Journal of Information Technology Education: Innovations in Practice

An Official Publication of the Informing Science Institute
InformingScience.org

JITEiip.org

Volume 24, 2025

WHAT INTERVENTIONS IMPROVE THE EFFECTIVENESS OF TECHNOLOGY-FACILITATED PEER ASSESSMENT? A META-ANALYSIS

Kistantia Elok Mumpuni*	Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; and Universitas Sebelas Maret, Surakarta, Indonesia	kistantiaelok.2022@student.uny.ac.id; kistantiaelok@staff.uns.ac.id
Samsul Hadi	Universitas Negeri Yogyakarta, Yogyakarta, Indonesia	samsul_hd@uny.ac.id
Slamet Suyanto	Universitas Negeri Yogyakarta, Yogyakarta, Indonesia	slamet_suyanto@uny.ac.id

* Corresponding author

ABSTRACT

Aim/Purpose	This meta-analysis aims to examine the effect size of various interventions in technology-facilitated peer assessment of learning performance. Furthermore, it aims to identify and recommend effective peer assessment interventions that highlight opportunities for enhancing peer assessment practices and support the strategic implementation of interventions to optimize learning performance.
Background	Peer assessment is a pivotal pedagogical tool, fostering students' critical judgment and self-assessment skills and providing educators with a valuable understanding of individual progress. However, concerns among teachers and students persist regarding the effectiveness of peer assessment. These challenges can be addressed through intervention settings and technological facilities.
Methodology	This meta-analysis examines how different intervention settings influence learning outcomes in technology-facilitated peer assessment. Using the PRISMA framework, 24 eligible studies comprising 79 data sets were systematically identified based on predefined inclusion and exclusion criteria. Extracted data were organized into four principal categories: (1) participant characteristics, (2) intervention moderators, (3) research outcomes, and (4) study records. A random-

Accepting Editor Stamatis Papadakis | Received: May 25, 2025 | Revised: July 11, July 19, 2025 | Accepted: July 29, 2025.
Cite as: Mumpuni, K. E., Hadi, S., & Suyanto, S. (2025). What interventions improve the effectiveness of technology-facilitated peer assessment? A meta-analysis. *Journal of Information Technology Education: Innovations in Practice, 24,* Article 16. https://doi.org/10.28945/5606

	effects model was employed to synthesize the findings. All analyses were conducted in R, primarily using the {meta} package, with additional support from {dmetar} and {metafor} for outlier detection and visualization.
Contribution	This study offers a synthesized overview of current research on technology-facilitated peer assessment, contributing to a clearer understanding of its setting and implementation. By identifying relevant trends and effective intervention areas, the findings provide useful guidance for researchers and educators in developing more informed and context-appropriate peer assessment strategies. The review also highlights how technology-facilitated peer assessment can be used for student learning and teacher facilitation through more efficient monitoring and feedback delivery.
Findings	The analysis revealed an effect size of 0.31 [95% CI: 0.230–0.413], confirming the positive impact of well-designed interventions in technology-facilitated peer assessment compared to alternative or minimal intervention. These interventions enhance key elements such as rubric comprehension, feedback quality, evaluative autonomy, and self-reflection, leading to improved learning outcomes. Effective settings prioritize process-oriented strategies, reciprocal roles, individually allocated feedback, and input from a single peer, while allowing for adaptation based on the characteristics of students, instructors, school contexts, and instructional content.
Recommendations for Practitioners	Teachers should first determine whether the peer assessment targets written work, products, or performance, ensuring alignment with learning objectives. Subsequently, the setting of the PA intervention should include clear specifications regarding the type of feedback, assessment procedures, and the roles of participants. Providing structured training for both assessors and assessees is essential to enhance feedback quality. Additional factors warranting consideration include the level of anonymity in the assessment process, the duration of the assessment period, student characteristics, and the potential for friendship bias.
Recommendations for Researchers	While peer assessment has demonstrated validity and reliability across various educational contexts, researchers should continue to explore the conditions under which it most effectively supports learning. Future studies should investigate strategies to minimize potential social challenges, such as peer conflict or bias, through careful group formation and facilitation. The selection of technology should prioritize usability and minimize additional burden for both teachers and students. Additionally, research should consider the impact of voluntary versus mandatory participation, as fostering intrinsic motivation may enhance the quality and acceptance of peer assessment practices.
Impact on Society	The findings underscore that, while peer assessment holds potential for enhancing learning, many students initially struggle to provide meaningful feedback, and teachers may question the credibility of peer-generated evaluations. These challenges point to a broader need for structured interventions, such as training, scaffolding with rubrics, and thoughtful feedback design, to build students' confidence and competence in assessing peers. For educators and policymakers, this suggests that investing in supportive structures and technologies for peer assessment can lead to more engaging and effective classroom practices.
Future Research	Future research should focus on developing peer assessment platforms to meet the needs of educational contexts. Beyond refining the intervention setting, further studies should explore the integration of peer and teacher feedback for

more robust empirical findings. Expanding research across diverse subjects and educational levels is also essential.

Keywords peer assessment, intervention, technology-facilitated, learning performance, meta-analysis

INTRODUCTION

Peer assessment (PA) has gained significant recognition among scholars for its crucial role in fostering essential learning skills and competencies. Through PA, students engage in evaluating and providing feedback on their peers' work, including grading, assessing quality, and appraising performance. This process applies to a wide range of activities, such as science laboratory experiments, presentations, argumentative essays, and project outcomes (Topping, 2009). Engaging with peers also offers valuable benefits (van Helden et al., 2023), fostering the development of judgment, evaluative abilities, and self-assessment skills. This approach not only fosters the development of essential skills, such as delivering constructive feedback and formulating critical judgments, but also enhances students' abilities to analyze past experiences and refine future decision-making from both self-regulatory and outcome-oriented perspectives.

Despite the many benefits of peer assessment, some teachers remain uncertain about the validity and usefulness of student-generated feedback and scores (Rotsaert, Panadero, & Schellens, 2018). Initial attempts at peer assessment often result in low-quality feedback, as students find the process challenging (Hovardas et al., 2014). Some students also reported that peer comments tend to be too brief or vague to understand (Hsia et al., 2016). Another issue in peer assessment is the privacy of the assessor's identity (Barahona et al., 2023), temporal and logistical barriers (Broadbent et al., 2018), and difficulty following up on received feedback (Gielen & De Wever, 2015), especially in the implementation of traditional peer assessment, face-to-face, and paper-based.

These indicate that guidance, training, and well-defined rubrics are needed to increase students' self-confidence and self-regulation (Ghahari & Sedaghat, 2018; Hsia et al., 2016). Therefore, interventions in peer assessment are implemented as targeted actions, referring to deliberate strategies or measures designed to enhance their effectiveness and facilitate the assessment process. Teachers can carry out various interventions to maximize the quality of peer assessment. The results of this intervention have been empirically proven to improve learning performances, including anonymity (G. Y. Lin, 2018), providing rubric and modified feedback (Gielen & De Wever, 2015; Hsia et al., 2016; Latifi et al., 2021), assessor training (Ocampo et al., 2023; Sluijsmans et al., 2002), interactive feedback (H. C. Lin et al., 2021; Zheng et al., 2018), and interactive rubric development (Lai & Hwang, 2015), among others.

Technology also offers viable solutions to these issues. For instance, the use of network devices (e.g., tablets or laptops) to protect students' anonymity, allowing them to provide feedback more freely based on rubrics (Hsia & Hwang, 2021). Additionally, technology-facilitated assistance assists task assignments, provides anonymous feedback, and supports revision based on peer feedback, thereby enhancing critical thinking skills (Barahona et al., 2023). Digitization of the assessment process also affects the preferences and results of assessments, as they are more accustomed to using digital tools. Peer assessment based on computers or gadgets also showed more significant effects and quality than using paper (H. Li et al., 2019) and has a higher value of flexibility, efficiency, and accessibility (Chen, 2016).

To further estimate the effectiveness of these interventions in technology-facilitated PA (TFPA), a meta-analysis is needed to determine the overall and subgroup effect sizes of TFPA on learning performance. This analysis will provide a quantitative synthesis of existing studies, offering valuable

guidance for teachers, students, and policymakers on effectively implementing peer assessment for optimal outcomes.

Research Questions

This study aims to examine the effect sizes of different intervention settings in TFPA and their impact on student learning outcomes. The overarching goal is to identify and recommend effective approaches that enhance the quality and impact of PA practices. Five research questions were formulated to guide this investigation.

RQ1. Which types of interventions yield the most significant effects when implemented in technology-facilitated peer assessment?
RQ2. Which approach – reciprocal or non-reciprocal peer assessment – results in more effective learning performances?
RQ3. How should feedback and scoring be delivered – individually or in groups – to maximize the effectiveness of peer assessment?
RQ4. What is the optimal number of peer feedback provider instances in peer assessment to enhance learning performances?
RQ5. What is an effective peer assessment setting that involves appropriate interventions to enhance learning performances?

Literature Review

This literature review provides an overview of peer assessment interventions, the integration of technology in peer assessment practices, and previous meta-analyses related to TFPA. By organizing these discussions, this section aims to contextualize the current study and highlight the rationale for conducting a meta-analysis in this area.

Peer Assessment Intervention

The first concept of peer assessment (PA) is to improve learning performance with the help of peers (peer-assisted learning), which results in much better results than learning alone (Topping & Ehly, 1998). In PA, students with comparable levels of knowledge or backgrounds engage not only in learning and completing assignments but also in the role of a tutor, evaluating their peers' work, receiving feedback, and refining their assignments (Topping, 1998). van Helden et al. (2023) state three distinct types of PA – peer review, peer grading, and peer evaluation – each serving different educational purposes. Peer review involves students assessing each other's written output and providing feedback. Peer grading is when students evaluate each other's work, either formative or summative, based on specified criteria. Peer evaluation encompasses students assessing each other within group dynamics, reflecting on transversal skills, like teamwork, effort, or intellectual contributions related to their assignment, and offering feedback within this context. The selection of PA articles used in this paper refers to all of van Helden's types.

In the development of PA implementation, different intervention settings are applied, offering various opportunities for improvement and adaptation. Interventions involve strategies or actions aimed at improving the quality of peer assessment. While some elements, such as training or the use of rubrics, may be considered standard components of any peer assessment activity, in quasi-experimental studies where these elements were treated as variables between control and experimental groups, we classified them as interventions. In this study, interventions include assessors' anonymity versus identification, providing or abstention rubric, training or no training for assessors, group or individual assessor, and variation in feedback type. Furthermore, intervention studies by Hoo et al. (2022) were designed to improve students' feedback literacy, where instructors were occasionally involved in the PA process and even provided suggestions regarding follow-up received as a form of scaffolding. Various effects of intervention in PA will be discussed in this study to identify the most effective or appropriate intervention settings to meet students' needs.

Evaluating the suitability of peer assessment interventions in meeting students' needs requires an understanding of their impact on learning performance. In this study, learning performance refers to both the outcomes and processes students demonstrate during learning (Moccozet, 2012). It reflects their ability to apply acquired knowledge and skills (Yin & Yuan, 2021) and is influenced by teacher competencies and student interactions (Costa et al., 2015). It also considers factors such as learning attitude, motivation, cognitive load, metacognitive awareness, and attitudes toward giving and receiving peer feedback.

Technology-Facilitated Peer Assessment

Technology-facilitated peer assessment (TFPA) is an effort to make it easier for students and teachers to organize, provide assessments, and follow up. It was conducted to support educators in organizing their teaching plans, especially focusing on peer assessment (Ocampo & Panadero, 2023). Technology in this study refers to computing devices (e.g., laptops, tablets, smartphones, and video recorders) and digital systems and applications (e.g., digital platforms, assessment software, and web-based peer assessment). The use of technology, like mobile peer assessment, improved innovative competence and student skills (Hsia & Hwang, 2021).

TFPA serves to support both students and teachers in organizing, delivering, and following up on assessments. Technology devices such as smartphones, tablets, laptops, and video recorders are commonly used to access or develop platforms designed for peer assessment, enabling more student-centered evaluation approaches (Ocampo & Panadero, 2023). TFPA not only aids teachers in streamlining instructional planning but also reflects advancements in web-based assessment technologies. Notably, mobile peer assessment has been shown to enhance students' innovative competencies and academic skills (Hsia & Hwang, 2021). Student feedback further indicates positive impacts on social skills, intrinsic motivation, and the ability to understand evaluation criteria, engage in objective self-reflection, and appreciate the quality of peers' work.

The selection of appropriate technology for peer assessment should be guided by the digital competencies of both teachers and students, as well as the availability of institutional support (van Helden et al., 2023). Higher levels of digital literacy and adequate infrastructure allow for the integration of more advanced tools. Equally important is ensuring that the chosen technology aligns with the pedagogical design of the peer assessment. When tool functionality and instructional goals are well-matched, the implementation of peer assessment becomes more effective and pedagogically meaningful.

The positive impact of technology on learning performance has been well-documented. Mulet et al. (2019) found that technology integration contributes to improved student outcomes. Similarly, Wang et al. (2021) reported that mobile-based interventions enhanced students' learning motivation, self-efficacy, and group efficacy. These findings offer valuable guidance for educators and platform developers in designing technology-facilitated peer assessment systems that effectively support student learning and engagement.

Meta-Analysis for TFPA

Meta-analysis, introduced by Glass and Smith (1979), offers a systematic approach to synthesizing empirical findings across studies and has been widely applied in education and social sciences. This method provides a systematic and quantitative approach for integrating results from multiple empirical studies, offering deeper insights than traditional literature reviews. Meta-analysis focuses on calculating effect sizes rather than relying solely on p-values, enabling a more comprehensive and statistically robust synthesis of findings (Borenstein et al., 2009; Lipsey & Wilson, 2001).

To estimate the overall impact of TFPA on learning performance, meta-analysis computes a summary effect size while accounting for variability within and between studies using a random-effects model. Effect sizes, commonly expressed as standardized mean differences (e.g., Cohen's d or Hedges' g), allow for comparisons across diverse outcome measures. When substantial heterogeneity

exists, subgroup analyses using moderator variables help identify meaningful patterns and reduce unexplained variance across studies (Borenstein et al., 2009; Hunter & Schmidt, 2006). This analytical rigor supports the aim of this study: to generate evidence-based guidance that informs effective implementation of technology-facilitated peer assessment.

Previous meta-analysis studies have consistently highlighted the positive impact of technology-facilitated peer assessment (TFPA) on learning performance. C.-Y. Chang et al. (2021) highlighted significant improvements in diverse settings when PA is supported by technology. Zhan et al. (2023) found that online PA enhances higher-order thinking, especially convergent thinking. Yan et al. (2022) reported academic gains from both self- and peer-assessment, with stronger effects in online formats. Double et al. (2020) emphasized the need for rigorous experimental designs to isolate TFPA's effects, highlighting the importance of design over delivery mode. Collectively, these findings support TFPA's effectiveness and underscore the need for further meta-analysis research to guide practice.

Meta-analysis not only confirms TFPA's general effectiveness but also identifies specific intervention settings that are associated with improved outcomes. Understanding which implementation strategies and instructional designs are most effective provides valuable guidance for educators. Such evidence-based practices recommend effective intervention settings for improving the quality and impact of peer assessment.

METHOD

This study investigates the impact of different intervention settings within a technology-facilitated peer assessment context. The experimental groups, classified as implementing stronger interventions, showed more positive effects on learning performance compared to control groups receiving alternative or minimal interventions. A meta-analysis is conducted to synthesize the findings. The data used in the meta-analysis were sourced from studies relevant to the objectives. Only studies meeting specific criteria and containing similar variables were selected to ensure accuracy (Borenstein et al., 2009). It can be done in the context of a systematic review, which is the process of systematically searching, assessing, and synthesizing data from a large number of studies.

There are three essential stages in this meta-analysis study as follows.

STUDY SELECTION

Meta-analysis refers to the statistical synthesis of results from a set of studies. Although the statistical methods used in meta-analyses can be applied to any data, synthesis is only meaningful if research is collected systematically (Borenstein et al., 2009). In this study, the characteristics of the data were restricted to interventions in peer assessment settings, peer assessments carried out online or facilitated by technology, quasi-experimental studies, sufficient statistical information, reputable articles, and written in English.

Relevant and reliable studies for this meta-analysis were identified through Scopus databases and platforms of reputable academic publishers. These sources were selected due to their broad usage and inclusion of high-quality content. The timespan for publication was from 2010 to 2024. To identify suitable studies, we searched with possible combinations of keywords grouped into two distinct categories: those related to peer assessment such as "peer assessment" OR "peer feedback" OR "peer review" OR "peer" OR "online peer assessment" OR "peer evaluation" OR "peer scoring", and related technology-facilitate such as "technology-facilitate" OR "technology-supported" OR "online peer". We reviewed existing systematic studies, examined their references, and searched for articles citing key peer assessment research. This approach identified relevant studies potentially undetected by keyword searches.

Selecting studies following the PRISMA flowchart (Moher et al., 2009) shows the four stages of determining which articles are eligible. First is identification, which is to get the initial information from

the article records. The second is screening to filter some article records based on general criteria. The third is eligibility, which is the stage where researchers will filter with specific criteria to get the expected information. The last included is the final number of records that will be used in the meta-analysis. In addition, researchers have identified new studies using other methods, such as PRISMA 2020 (Page et al., 2021). One method is to track articles through citations from credible articles, according to criteria. Thus, newer articles of this type can be obtained. The PRISMA flow diagram for this meta-analysis is shown in Figure 1.

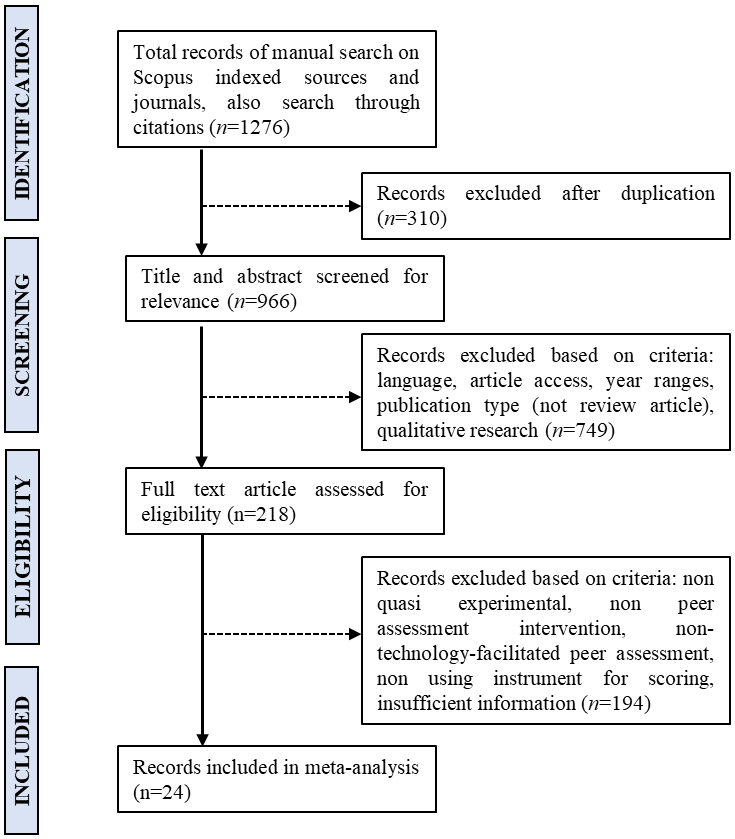

Figure 1. PRISMA flow diagram of the search and screening process

In the initial search, we found 1276 articles using keywords and citation searches. A total of 966 articles were obtained after clearing duplication from different databases. In the first screening, categories were used: (1) language, the article must be in English; (2) article access, articles can be downloaded for review; (3) time range from 2010 to 2024; (4) publication type, only research paper; and (5) quantitative approach for data analysis. So, as many as 749 articles were excluded. The full texts of 218 articles were examined at the eligibility stage with advanced criteria, namely whether the article includes quasi-experimental, peer assessment intervention, online or technology-facilitated peer assessment, using an instrument for scoring, and sufficient statistical information. We included only articles that implement peer assessment in both control and experimental groups, where at least one group receives an intervention, such as training, feedback requests, or anonymity. Studies comparing peer and teacher/expert assessment results were excluded. This research focuses on studies that integrate technology into peer assessment to explore its impact on learning. Only studies with well-defined assessment instruments and reliable data, including clear minimum and maximum scores, were considered. However, studies that lacked mean and standard deviation data for both control and experimental groups were excluded, despite meeting other criteria. Additionally, studies using quantita-

tive content analysis were not included. As a result, 194 articles were excluded based on eligibility criteria, resulting in 24 articles for meta-analysis. Some of these articles contained multiple data studies, as they examined control and multiple experimental group comparisons across different instructional settings and learning performance – the final comprised 24 studies encompassing 81 data sets.

CODING OF STUDIES

At the end of the selection study, 24 studies were used for analysis. Each study may contribute one or multiple data sets for effect size calculation. If a study includes one control group and multiple experimental groups, multiple data sets can be derived. Similarly, if a study evaluates multiple aspects of learning performance, additional data sets are extracted. From the 24 selected studies, a total of 81 data sets were obtained. To account for variations in effect sizes, a set of moderator variables was identified and systematically coded (Zhan et al., 2023). To better understand variations in effect sizes, key moderator variables were identified and systematically coded in a form.

The form was carefully developed and refined to include four essential components: (1) participant demographics, (2) intervention moderators, (3) research data, and (4) records. Article demographic information in the form of title, author, year, country, resource, age of participants, school level, and keywords was included. Intervention moderator information includes anonymity, peer assessment rubrics, number of assessors, feedback for groups or individually, reciprocal mode, feedback variation, and dialogical interaction. The coding and definition of moderators are presented in Table 1. Information on research data includes learning performance assessed, number of participants (n), mean, and standard deviation (SD) in each control and experimental class. If the study presents data in unit form or per aspect of specific learning performance, the researcher calculates the mean and standard deviation. Additionally, the researcher includes notes for each study, covering details on the assessment instrument (such as validity and reliability index), the technology used, key findings, and other relevant observations.

Table 1. The code and definition of moderators

Moderator code	Definition
PA intervention types	Intervention in peer assessment (PA) occurring in the experimental class. The PA interventions are categorized into three types: "participant" (treatment to students, such as training, anonymity, expert guidance, or role assignment), "process" (involving the use of rubrics, interactive feedback, free selection and invitations, or scaffolding facilitation), and "feedback" (types of feedback, including rating-comment, feedforward, structured feedback, or scripted feedback).
Reciprocal	Reciprocal peer assessment refers to a setting in which students take on both roles simultaneously: as assessors (giving feedback) and as assessees (receiving feedback). If students engage in both roles, it is considered "Yes"; if they only perform one role, it is "No".
Allocation	Allocation explains whether the feedback and scores provided by assessors are directed at "individual" or "group".
Number of PFB (peer feedback) providers	The number of PFB providers refers to the number of assessors who provide feedback to each individual or group. The number of PFB providers is classified into four, namely "1", "2", "3", and "more than 3".

STATISTICAL ANALYSIS

After tabulating the data, we calculated the effect size and determined the summary of the effect model. Based on the explanation of Borenstein et al. (2009), if the data are obtained from different studies and different instruments to assess learning outcomes, then the measurement scale will differ among the studies. Thus, we cannot combine the differences in raw averages. Ultimately, the mean

difference was divided within each study by the standard deviation. In this study, standard deviations were combined across studies to calculate the standardized mean difference (SMD) as a common effect size index. Given the heterogeneity among studies, a random effects model was employed to address variability across studies. Variations in this study include participant characteristics, interventions in peer conditioning, and observed learning performance.

Furthermore, the analysis included procedures to detect outliers and potential publication bias. This detection aims to identify studies with effect sizes that differ substantially from the overall distribution. After eliminating outliers and biased studies, the data were analyzed for effect size. In addition to calculating the overall effect size of the impact of peer assessment intervention on learning performance, we also explored factors that moderate the effect size. These moderators can be classified as shown in Table 1.

All analyses were performed in R version 4.2.1 (Posit Team, 2023). The main analysis was performed using {meta} packages (Schwarzer et al., 2015), which include effect size, forest plot, and publication bias. Detection of outliers and plots used {dmetar} packages (Harrer et al., 2022) and {metafor} packages (Viechtbauer, 2010).

RESULT

Overall, the findings indicate that variations in intervention type led to different impacts on learning performance, with experimental groups generally outperforming control groups. This section describes the characteristics of studies, detection of outliers and biases, overall effect, effect on moderator variables, and responses to the research questions, respectively.

OVERVIEW OF SELECTED STUDIES

Meta-analysis serves to examine empirical studies that previous researchers have published. The quantitative research reviewed has the criterion of carrying out peer assessment that has been intervened and facilitated with technology. A total of 24 studies containing 81 data sets were analyzed using R. The description of the characteristics of each study is presented in Table 2. The studies, conducted between 2012 and 2024 across various countries, reflect a wide range of educational levels, including primary, secondary, and university settings. The subject areas also vary, encompassing Social Sciences, Natural Sciences, and the Humanities. In terms of assessment objects in this study, peer assessment was applied to three main categories: performance (e.g., presentations, microteaching, or demonstrations), product (e.g., concept maps or posters), and writing (e.g., essays or reports).

Outlier detection

Outliers can significantly affect the accuracy of a meta-analysis by distorting the combined effect size (Harrer et al., 2022). To enhance the robustness of the findings, this study applied Influence Analysis and GOSH plot analysis using the {dmetar} and {metafor} packages. These methods were used to detect studies that had an unusually strong influence on the overall findings, allowing for a more balanced and valid interpretation of the data.

Outlier detection results are shown in Figure 2. Influence Analysis using Baujat plots (Figure 2a) shows the contribution of each study to overall heterogeneity (as measured by Cochran's Q) on the horizontal axis, and its effect on the size of the combined effect on the vertical axis (Baujat et al., 2002). Three study data points were located on the right and upper sections of the plot: Valero Haro et al. (2024) (1), Xie and Zhang (2024), and Latifi et al. (2023) (1). K-means clustering from the GOSH plot analysis identified two major outliers: Valero Haro et al. (2024) (1) and Xie and Zhang (2024). Figures 2b and 2c present the GOSH plot, highlighting that study data set 70 (Valero Haro et al., 2024) (1) and 81 (Xie & Zhang, 2024) exhibited both high heterogeneity and larger effect sizes.

Table 2. Characteristics of studies (n = 24)

Author(s) & year	Country	Educational level	Technologies tools	Material/ course	Intervention PA on	Object assessed	Research issues (variable)
Gielen and De Wever (2012)	Belgium	University	Wiki (computer-supported collaborative learning environment (CSCL))	Instructional Sciences	Structured feedback	Writing	Learning effect, quality of product, providing PFB, receiving PFB, received PFB, provided PFB
Gielen and De Wever (2015)	Belgium	University	Wiki-based computer-supported	Instructional Sciences class	Additional Feedback condition	Writing a scientific article	Perception towards the additional FB condition
Lee et al. (2016)	Taiwan	Secondary	Web-based peer assessment system	Computer course for PhotoCap6	PA with scaffolding	Creative works with PhotoCap6	Learning attitude
Lai and Hwang (2015)	Taiwan	Primary	Mobile learning system	Art course	Interactive rubric	Poster	Learning motivation, meta-cognitive awareness, cognitive load
Hsia et al. (2016)	Taiwan	University	Online peer feedback system (MyeDance); videotaped	Dance courses	Various feedback	Dance perform	Dance performance, intrinsic motivation, extrinsic motivation, self-efficacy
L. Li (2016)	USA	University	The self and peer assessment tool in Blackboard	Technology application course (for education)	Anonymity and training assessor	Developed a WebQuest	Value/usefulness, pressure/tension
van Ginkel et al. (2017)	Netherlands	University	Videotaped	Forest and nature conservation; nutrition and health	Guiding	Oral presentation	Specificity of FB, actual performance, structure of FB, intensity of FB, content-related arguments

Author(s) & year	Country	Educational level	Technologies tools	Material/ course	Intervention PA on	Object assessed	Research issues (variable)
Rotsaert, Panadero, Schellens, and Raes (2018)	Belgium	Secondary	Mobile response technology (MRT), free tool Socrative™ (Bèta Release)	Specific internship institute (e.g., a local library)	Guiding	Presentation	Perceived usefulness of PFB, students' PFB skills perception
Güler (2017)	Turkey	University	WhatsApp	Computer education and instructional technologies (education)	Anonymity	Present a workshop	Anonymity-scale scores, perceived rating fairness scale, positive attitude scale
G. Y. Lin (2018)	Taiwan	University	Online peer assessment and reflection (onpeer) system; Facebook-based learning application; videotaped	Teacher-training course	Anonymity	Microteaching performance	Perceived fairness, perceived overall learning, attitude toward the system
Zheng et al. (2018)	China	University	Web-based peer assessment	Essay about relaxation	Interactive feedback	Writing an essay	Writing performance, meta-cognitive awareness, self-efficacy
Huisman et al. (2018)	Netherlands	University	Virtual learning environment	Education and child development studies	Role as assessors and assessees	Writing an essay	Writing performance
van den Bos and Tan (2019)	Netherlands	University	Peergrade (free online peergrade)	English of the Common European Framework of Reference for languages	Anonymity	Writing	Writing performance

Author(s) & year	Country	Educational level	Technologies tools	Material/ course	Intervention PA on	Object assessed	Research issues (variable)
Hwang and Chang (2021)	Taiwan	Primary	Mobile concept mapping learning system (by tablet)	Yehliu geology	Interactive feedback	Concept map	Learning motivation, self-efficacy, critical thinking tendency, environmental identity, cognitive load, assessment FB quality, concept mapping score
H. C. Lin et al. (2021)	Taiwan	University	Online interactive peer-review system; videotaped	Health assessment training course	Interactive feedback	Demonstration health assessment	Critical thinking tendency, self-efficacy, reflective thinking, peer review
Latifi et al. (2021)	Iran	University	EduTech (online peer learning platform)	Applying Computer in the Educational Sciences course	Feedback and forward	Argumentative essay writing	Quality of peer learning processes, quality of argumentative essay writing, domain-specific knowledge learning
Dmoshinskaia et al. (2021)	Russia	Secondary	Special online peer assessment tool	Physics lesson	Providing rubric	Concept map	Knowledge test
Ocampo et al. (2023)	Philippines	University	Eduflow	Psychology classes (about online learning)	Training assessor	Writing an experimental research paper	Writing' peer score, perceptions of trust in one's self as an assessor, perceptions of comfort

Author(s) & year	Country	Educational level	Technologies tools	Material/ course	Intervention PA on	Object assessed	Research issues (variable)
Latifi et al. (2023)	Iran	University	EduTech	Applying computers in education (about "The use of mobile phones and tablets in the classroom should be banned")	Providing rubric	Argumentative essay	FB quality, quality of argumentative essay, domain-specific learning
Valero Haro et al. (2024)	Netherlands	University	Online educational classrooms	Introduction to Molecular Life Sciences and Biotechnology	Providing rubric	Essay about "Genetically Modified Organisms (GMOs)"	PFB quality, domain-specific knowledge
Y. Lin et al. (2023)	China	University	Cloud, classroom online learning system	Research Methods in Education course	Free selection and invitation	Research proposal	Participants' flow, participants' motivation
Stančić et al. (2024)	Serbia	University	Uploaded system	Systems for supporting children's linguistic (pedagogy)	Role as Peer experts and Peer users	Create learning material	Perceived competence, interest, effort, usefulness
Xiaomeng and Ravindran (2024)	China	University	Online platform	English class (whether to study at home or abroad)	Role as assessors and assessees	Argumentative essay	Writing performance
Xie and Zhang (2024)	Hongkong	University	Moodle and Google Docs	Linguistic course	Anonymity, rubric, and traceability	Oral presentation	Feedback quality

Note: PA: Peer Assessment, PFB: Peer Feedback; FB: Feedback

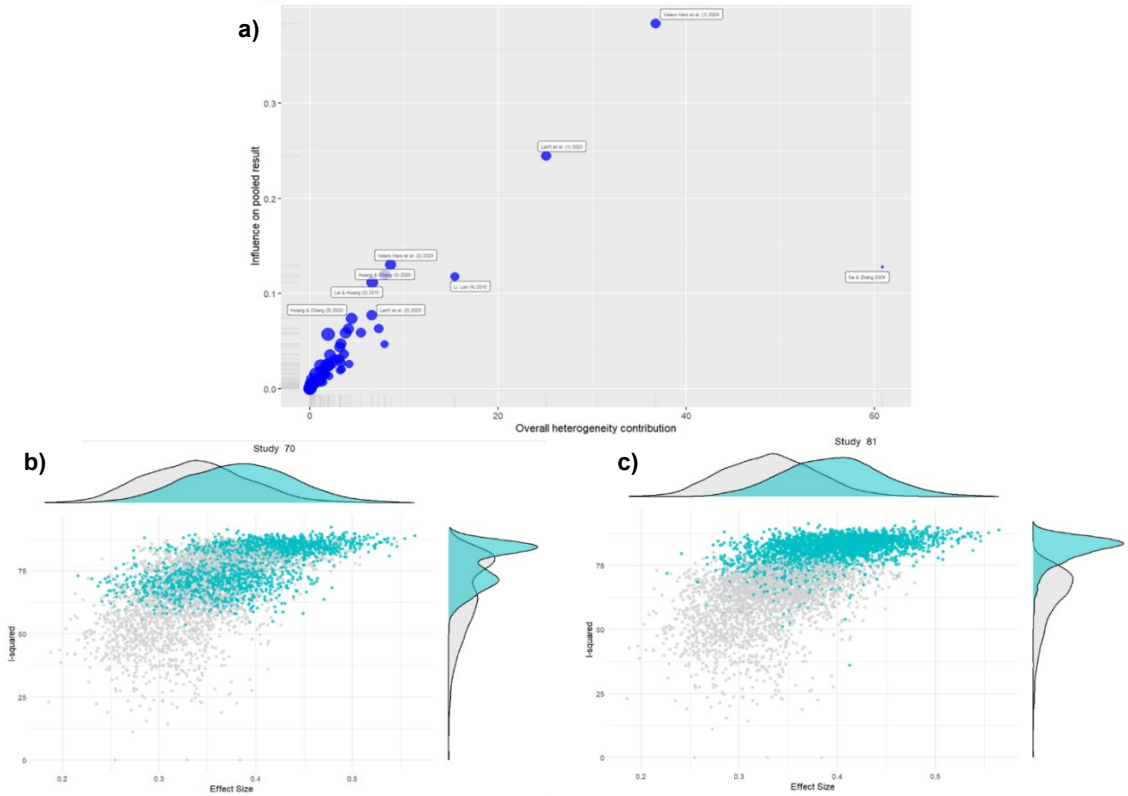

Figure 2. Outlier detection using (a) Baujat plot represents each study's impact on the overall heterogeneity and GOSH plot: (b) study data set 70 and (c) study data set 81 as potential outliers, evidenced by the asymmetry in the blue and gray areas.

A systematic trial was conducted by removing one outlier (study data set 81), two outliers (study data sets 70 and 81), and three outliers (study data sets 65, 70, and 81). Eliminating all three datasets reduced heterogeneity from 70.7% to 48.1%, but it also decreased the effect size from 0.3573 to 0.2938. This suggests that the excluded studies had substantial weight, leading to a lower effect size while improving heterogeneity reduction. To balance these factors, the analysis opted to remove two outliers (study data sets 70 and 81), which effectively lowered heterogeneity from 70.7% to 55.4% while maintaining a reasonable effect size reduction from 0.3573 to 0.3106. To maintain heterogeneity within a moderate threshold (below 60%), data sets 70 and 81 were excluded from the analysis. Additionally, the analysis was strengthened by conducting further publication bias tests.

Bias publication

Meta-analysis, while aggregating effect sizes across studies, may introduce bias. Publication bias was evaluated using a funnel plot, Rosenthal's classic fail-safe N test, and Egger's test (Megías et al., 2018). Adopting the most commonly used approaches, this study employed a funnel plot and Egger's test to assess potential bias. Figure 3 illustrates a nearly symmetrical funnel plot, indicating no apparent bias in the analysis. This result is supported by the results of Egger's test, which shows intercept values of 0.872, $t = 0.946$, and $p = 0.347$, and there is a statement "Egger's test does not indicate the presence of funnel plot asymmetry". Both methods confirmed that the 79 data sets were unbiased, and therefore, no data needed to be excluded from the analysis.

Following outlier detection and publication bias analysis, 79 data sets from 24 studies were retained for further analysis. The main analysis focused on examining the effects of intervention settings within technology-facilitated peer assessment. All statistical procedures were conducted using R.

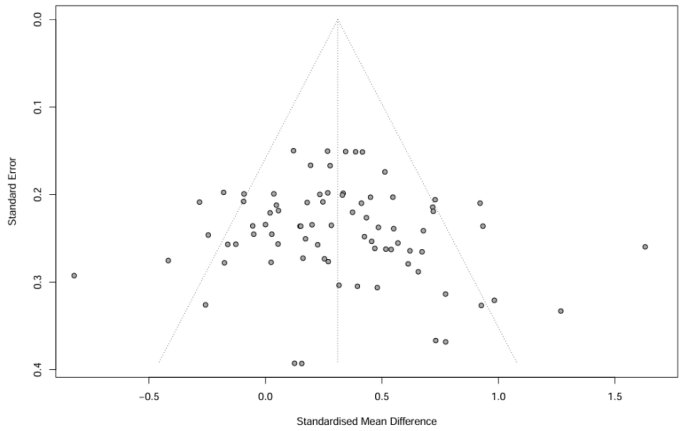

Figure 3. The funnel plot indicates no detectable publication bias

Overall Effect of Intervention in Peer Assessment

The analysis in this study uses a random effects model. The random effect model assumes that this variation is due to differences in true effect sizes from various studies (Borenstein et al., 2009). Therefore, this model involves calculating the estimation of $\tau 2$, which quantifies the variability in true effect sizes. By accounting for this variation, the combined effect can be calculated as the mean of the true effect size distribution.

In the analysis using {update.meta} function (by eliminating outliers), the variance of heterogeneity between studies was estimated at $\tau 2$= 0.0685 (95% CI [0.0418, 0.1373]), not containing the number 0.0, which indicates that there is some heterogeneity between studies. The I2 value is 55.4% (95% CI [42.5%, 65.4%]). Using the rule of thumb (Higgins & Thompson, 2002), this heterogeneity is classified as moderate. From the test of heterogeneity, the results of Q = 174.81; df = 78; and p-value < 0.0001. The Q value is considerably higher than the expected degrees of freedom, indicating significant heterogeneity ($p<0.001$). As highlighted by Y. Chang et al. (2022), when heterogeneity is detected, subgroup analyses, such as categorizing intervention, can be conducted to account for unexplained variability across study estimates.

The forest plot (Figure 4) illustrates the standardized mean differences (SMD), summarizing the number of studies, total participants, means, and standard deviations (SD) for both experimental and control groups. It also presents effect sizes, confidence intervals (CI), and each study's relative weight in the overall analysis. Figure 4 indicates considerable variation among the 79 data sets, with effect sizes ranging from a minimum of 0.2323 to a maximum of 0.3889. This effect size describes an index to measure the difference between the experimental and control groups in the peer assessment intervention. This effect size represents a standardized index used to compare the outcomes of experimental and control groups within peer assessment interventions. In this meta-analysis, several studies reported negative effect sizes (such as Gielen and De Wever (2012) (2), Gielen and De Wever (2015) (3), and L. Li (2016) (4)), indicating that the intervention did not produce greater learning performance in the experimental groups compared to the controls. In contrast, studies like Lai and Hwang (2015) (2), van den Bos and Tan (2019), Ocampo et al. (2023) (5), and Y. Lin et al. (2023) (2) demonstrated positive effect sizes, suggesting that the intervention was more effective than the control condition. Including both positive and negative findings provides a balanced view of the overall impact and strengthens the credibility of the overall analysis.

Overall, the results of data analysis obtained an effect size of 0.3106 with a p-value of <0.0001 (see Figure 4). Based on Kraft's (2020) recalibrated interpretation of effect sizes in education research, the overall effect size of 0.3106 found in this meta-analysis can be considered moderate and substantively important.

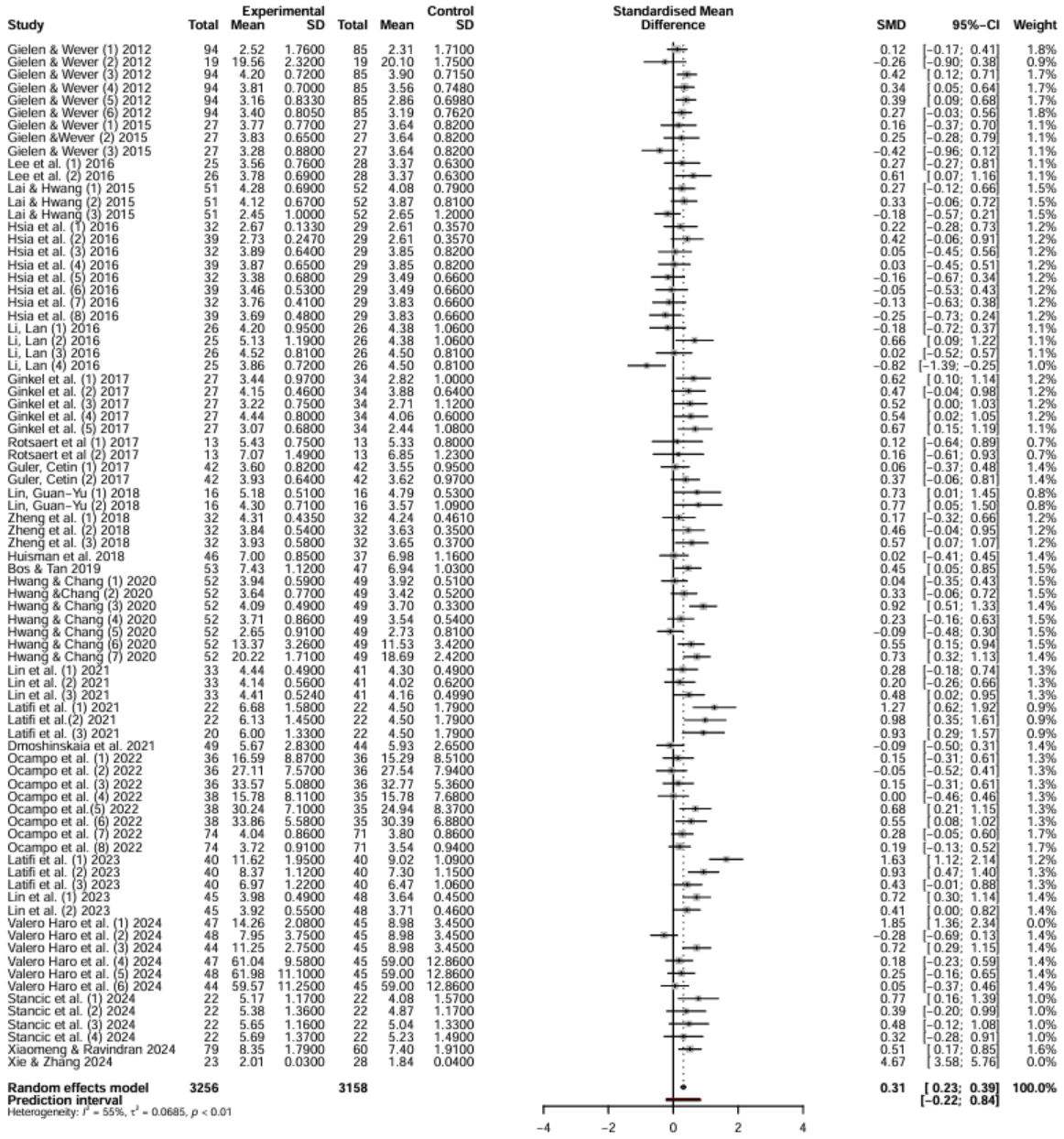

Figure 4. Overall forest plots summarizing effect sizes (0.23–0.39), confidence intervals, and study weights across included studies

A value of 0.3106 indicates that implementing stronger interventions in experimental groups yields more positive effects than alternative or minimal ones in control groups. The hypothesis that the experimental group outperformed the control group has been confirmed. These findings provide a foundation for future research on peer assessment interventions to strengthen empirical evidence.

Sub-Group Analysis of the Moderator Variable in Peer Assessment Intervention

The impact of PA interventions on learning performance is influenced by a moderating variable, referred to as a subgroup. Additionally, subgroup analysis is essential for identifying the sources of heterogeneity and evaluating whether statistical significance varies across different conditions. Subsequently, a Q test is conducted on the overall subgroup results to assess whether there are significant

differences among the groups (Harrer et al., 2022). The results of the sub-group analysis are presented in Table 3, and one of the forest plots of the moderator variable in PA Intervention in Figure 5.

Table 3. Results of the sub-groups analysis moderators

Moderator	k	Effect size (g)	95%-CI	I^2	p	X^2 sub-group	p-sub-group
PA intervention types:						2.20	0.3331
Feedback	20	0.2167	0.0647- 0.3687	55.2%	<0.01		
Process	29	0.3722	0.2330- 0.5113	66.4%	<0.01		
Participant	30	0.3100	0.1986- 0.4214	36.0%	0.03		
Reciprocal						0.49	0.4835
Yes	66	0.3197	0.2298- 0.4096	58.4%	<0.01		
No	13	0.2590	0.1148- 0.4031	32.9%	0.12		
Allocation						0.46	0.4982
Group	18	0.2803	0.1888- 0.3717	4.3%	0.40		
Individual	61	0.3272	0.2269- 0.4274	61.7%	<0.01		
Number of PFB Providers						5.45	0.1418
more than 3	33	0.2459	0.1531- 0.3387	31.2%	0.05		
3	16	0.2943	0.1822- 0.4065	0.0%	0.64		
2	5	0.0391	-0.4588 - 0.5369	77.4%	<0.01		
1	25	0.4616	0.2843- 0.6389	72.6%	<0.01		

Table 3 identifies the most effective intervention type in the peer assessment (PA) process, with the highest effect size observed for process-oriented interventions (g = 0.3722). Reciprocal PA proves more effective than non-reciprocal approaches. Feedback and scoring are also more impactful when provided individually rather than in groups. Interestingly, assigning a single peer assessor yields the highest effect size (g = 0.4616), likely due to reduced variability in feedback. In contrast, using two assessors may introduce conflicting perspectives, potentially causing ambiguity in interpretation.

Figure 5 illustrates the contribution of each study to the subgroup analysis of PA intervention types. All three subgroups demonstrate positive effects (as shown by diamonds positioned to the right of the vertical line). The highest effect size appears in the process-oriented subgroup (g = 0.37), with Latifi et al. (1) (2023) contributing most significantly.

Responses to the Research Questions

Based on the results of the study, the following sections address the research questions:

RQ1. Intervention settings in the experimental class were grouped into three types, namely participant, process, and feedback. Figure 5 shows the subgroup analysis for the PA intervention. A slight difference was observed among the subgroups (p-subgroup = 0.1821), while each subgroup showed statistically significant results ($p<0.05$). The impact of the process (g = 0.3722) was greater than that of feedback (g = 0.2167) and participant (g = 0.3100). Process-related interventions involved the use of specialized rubrics, interactive feedback or rubrics, free selection and invitation, or facilitated with scaffolding.

RQ2. The moderation effects across the reciprocal and allocation sub-groups were relatively similar. When students acted in both roles as assessor and assessee, the effect size (g = 0.3197, $p < 0.01$) was greater than when they participated in only one role (g = 0.2590, p = 0.12). It shows that reciprocity is more suggested in peer assessment.

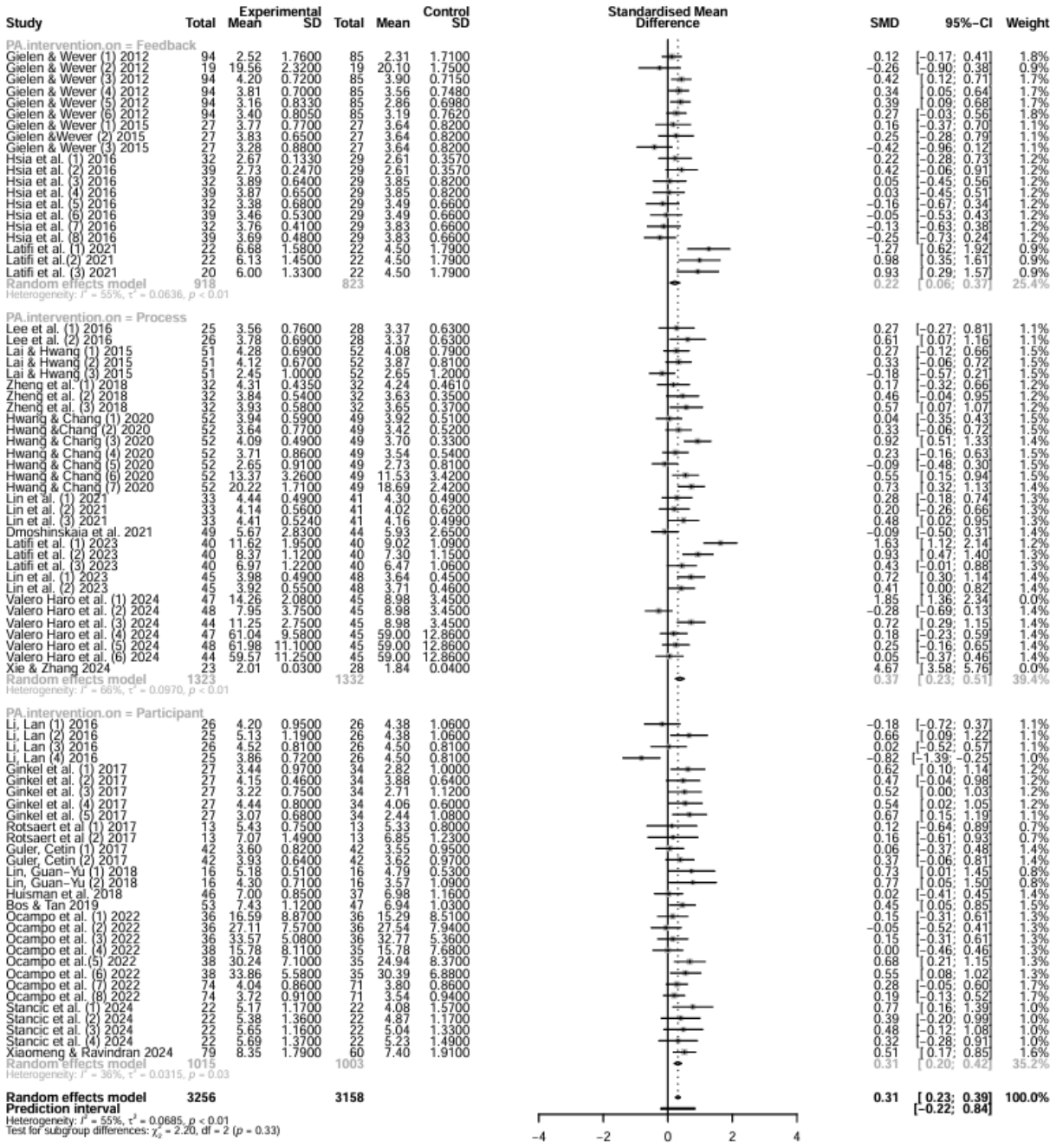

Figure 5. Forrest plot of the moderator variable in PA intervention types

RQ3. The effect size was slightly higher when assessment was allocated to individuals (g = 0.3272, p < 0.01) compared to groups (g = 0.2803, p = 0.40). Providing individually allocated feedback is recommended for more effective peer assessment.

RQ4. Intervention in the number of PFB providers was most effective when students received feedback from a single peer (g = 0.4616, p < 0.01). In contrast, having three or more peers resulted in nearly identical effect sizes (g = 0.2943 and g = 0.2459, respectively). The lowest effect was found when two peers were involved (g = 0.0391; 95% CI: -0.4588 - 0.5369; p < 0.01). As a result, recommended that teachers avoid using two assessors.

RQ5. Based on the findings, effective peer assessment settings prioritize process-oriented interventions, adopt a reciprocal structure, provide individually allocated feedback, and favor feedback received from a single peer. Certain adjustments should also be considered based on the characteristics of students, teachers, the school context, and the specific learning content.

DISCUSSION

INTERVENTION IN PEER ASSESSMENT IMPROVES LEARNING PERFORMANCE

In general, well-designed interventions have a more positive impact on learning performance compared to alternative or minimal interventions; in sub-group analysis for PA intervention, the highest lies in the process at 0.3722 (95% CI [0.2330, 0.5113]). Accordingly, teachers should place greater focus on students' processes during peer assessment, such as the importance of preparing clear rubrics (Dmoshinskaia et al., 2021; Lai & Hwang, 2015; Latifi et al., 2023; Valero Haro et al., 2024), the presence of interactive feedback (Hsia & Hwang, 2021; G. Y. Lin, 2018; H. C. Lin et al., 2021), scaffolding process during learning process (Lee et al., 2016), and free selection and invitation assessor according to student needs (Y. Lin et al., 2023).

In addition, intervention in feedback also gives positive and significant results. An attempt to modify the type of feedback, including feedback, feedforward, a combination of feedback and feedforward, and undirected feedback. According to Valero Haro et al. (2024), the four types of feedback differ significantly in the quality of feedback produced, especially in directed feedback and combination. Yu and Schunn (2023) also recommended designing interventions that enhance feedback provision and increase the length of peer feedback comments. Therefore, offering peer feedback, regardless of its structure or format, serves as an effective and impactful learning strategy that improves students' learning performance.

Participants in peer assessment studies have been categorized based on their roles and conditions, such as peer users vs. peer experts (Stančić et al., 2024), training vs. no training (Ocampo et al., 2023), and anonymous vs. identified settings (G. Y. Lin, 2018). These participant-related interventions significantly influenced peer assessment outcomes, with peer experts (those with prior experience), trained participants, and those in anonymous conditions demonstrating more positive effects. Similarly, L. Li (2016) found that both the training and anonymity groups outperformed the identity group, with the training group expressing greater appreciation for peer assessment and experiencing less pressure.

Reciprocal peer assessment, in which students serve as both assessors and assessees, has been found more effective than assuming a single role. This dual engagement encourages students to analyze and evaluate peer work, thereby improving their own task performance (Huisman et al., 2018; Latifi et al., 2021). Both roles offer distinct benefits, as providing constructive feedback promotes cognitive processing and self-reflection (Fu et al., 2019).

Feedback is most effective when given to students individually. Peer feedback comments or scoring-based feedback is more appropriate for assessing an entire group, each student (Hsia et al., 2016). These individualized comments help students better understand their strengths and areas for improvement, ultimately enhancing their performance. Providing individual feedback through online interaction has been shown to produce higher-quality peer reviews than conventional group-based peer assessment, particularly in terms of cognitive depth and reflective learning (H. C. Lin et al., 2021).

As previously discussed, Table 3 shows that the highest effect sizes for peer feedback (PFB) were observed when students received feedback from a single provider, often implemented in paired settings. Such arrangements enhance constructive dialogue, as meaningful argumentative discussions are more

likely to emerge in more intimate settings (Valero Haro et al., 2024). However, there is a potential risk of imbalance, where one student may dominate the process, limiting the cognitive benefits for the other (Latifi et al., 2023). To mitigate this, students should be grouped according to their characteristics to ensure balanced participation.

This discussion highlights intervention settings associated with the highest effect sizes. However, this does not imply that other interventions are ineffective. Specific interventions should be designed with careful consideration of the learning context and objectives. Emphasizing the importance of planning peer assessment appropriately ensures its alignment with the intended learning situation (Barahona et al., 2023). Rather, we aim to offer an overview of PA intervention settings tailored to specific needs and conditions, which is presented in the following discussion.

Importance of Technologies for PA

This study categorizes the technologies used to facilitate peer assessment into four types, with some articles employing more than one. Most are online or web-based (21 articles), although these cover a wide variety of different products, which might be considered separate categories. Videotaping, as many as four articles, is also widely used as a medium for students to give assessments. The next order is to use a wiki (2 studies), because they are readily available without cost and easily accessible, enabling peers to collaborate in groups by contributing, commenting, and further editing one another's work.

A study investigated the use of Facebook as a platform for peer assessment. The use of Facebook, which is a social media, has advantages and disadvantages. The advantage is that students feel comfortable and familiar with Facebook, and it is easier to create a group through Facebook groups (G. Y. Lin, 2016). On the other hand, Facebook, as an open platform, does not provide support for anonymous interactions (G. Y. Lin, 2016) and restrictions on using social media in schools (Topping, 2023). WhatsApp is a mobile messaging app that has also been implemented as a medium for providing feedback. Utilizing a common platform enables students to engage with each other's comments during the assessment process, fostering better interaction between assessors and assessees. However, anonymity cannot be maintained unless the assessor submits feedback privately to the teacher (Güler, 2017).

Currently, mobile phones are becoming a more widely developed technology to facilitate peer assessment, particularly in supporting online and web-based platforms. Their widespread availability among teachers and students makes them a common resource in schools. Additionally, students demonstrate a positive attitude when using mobile devices for assessments (Clase et al., 2010; H. Li et al., 2019). Technology-based approaches for formative assessment show students' willingness to use technology for peer feedback, self-reflection, and engagement (Hsia et al., 2016; Wylie & Lyon, 2019). Teachers effectively utilized these systems to identify areas where students required additional support in peer assessment processes. Integrating formative assessment practices within engaging tasks empowered teachers to recognize better ways of supporting students. The use of technology also helps teachers provide feedback to students individually and quickly.

Technology-facilitated PA is expected to become more important than analog PA to support universities in online learning (Topping, 2023). More than three-quarters of this study was conducted at universities, where students have the ability to engage in peer assessment (Mumpuni et al., 2022). However, peer assessment is also valuable in primary and secondary schools, as it positively impacts student learning. Providing guidance and training for teachers at these levels is essential to effectively implementing interventions in peer assessment. In summary, there is great potential for developing peer assessment platforms that enhance learning outcomes and support teachers in applying peer assessment practices more effectively (Ocampo & Panadero, 2023), such as the web-based e-assessment tool WebAVALIA (Babo et al., 2020).

In selecting technologies for PA, it is essential to align the selection with the student-teacher's ability to use technology and the available support within the educational level (van Helden et al., 2023). Higher digital literacy among both teachers and students, along with sufficient support, facilitates the integration of more advanced technologies. Furthermore, referencing prior designs can inform the selection of the right technologies by ensuring they align with the intended peer assessment format and meet the necessary criteria and functions.

PA INTERVENTION SETTING

After confirming through the meta-analysis that PA interventions positively impact learning performance, the next step is to explore effective PA settings. The studies included in this meta-analysis cover diverse subject areas, object assessed, education level, and intervention settings. Sub-group analysis indicates that all these factors contribute positively. The following is a recommendation (see Table 4) and review of key considerations in designing peer assessment interventions to optimize learning outcomes.

Table 4. Recommended peer assessment intervention settings

Type intervention	Primary intervention	Alternative intervention
PA intervention types in:	Process-oriented (rubrics, interactive feedback or rubrics, or facilitated with scaffolding)	Feedback-oriented (feedback, feedforward, a combination of feedback and feedforward, and undirected feedback, structured feedback)
		Participant-oriented (peer users vs. peer experts, training vs. no training, and anonymous vs. identified settings)
Reciprocal	Yes (Students serve as both assessor and assesse)	No (students assigned only one role, assessor or assessee)
Allocation	Individual (Individual-directed feedback, for personalized tasks)	Group (feedback for group, for collaborative project work)
Number of PFB Providers	1 (using single-peer feedback, especially in pairs)	3 or more than 3 (use one or more than three; avoid assigning exactly two providers)

Process-oriented interventions are generally considered a well-supported foundation for effective peer assessment, involving the use of clear rubrics, interactive feedback, and scaffolding. Hwang and Chang (2021) found that interactive feedback enhanced students' critical thinking. Combining process-oriented with participant- and feedback-oriented may further improve learning outcomes. For instance, alongside clear rubrics, teachers can introduce structured feedback forms, provide assessor training, and apply anonymity to minimize bias.

The anonymity of the assessor should be considered. Feedback, both quantitative and qualitative, has been shown to improve the quality of peer feedback, perceived learning, perceived fairness, and attitudes toward the system (G. Y. Lin, 2018). Furthermore, the anonymity feature in peer feedback allows for more authentic and constructive evaluations (Hsia & Hwang, 2021). However, some studies compare their effectiveness, potentially spanning over time, as anonymity might be more effective in the short term while identification may prove better in the long term (Topping, 2023). Therefore, the decision of an anonymous or identified assessor is based on the needs of the teacher.

Peer assessment interventions can be implemented in groups or pairs. To maximize the cognitive and reflective benefits of reciprocal feedback, it is recommended that students alternate roles as both feedback providers and recipients. When forming pairs or groups, educators should avoid random assignment and instead consider student characteristics, such as ability levels and communication styles. One recommended approach for pairing assessors and assessees is the use of a Balanced Allocation algorithm (Alkhalifa et al., 2022). Thoughtful pairing or grouping strategies can help prevent imbalances in feedback quality – where some students may receive rich, constructive input while others receive minimal feedback – and promote a more equitable and effective peer assessment process.

Individual-directed feedback and using a single PFB provider are also strongly recommended in peer assessment. Regardless of the feedback design, delivering feedback individually tends to yield positive effects (Gielen & De Wever, 2015), as it allows each student to reflect and make meaningful improvements to their work. However, many studies in this meta-analysis involved more than three assessors per assessee; the most positive effects were observed when a single peer provided feedback. When PA involves multiple feedback providers who identify the same issue in a student's performance, it highlights a critical area for improvement – recommended for the teacher to involve an odd assessor. This may be due to the difficulty the assessee faces in determining the appropriate follow-up when two assessors provide opposite feedback or scores.

The assessed object influences the selection of appropriate interventions. Writing tasks are more frequently selected for peer assessment than product-based or performance-based assignments. This may be due to the challenges in assessing performance, such as difficulty reviewing students' actions when assessors are uncertain about the grades assigned. Peer assessment is more commonly applied to writing because it is more practical and easier to implement. These findings underscore the need for more innovative intervention strategies to improve the effectiveness of performance assessment, including the use of well-defined rubrics (Robles et al., 2024; Thite et al., 2024), structured observation, and assessor training.

The effective implementation of PA should be supported by appropriate technology. As previously discussed, teachers select technological tools that align with students' characteristics, their own digital competence, the availability of resources, and the nature of the subject matter. Technology serves not only to facilitate the PA process but also to enhance accessibility, interaction, and the quality of feedback.

While peer assessment has been shown to improve student performance and foster critical thinking and reflection (van Helden et al., 2023), its success depends on the teacher's ability to adapt interventions to specific classroom contexts. When primary interventions are not feasible, alternative strategies may be more appropriate. Given the diversity in subject content, student profiles, school environments, and technological access, it is essential for teachers to tailor peer assessment interventions to ensure meaningful and effective implementation. The ultimate goal is to ensure that peer assessment is not only conducted but also delivers optimal benefits for student learning.

IMPLICATIONS FOR RESEARCH AND EDUCATION

In conclusion, the first step in implementing peer assessment is to determine the specific aspects that will be evaluated (van Helden et al., 2023). Teachers need to decide whether the assessment will focus on written work, products, or performance, ensuring alignment with the subject matter. Following this, it is essential to create the parameters of the PA intervention, including feedback types, assessment processes, and participant roles. Additionally, training and guidance for assessors and assessees are crucial in enhancing feedback quality. Other factors to consider include whether assessments will be conducted anonymously or with identified assessors, the duration of the assessment period, student characteristics, and the potential for friendship bias.

The validity and reliability of peer assessment have been widely supported by research. Studies have demonstrated a strong correlation between feedback provided by students and that of experts or teachers (Hovardas et al., 2014; Vander Schee & Birrittella, 2021), including in elementary education settings (Hung, 2018). Teachers play a critical role in offering professional input to enhance the meaningfulness of feedback (Prilop et al., 2021).

Although few negative effects on learning outcomes have been reported (van Helden et al., 2023), potential social challenges – such as jealousy or retaliation among students – should be anticipated through thoughtful forming of a peer group. Additionally, the selected technology should be practical and not impose excessive burdens on teachers or students. As noted by Topping (1998), peer assessment is often perceived by students as a compulsory task rather than a meaningful learning activity. However, when students understand its benefits and engage in the process voluntarily, its effectiveness and overall impact are expected to improve.

LIMITATIONS

This study has two main limitations. First, a relatively small number of studies met the strict selection criteria. Although research on technology-facilitated peer assessment with interventions is relatively abundant, only a limited number reported sufficient statistical information – such as sample sizes, means, and standard deviations for both control and experimental groups – necessary for meta-analysis. Limitations of the study may stem from the search strategy or the scope of databases used, which may have excluded relevant literature not captured through the selected sources. The small number of eligible studies may have constrained the strength of the conclusions. Second, the interpretation of the effect size presents a challenge. Although the overall effect size ($g = 0.3106$) meets the threshold for substantive importance in education research proposed by Kraft (2020), it may still be perceived as small by readers more familiar with Cohen's benchmarks. This discrepancy underscores the ongoing challenge regarding appropriate standards for interpreting effect sizes in educational research.

CONCLUSIONS

Our study achieved its primary objective of identifying which TFPA interventions and implementation strategies most significantly influence students' learning performance. By examining the moderating variables, the findings offer evidence-based guidance for designing more effective and educationally appropriate PA practices. A meta-analysis of 24 studies (79 data sets), using contrast groups without publication bias and excluding extreme outliers, revealed that intervention settings in PA enhanced learning performance. The findings suggest that effective PA settings emphasize process-oriented interventions, incorporate reciprocal roles, allocate feedback to individuals, and prefer feedback from a single peer. However, these settings should adopt an adaptive approach to align with the specific characteristics of learners, instructors, school environments, and instructional content.

This study also demonstrates the methodological value of meta-analysis in educational research, allowing researchers to identify patterns, estimate effect sizes, and examine the influence of moderating variables. The findings confirm the effectiveness of TFPA in enhancing learning performance while offering valuable insights for refining theoretical models of assessment. The analysis highlights key intervention areas and demonstrates that TFPA improves students' understanding of rubrics, feedback skills, self-reflection, and learning outcomes. It also supports teachers in monitoring progress and delivering individualized feedback more effectively.

TFPA supports the development of essential digital skills among students – such as communication, collaboration, and critical reflection – while assisting educators in implementing effective, technology-enhanced assessment strategies. These outcomes align with key dimensions of global digital education frameworks, particularly the Digital Competence Framework (DigComp) and Digital Competence Framework for Educators (DigCompEdu), which emphasize both learners' digital competence and educators' pedagogical integration of digital tools. In addition to its pedagogical contributions,

this study offers policy-relevant recommendations for strengthening digital learning environments. The findings can inform curriculum development, teacher professional development, and institutional decision-making, supporting the strategic integration of adaptive, student-centered peer assessment models into broader digital education.

Future research should focus on developing PA platforms to meet the needs of educational contexts. Beyond refining the intervention setting, further studies should explore the integration of peer and teacher feedback for more robust empirical findings. Expanding research across diverse subjects and educational levels is also essential.

ACKNOWLEDGMENTS

This article is part of an ongoing doctoral study at the Graduate School of Educational Research and Evaluation (PEP), which the first author (K.E.M.) is currently pursuing at Universitas Negeri Yogyakarta (UNY), Indonesia. We would like to express our deepest gratitude to The Ministry of Higher Education, Science, and Technology of Republic Indonesia (KEMENDIKTISAINTEK), The Center of Higher Education Funding and Assessment (PPAPT), and The Indonesia Endowment Funds for Education (LPDP) of The Republic of Indonesia for providing the Indonesia Educational Scholarships (BPI), enabling the first author (K.E.M.) to conduct this study.

REFERENCES

Alkhalifa, A., Devlin, M., & Alkhattabi, M. (2022). Matching authors and reviewers in peer assessment based on authors' profiles. *Journal of Information Technology Education: Innovations in Practice*, *21*, 135–153. https://doi.org/10.28945/5038

Babo, R., Suhonen, J., & Tukiainen, M. (2020). Improving workgroup assessment with WebAVALIA: The concept, framework and first results. *Journal of Information Technology Education: Innovations in Practice*, *19*, 157–184. https://doi.org/10.28945/4627

Barahona, C., Nussbaum, M., Martin, V., Meneses, A., Arriagada, S., Di Serio, A., & Hilliger, I. (2023). Technology-scaffolded peer assessment for developing critical thinking in pre-service teacher training: The importance of giving feedback. *Educational Technology Research and Development*, *71*(2), 667–688. https://doi.org/10.1007/s11423-022-10173-1

Baujat, B., Mahé, C., Pignon, J.-P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine*, *21*(18), 2641–2652. https://doi.org/https://doi.org/10.1002/sim.1221

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

Broadbent, J., Panadero, E., & Bouda, D. (2018). Implementing summative assessment with a formative flavour: A case study in a large class. *Assessment and Evaluation in Higher Education*, *43*(2), 307–322. https://doi.org/10.1080/02602938.2017.1343455

Chang, C.-Y., Lee, D.-C., Tang, K.-Y., & Hwang, G.-J. (2021). Effect sizes and research directions of peer assessments: From an integrated perspective of meta-analysis and co-citation network. *Computers and Education*, *164*, 104123. https://doi.org/10.1016/j.compedu.2020.104123

Chang, Y., Phillips, M. R., Guymer, R. H., Thabane, L., Bhandari, M., Chaudhary, V., Wykoff, C. C., Sivaprasad, S., Kaiser, P., Sarraf, D., Bakri, S., Garg, S. J., Singh, R. P., Holz, F. G., & Wong, T. Y. (2022). The 5 min meta-analysis: Understanding how to read and interpret a forest plot. *Eye*, *36*(4), 673–675. https://doi.org/10.1038/s41433-021-01867-6

Chen, T. (2016). Technology-supported peer feedback in ESL/EFL writing classes: A research synthesis. *Computer Assisted Language Learning*, *29*(2), 365–397. https://doi.org/10.1080/09588221.2014.960942

Clase, K. L., Gundlach, E., & Pelaez, N. J. (2010). Calibrated peer review for computer-assisted learning of biological research competencies. *Biochemistry and Molecular Biology Education*, *38*(5), 290–295. https://doi.org/10.1002/bmb.20415

Costa, C., Cardoso, A. P., Lima, M. P., Ferreira, M., & Abrantes, J. L. (2015). Pedagogical interaction and learning performance as determinants of academic achievement. *Procedia – Social and Behavioral Sciences*, *171*, 874–881. https://doi.org/10.1016/j.sbspro.2015.01.203

Dmoshinskaia, N., Gijlers, H., & de Jong, T. (2021). Giving feedback on peers' concept maps in an inquiry learning context: The effect of providing assessment criteria. *Journal of Science Education and Technology*, *30*(3), 420–430. https://doi.org/10.1007/s10956-020-09884-y

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, *32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

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. https://doi.org/10.1007/s40692-019-00131-x

Ghahari, S., & Sedaghat, M. (2018). Optimal feedback structure and interactional pattern in formative peer practices: Students' beliefs. *System*, *74*, 9–20. https://doi.org/10.1016/j.system.2018.02.003

Gielen, M., & De Wever, B. (2012). Peer assessment in a Wiki: Product improvement, students' learning and perception regarding peer feedback. *Procedia - Social and Behavioral Sciences*, *69*, 585–594. https://doi.org/10.1016/j.sbspro.2012.11.450

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 and Education*, *88*, 370–386. https://doi.org/10.1016/j.compedu.2015.07.012

Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2-16. https://doi.org/10.3102/01623737001001002

Güler, Ç. (2017). Use of WhatsApp in higher education: What's up with assessing peers anonymously? *Journal of Educational Computing Research*, *55*(2), 272–289. https://doi.org/10.1177/0735633116667359

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2022). *Doing meta-analysis with R: A hands-on guide*. CRC Press. https://www.routledge.com/Doing-Meta-Analysis-with-R-A-Hands-On-Guide/Harrer-Cuijpers-Furukawa-Ebert/p/book/9780367610074

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/https://doi.org/10.1002/sim.1186

Hoo, H. T., Deneen, C., & Boud, D. (2022). Developing student feedback literacy through self and peer assessment interventions. *Assessment and Evaluation in Higher Education*, *47*(3), 444–457. https://doi.org/10.1080/02602938.2021.1925871

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers and Education*, *71*, 133–152. https://doi.org/10.1016/j.compedu.2013.09.019

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 and Education*, *96*, 55–71. https://doi.org/10.1016/j.compedu.2016.02.004

Hsia, L. H., & Hwang, G. J. (2021). Enhancing students' choreography and reflection in university dance courses: A mobile technology-assisted peer assessment approach. *British Journal of Educational Technology*, *52*(1), 266–287. https://doi.org/10.1111/bjet.12986

Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2018). Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment and Evaluation in Higher Education*, *43*(6), 955–968. https://doi.org/10.1080/02602938.2018.1424318

Hung, Y. (2018). Group peer assessment of oral English performance in a Taiwanese elementary school. *Studies in Educational Evaluation*, *59*, 19–28. https://doi.org/10.1016/j.stueduc.2018.02.001

Hunter, J. E., & Schmidt, F. L. (2006). Methods of meta-analysis: Correcting error and bias in research findings (2nd ed.). Sage.

Hwang, G. J., & Chang, S. C. (2021). Facilitating knowledge construction in mobile learning contexts: A bi-directional peer-assessment approach. *British Journal of Educational Technology*, *52*(1), 337–357. https://doi.org/10.1111/bjet.13001

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Lai, C. L., & Hwang, G. J. (2015). An interactive peer-assessment criteria development approach to improving students' art design performance using handheld devices. *Computers and Education*, *85*, 149–159. https://doi.org/10.1016/j.compedu.2015.02.011

Latifi, S., Noroozi, O., & Talaee, E. (2021). Peer feedback or peer feedforward? Enhancing students' argumentative peer learning processes and outcomes. *British Journal of Educational Technology*, *52*(2), 768–784. https://doi.org/10.1111/bjet.13054

Latifi, S., Noroozi, O., & Talaee, E. (2023). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments*, *31*(2), 655–669. https://doi.org/10.1080/10494820.2020.1799032

Lee, C. I., Yang, Y. F., & Mai, S. Y. (2016). The impact of a scaffolded assessment intervention on students' academic achievement in web-based peer assessment activities. *International Journal of Distance Education Technologies*, *14*(4), 41–54. https://doi.org/10.4018/IJDET.2016100104

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2019). Does peer assessment promote student learning? A meta-analysis. *Assessment and Evaluation in Higher Education*, *45*(2), 193–211. https://doi.org/10.1080/02602938.2019.1620679

Li, L. (2016). The role of anonymity in peer assessment. *Assessment and Evaluation in Higher Education*, *42*(4), 645–656. https://doi.org/10.1080/02602938.2016.1174766

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 and Technology Education*, *12*(9), 2295–2307. https://doi.org/10.12973/eurasia.2016.1280a

Lin, G. Y. (2018). Anonymous versus identified peer assessment via a Facebook-based learning application: Effects on quality of peer feedback, perceived learning, perceived fairness, and attitude toward the system. *Computers and Education*, *116*, 81–92. https://doi.org/10.1016/j.compedu.2017.08.010

Lin, H. C., Hwang, G. J., Chang, S. C., & Hsu, Y. D. (2021). Facilitating critical thinking in decision making-based professional training: An online interactive peer-review approach in a flipped learning context. *Computers and Education*, *173*, 104266. https://doi.org/10.1016/j.compedu.2021.104266

Lin, Y., Zhang, Y., Yang, Y., Lu, Y., Zhou, P., & Wang, Y. (2023). "Free selection and invitation" online peer assessment of undergraduates' research competencies, flow, motivation and interaction in a research methods course. *Journal of Computing in Higher Education*, *36*, 727–772. https://doi.org/10.1007/s12528-023-09374-1

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

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 and Biobehavioral Reviews*, *84*, 198–203. https://doi.org/10.1016/j.neubiorev.2017.12.003

Moccozet, L. (2012, July). Introducing learning performance in personal learning environments. *Proceedings of the IEEE 12th International Conference on Advanced Learning Technologies*, *Rome, Italy,* 702–703. https://doi.org/10.1109/ICALT.2012.75

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology, 62*(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005

Mulet, J., van de Leemput, C., & Amadieu, F. (2019). A critical literature review of perceptions of tablets for learning in primary and secondary schools. *Educational Psychology Review*, *31*, 631–662. https://doi.org/10.1007/s10648-019-09478-0

Mumpuni, K. E., Priyayi, D. F., & Widoretno, S. (2022). How do students perform a peer assessment? *International Journal of Instruction*, *15*(3), 751–766. https://doi.org/10.29333/iji.2022.15341a

Ocampo, J. C. G., & Panadero, E. (2023). Web-based peer assessment platforms: What educational features influence learning, feedback and social interaction? In O. Noroozi & B. De Wever (Eds.), The power of peer learning: Fostering students' learning processes and outcomes (pp. 165–182). Springer. https://doi.org/10.1007/978-3-031-29411-2_8

Ocampo, J. C. G., Panadero, E., & Díez, F. (2023). Are men and women really different? The effects of gender and training on peer scoring and perceptions of peer assessment. *Assessment and Evaluation in Higher Education*, *48*(6), 760–776. https://doi.org/10.1080/02602938.2022.2130167

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. The BMJ, *372*(71). https://doi.org/10.1136/bmj.n71

Posit Team. (2023). *RStudio: Integrated development environment for R*. Posit Software, PBC.

Prilop, C. N., Weber, K. E., & Kleinknecht, M. (2021). The role of expert feedback in the development of pre-service teachers' professional vision of classroom management in an online blended learning environment. *Teaching and Teacher Education*, *99*, 103276. https://doi.org/10.1016/j.tate.2020.103276

Robles, V. C., Ruíz-Rey, F. J., Cebrián-de-la-Serna, M., & Lourenço-Martins, F. M. (2024). Formative evaluation with technologies among teachers for the development of educational innovation projects. *International Journal of Educational Research and Innovation*, *22*, 1–9. https://doi.org/https://doi.org/10.46661/ijeri.7861

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Peer assessment use, its social nature challenges and perceived educational value: A teachers' survey study. *Studies in Educational Evaluation*, *59*, 124–132. https://doi.org/10.1016/j.stueduc.2018.07.001

Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2018). "Now you know what you're doing right and wrong!" Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education*, *33*(2), 255–275. https://doi.org/10.1007/s10212-017-0329-x

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Springer. https://doi.org/10.1007/978-3-319-21416-0

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Bastiaens, T. J. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation*, *29*(1), 23–42. https://doi.org/10.1016/S0191-491X(03)90003-4

Stančić, M., Milin, V., & Gašić, A. (2024). Peer feedback in higher education: How students' roles affect providing and receiving feedback. *Studies in Educational Evaluation*, *83*, 101408. https://doi.org/10.1016/j.stueduc.2024.101408

Thite, S., Ravishankar, J., Tomeo-Reyes, I., & Ortiz, A. M. (2024). Design of a simple rubric to peer-evaluate the teamwork skills of engineering students. *European Journal of Engineering Education*, *49*(4), 623–646. https://doi.org/10.1080/03043797.2024.2338239

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Topping, K. J. (2009). Peer assessment. *Theory into Practice*, *48*(1), 20–27. https://doi.org/10.1080/00405840802577569

Topping, K. J. (2023). Digital peer assessment in school teacher education and development: a systematic review. *Research Papers in Education*, *38*(3), 472–498. https://doi.org/10.1080/02671522.2021.1961301

Topping, K., & Ehly, S. (Eds.). (1998). *Peer-assisted learning*. Routledge. https://doi.org/10.4324/9781410603678

Valero Haro, A., Noroozi, O., Biemans, H. J. A., Mulder, M., & Banihashem, S. K. (2024). How does the type of online peer feedback influence feedback quality, argumentative essay writing quality, and domain-specific learning? *Interactive Learning Environments*, *32*(9), 5459–5478. https://doi.org/10.1080/10494820.2023.2215822

van den Bos, A. H., & Tan, E. (2019). Effects of anonymity on online peer review in second-language writing. *Computers and Education*, *142*, 103638. https://doi.org/10.1016/j.compedu.2019.103638

Vander Schee, B. A., & Birrittella, T. D. (2021). Hybrid and online peer group grading: Adding assessment efficiency while maintaining perceived fairness. *Marketing Education Review*, *31*(4), 275–283. https://doi.org/10.1080/10528008.2021.1887746

van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2017). Fostering oral presentation performance: does the quality of feedback differ when provided by the teacher, peers or peers guided by tutor? *Assessment and Evaluation in Higher Education*, *42*(6), 953–966. https://doi.org/10.1080/02602938.2016.1212984

van Helden, G., van der Werf, V., Saunders-Smits, G. N., & Specht, M. M. (2023). The use of digital peer assessment in higher education – An umbrella review of literature. *IEEE Access*, *11*, 22948–22960. https://doi.org/10.1109/ACCESS.2023.3252914

Viechtbauer, W. (2010). Conducting meta-analyses with metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wang, C., OuYang, J., & Wu, F. (2021). Subgroups of assessor and assessee: The relationship between students' peer assessment roles and perceptions of MSCL in science education. *Journal of Science Education and Technology*, *30*(6), 816–828. https://doi.org/10.1007/s10956-021-09922-3

Wylie, E. C., & Lyon, C. J. (2019). The role of technology-enhanced self- and peer assessment in formative assessment. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 170–191). Routledge. https://doi.org/10.4324/9780429507533-10

Xiaomeng, Z., & Ravindran, L. (2024). The effects of peer feedback provision and reception on lower-proficiency EFL learners' writing development. *Studies in Educational Evaluation*, *83*, 101422. https://doi.org/https://doi.org/10.1016/j.stueduc.2024.101422

Xie, Q., & Zhang, C. (2024). Online peer feedback via Moodle forum: Implications for longitudinal feedback design and feedback quality. *Computers & Education*, *223*, 105167. https://doi.org/10.1016/j.compedu.2024.105167

Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, *37*, 100484. https://doi.org/10.1016/j.edurev.2022.100484

Yin, B., & Yuan, C. H. (2021). Precision teaching and learning performance in a blended learning environment. *Frontiers in Psychology*, *12*, 631125. https://doi.org/10.3389/fpsyg.2021.631125

Yu, Q., & Schunn, C. D. (2023). Understanding the what and when of peer feedback benefits for performance and transfer. *Computers in Human Behavior*, 147, 107857. https://doi.org/10.1016/j.chb.2023.107857

Zhan, Y., Yan, Z., Wan, Z. H., Wang, X., Zeng, Y., Yang, M., & Yang, L. (2023). Effects of online peer assessment on higher-order thinking: A meta-analysis. *British Journal of Educational Technology*, *54*(4), 817–835. https://doi.org/10.1111/bjet.13310

Zheng, L., Cui, P., Li, X., & Huang, R. (2018). Synchronous discussion between assessors and assessees in web-based peer assessment: Impact on writing performance, feedback quality, meta-cognitive awareness and self-efficacy. *Assessment and Evaluation in Higher Education*, *43*(3), 500–514. https://doi.org/10.1080/02602938.2017.1370533

AUTHORS

Kistantia Elok Mumpuni is a lecturer in Biology Education at Universitas Sebelas Maret (UNS), Indonesia. She is currently pursuing her doctoral studies in Educational Research and Evaluation at Universitas Negeri Yogyakarta (UNY) and holds a master's degree in education from Universitas Negeri Malang. Her research focuses on teaching strategies, educational evaluation, biology education, and environmental science learning (Envi-sci), with particular attention to performance assessment, including peer assessment and teaching performance. She has published in national and international journals in these areas and actively serves as a reviewer for scholarly publications aligned with her expertise.

Samsul Hadi is a Professor of Vocational Education Assessment at Universitas Negeri Yogyakarta (UNY), Indonesia. He holds a doctoral degree in Research and Evaluation Education from UNY. His expertise spans statistical data analysis, Item Response Theory, and Computerized Adaptive Testing, specifically applied to vocational and technical education. Professor Hadi's research centers on developing valid, reliable assessment models for vocational programs, especially those integrating local-content curricula and adaptive testing technologies. He has published extensively, and his work at international conferences highlights his ongoing commitment to enhancing vocational learning quality in the era of Industry 4.0.

Slamet Suyanto is a Professor of Biology Education at Universitas Negeri Yogyakarta (UNY), Indonesia. He earned his doctoral degree in Educational Technology from Universitas Negeri Jakarta and his master's in education from the University of Houston, USA. His expertise includes curriculum design, STEM-based learning, formative assessment, and the integration of local wisdom into sustainable science education. He has contributed to the field through authorship of educational resources and scholarly publications. He is passionate about ecopedagogy and 21st-century learning and actively engages in research, academic mentoring, and scientific publication.