		2. By teachers.
		3. By students.
	Round	1. One round.
		2. Two rounds or more.
	Reward mechanism	1.With reward (course credit, reinforcement for participation).
		2. Without reward.
Criteria	Criteria development	1. By teachers.
		2. By students.
	Assessment method	1. Quantitative only.
		2. Qualitative only.
		3. Both quantitative and qualitative.
	Quantitative feedback format	1. Score.
		2. Likert scale.
	Qualitative feedback format	1. Structured feedback.
		2. Unstructured feedback.
	Feedback method	1.Written feedback.
		2. Speaking feedback.
		3. Video feedback.
		4. Mixed feedback.
Division of labor	Grouping type	1. Individual.
		2. Collaborative.
	Number of assessors per assignment	1. Less than 5.
		2. Between 5–10.
		3. More than 10.
	Number of assignments per assessor	1. Less than 5.
		2. Between 5–10.
		3. More than 10.

Object	Learning domain	1. Natural science (science, mathematics, physics, biology, geography, medicine, and so on).
		2. Social science (politics, history, education, psychology, linguistics, art, and so on).
		3. Engineering and technological science (engineering, computer science, educational technology, and so on).
	Assignment type	1. Writing essay.
		2. Project proposal (for example, WebQuest project, training plan, research report, and so on).
		3. Artefact (for example, poster, website, multimedia video, course material, and so on).
		4. Acting (for example, oral presentation, performance, and so on).
	Learning outcome	1. Cognitive outcome.
		2. Attitude or perception.
		3. Mixed.
Tools	System	1. Dedicated Web-based peer assessment system.
		2. General learning management system.
		3. Social media.
		4. Mobile application.
	Functionalities	1. Basic (assignment submission, peer grading, and making comments).
		2. Advanced (assignment submission, peer grading, making comments, discussing with reviewers, and criterion development).
	Scaffolding	1. With scaffolding.
		2. Without scaffolding.

Validity and Inter-Rater Reliability

To ensure the validity of the coding, two experienced domain experts were asked to confirm the suitability of the coding scheme and the accuracy of the coding results. To evaluate inter-rater reliability of coding, the Cronbach alpha test was conducted and the alpha coefficient achieved was 0.95, showing good reliability. All discrepancies were discussed face-to-face and solved by the three coders.

Results and Discussion

RQ1: What School Levels Participated in the Technology-Supported Peer Assessment Research?

Table 2 presents the number of schools that participated in the 134 technology-supported peer assessment studies. With regard to the school level, it was found that most peer assessment studies were conducted in higher education (81%). Few studies (19%) were conducted in K-12 settings. However, it is necessary to provide quality training and structured guidelines or scaffolding for assessors when peer assessment is implemented in K-12 schools. It is recommended to allow university students to develop criteria and provide high quality feedback when peer assessment is conducted in higher education institutions. With respect to the sample size of studies reviewed, 40% of studies involved less than 50 participants, 35% of studies involved more than 100 participants, and 25% included between 51 to 100 participants.

Table 2

Descriptive Data for Peer Assessment Subject

Category	Subcategory	Total n (%)
School level	Primary school	8(6)
	Junior and senior high school	17(13)
	Higher education	109(81)
Sample size	1-50	54(40)
	51-100	33(25)
	More than 100	47(35)

RQ2: What Kinds of Rules Were Adopted in the Technology-Supported Peer Assessment Research?

Table 3 shows the peer assessment rules adopted in the technology-supported peer assessment studies. Many interesting results were found from the analyzed data. The following sections explain and discuss the findings one by one.

First, the results indicate that most technology-supported studies conducted peer assessment anonymously (69%). The main reasons for this, are that anonymity can reduce scoring bias (Magin, 2001) as well as protect learners' privacy with the support of technologies (Lin, 2016). Non-anonymous peer assessment may lead to the inflation of scores (Panadero & Brown, 2017). In fact, the choice of using anonymous versus non-anonymous peer assessment mainly depends on learning domains. A detailed analysis between anonymity and learning domains can be found in the results of RQ7.

Second, for peer assessment duration, around 54% of studies implemented peer assessment for more than 10 weeks, 22% of studies implemented peer assessment for 6-10 weeks, 14% of studies

implemented peer assessment less than one week, and 10% of studies implemented peer assessment for 2-5 weeks. It is suggested that practitioners should select appropriate assessment durations based on assignment types and learning domains. A further analysis on relationships among assessment durations, assignment types, and learning domains can be found in the results of RQ7 and RQ8.

Third, this study also investigated how assessors and assessees were matched in the technology-supported peer assessment studies. The results revealed that most studies randomly matched assessors and assesses by peer assessment supporting systems (65%), by teachers (23%), and by students themselves (12%). The tendency of studies to use systems to randomly match assessors and assessees may be due to previous research suggests that random assignment leads to less assessment bias (Li et al., 2016). In addition, students' products can be randomly distributed to peers for assessment by using the convenience of technologies, such as online peer assessment systems or social media. Therefore, it is suggested that random matching of assessors and assesses with the aid of technologies should be adopted in future studies.

Fourth, it was found that 78% of studies implemented only one round of peer assessment, the rest conducted peer assessment in two rounds or more. However, the internal reliability of peer assessment, namely the consistency within one assessor, can only be calculated after at least two rounds of peer assessment. In one study in which students took part in three rounds of peer assessment, it was discovered that students were engaged in cognitive processing in the first round, meta-cognitive processing in the second round, and providing affective feedback in the third round (Tsai & Liang, 2009). Thus, it is suggested that at least two-round peer assessment exercises should be conducted.

Finally, the results revealed that 73% of the studies did not provide reward for assessors. It is strongly suggested that learners who carefully participate in peer assessment activities and provide accurate feedback be rewarded through course credits, class participation points, extra points, bonus grades, books, and excursions in a timely manner so as to improve learners' engagement in peer assessment.

Table 3

Descriptive Data for Peer Assessment Rules

Category	Subcategory	Total n (%)
Anonymity	Anonymous	92(69)
	Non-anonymous	42(31)
Assessor training	Received training	73(55)
	No training	61(45)
Assessment duration	Less than one week	19(14)
	2-5 weeks	14(10)
	6-10 weeks	29(22)
	More than 10 weeks	72(54)
Assessor assignment	By system	87(65)
	By teachers	31(23)
	By students	16(12)
Round	One round	104(78)
	Two rounds or more	30(22)
Reward mechanism	With reward	36(27)
	Without reward	98(73)

RQ3: What Kinds of Evaluation Criteria Were Adopted in the Technology-Supported Peer Assessment Research?

Table 4 shows the descriptive data for the peer assessment criteria present across the studies reviewed. The results indicated that peer assessment criteria were developed by teachers among 94% of studies. Furthermore, it was found that most studies adopted both quantitative and qualitative feedback (61%). The rest of studies only adopted either quantitative (14%) or qualitative feedback (25%). This finding is in line with previous studies by Gielen and Wever (2015) who found that both scores and comments had the greatest effect on product quality. From this, it is suggested that peer assessment criteria should integrate both the quantitative and qualitative feedback in practice. Employing quantitative feedback may be most appropriate when the peer assessment is intended to evaluate peers' works in a summative way. Adopting qualitative feedback may be most suitable when the peer assessment intends to get detailed comments, constructive suggestions, or solutions.

With respect to the qualitative feedback format, the studies that adopted unstructured feedback (60%) are more than those that used structured feedback (40%). It is suggested that future studies adopt both structured qualitative feedback and quantitative feedback in order to provide guidelines, prompts, and templates to facilitate peer assessment.

Regarding the feedback method, most studies (97%) adopted written feedback. Few studies used a mixed feedback method (2%) or video feedback (1%). However, video or audio feedback can enable

better understanding of peer comments or suggestions. Thus, a mixed feedback format is recommended for future studies.

Table 4

Descriptive Data for the Peer Assessment Criteria

Category	Subcategory	Total n(%)
Criteria development	By teachers	126(94)
	By students	8(6)
Assessment method	Quantitative only	19(14)
	Qualitative only	33(25)
	Both quantitative and qualitative	82(61)
Quantitative feedback format	Score	61(60)
	Likert scale	40(40)
Qualitative feedback format	Structured feedback	46(40)
	Unstructured feedback	69(60)
Feedback method	Written feedback	130(97)
	Video feedback	1(1)
	Mixed feedback	3(2)

RQ4: How Were Labors Divided in the Technology-Supported Peer Assessment Research?

Table 5 demonstrates the division of labor in 134 technology-supported peer assessment studies. Regarding the grouping type, it was found that most studies adopted individual peer assessment – only 5% of studies conducted collaborative peer assessment. Collaborative peer assessment can increase the reliability and validity of peer assessment of the same student product, as it enables group members to discuss the quantitative and qualitative feedback before submitting final assessment results. Thus, bias will be decreased, and the accuracy of peer assessment will be increased. From this, it is suggested that collaborative peer assessment be adopted more frequently in future studies.

In addition, it was found that less than five assignments were evaluated by one assessor in 52% of studies. Among 26% of studies, 5-10 assignments were evaluated by one assessor. Among 22% of studies, more than 10 assignments were evaluated by one assessor. The reason for this might be that too many assignments may increase cognitive load for assessors. Concerning the number of assessors per assignment, it was found that 52% of studies invited less than five assessors to evaluate one assignment, 25% of studies invited more than 10 assessors, and 23% of studies invited 5-10 assessors. Within these studies, it may have been difficult to find assessors to evaluate assignments, thus leading to the relatively low number of individuals invited to assess assignments. It is suggested that in future studies, the number of assessors should be an odd number, and that at least three assessors should be required per assignment.

Table 5

Descriptive Data for the Division of Labor

Category	Subcategory	Total n (%)
Grouping type	Individual	127(95)
	Collaborative	7(5)
Number of assignments per assessor	Less than 5	70(52)
	Between 5–10	35(26)
	More than 10	29(22)
Number of assessors per assignment	Less than 5	69(52)
	Between 5–10	31(23)
	More than 10	34(25)

RQ5: What Were the Learning Objectives in the Technology-Supported Peer Assessment Research?

Table 6 presents the learning objectives of the 134 technology-supported peer assessment studies, including learning domains, assignment types, and learning outcomes. It was found that most peer assessment studies are conducted in social science (49%), followed by natural science (26%), and engineering and technological science (25%). In addition, the results indicated that the acting assignment is the least adopted type among four different assignment types (7%). Most studies focused on writing essays, project proposals, or artefacts. This implies that acting assignments such as performance, oral presentations, and speaking did not get enough attention in the technology-supported peer assessment studies. This reveals a mismatched to on the current educational trend, which is emphasizing on cultivating students' competence. It is suggested that future studies should pay more attention to assessing acting assignments with the aid of technologies. Furthermore, the findings also revealed that very few studies investigated students' attitudes or perceptions (9%) as learning outcomes. This suggests that future studies should focus on learning outcomes such as learning attitude, learning experience, satisfaction, and so on.

Table 6

Descriptive Data for the Learning Objectives

Category	Subcategory	Total n (%)
Learning domain	Natural science	35(26)
	Social science	66(49)
	Engineering and technological science	33(25)
Assignment type	Writing essays	47(35)
	Project proposals	31(23)
	Artefacts	47(35)
	Acting	9(7)
Outcome type	Cognitive outcomes	25(19)
	Attitudes or perceptions	13(9)
	Mixed	96(72)

RQ6: What Kinds of Tools Were Used in the Technology-Supported Peer Assessment Research?

Table 7 shows the tools adopted in the technology-supported peer assessment research. The results show that 42% of studies used a general learning management system, 35% of studies used dedicated peer assessment, 20% used social media, and 3% used mobile applications. Therefore, mobile applications were the least adopted tool used in the last 12 years within peer assessment studies. However, mobile devices such as mobile phones, and iPads have been widely used in the field of education. Mobile technologies enable learners to receive real-time feedback from peers and instructors, interact with peers instantly, and share information conveniently (Lai & Hwang, 2015). Therefore, it is strongly suggested that mobile-supported peer assessment should be adopted in practice so as to improve peer assessment efficiency and effectiveness.

In terms of the functionality of peer assessment system, most systems included basic functionalities such as assignment submission, peer grading, and making comments. Only 19% of studies developed dedicated systems with advanced functionalities, such as functions for supporting discussing with reviewers and criterion development.

Concerning scaffolding, the studies that providing scaffolding (17%) were less than those without scaffolding (83%). Hence, it is suggested that scaffolding should be embedded in peer assessment tools in order to facilitate peer assessment.

Table 7

Descriptive Data for the Peer Assessment Tools

Category	Subcategory	Total n (%)
System	Dedicated Web-based peer assessment system	47(35)
	General learning management system	56(42)
	Social media	27(20)
	Mobile application	4(3)
Functionalities	Basic	109(81)
	Advanced	25(19)
Scaffolding	With scaffolding	23(17)
	Without scaffolding	111(83)

RQ7: Are There any Significant Relationships Among Anonymity, Learning Domains, and Assessment Durations?

The present study also investigated how to choose anonymity and assessment durations based on learning domains as well as how to select learning domains and assessment durations based on assignment types. Table 8 shows the relationships discovered between learning domains and anonymity. The results indicated that there was a significant association between learning domains and anonymity ($\chi^2 = 8.47$, $p = 0.014$). For the studies whose learning domains were social science, non-anonymous assessment was adopted more than anonymous assessment (AR = 2.4). On the contrary, anonymous assessment was employed more than non-anonymous in engineering and technological science domains (AR = 2.7).

In addition, a significant relationship between learning domains and assessment durations was found ($\chi^2 = 17.88$, $p = 0.007$) (see Table 9). For the studies that centered on natural science domains, the assessment duration of less than one week demonstrated a growing tendency (AR = 3.4) and the assessment duration of more than 10 weeks showed a declining tendency (AR = -3.1). On contrast, if the studies were related to social science domains, the assessment duration of more than 10 weeks showed an increasing trend (AR = 2.6) and the assessment duration of less than one week displayed a decreasing trend (AR = -3.1).

Table 8

The Relationships Between Learning Domains and Anonymity

Learning domains	Anonymity
Natural science	1. Anonymous (AR = 0.0)
	2. Non-anonymous (AR = 0.0)
Social science	1. Anonymous (AR = -2.4)
	2. Non-anonymous (AR = 2.4)
Engineering and technological science	1. Anonymous (AR =2.7)
	2. Non-anonymous (AR = -2.7)

Note. AR: Adjusted residual values (AR with absolute values larger than 1.96 are significant).

Table 9

The Relationships Between Learning Domains and Assessment Durations

Learning domains	Assessment durations
Natural science	1. Less than one week (AR = 3.4)
	2. 2-5 weeks (AR = 0.9)
	3. 6-10 weeks (AR = 0.2)
	4. More than 10 weeks (AR = -3.1)
Social science	1. Less than one week (AR = -3.1)
	2. 2-5 weeks (AR = -1.1)
	3. 6-10 weeks (AR = 0.3)
	4. More than 10 weeks (AR = 2.6)
Engineering and technological science	1. Less than one week (AR = 0.2)
	2. 2-5 weeks (AR = 0.4)
	3. 6-10 weeks (AR = -0.6)
	4. More than 10 weeks (AR = 0.1)

RQ8: Are There any Significant Relationships Among Assignment Types, Learning Domains, and Assessment Durations?

The relationships between assignment types to be assessed and learning domains are shown in Table 10. As shown in Table 10, there was a positive relationship between assignment types and learning domains in the past 12 years (χ^2 = 30.96, p = 0.000). The adjusted residual value indicated that writing essays in social science subject domains had the fastest increasing trend (AR = 5.0), followed by artefacts in engineering and technological science (AR = 2.6) and natural science (AR = 2.2). On the other hand, writing essays in engineering and technological science domains (AR = -3.2) and in

natural science domains (AR = -2.6) as well as artefacts in social science domains (AR = -4.2) demonstrated a decreasing trend in the past 12 years. Therefore, teachers can select writing essays as the assignment type when the learning domain belongs to social science. When the learning domain focuses on engineering and technological science, teachers can engage students to design and assess artefacts, such as posters, websites, videos, course material, and so on. When the learning domain centers on natural science, project proposals such as WebQuest projects, training plans, research reports, and so on can be adopted as the assignment types.

Table 11 shows the relationship between assignment types and assessment durations. It was found that there was a significant association between assignment types and assessment durations (χ^2 = 18.61, p = 0.029). The adjusted residual value revealed that the assessment duration of time it typically took to assess written essays was 10 or more weeks (AR = 3.5), and that the assessment duration of time it typically took to assess project proposals was 6-10 weeks (AR = 2.1).   It is recommended that assessment duration differ depending on the type of assignment. Usually, assessing written essays should take more than 10 weeks. Assessing project plans should take 6-10 weeks.

Table 10

Relationships Between Assignment Types and Learning Domains

Assignment types	Learning domains
Writing essays	1. Natural science (AR = -2.6)
	2. Social science (AR = 5.0)
	3. Engineering and technological science (AR = -3.2)
Project proposal	1. Natural science (AR = 0.8)
	2. Social science (AR = -1.5)
	3. Engineering and technological science (AR = 1.0)
Artefacts	1. Natural science (AR = 2.2)
	2. Social science (AR = -4.2)
	3. Engineering and technological science (AR = 2.6)
Acting	1. Natural science (AR = -0.7)
	2. Social science (AR = 1.2)
	3. Engineering and technological science (AR = -0.7)

Note. AR: Adjusted residual values (AR with absolute values larger than 1.96 are significant).

Table 11

Relationships Between Assignment Types and Assessment Durations

Assignment types	Assessment durations
Writing essays	1. Less than one week (AR = -1.9)
	2. 2-5 weeks (AR = -1.7)
	3. 6-10 weeks (AR = -1.4)
	4. More than 10 weeks (AR = 3.5)
Project proposals	1. Less than one week (AR = 0.4)
	2. 2-5 weeks (AR = -0.2)
	3. 6-10 weeks (AR = 2.1)
	4. More than 10 weeks (AR = -1.9)
Artefacts	1. Less than one week (AR = 1.1)
	2. 2-5 weeks (AR = 1.1)
	3. 6-10 weeks (AR = -0.3)
	4. More than 10 weeks (AR = -1.2)
Acting	1. Less than one week (AR = 1.1)
	2. 2-5 weeks (AR = 1.6)
	3. 6-10 weeks (AR = -0.5)
	4. More than 10 weeks (AR = -1.4)

Conclusions

The present study extended the previous reviews on peer assessment by investigating subjects, objects, tools, rules, criteria, and division of labor for 134 technology-supported peer assessment studies published from 2006 to 2017. The main findings are summarized as follows.

First, it was found that most peer assessment activities were implemented in higher education. Usually, there were less than 50 participants who engaged in peer assessment activities. In addition, most peer assessment studies focused on mixed learning outcomes, rather than cognitive outcomes, attitudes, and perceptions. Acting assignments such as performance, oral presentations, or speaking were the least common type of assignments assessed across the studies reviewed. Second, it was found that anonymous assessment was more prominent in the studies than non-anonymous assessment. Most studies matched assessors and assesses randomly and implemented only one round of peer assessment. There were less studies that provided rewards to assessors than studies in which assessors were not provided rewards. Third, the results revealed that most studies developed peer assessment criteria by teachers rather than by students. Most studies adopted unstructured feedback

rather than structured feedback. There were more studies that adopted numeric scores than studies that employed Likert scales. Fourth, most studies conducted peer assessment individually rather than collaboratively. The analyzed result shows that in over half of the studies reviewed, less than five assessors were invited to evaluate an assignment; and less than five assignments were assigned to one assessor. Fifth, it was found that there were more studies that adopted a general learning management system than those that used a dedicated peer assessment system. There were less studies that provided scaffolding for peer assessment tasks than studies without scaffolding. Peer assessment tools with basic functionalities were more prominent within studies than peer assessment tools with advanced functionalities. Sixth, across studies, there were significant associations between learning domains and anonymity as well as assessment durations. Significant relationships between assignment types and learning domains as well as assessment durations were also found in this study.

Implications

The present study had several implications for practitioners and researchers. First, peer assessment is a very effective strategy that can be adopted in both small scale courses (Hsia et al., 2016) and massive open online courses (MOOCs; Formanek, Wenger, Buxner, Impey, & Sonam, 2017). In order to achieve better peer assessment results, teachers and practitioners should design peer assessment activities based on the following six components, namely subjects, objects, tools, rules, criteria, and division of labor. Second, the rules of peer assessment are very important for successful peer assessment activities. It is suggested at least two-rounds of peer assessment should be conducted. In addition, assessors who participated in peer assessment activities should be rewarded to stimulate motivations and improve feedback quality. Third, peer assessment criteria is another crucial element. The criteria development, assessment method, and feedback format (qualitative, quantitative, combination of both) should be designed elaborately before implementation. It is recommended that various formats of quantitative and qualitative feedback be integrated in peer assessment practice. Finally, peer assessment tools should be developed to facilitate the implementation of peer assessment. Different types of systems, advanced functionalities, and scaffolding should be developed in advance. Appropriate use of technologies to develop learners' positive attitudes, cognitions, metacognitions, emotions, behaviors, and values should be an ultimate goal of peer assessment.

Limitations and Future Studies

This study was constrained by two limitations. First, this study only included 134 studies published in related journals from 2006 to 2017. Therefore, cautions should be made when generalizing the results due to the small sample size and descriptive analysis. More data sources should be included in future study to conduct advanced statistical analysis. Second, this study mainly investigated six components of peer assessment activities. It is suggested that future studies should analyze students' behaviors and high-order thinking skills during technology-supported peer assessment activities. The following strategies are also recommended for peer assessment studies in the future.

Only few studies adopted mobile apps to conduct peer assessment. It is suggested that mobile-supported peer assessment should be adopted in future studies. In addition, learners' perceptions of peer assessment, behavior patterns, and higher-order thinking skills should also be investigated in future studies. In the study at hand, researchers did not explore how emotions may impact peer assessment. However, previous research suggests that emotional states have a great

impact on peer assessment quality (Cheng, Hou, & Wu, 2014). For example, learners who are experiencing positive emotions during assessment usually provide positive feedback, and learners who are experiencing negative emotions may tend to make negative comments. Therefore, it may be worth to explore how to promote positive emotions during peer assessment activities. Lastly, new techniques can be applied to automatically evaluate the quality of feedback, and intervention can be designed to help assessors provide high quality feedback.

Acknowledgement

References

Adachi, C., Tai, J., & Dawson, P. (2018). A framework for designing, implementing, communicating, and researching peer assessment. *Higher Education Research & Development, 37*(3), 453–467. https://doi.org/10.1080/07294360.2017.1405913

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: Developing the assessment design decisions framework. *Teaching in Higher Education, 21*(5), 545–556. https://doi.org/10.1080/13562517.2016.1160217

Chang, Y. H. (2016). Two decades of research in L2 peer review. *Journal of Writing Research, 8*(1), 81–117. https://doi.org/10.17239/jowr-2016.08.01.03

Cheng, K. H., Hou, H. T., & Wu, S. Y. (2014). Exploring students' emotional responses and participation in an online peer assessment activity: A case study. *Interactive Learning Environments, 22*(3), 271–287. https://doi.org/10.1080/10494820.2011.649766

Ching, Y. H., & Hsu, Y. C. (2016). Learners' interpersonal beliefs and generated feedback in an online role-playing peer-feedback activity: An exploratory study. *International Review of Research in Open & Distance Learning, 17*(2), 105–122. https://doi.org/10.19173/irrodl.v17i2.2221

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

Cheng, K. H., Liang, J. C., & Tsai, C. C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education, 25,* 78–84. https://doi.org/10.1016/j.iheduc.2015.02.001

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engeström, R. Miettinen, & R.-L. Punamäki (Eds.), *Perspectives on activity theory* (pp. 19–38). New York, NY: Cambridge University Press.

Engeström, Y. (2001). Expansive learning at work: Toward an activity-theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133–156. https://doi.org/10.1080/13639080020028747

Formanek, M., Wenger, M. C., Buxner, S. R., Impey, C. D., & Sonam, T. (2017). Insights about large-scale online peer assessment from an analysis of an astronomy MOOC. *Computers & Education, 113*, 243–262. https://doi.org/10.1016/j.compedu.2017.05.019

Gielen, M., & De Wever, B. (2015). Scripting the role of assessor and assessee in peer assessment in a Wiki environment: Impact on peer feedback quality and product improvement. *Computers & Education, 88,* 370–386. https://doi.org/10.1016/j.compedu.2015.07.012

Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137–155.

https://doi.org/10.1016/j.compedu.2015.07.012

Hsia, L. H., Huang, I., & Hwang, G. J. (2016). Effects of different online peer-feedback approaches on students' performance skills, motivation, and self-efficacy in a dance course. *Computers & Education, 96,* 55–71. https://doi.org/10.1016/j.compedu.2016.02.004

Hsu, T. C. (2016). Effects of a peer assessment system based on a grid-based knowledge classification approach on computer skills training. *Educational Technology & Society*, *19*(4), 100–111. Retrieved from https://www.j-ets.net/ETS/journals/19_4/9.pdf

Hwang, G. J., & Wu, P. H. (2014). Applications, impacts, and trends of mobile technology-enhanced learning: A review of 2008–2012 publications in selected SSCI journals. *International Journal of Mobile Learning and Organisation, 8*(2), 83–95. https://doi.org/10.1504/IJMLO.2014.062346

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. https://doi.org/10.1016/j.compedu.2015.02.011

Li, H., Xiong, Y., Zang, X., Kornhaber, M. L., Lyu, Y., Chung, K. S., & Suen, H. K. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education, 41*(2), 245–264. https://doi.org/10.1080/02602938.2014.999746

Lin, G. Y. (2016). Effects that facebook-based online peer assessment with micro-teaching videos can have on attitudes toward peer assessment and perceived learning from peer assessment. *Eurasia Journal of Mathematics, Science, & Technology Education, 12*(9), 2295–2307. https://doi.org/10.12973/eurasia.2016.1280a

Magin, D. (2001). Reciprocity as a source of bias in multiple peer assessment of group work. *Studies in Higher Education, 26*(1), 53–63. https://doi.org/10.1080/03075070123503

Mostert, M., & Snowball, J. D. (2013). Where angels fear to tread: Online peer-assessment in a large first-year class. *Assessment & Evaluation in Higher Education, 38*(6), 674–686. https://doi.org/ 10.1080/02602938.2012.683770

Panadero, E., & Brown, G. T. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education, 32*(1), 133–156. https://doi.org/10.1007/s10212-015-0282-5

Park, Y., & Jo, I. H. (2017). Using log variables in a learning management system to evaluate learning activity using the lens of activity theory. *Assessment & Evaluation in Higher Education*, *42*(4), 531–547. https://doi.org/10.1080/02602938.2016.1158236

Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment*, *11*(1), 1–31. https://doi.org/10.1207/s15326977ea1101_1

Shih, R. C. (2011). Can Web 2.0 technology assist college students in learning English writing? Integrating Facebook and peer assessment with blended learning. *Australasian Journal of Educational Technology, 27*(5), 829–845. https://doi.org/10.14742/ajet.934

Speyer, R., Pilz, W., Van Der Kruis, J., & Brunings, J. W. (2011). Reliability and validity of student peer assessment in medical education: A systematic review. *Medical Teacher, 33*(11), 572–585. https://doi.org/10.3109/0142159X.2011.610835

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. https://doi.org/10.2307/1170598

Topping, K. (2017). Peer assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology*, *1*(1), 1–17. doi: 10.31532/InterdiscipEducPsychol.1.1.007

Tsai, C. C. (2009). Internet-based peer assessment in high school settings. In L. T. W. Hin & R. Subramaniam (Eds.), *Handbook of research on new media literacy at the K-12 level: Issues and challenges* (pp. 743-754). Hershey, PA: Information Science Reference (IGI Global).

Tsai, C. C., & Liang, J. C. (2009). The development of science activities via on-line peer assessment: The role of scientific epistemological views. *Instructional Science*, *37*(3), 293–310. https://doi.org/10.1007/s11251-007-9047-0

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.

Wulf, J., Blohm, I., Leimeister, J. M., & Brenner, W. (2014). Massive open online courses. *Business & Information Systems Engineering, 6*(2), 111–114. https://doi.org/10.1007/s12599-014-0313-9

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education, 11*(3-4), 186–193. https://doi.org/10.1016/j.iheduc.2008.06.005

Yu, F. Y., & Wu, C. P. (2011). Different identity revelation modes in an online peer-assessment learning environment: Effects on perceptions toward assessors, classroom climate, and learning activities. *Computers & Education, 57,* 2167–2177. https://doi.org/10.1016/j.compedu.2011.05.012

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. https://doi.org/10.1504/IJMLO.2016.079503

Zheng, L., Huang, R., & Yu, J. (2014). Identifying computer-supported collaborative learning (CSCL) research in selected journals published from 2003 to 2012: A content analysis of research topics and issues. *Educational Technology & Society, 17*(4), 335–351. Retrieved from

https://www.j-ets.net/ETS/journals/17_4/23.pdf

Athabasca
University