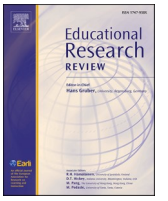
Educational Research Review 53 (2026) 100848

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev



Review

Peer feedback quality matters for post-feedback outcomes: A meta-analysis

Ecenaz Alemdag [a,b,*], Anastasiya Lipnevich [c,d], Lisa Bardach [e,f], Marlit Annalena Lindner [a,b]

[a] EUF - Chair of Digital Education, Europa-Universität Flensburg, Germany
[b] IPN - Leibniz Institute for Science and Mathematics Education, Kiel, Germany
[c] National Board of Medical Examiners, USA
[d] Queens College and the Graduate Center, City University of New York, USA
[e] Department of Psychology, University of Giessen, Germany
[f] Kaunas University of Technology, Lithuania

ARTICLE INFO

Keywords:
Peer feedback
Peer assessment
Feedback quality
Feedback uptake
Feedback receptivity

ABSTRACT

Peer feedback quality is key to enhancing learning through peer assessment. However, varied operationalizations of feedback quality across studies have led to inconsistencies in the literature. There are also contradictory findings regarding the relationship between feedback quality and students' use of feedback in current and future performance tasks. In this meta-analysis, we described peer feedback quality across five dimensions (i.e., amount, content, accuracy, features, and perceptions) and its relationship with other related feedback elements, drawing on recent conceptual frameworks and an integrative model of peer feedback. We then aggregated studies that examined the correlation between multifaceted peer feedback quality and post-feedback outcomes (i.e., revision, performance, learning, and course achievement) and analyzed potential moderating effects of several variables. Following a systematic literature search, 49 eligible papers with 53 independent samples and 241 effect sizes were identified. Using the robust variance estimation method, the synthesis of these effect sizes revealed a significant, positive correlation between peer feedback quality and post-feedback outcomes ($r = .20$). Moderator analyses showed that the strength of the correlations varied across post-feedback outcomes. Furthermore, post-feedback outcomes were more strongly associated with the quality of feedback students provided, rather than the quality of the feedback they received. Additionally, giving training prior to peer assessment strengthened the correlation between peer feedback quality and post-feedback outcomes. Overall, this meta-analysis corroborates the relationship between peer feedback quality and students' use of feedback and highlights the importance of operationalizations, training, and peer feedback roles in this connection.

1. Introduction

Peer assessment for formative purposes is an assessment method in which students give and receive feedback on tasks to improve their own and their peers' learning processes and final works. Meta-analyses on peer assessment have consistently demonstrated its

* Corresponding author. Europa-Universität Flensburg, Auf dem Campus 1a, 24943 Flensburg, Germany.
E-mail address: ecenazalemdag@gmail.com (E. Alemdag).

https://doi.org/10.1016/j.edurev.2026.100848
Received 15 August 2025; Received in revised form 3 September 2026; Accepted 5 September 2026
Available online 6 September 2026
1747-938X/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

positive effect on academic performance (Double et al., 2020; Li et al., 2020; Yan et al., 2022; Zheng et al., 2020) and further academic outcomes such as learning strategies and academic perseverance (Li et al., 2021). The contributions of peer assessment to the development of higher-order thinking skills, including critical thinking, reasoning, metacognitive thinking, argumentation, and reflection, were also revealed (Zhan et al., 2023). However, despite these benefits, not all peer assessment activities lead to meaningful student actions that improve learning and performance.

One key factor causing unproductive peer assessment practices is the low quality of feedback. Assessment and feedback models (e.g., Narciss, 2008; Nicol & Macfarlane-Dick, 2006; Panadero & Lipnevich, 2022) have identified feedback quality as a crucial condition for supporting self-regulation and learning. However, empirical research on peer assessment has produced mixed findings regarding the role of peer feedback quality. Some studies have revealed significant positive correlations between peer feedback quality, performance, and learning (e.g., Jin et al., 2024; Liu & Li, 2014; Wu & Schunn, 2021b; Yu & Wu, 2016), whereas others have reported non-significant or even negative correlations (e.g., Huisman et al., 2017; Ludemann & McKain, 2014; Novakovich & Long, 2013). In addition, prior meta-analyses on peer assessment (e.g., Double et al., 2020; Li et al., 2020, 2021; Zhan et al., 2023; Zheng et al., 2020) have generally overlooked the role of feedback quality in their moderator analyses. This gap may be due to the various approaches used to define and measure peer feedback quality, resulting in inconsistencies across study frameworks and findings.

In this study, we define peer feedback quality as a multidimensional construct, drawing on recent conceptual frameworks (e.g., Zhang et al., 2024a; Zhang & Schunn, 2023). We focus on five key dimensions of peer feedback quality: (1) amount, (2) content, (3) accuracy, (4) features, and (5) perceptions. An example of high-quality feedback based on these dimensions would be feedback that offers a substantial number of comments, accurately identifies high-level problems, provides justifications and suggestions, and is perceived as helpful and agreeable by the feedback receiver.

Furthermore, in this study, we assume that high-quality peer feedback increases the likelihood that students take constructive actions after providing and receiving feedback by revising their initial performance, improving their learning, or demonstrating greater engagement in future tasks. This assumption aligns with the ultimate aims of self-regulated learning (Butler & Winne, 1995) and feedback processes (Panadero & Lipnevich, 2022). Yet no meta-analysis has systematically examined whether and how peer feedback quality relates to post-feedback outcomes.

Analyzing the relationship between peer feedback quality and student outcomes is essential for establishing validity evidence for the peer feedback quality dimensions proposed in recent conceptual frameworks, which are still undergoing empirical validation. Therefore, in this meta-analysis, we aimed to examine the correlations between peer feedback quality and post-feedback outcomes.

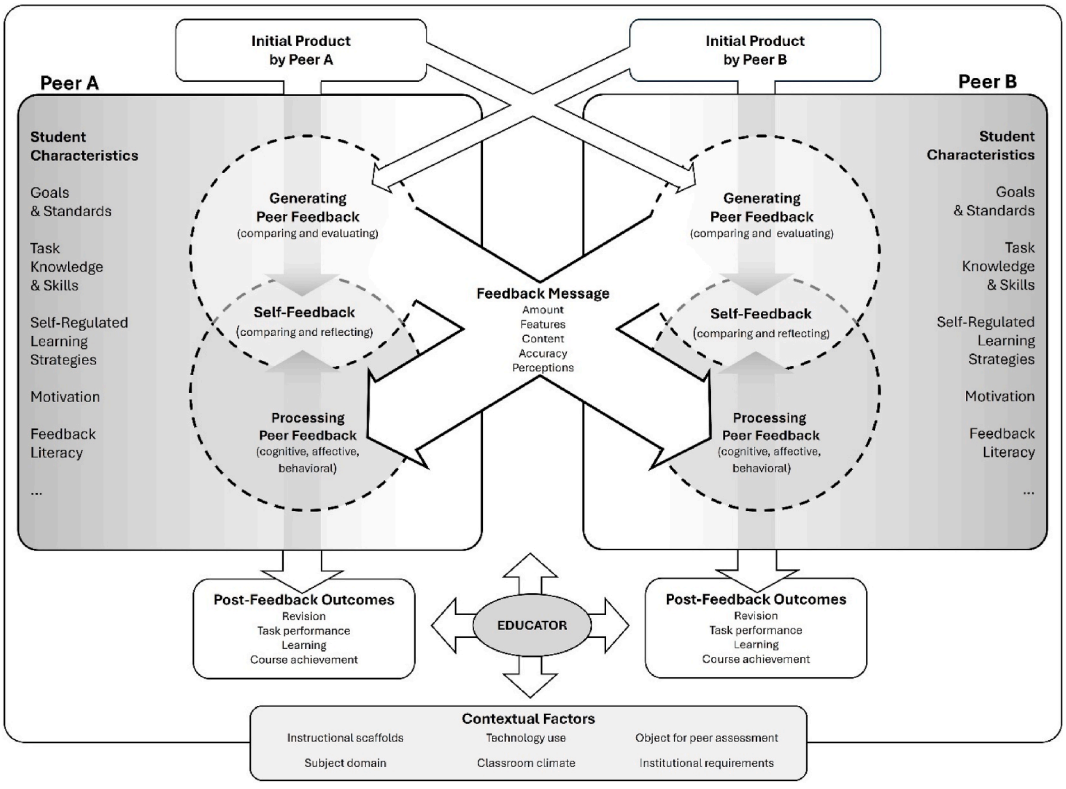


Fig. 1. Peer Feedback Model

Note. The gray areas associated with Peers A and B represent internal processes, including the generation and processing of peer and self-feedback, whereas the white areas represent external actions, such as the exchange of initial products, the provision of peer feedback, and the development of post-feedback outcomes.

Additionally, we sought to examine how this relationship varies across different operationalizations and conditions. These findings can provide insights into the relationship between two multifaceted variables and specify contexts for productive peer assessment practices that lead to enhanced learning and performance.

1.1. Towards a peer feedback model

Several theoretical models have been proposed to explain feedback processes and influential factors, but only a few illustrate interactions among different feedback agents or sources (Lipnevich & Panadero, 2021). In some models, peers are considered an external source of feedback for learners (e.g., Andrade, 2013; Butler & Winne, 1995; Evans, 2013; Lipnevich & Smith, 2022). However, these models do not specify how learners interact as both feedback providers and receivers, nor do they explain the internal processes associated with these dual roles. To address the complexity inherent in formative peer assessment activities, the models proposed by Narciss (2013) and Alemdag and Yildirim (2022b) explain learner interactions and processes. Nevertheless, because these models incorporate numerous interconnected elements, there remains a need for more parsimonious models that provide a clearer account of peer feedback interactions and processes.

We build on existing models and theoretical considerations to propose a new model of the core processes underlying peer feedback, illustrated through a one-on-one feedback exchange between two peers (Fig. 1). Our model considers both the internal (i.e., latent) generation and processing of peer feedback and self-feedback within each learner and the external (i.e., manifest) exchange of peer feedback messages, along with observable post-feedback outcomes. As shown in Fig. 1, first, each peer reviews the other's initial product, generating peer feedback through a process of comparison and evaluation. Critically, generating peer feedback is connected to self-feedback within the same learner, as comparing one's work with a peer's naturally stimulates reflection on one's own performance. The delivered feedback message, characterized by its quality dimensions (amount, features, content, accuracy, and perceptions), then reaches the receiving peer, who processes it cognitively, affectively, and behaviorally before producing observable post-feedback outcomes. Student characteristics shape this entire process, while the educator and broader contextual factors operate at the bottom of the figure, influencing the environment in which the exchange takes place. In this section, we describe the model in detail with the feedback elements of the MISCA model (i.e., Message, Implementation, Student, Context, and Agents; Panadero & Lipnevich, 2022).

Agents. Unlike general feedback models, which position peers as one of several external sources of information for the learner, our model places learners themselves at the center of the feedback process, simultaneously as providers of feedback to peers and as receivers of feedback from peers. This reciprocal relationship is illustrated through feedback exchanges between Peer A and Peer B (Fig. 1), with the educator contributing to the broader context, for example by fostering a positive error climate and providing training on effective peer feedback, rather than serving as the primary feedback source.

As feedback providers, students should consider task requirements and standards and compare them with their peers' products to make evaluative judgments. Based on these judgments, they design evaluative and tutoring actions to help their peers reflect on and, ideally, improve their work (see also Narciss, 2008; 2013; van Popta et al., 2017). Overall, generating peer feedback can be conceptualized as a constructive learning activity that entails integrating prior and current knowledge to construct new knowledge (Wu & Schunn, 2023). It can also be viewed as an active problem-solving process that involves detecting problems in peers' work, diagnosing them, and generating and communicating potential solutions (see also Cho & MacArthur, 2011).

A distinctive feature of our model is the explicit connection between peer feedback generation and self-feedback (Fig. 1). While providing feedback to peers, students can simultaneously compare their work with peers' work and reflect on their own performance, which can improve their understanding of task requirements and standards and may stimulate self-feedback (Nicol, 2021). Moreover, self-feedback emerges when students process the peer feedback they receive. They can compare feedback from multiple peers and compare peer feedback with their internal standards and self-feedback, which helps them decide what they need to do to enhance their task performance (Narciss, 2008, 2013; Nicol, 2021). By processing self-feedback that emerges during both the generation and uptake of peer feedback, students can adjust their goals, standards, task knowledge, skills, strategies, and motivation (Andrade, 2013; Butler & Winne, 1995; Narciss, 2008, 2013).

Message. The feedback message is the externally visible product of the peer feedback generation process, representing the written or spoken comment a student delivers to a peer. It can be classified in different ways, related to its content, function, and presentation (Narciss, 2008, 2013). One critical criterion in the analysis of feedback messages is their information quality, which can impact students' use of feedback (Panadero & Lipnevich, 2022). To conceptualize the quality of peer feedback, we draw on a recent framework by Zhang et al. (2024a), which is described in detail in Section 1.2. This provides an important basis for the present meta-analysis, as well as for the theoretical conceptualization of the characteristics of exchanged messages in our model and for how feedback quality may influence feedback implementation.

Implementation. The implementation of feedback involves both internal processing of external peer feedback and the work on post-feedback outcomes. Processing of peer feedback can occur at cognitive, affective, and behavioral levels (Lipnevich & Smith, 2022). At the cognitive level, students decode, analyze, and understand the feedback message, with possible transitions to self-feedback to compare and integrate external and internal information, reflect on their performance, and decide how to improve it. Emotional reactions students show to peer feedback messages comprise affective processing. Finally, behavioral processing involves the activities students engage in after receiving peer feedback (e.g., rereading peer feedback and asking peers for elaboration). Following the processing of peer feedback, internal changes in students' task knowledge, skills, strategies, and motivation, as well as observable changes in current and subsequent task performance, can occur. We define observable changes as post-feedback outcomes that may emerge either during or after peer assessment (see Section 1.3 for details).

Student characteristics. Student characteristics occupy a distinctive position in the model: they simultaneously shape and are shaped by every other element. For instance, a student's goals and standards, task knowledge and skills, self-regulated learning strategies, motivation, and feedback literacy may influence how they generate peer feedback, how they process the feedback they receive, and what post-feedback outcomes result, as outlined in previous models (e.g., Alemdag & Yildirim, 2022b; Narciss, 2013). At the same time, engaging in peer feedback activities can itself change these characteristics, for example by shifting self-efficacy beliefs or developing self-regulatory skills (for a review, see Yan et al., 2026). To represent this bidirectional relationship between student characteristics and the feedback process, we used dashed outlines for the internal peer feedback activities (Fig. 1). Additional characteristics such as gender and educational level, which are stable variables in the peer feedback process, are not depicted but may also play moderating roles, as explained in Section 1.4.

Context. The context-related factors are the final element of our model and can affect student characteristics and peer feedback processes. They involve instructional scaffolds (e.g., training, rubrics provided during feedback provision, anonymity; as explained in Section 1.4), technologies integrated into the peer assessment context, types of objects for peer assessment, disciplines or subject domains, classroom climate (e.g., error climate), and institutional requirements. Educators also play an important role in creating a classroom context conducive to learning from peer feedback (Panadero & Lipnevich, 2022). As indicated in our model (Fig. 1), teachers' contributions to the peer feedback process are directed towards all external areas of the model, including the delivery of peer feedback messages (e.g., teacher feedback on students' peer feedback) and post-feedback outcomes (e.g., by providing revision opportunities, teacher feedback, or new tasks for transfer or learning).

Taken together, these elements describe a system in which peer feedback quality can shape post-feedback outcomes through the message exchange, but may also be moderated by agents, student characteristics, and context. The following sections examine each of these components, beginning with peer feedback quality (Section 1.2), followed by post-feedback outcomes (Section 1.3) and potential moderating factors (Section 1.4).

1.2. Conceptualizing and measuring peer feedback quality

Several descriptions have been provided to explain what constitutes quality feedback. One of the most frequently cited is by Hattie and Timperley (2007), who define quality feedback in terms of three components that address the following key questions: "*Where am I going?* (What are the goals?), *How am I going?* (What progress is being made toward the goal?), and *Where to next?* (What activities need to be undertaken to make better progress?)" (p. 86). Additionally, Hattie and Timperley (2007) described four levels of feedback (i.e., self, task, process, and self-regulation) and argued that feedback at the process and self-regulation levels is particularly effective in enhancing students' understanding, task performance, and competence. Although this framework remains foundational for understanding the functions and levels of feedback, a more fine-grained conceptualization that accounts for the specific qualities of feedback messages may provide a more complete picture of peer feedback as a distinct feedback source.

In this regard, Zhang et al. (2024a) introduced a comprehensive conceptual framework. According to their model, peer feedback quality comprises six dimensions: amount, actual comment quality (or feedback features), feedback content, rating accuracy, perceived comment quality, and the reviewing process. Zhang et al. (2024a) also empirically demonstrated the multidimensional nature of this construct, although they noted some overlap among indicators across dimensions. The present study draws on five of these six dimensions, grouping them as amount, content, accuracy, features, and perceptions. The reviewing process dimension is excluded because its indicators are subject to interpretive ambiguities. Specifically, it involves time- and count-based process measures (e.g., review latency and the number of comment revisions) that can reflect multiple underlying factors. For example, a higher number of comment revisions may indicate either greater effort to provide high-quality feedback or uncertainty and confusion on the part of the student.

As the first dimension of peer feedback quality, *feedback amount* refers to the quantity of feedback, including the number of comments, reviews, and words or sentences within the feedback. Prior studies (e.g., Cho & MacArthur, 2011; Wu & Schunn, 2021b) have shown positive correlations between the amount of feedback provided or received and improvements in task revisions and performance. However, this relationship may not always be linear, as suggested by Zhang and Schunn (2023). For instance, Patchan et al. (2016) found that when the number of feedback comments received exceeded the average feedback comments in the class, students were less likely to implement the feedback, and they attributed this finding to the increased cognitive load of evaluating and selecting actionable comments.

Second, the dimension of *actual feedback quality*, as proposed by Zhang et al. (2024a), encompasses several measures, including whether comments are implemented or processed, the impact of implementation on performance, and feedback features. Actual peer feedback quality is typically assessed by external raters who code the feedback messages according to predefined criteria. The present meta-analysis specifically focuses on feedback features, also known as the structural components of feedback (e.g., problem identification, explanation, suggestions or solutions, localization, and mitigating praise; He & Gao, 2023). These features are typically separated into two broad categories: verification and elaboration (Gielen & De Wever, 2015; Narciss, 2008). Verification includes positive, negative, or neutral statements about whether a learner performed a task correctly, while elaboration provides justifications, explanations, or suggestions for improvement. Gielen and De Wever (2015) recommended providing more elaboration than verification to enhance the quality of feedback.

However, empirical studies have yielded mixed results regarding the impact of feedback features for feedback providers and receivers. On the one hand, elaborated feedback, particularly justifications (e.g., Gielen et al., 2010; Illana-Mahiques, 2021; Wu & Schunn, 2021a) and suggestions (e.g., Cho & MacArthur, 2011; Nelson & Schunn, 2009) in provided and received feedback, has been positively correlated with the implementation of feedback and post-feedback task performance. On the other hand, others found

null or negative effects of either informative or suggestive elaborations in provided and received feedback (e.g., He & Gao, 2023; Lu & Law, 2012; Zhang et al., 2025). Some peer feedback studies also reported positive effects of verifications in the received feedback on feedback implementation, performance, or learning (e.g., Lu et al., 2023; Lu & Law, 2012; Patchan et al., 2016). Thus, analyzing the overall effects of verifications and elaborations within peer feedback remains important.

Third, the *feedback content* dimension concerns the meaning and focus of the comments. It assesses whether feedback addresses critical, high-level issues (e.g., meaning-related, global problems/solutions), covers assessment criteria from a rubric, refers to the entire work or specific sections, or includes original comments that are not copied from other reviewers' responses or external sources (Zhang et al., 2024a). Although revisions based on feedback addressing high-level issues can lead to significant performance gains, some studies found that students were less likely to implement such feedback when they received (e.g., He & Gao, 2023; Patchan et al., 2016).

Fourth, *rating accuracy* refers to the degree to which student ratings align with expert ratings (Zhang et al., 2024a). In the context of formative peer assessment or qualitative peer feedback, this term corresponds to content accuracy. Content accuracy gauges the degree to which students' comments on peers' strengths, weaknesses, and suggestions align with expert evaluations. Assessments of content accuracy are hence usually conducted by experts knowledgeable about the task and evaluation criteria. Although only a few quantitative studies have examined the influence of content accuracy in peer feedback, some evidence highlights its importance. For example, Gielen et al. (2010) found a significant and positive effect of the accuracy of criticisms in received peer feedback (i.e., those indicating weaknesses or suggestions for improvement) on students' final writing performance after controlling for initial performance. Similarly, He and Gao (2023) identified accurate problem detection in received peer feedback as a significant predictor of feedback implementation, with accuracy showing the largest effect size among the variables examined.

Fifth, the dimension of *perceived feedback quality* focuses on how students perceive the feedback they give or receive. This includes evaluations of feedback helpfulness, understandability, and their willingness to act on the feedback. Unlike the other dimensions that rely on the content analysis of written feedback, perceived feedback quality is assessed through self-reported questionnaires. For instance, Alqassab et al. (2019) included a questionnaire item assessing students' perceptions of the usefulness of the feedback they provided: "How useful do you think your feedback will be for this student?" (p. 145). Findings in this area are also mixed. For instance, Lu et al. (2023) reported significant relationships between acceptance of received feedback and peer-initiated writing revisions. However, both Lu et al. (2023) and Huisman et al. (2018) found no significant correlations between perceived quality of received feedback and final writing performance. Notably, Ludemann and McMakin (2014) reported a large, negative correlation between the perceived helpfulness of provided feedback and writing performance.

Finally, the dimensions of peer feedback quality may be measured individually or combined into an integrative measure. Using integrative measures allows for a more holistic evaluation of feedback quality (He & Gao, 2023). For example, Dmoshinskaia et al. (2022) combined measures of feedback usefulness (justification and suggestions) with feedback accuracy to create a single composite measure. Similarly, Wu and Schunn (2023) analyzed the frequency of comments addressing high-level issues in specific feedback features such as praise, explanation, and suggestion.

Overall, the dimensions proposed by Zhang et al. (2024a) offer a valuable conceptual framework for future peer assessment studies. However, the contradictory findings regarding their impacts on the improvement of students' current task performance and learning underscore the need for a systematic synthesis to clarify the conditions under which these peer feedback quality dimensions serve as significant predictors of post-feedback outcomes, given the complex interplay of factors such as message content, implementation, context, and student roles (Panadero & Lipnevich, 2022). This meta-analysis addresses that need directly.

1.3. Post-feedback outcomes

Having defined peer feedback quality, we now turn to the outcomes that quality feedback is expected to produce. In the current study, we refer to the observable indicators of implementation of peer feedback as post-feedback outcomes (Fig. 1) and focus on their relation with peer feedback quality. Post-feedback outcomes associated with peer feedback quality can be grouped into four main categories: revision, performance, learning, and course achievement. These categories differ in terms of their temporal proximity to the provision or reception of peer feedback, as well as the similarity between the tasks in which the feedback was given and those in which post-feedback outcomes were measured.

Revision refers to the changes students make to their initial task performance after providing or receiving peer feedback. These changes can be further categorized as peer-initiated or self-initiated (Lu et al., 2023; Wu & Schunn, 2023). Peer-initiated changes result directly from implementing peer feedback, while self-initiated changes may emerge from insights gained either while providing feedback to others or from further reflection on the feedback received. Revision is often seen as a transitional stage that links feedback processes with subsequent outcomes (Lu et al., 2021). *Performance,* as another post-feedback outcome, refers to "the knowledge gain observed on repeated tasks (i.e., pre-post multiple-choice questions; multiple drafts of the same writing assignment)" (Nelson & Schunn, 2009, p. 376). Students' revised or final task performance following peer assessment represents a specific and common example of this outcome. *Learning* involves knowledge gains demonstrated in subsequent transfer tasks, such as answering application questions on new tests or completing future writing assignments (Nelson & Schunn, 2009). Revision and performance represent more proximal outcomes that are assessed in relation to the task involving peer feedback, whereas learning reflects a more distal outcome that becomes evident after the peer feedback process for the current task is complete. *Course achievement* may represent an even more distal outcome, as it is typically assessed after a longer period of instruction and across multiple tasks for peer assessment.

The type of post-feedback outcome may influence its relationship with peer feedback quality. For example, while high-quality peer feedback may increase the likelihood of or the number of revisions, it may not always lead to significant improvements in students'

final task performance or learning (e.g., Lu et al., 2021, 2023; Patchan et al., 2016). Students can also improve their task understanding and learning without revising their current task, particularly when they provide high-quality feedback (e.g., Lu & Law, 2012; Wu & Schunn, 2021b, 2023). Overall, post-feedback outcomes are multifaceted, and their relationship with peer feedback quality is likely to vary depending on students' peer feedback roles (i.e., feedback provision or reception) as well as individual and contextual factors. The following section examines the moderating variables considered in this meta-analysis.

1.4. Potential moderating factors of peer feedback quality and post-feedback outcomes

To provide a more nuanced understanding of the variations in the correlation between peer feedback quality and post-feedback outcomes in this study, we consider the roles of different elements in our peer feedback model (Fig. 1). They include gender and educational level as student characteristics; the peer feedback role as a feedback-agent factor; training, on-task instructional support, and anonymity as contextual factors.

Gender has been identified as an influential demographic factor in peer feedback. A recent systematic review (Kerman et al., 2024) and prior empirical studies (Almuadi & Barr, 2023; Noroozi et al., 2020, 2023; Ocampo et al., 2024) have shown that female students tend to provide higher quality peer feedback on specific topics or documents. Regarding post-feedback outcomes, Wu and Schunn (2021a) found that girls were more likely to apply peer feedback than boys. A similar finding emerged when students received feedback on the same topic more frequently (Wu & Schunn, 2020). However, evidence regarding performance improvement after receiving peer feedback is mixed. For example, Noroozi et al. (2023) found greater improvements among females, while Noroozi et al. (2020) reported higher gains among males.

Educational level may also influence the association between peer feedback quality and post-feedback outcomes. Older or more highly educated students often have more experience with peer assessment activities and may know better how to benefit from high-quality feedback. At the same time, prior negative experiences, such as receiving inaccurate or superficial feedback, may lead some students to devalue peer assessment, diminishing its impact. Previous meta-analyses have also yielded contradictory findings regarding educational level as a moderator. To illustrate, Double et al. (2020) found that peer assessment with grading was effective in higher education but not in primary or secondary education settings. Similarly, Yan et al. (2022) reported a significant moderating effect of educational level on the effectiveness of peer assessment combined with self-assessment, with a greater effect size found in higher education. However, they indicated no such effect when peer assessment was implemented on its own. Other meta-analyses (e.g., Li et al., 2021; Zhan et al., 2023) identified educational level as a non-significant moderator.

Peer feedback roles in peer assessment include feedback provision and reception that may contribute to post-feedback outcomes to varying degrees. Some studies have shown that the quality of both provided and received feedback correlates with learning outcomes, while others have found more substantial effects for feedback provision compared to reception when both were analyzed simultaneously (e.g., Chen et al., 2024; Cui & Schunn, 2024; Illana-Mahiques, 2021; Li et al., 2010). Cho and MacArthur (2011) proposed the "learning by reviewing" hypothesis, suggesting that providing feedback is an active problem-solving process involving problem detection, diagnosis, and solution generation. Van Popta et al. (2017) also highlighted the value of feedback provision, noting its contributions to higher-order thinking, critical reflection on both peer and personal work, and knowledge building through various cognitive processes (e.g., comparing ideas, questioning, evaluating, and suggesting revisions). While feedback provision often appears more beneficial than reception, systematic investigations comparing the two are still needed (Zong et al., 2021).

Training may be another potential moderator of the correlation between peer feedback quality and post-feedback outcomes. Training clarifies expectations for assessors and assessees, explains assessment criteria, demonstrates how to assess peer work and process feedback, and provides practice opportunities before actual peer assessment (Topping, 2009; Winstone et al., 2017). Empirical studies (e.g., Alemdag & Yildirim, 2022a; Alqassab et al., 2018; Li, 2017; Saito, 2008) support the benefits of training, reporting improvements in student performance, feedback quality, feedback implementation, and perceptions of peer assessment's usefulness. Consistent with these findings, previous meta-analyses and systematic reviews (e.g., Fleckney et al., 2025; Li et al., 2020; Zheng et al., 2020) have also revealed significantly greater effects of technology-facilitated peer assessment on learning when students received training and have recommended training to optimize the effectiveness of peer assessment.

In addition to training, specific *on-task instructional* support can help students provide high-quality peer feedback and better process and use the received peer feedback (Hornstein et al., 2025). Rubrics, prompts, sentence starters, structured templates, back-feedback forms, and dialogue tools are some examples of instructional support for peer feedback provision and reception. In the most recent meta-analysis, Hornstein et al. (2025) classified instructional support based on the peer feedback stage (feedback provision or reception) into which it was integrated and examined the separate effects of these types of support on process and outcome measures. They reported positive effects of instructional support for peer feedback provision on the quality of both feedback provision and reception, but not on subject-matter-related knowledge. Moreover, they found that instructional support for feedback reception had no significant impact on feedback reception or increased knowledge. These mixed findings raise questions about the potential role of instructional supports for peer feedback provision and reception in the relationship between peer feedback quality and post-feedback outcomes.

Anonymity in peer assessment, also considered a kind of instructional support, has been widely debated. It may lessen students' fear of being evaluated, encourage students to provide more critical feedback, and reduce their bias related to their peers' abilities (e.g., Lin, 2018; Vanderhoven et al., 2015). For instance, students aware of an assessor's competency in non-anonymous conditions may judge peer feedback differently, which can influence their revisions and learning. Strijbos et al. (2010) found that graduate students perceived feedback from competent peers as more adequate. However, they determined that learners receiving feedback from highly competent peers performed worse on the post-test than those assessed by less competent peers. In their systematic review, Panadero

and Alqassab (2019) also concluded that although anonymity may increase the provision of critical feedback, it enhances performance and achievement only marginally. Finally, meta-analyses have shown anonymity to be a non-significant moderator of peer assessment effects on learning achievement (Zheng et al., 2020), non-cognitive outcomes (Li et al., 2021), and higher-order thinking (Zhan et al., 2023).

1.5. The present study

Despite a growing body of systematic reviews and meta-analyses on peer assessment (e.g., Alqassab et al., 2023; Double et al., 2020; Fleckney et al., 2025; Kerman et al., 2024; Li et al., 2020, 2021; Panadero et al., 2023; Zhan et al., 2023; Zheng et al., 2020), none has specifically examined the nature of qualitative peer feedback messages and their role in shaping post-feedback outcomes. A recent systematic review on online peer feedback in higher education (Gao et al., 2024) categorized peer feedback content and functions and reported their frequencies. Nonetheless, it did not include analyses to calculate correlations between these categories that correspond to peer feedback quality and post-feedback outcomes in the current study. Furthermore, the latest meta-analysis on peer feedback (Hornstein et al., 2025) examined the effect of instructional scaffolds on peer feedback processes and outcomes; however, it did not involve investigating the correlations among feedback provision, feedback reception, and outcomes. Overall, the definition of quality peer feedback and its relation to post-feedback outcomes, as well as the effects of specific feedback quality dimensions, remain unclear in the literature (He & Gao, 2023).

Given these gaps in existing literature, there is a need for a meta-analysis that operationalizes the quality of peer feedback and provides validity evidence for this operationalization. Although Zhang et al. (2024a) proposed a multidimensional conceptual framework for describing quality peer feedback, further research is needed to validate this framework. Considering the multifaceted post-feedback outcomes, it is also crucial to examine how peer feedback quality correlates with various outcomes. Lastly, examining the moderating effects of contextual factors on the role of peer feedback quality is essential for explaining the variations in reported effect sizes (Zong et al., 2021).

The current meta-analysis addresses these needs by synthesizing studies on peer feedback quality drawing on Zhang et al.'s (2024a) conceptual framework and quantifying the relationship between peer feedback quality and post-feedback outcomes. Further analyses are conducted to investigate how operationalizations of peer feedback quality and post-feedback outcomes, as well as some student-related and contextual variables, affect these correlations. In doing so, this meta-analysis provides empirical validation for the peer feedback quality dimensions proposed by Zhang et al. (2024a). It further identifies the conditions under which peer feedback quality is most strongly associated with post-feedback outcomes and advances understanding of when and for whom high-quality peer feedback is most effective. Specifically, this study aims to answer the following research questions:

(1) What is the overall strength of the relationship between peer feedback quality and post-feedback outcomes?
(2) Does the strength of the relationship between peer feedback quality and post-feedback outcomes vary across different operationalizations of these variables?
(3) Does the strength of the overall relationship between peer feedback quality and post-feedback outcomes vary according to the following moderators: gender, educational level, peer feedback role, training, on-task instructional support, and anonymity?

2. Method

2.1. Eligibility criteria

The primary eligibility criteria used to select studies for this meta-analysis are listed in Table 1. Additionally, we included studies published as journal articles, conference proceedings, dissertations, or theses, and available in full text. The search was limited to reports in English. There were no restrictions regarding the year of publication. Finally, studies were only considered eligible if they reported sufficient data to calculate effect sizes for the correlation between peer feedback quality and post-feedback outcomes.

2.2. Search strategy and selection process

To identify studies meeting these criteria, a comprehensive search was conducted across six databases on October 18, 2024, and April 4, 2025: Web of Science, Scopus, ERIC, PsycArticles, APA PsycInfo, and ProQuest. The complete search query is provided in Table 2. It includes terms related to the key eligibility criteria outlined in Table 1. While document type was selected as dissertations and theses when searching ProQuest, it was classified as (early access) journal articles and conference/proceeding papers in other databases. Titles, abstracts, and keywords were screened using this strategy. The database search yielded 2403 unique results, excluding Zhang et al. (2024a), which was used solely to inform the conceptual framework of this study.[1] Following a review of titles and abstracts, 134 reports were selected for further consideration. To ensure thoroughness, ASReview, an open-source machine learning tool designed to assist with title and abstract screening in literature reviews (van de Schoot et al., 2021), was used to identify potential false negatives that may have been missed during the initial screening phase. After providing ASReview with training data

[1] However, the following paper from the same authors (Zhang et al., 2024b), which also used the similar dataset (i.e., 20,912 feedback comments on higher-level writing aspects produced by 844 secondary school students), was included in the meta-analysis.

Table 1
Eligibility criteria followed in the study.

Criterion	Explanation
Population	Students in K-12 or higher education
Intervention	Students' provision/reception of peer feedback that gives qualitative comments about a student's performance in a task
Peer feedback quality	Quality of peer feedback, individually written by students, evaluated with regard to amount, features, accuracy, content, or perceptions, and analyzed at feedback provider or receiver levels
Post-feedback outcome	Revision of initial individual task performance, revised task performance, performance in other tasks or tests, or course performance
Study design	Quantitative studies that report correlations between peer feedback quality and post-feedback outcomes

Table 2
Terms included in search query.

Population	Intervention	Peer feedback quality	Post-feedback outcome	Study design/ analysis
school OR student* OR universit* OR k-12 OR k12 OR "higher education" OR undergraduat* OR graduat* OR college OR "post-secondary" OR postsecondary OR tertiary OR "pre-service teacher*"	"peer feedback" or "peer review" or "peer assessment" or "peer evaluation"	amount or length or quantity or content or quality or feature* or style or type* or nature or accuracy or accurate or perception* or perceived or understanding or agreement or helpfulness or willingness	performance or improvement or progression or growth or learning or uptake or implementation or implemented or revision or revised or "feedback use"	relation* or associat* or correlat* or predict* or regression

Notes. Keywords related to the five main criteria were connected with the AND Boolean operator.

(the selected reports labeled as relevant and five randomly selected reports labeled as irrelevant), the remaining unlabeled reports were ranked from most to least relevant. Following the stopping rule by Boetje and van de Schoot (2024), screening continued until 50 consecutive irrelevant records were found, which occurred after examining approximately 150 additional records. This process resulted in three more potential reports being identified. In total, 137 reports were selected after the title and abstract review stage. To evaluate the reliability of title and abstract screening while maintaining a feasible review process, an independent second coder assessed a randomly selected subset of 367 records (approximately 15% of the total). The first author and the second coder reached 96% agreement ($\kappa = .59$). The imbalanced distribution of ratings (i.e., 351 papers excluded and 16 papers included) may have affected the κ value (Feinstein & Cicchetti, 1990). The two coders resolved each disagreement through discussion, considering all eligibility criteria (Table 1).

Out of 137 reports, full-text access was obtained for 132 reports. After the first author applied the eligibility criteria to the remaining full texts, 50 reports were included in the meta-analysis. Two additional studies were identified by screening the reference lists of the included papers and previous related reviews, bringing the total number of eligible reports to 52. Intercoder reliability for full-text screening was assessed using a randomly selected sample of 20 reports (15% of the 132 potentially eligible reports). The first author and a trained second coder achieved 85% agreement ($\kappa = .69$), indicating substantial agreement (Landis & Koch, 1977). Discrepancies between the two coders were later discussed and resolved. For example, following discussion, the two papers initially included by the second coder were excluded because they did not report a measure of peer feedback quality or a correlation between peer feedback quality and post-feedback outcomes. The entire selection process, including reasons for exclusion during full-text screening, is illustrated in the PRISMA flow diagram (Page et al., 2021) shown in Fig. 2.

2.3. Data extraction and coding

Quantitative data required to calculate effect sizes for the correlations between peer feedback quality and post-feedback outcomes were extracted from the 52 reports. These data included Pearson or Spearman correlation coefficients, Cohen's d, η^2, standardized or unstandardized regression coefficients, and descriptive statistics. In addition, data related to feedback quality dimensions, post-feedback outcomes, and moderator variables were collected for this meta-analysis. Explanations of these variables are provided in Table 3. The entire coding scheme with additional descriptive variables (e.g., source type, subject domain, object of peer assessment, and technology use) is presented in Supplementary Table S1. Furthermore, Supplementary Table S2 lists all the studies coded according to these variables.

During the coding phase, the first author applied the study's coding scheme to all eligible reports. A master's level student assistant was then trained to apply this coding scheme and extract data from a sample of studies as a second coder. To this end, the first author and the student assistant coded six papers, discussing and refining the coding scheme. The student assistant then coded 13 additional papers (25% of 52 reports) alone. Inter-rater reliability analyses revealed substantial agreement between the two coders ($\kappa > .60$; see Table 3). Discrepancies were resolved through discussion, after which the first author revised the explanations in the coding scheme to make them clearer for readers.

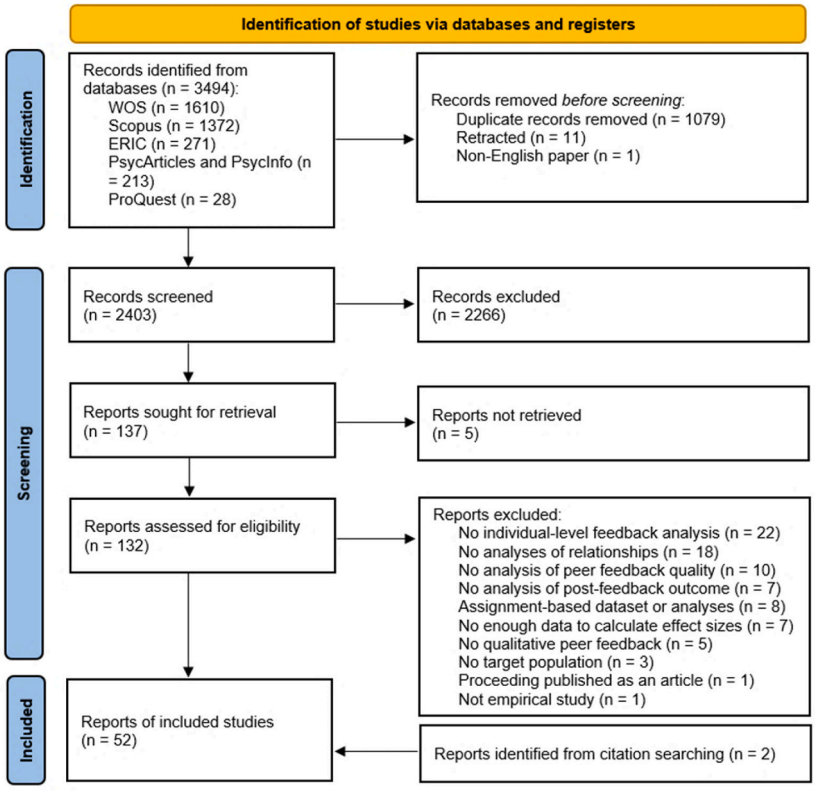


Fig. 2. PRISMA Flow Diagram of this Meta-Analysis.

2.4. Effect measures

The Pearson product-moment correlation coefficient (*r*) was selected as the primary effect size measure. For studies that reported statistics other than *r*, appropriate conversion formulas (Supplementary Table S3) were applied. Specifically, standardized regression coefficients were calculated from unstandardized regression coefficients, following Bowman (2012). All standardized regression coefficients were then converted to *r* using Peterson and Brown's (2005) formula. In addition, effect sizes reported as Cohen's *d* (Borenstein et al., 2009), Spearman's correlations (Croux & Dehon, 2010; Pearson, 1907), and η^2 (Lakens, 2013) were also transformed into *r*. Because the variance of *r* is dependent on its magnitude, all Pearson correlation coefficients were converted to Fisher's *z* values for the statistical analyses (Borenstein et al., 2009). However, for ease of interpretation, the results, including overall effect sizes and confidence intervals, were converted back to Pearson correlations.

2.5. Synthesis methods

Before synthesizing the effect sizes, we noted that some studies reported multiple correlation coefficients. These reflected different metrics of peer feedback quality, various post-feedback outcomes, and correlations arising when the same students served as both feedback providers and receivers, thereby violating the assumption of independence among effect sizes. To address this dependency, we applied robust variance estimation in this meta-analysis (Hedges et al., 2010; Tanner-Smith & Tipton, 2014). The analyses were conducted in R using the robumeta package (Fisher et al., 2023), selecting the correlated effects model with small-sample corrections, and setting the rho value at .8, as recommended by Hedges et al. (2010) and Tanner-Smith and Tipton (2014). Effect sizes were weighted by inverse variance in the robust variance estimation method, such that studies with larger sample sizes, and thus greater precision, received greater weight in the analyses (Hedges et al., 2010). To assess the weighted average effect sizes from the meta-analysis, we considered $r = .10$ a small effect, $r = .20$ a medium effect, and $r = .30$ a large effect (Gignac & Szodorai, 2016).

To assess heterogeneity, both I^2 statistics and prediction intervals were calculated. Following Higgins et al.'s (2003) guidelines, I^2 values were interpreted as low (25%), moderate (50%), or high (75%) proportions of variance attributable to true effect size differences. Prediction intervals provided additional information for the expected variation in true effect sizes across studies (Borenstein, 2024).

Given the high observed heterogeneity, separate meta-regression analyses were conducted for each moderator to explore potential sources of variation in effect sizes across studies. Moderator subcategories with fewer than four degrees of freedom were excluded from these analyses to ensure adequate statistical power (Tanner-Smith et al., 2016). For moderators with more than two subcategories, a

Table 3
Description of main and moderator variables in the study.

Variable	Category	Explanation
Dimension of peer feedback quality (κ = .64, 75%)	1. Amount	This dimension includes the following measures obtained from analyzing peer feedback: the number of feedback comments, reviews, and words or sentences in feedback comments.
	2. Content	This dimension includes the following measures obtained from analyzing peer feedback: addressing critical issues (e.g., meaning-related, global problems/solutions, high-level) in performance, giving feedback related to assessment criteria or criteria use, and giving feedback about some particular topics or errors.
	3. Accuracy	This dimension relates to how well the feedback accurately addresses strengths or problems in performance or improvement strategies. The measures are obtained by analyzing peer feedback.
	4. Features	This dimension includes three subcategories: (1) verification, (2) elaboration, and (3) verification and elaboration. The measures are obtained from analyzing peer feedback. (1) Verification: The feedback expressed positive (e.g., praise), negative evaluations (e.g., problems), or neutral statements (i.e., summary) without providing further information or elaborations. (2) Elaboration: The feedback provided suggestions/solutions, justified strengths, weaknesses, or suggestions/solutions, or indicated places of problems. (3) Verification and elaboration: One feedback quality measure took into account the existence of components regarding both verification and elaboration. In other words, some components of verification and elaboration were combined into a single measure.
	5. Perceptions	This dimension includes the following measures obtained from analyzing questionnaire responses: perceived feedback adequacy/quality, willingness to improve, or perceived helpfulness/usefulness/accuracy of feedback.
	6. Integrative	The feedback quality measure obtained from analyzing qualitative peer feedback takes into account more than one dimension listed here (e.g., content + features, accuracy + features). For example, it checks whether the feedback comment addresses a high-level issue in terms of content and includes a specific feedback feature.
Post-feedback outcome (κ = .66, 76%)	1. Revision	Students' implementations of peer feedback or revisions were analyzed (e.g., the number of revisions after peer feedback reception or the amount of implemented peer feedback).
	2. Task performance	Students' revised task performance after peer assessment, or the improvement in their performance on a particular task from the first version to the revised version, was analyzed.
	3. Learning	Students' performance in the subsequent relevant tasks or tests was analyzed.
	4. Course achievement	Students' overall course achievement was analyzed.
Gender (ICC = .999)	-	The percentage of females in the sample who provided/received peer feedback
Educational level (κ = 1.00)	1. K-12	Primary and secondary grade students
	2. Higher education	Undergraduate and graduate students in higher education
Peer feedback role (κ = .62, 77%)	1. Feedback provision	The quality of the feedback provided was associated with feedback providers' post-feedback outcomes.
	2. Feedback reception	The quality of the feedback received was associated with feedback receivers' post-feedback outcomes.
Training (κ = .64, 77%)	1. Yes	Training for peer assessment was provided for the participants. It might be explicitly stated by authors, OR it can be inferred if there was a structured preparatory stage explaining assessment criteria of a specific task, example works or comments, or how to effectively give or receive peer feedback before peer assessment.
	2. No	Before the peer assessment, training was not provided.
On-task support for peer feedback provision (κ = .26, 77%)[a]	1. Yes	The feedback provision stage was facilitated with cognitive support (e.g., assessment criteria, rubrics, prompts or instructions, templates, or oral teacher guidance).
	2. No	Support for the feedback provision stage was not available.
On-task support for peer feedback reception (κ = .65, 85%)	1. Yes	The feedback reception stage was facilitated with cognitive support (e.g., back-feedback forms to evaluate the received feedback, teacher explanations, dialogs, and discussions between feedback providers and receivers).
	2. No	Support for the feedback reception stage was not available.
Anonymity (κ = .63, 77%)	1. Anonymous	Feedback providers did not know to whom they were giving feedback. OR feedback receivers did not know who gave them feedback.
	2. Non-anonymous	Feedback providers knew to whom they were giving feedback. OR feedback receivers knew who gave them feedback.
	3. Not available	There is not enough information about anonymity in the paper.

[a] The low kappa value may be attributable to the imbalanced distribution of data.

reference category was selected, and omnibus *F*-tests or Wald tests were performed to assess the linear contrasts of the models (Tipton & Pustejovsky, 2015) using the clubSandwich package in R (Pustejovsky, 2024).

Additionally, sensitivity analyses were performed. Potential outliers were identified by examining whether the studentized deleted residuals for individual effect sizes exceeded 1.96 (Viechtbauer & Cheung, 2010). Because such outliers can occur by chance, influence diagnostics were also run using the influence function in R to detect cases that substantially impacted the fitted model (Viechtbauer & Cheung, 2010). Based on this analysis, one study (Jin et al., 2024) and one effect size from each of two studies (Costley et al., 2025; Ludemann & McMakin, 2014) were excluded, leaving 51 studies in the final meta-analysis. Finally, sensitivity analyses were conducted using different rho values (0, .2, .4, .6, and 1) for robust variance estimation to test the robustness of the results. The overall effect size

remained stable across these variations.

2.6. Bias assessment

We assessed both the potential biases in the findings of the included studies and the risk of bias in the results of our meta-analysis. To evaluate bias in the individual studies, we used the checklist developed by Liu et al. (2024), which was adapted from the Joanna Briggs Institute Appraisal Checklist for Analytical Cross-Sectional Studies (Moola et al., 2020). In this meta-analysis, five questions from the checklist were rated using a three-point scale. Detailed descriptions for each rating point were provided for all questions (see Supplementary Table S4). Based on the results of this assessment, two studies that scored below 60 out of 100 were excluded. This left a final sample of 49 studies included in the meta-analysis.

The risk of bias in the meta-analysis itself was evaluated by examining the funnel plot of the aggregated effect sizes and by applying Duval and Tweedie's trim and fill method, as well as Egger's regression tests. Specifically, we analyzed whether the effect sizes were symmetrically distributed around the average effect size in the funnel plot, which would indicate a low likelihood of publication bias (Borenstein et al., 2009). The trim and fill method was also used to estimate the number of potentially missing studies and to calculate the adjusted overall effect size if those studies were included. Furthermore, Egger's regression test (Egger et al., 1997) was performed to quantitatively assess the symmetry of the funnel plot with aggregated effect sizes. Additionally, we conducted a meta-regression with the robust variance estimation method, incorporating the standard error as the predictor variable, to evaluate publication bias. Finally, the genuine, true effect beyond publication bias was evaluated using the precision-effect test and precision-effect estimate with standard errors (PET-PEESE) method (Stanley & Doucouliagos, 2014), which incorporates standard errors or sampling variance as predictors in two meta-regression models. Specifically, PET is recommended in the absence of a genuine effect, whereas PEESE is used when a genuine effect is present (Stanley & Doucouliagos, 2014).

3. Results

3.1. Preliminary analyses

The funnel plot in Fig. 3 displayed an almost symmetrical distribution of aggregated effect sizes around the overall effect size. The trim and fill method also suggested that no studies were missing on either the left or right side of the plot. Additionally, Egger's regression test, performed using the aggregated effect sizes, indicated no publication bias ($z = -.56$, $p = .576$). Similarly, when the meta-regression was conducted using the robust variance estimation method, with the standard error as the predictor variable, the results indicated nonsignificant asymmetry in the funnel plot ($\beta = -.23$, 95% CI [-.26, .81], $p = .658$). Finally, following the PET-PEESE method, we detected a significant intercept or bias-corrected effect in the PET results ($\beta = .23$, 95% CI [.10, .36], $p = .001$), which directed us to use PEESE (Stanley & Doucouliagos, 2014). The PEESE results revealed a significant true effect size ($\beta = .22$, 95% CI [.14, .31], $p < .001$). This corresponds to Pearson's $r = .22$, which is comparable to the original overall effect size ($r = .20$) presented in Section 3.2, indicating robustness of the estimated effect.

3.2. Study characteristics and overall effect

This meta-analysis included 241 effect sizes derived from 53 samples that were reported in 49 papers (47 journal articles and two conference proceedings). Study participants came from various continents ($n_{\text{Asia}} = 25$, $n_{\text{North America}} = 14$, $n_{\text{Europe}} = 9$, $n_{\text{Asia and Europe}} = 2$, $n_{\text{Australia}} = 2$, $n_{\text{South America}} = 1$). Among the studies that reported gender percentages, the mean proportion of female participants was 62%. Higher education students ($n = 39$) were included in more studies than K-12 students ($n = 14$). The subject domains of the peer assessment tasks were primarily arts and humanities ($n = 26$), natural sciences, mathematics, and statistics ($n = 9$),

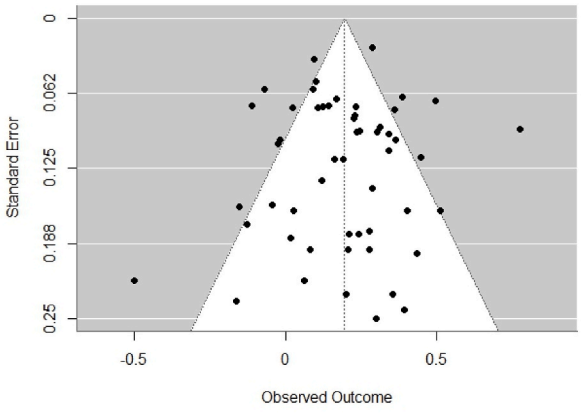


Fig. 3. Funnel plot of aggregated effect sizes.

and education ($n = 7$). Written assignments ($n = 41$) were the most frequently used objects of peer assessment in these studies. Additionally, almost all studies incorporated technological tools to facilitate peer assessment ($n = 50$) and applied bidirectional peer assessment where students were both feedback providers and receivers ($n = 49$).

As a primary finding, the meta-analysis revealed a statistically significant and positive correlation between peer feedback quality and post-feedback outcomes (Pearson's $r = .20$, 95% CI [.15, .25], $p < .001$). In other words, higher peer feedback quality was associated with better post-feedback outcomes. This correlation falls between the small and medium benchmarks according to Cohen (1988) and corresponds to a medium effect size according to Gignac and Szodorai (2016). There was also substantial heterogeneity among the effect sizes ($I^2 = 89\%$, $\tau^2 = .06$, PI [–.30, .70]).

3.3. Effect sizes based on different operationalizations of peer feedback quality and post-feedback outcomes

Further analyses were conducted to examine how the correlation varied across different operationalizations of two core variables of this study: peer feedback quality and post-feedback outcomes. Table 4 reports effect sizes for each group of these variables, as well as for other moderator variables.

First, the measures of *peer feedback quality* were categorized into six groups: amount, content, accuracy, features, perceptions, and integrative measures. However, except for Zhang et al.'s (2024a) paper used for the conceptual framework of this study, there was only one study investigating the correlation between overall peer feedback accuracy and post-feedback outcomes (Alqassab et al., 2019).[2] Hence, it was not possible to calculate an overall effect size for the accuracy dimension. The studies (e.g., Dmoshinskaia et al., 2022; Wu et al., 2022) that analyzed accuracy with other dimensions (e.g., feedback features) within a single measure were classified under the integrative category. Aside from students' perceptions ($r = .15$, $p = .078$), four feedback quality dimensions showed significant positive correlations with post-feedback outcomes (Fig. 4). Specifically, feedback amount ($r = .24$, $p = .003$), content ($r = .18$, $p = .042$), features ($r = .17$, $p = .003$), and integrative measures of feedback quality ($r = .22$, $p = .002$) all demonstrated significant associations with post-feedback outcomes. However, the meta-regression results indicated that the perception dimension did not differ significantly from other dimensions in terms of its correlation with post-feedback outcomes (Table 5). The omnibus F test also revealed that the correlations were not significantly affected by the feedback quality dimension ($F = .30$, $p = .875$).

Additional analyses were performed for the *feedback features* dimension, focusing on 110 effect sizes that were categorized into three subcategories: verification, elaboration, and the combination of verification and elaboration. Statistically significant correlations were found between feedback quality and post-feedback outcomes when the measured feedback feature was the frequency of elaborations ($r = .10$, $p = .01$) and the combination of verifications and elaborations ($r = .30$, $p = .003$; Table 4). Meta-regressions using the combination category (verifications + elaborations) as the reference group revealed that including both features in one measure was more strongly associated with post-feedback outcomes, compared to the measures that focused only on verifications ($\beta = -.24$, $p = .025$) or elaborations ($\beta = -.21$, $p = .013$; Table 5). An additional pairwise comparison showed that elaboration and verification measures did not differ significantly from each other ($\beta = -.03$, $p = .611$). Omnibus F test results indicated that the correlations between peer feedback quality and post-feedback outcomes did not significantly differ in terms of feedback features ($F = 3.74$, $p = .0504$), although there was a tendency for the measures that included both verification and elaboration to show stronger associations with outcomes compared to measures that included either verification or elaboration alone.

The correlations between peer feedback quality and different *post-feedback outcomes* were also analyzed. Due to the low frequency of studies focusing on course achievement, this category was not included in the analyses. Peer feedback quality was significantly correlated with revision ($r = .18$, $p = .025$), task performance ($r = .16$, $p < .001$), and learning ($r = .26$, $p < .001$; Table 4). Meta-regression analysis, with revision as the reference group, revealed that individual comparisons with learning and task performance were not statistically significant (Table 5). An additional pairwise comparison showed that compared with learning outcome, task performance outcome had significantly smaller effects ($\beta = -.11$, $p = .015$). The omnibus F test (Table 5) also indicated significant differences among outcome categories ($F = 3.81$, $p = .042$).

Moreover, in an exploratory analysis, we examined the correlations between peer feedback quality and post-feedback outcomes separately for both provided and received feedback. For provided feedback, the effect sizes of revision ($r = .12$, $p = .060$), task performance ($r = .25$, $p < .001$), and learning ($r = .31$, $p < .001$) were ordered from lowest to highest. In meta-regressions that excluded the revision category due to its degrees of freedom being lower than 4, the correlations between provided peer feedback quality and performance and learning did not differ significantly (Table 5). For received feedback, the revision category demonstrated the highest effect size ($r = .21$, $p = .047$), followed by learning ($r = .15$, $p = .069$) and task performance ($r = .13$, $p = .002$). However, for received feedback, the type of post-feedback outcome did not significantly influence the strength of the correlation.

3.4. Moderator analyses

The first moderator variable that we examined was *gender*, specifically the percentage of female students in each sample. Due to missing data in 14 samples, a total of 39 samples with 177 effect sizes were included in this meta-regression. The results showed that gender was not a significant moderator of the relationship between peer feedback quality and post-feedback outcomes (Table 5). Similarly, *educational level* did not significantly moderate the relationship between feedback quality and outcomes. Peer feedback

[2] Both Zhang et al. (2024a) and Alqassab et al. (2019) found significant positive correlations between peer feedback accuracy and post-feedback outcomes, with correlation coefficients of $r = .48$ and $r = .64$, respectively.

Table 4
Results of subgroup analyses.

Variable	*n*	*k*	Estimate	95% CI	*t*	*df*	*p*
Quality dimension[a]							
Amount	14	28	.24	[.10, .37]	3.77	11.43	.003
Content	5	25	.18	[.01, .34]	2.95	3.99	.042
Features	19	110	.17	[.07, .27]	3.45	16.29	.003
Integrative	17	58	.22	[.10, .27]	3.97	12.64	.002
Perception	8	19	.15	[-.02, .32]	2.12	6.08	.078
Feedback features							
Verification + elaboration	10	25	.30	[.14, .45]	4.44	6.94	.003
Elaboration	12	42	.10	[.03, .16]	3.15	10.03	.010
Verification	14	43	.07	[-.06, .19]	1.15	11.44	.273
Post-feedback outcomes[b]							
Revision	10	36	.18	[.03, .32]	2.78	7.66	.025
Learning	14	54	.26	[.22, .31]	13.02	11.68	< .001
Task performance	35	148	.16	[.09, .23]	4.45	30.80	< .001
Post-feedback outcomes and provided feedback[c]							
Revision	5	15	.12	[-.01, .24]	2.82	3.28	.060
Learning	12	40	.31	[.24, .38]	9.58	9.97	< .001
Task performance	18	54	.25	[.14, .35]	4.71	14.53	< .001
Post-feedback outcomes and received feedback[b]							
Revision	8	21	.21	[.00, .39]	2.46	6.23	.047
Learning	7	14	.15	[-.02, .31]	2.25	5.45	.069
Task performance	27	94	.13	[.05, .20]	3.43	24.29	.002
Educational level							
Higher education	39	139	.21	[.14, .27]	6.34	37.0	< .001
K-12	14	102	.18	[.08, .27]	3.92	12.4	.002
Peer feedback role							
Provided	32	111	.29	[.22, .36]	8.01	27.3	< .001
Received	37	130	.13	[.06, .19]	3.79	31.4	< .001
Training							
No	22	101	.12	[.05, .20]	3.36	20.1	.003
Yes	32	140	.26	[.19, .32]	7.66	29.5	< .001
On-task support for feedback provision							
No	6	12	.08	[-.21, .36]	.69	4.84	.525
Yes	47	229	.22	[.16, .27]	8.09	44.42	< .001
On-task support for feedback reception							
No	38	163	.20	[.13, .27]	5.89	36.1	< .001
Yes	15	78	.20	[.12, .28]	5.21	13.4	< .001
On-task support for feedback reception (received feedback)							
No	28	93	.13	[.06, .20]	3.66	25.64	.001
Yes	9	37	.17	[.05, .27]	3.37	7.56	.011
Anonymity							
Anonymous	27	161	.21	[.14, .28]	5.66	25.0	< .001
Not available	13	30	.16	[.02, .30]	2.42	11.7	.033
Non-anonymous	13	50	.22	[.12, .31]	4.95	11.9	< .001

Note. Estimates and confidence intervals are based on Pearson correlations.
[a] Only one study was related to peer feedback accuracy; therefore, it was not included in the subgroup analysis.
[b] Only two studies were related to the course achievement; therefore, they were not included in the subgroup analysis.
[c] Only one study was related to the course achievement; therefore, it was not included in the subgroup analysis.

quality was significantly correlated with post-feedback outcomes in both K-12 education ($r = .18$, $p = .002$) and higher education ($r = .21$, $p < .001$; Table 4).

Students' *peer feedback role* (i.e., feedback provision or reception) had a significant influence on the magnitude of the relationship (Omnibus $F = 11.4$, $p = .001$). Specifically, the quality of peer feedback provided was more strongly associated with post-feedback outcomes ($r = .29$, $p < .001$) than the quality of feedback received ($r = .13$, $p < .001$).

Regarding instructional support in peer assessment (i.e., training, on-task support for feedback provision and reception, and anonymity), the only significant moderator was *training* (Omnibus $F = 7.12$, $p = .011$). The correlation between peer feedback quality and post-feedback outcomes was significantly higher when students received training before peer assessment ($r = .26$, $p < .001$), compared to when no training was provided ($r = .12$, $p = .003$).

On-task support *for feedback provision* was reported for most samples ($n = 47$ out of 53). While the correlation between peer feedback quality and post-feedback outcomes was statistically significant when this kind of support was available ($r = .22$, $p < .001$), the correlation became nonsignificant in the absence of this support ($r = .08$, $p = .525$). However, on-task scaffolding for feedback provision was not a significant moderator (Omnibus $F = 1.45$, $p = .273$). Similarly, *on-task* support *for feedback reception* did not significantly affect the correlations (Omnibus $F = .00$, $p = .968$) even when we restricted our analyses to the relationship between the quality of received feedback and post-feedback outcomes (Omnibus $F = .39$, $p = .541$).

Finally, a*nonymity* did not significantly moderate the correlations either (Table 5). Significant correlations were found between

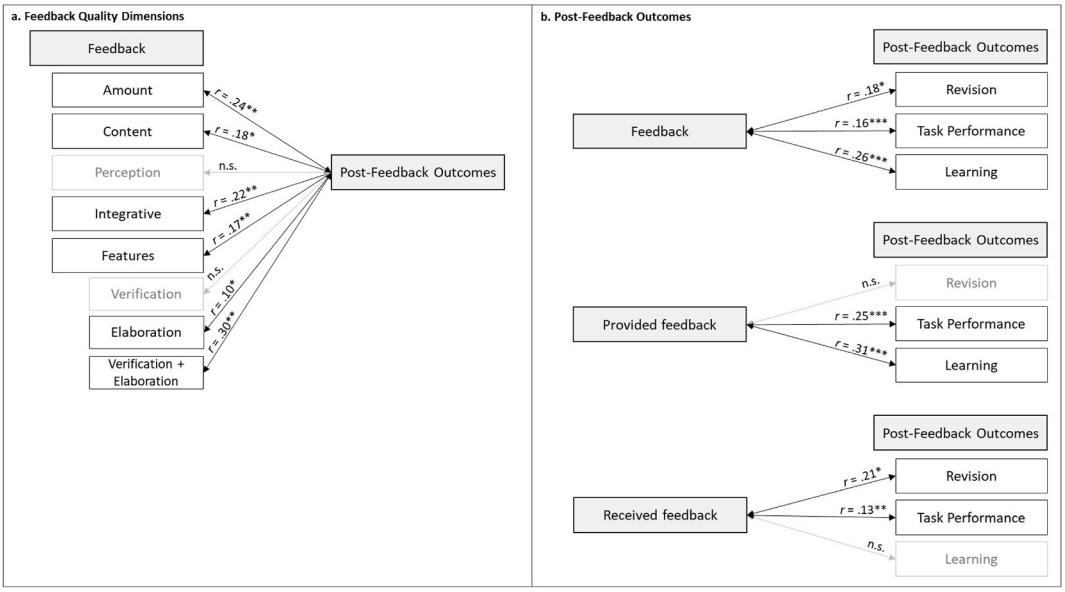


Fig. 4. Pearson Correlation Coefficients between Subcategories of Peer Feedback Quality and Post-Feedback Outcomes
Note. Gray arrows and n.s. indicate not statistically significant correlations.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

peer feedback quality and post-feedback outcomes under both anonymous ($r = .21$, $p < .001$) and non-anonymous ($r = .22$, $p < .001$) feedback conditions.

4. Discussion

In this meta-analysis, our primary goal was to synthesize existing research to determine the overall strength of the relationship between peer feedback quality and post-feedback outcomes. Additionally, we sought to explore whether this relationship varies depending on how key variables are defined and measured. Finally, we investigated a range of potential moderating factors (i.e., students' gender, educational level, the role they played in providing or receiving feedback, the presence of training, instructional support during feedback activities, and anonymity) to better understand under which conditions the correlation between peer feedback quality and outcome variables is most pronounced.

The meta-analysis revealed a significant positive correlation between peer feedback quality and post-feedback outcomes ($r = .20$). In other words, as the quality of peer feedback increases, students' subsequent outcomes (e.g., acting upon feedback they provided or received) also improve. Previous meta-analyses have demonstrated the overall positive effects of peer assessment on various learning outcomes based on experimental studies (Double et al., 2020; Li et al., 2020, 2021; Zhan et al., 2023; Zheng et al., 2020). However, the relationship between the qualitative peer feedback messages and their implementation, as well as their complex interplay with other feedback components (i.e., context and feedback agents) (Panadero & Lipnevich, 2022), had not yet been synthesized quantitatively. Contradictory findings in the literature, along with recent conceptualizations of peer feedback quality (e.g., Zhang et al., 2024a; Zhang & Schunn, 2023), have underscored the need for systematic reviews and meta-analyses to provide clearer conclusions about this relationship and to validate interpretations based on proposed quality dimensions. In this meta-analysis, we addressed these gaps by identifying positive correlations between peer feedback quality, measured in terms of amount, content, accuracy, features, or perceptions, and post-feedback outcomes (i.e., revision, task performance, learning, and course achievement). In addition to these empirical findings, this review proposes a model of peer feedback that may help explain why the quality-outcome relationship is not always direct. The model suggests that peer feedback quality operates through the message element, and that the relationship between the feedback message and implementation may change depending on their operationalizations, student characteristics, students' roles as main agents in this process, and contextual factors. This may account for the heterogeneity observed in the present findings and for the mixed results reported in the primary literature.

4.1. Effect sizes based on different operationalizations of peer feedback quality and post-feedback outcomes

The main variables of this meta-analysis (i.e., peer feedback quality and post-feedback outcomes) were categorized to examine how different operationalizations of these variables affect the correlations between them.

First, we investigated the correlations between *dimensions of peer feedback quality* and post-feedback outcomes. Although the correlations were not significantly affected by the feedback quality dimension, some noteworthy findings emerged. Feedback amount, content, features, and integrative measurements of feedback quality were all significantly correlated with post-feedback outcomes,

Table 5
Results of meta-regressions.

Variable	Omnibus *F*	Estimate	95% CI	*t*	*df*	*p*
Quality dimension	**.30 ($p = .875$)**					
Intercept (perception)		.15	[-.02, .32]	2.12	6.08	.078
Integrative		.06	[-.14, .26]	.71	10.93	.495
Amount		.09	[-.12, .30]	.95	11.11	.362
Content		.03	[-.18, .24]	.29	9.67	.776
Features		.02	[-.18, .21]	.21	10.24	.842
Feedback features	**3.74 ($p = .0504$)**					
Intercept (verification + elaboration)		.30	[.14, .45]	4.44	6.94	.003
Elaboration		−.21	[-.36, −.05]	−2.82	15.62	.013
Verification		−.24	[-.43, −.03]	−2.48	15.31	.025
Post-feedback outcomes[a]	**3.81 ($p = .042$)**					
Intercept (revision)		.18	[.03, .32]	2.78	7.66	.025
Learning		.09	[-.06, .23]	1.28	15.54	.220
Task performance		−.02	[-.19, .14]	−.31	11.73	.765
Post-feedback outcomes and provided feedback[b]	**.66 ($p = .425$)**					
Intercept (learning)		.30	[.23, .37]	9.03	10.0	< .001
Task performance		−.05	[-.18, .08]	−.81	22.1	.425
Post-feedback outcomes and received feedback[a]	**.37 ($p = .703$)**					
Intercept (revision)		.21	[.00, .39]	2.46	6.23	.048
Learning		−.06	[-.29, .18]	−.56	10.01	.591
Task performance		−.08	[-.29, .13]	−.89	9.29	.396
Gender						
Intercept		.16	[.00, .31]	2.17	12.5	.050
Percentage of females		.00	[-.002, .003]	.52	17.0	.613
Educational level	**.41 ($p = .528$)**					
Intercept (higher education)		.21	[.14, .27]	6.34	37.0	< .001
K12		−.36	[-.15, .08]	−.64	22.2	.528
Peer feedback role	**11.4 ($p = .001$)**					
Intercept (provided)		.29	[.22, .36]	8.01	27.3	<.001
Received		−.17	[-.27, −.07]	−3.37	48.0	.001
Training	**7.12 ($p = .011$)**					
Intercept (no)		.13	[.05, .20]	3.36	20.1	.003
Yes		.14	[.03, .23]	2.67	44.3	.011
On-task support for feedback provision	**1.45 ($p = .273$)**					
Intercept (no)		.08	[-.21, .36]	.69	4.84	.525
Yes		.14	[-.14, .40]	1.21	6.08	.273
On-task support for feedback reception	**.00 ($p = .968$)**					
Intercept (no)		.20	[.13, .27]	5.89	36.1	< .001
Yes		−.00	[-.11, .11]	−.04	23.7	.968
On-task support for feedback reception (received feedback)	**.39 ($p = .541$)**					
Intercept (no)		.13	[.06, .20]	3.66	25.6	.001
Yes		.04	[-.09, .17]	−.63	12.5	.541
Anonymity	**.28 ($p = .76$)**					
Intercept (anonymous)		.21	[.13, .28]	5.66	25.0	< .001
Not available		−.05	[-.21, .11]	−.65	24.2	.525
Non-anonymous		.01	[-.11, .13]	.17	24.7	.863

Note. Estimates and confidence intervals are based on Pearson correlations.
For categorical moderators, intercepts represent the estimated effect size for the reference category, and other coefficients represent differences relative to the reference category.

[a] The category of course achievement was not included in the moderator analysis as its *df* was lower than 4.

[b] The categories of revision and course achievement were not included in the moderator analysis as their dfs were lower than 4.

which supported the validity of conceptual frameworks suggesting the use of these dimensions as feedback quality indicators (e.g., Zhang et al., 2024a; Zhang & Schunn, 2023). However, student perceptions were not significantly correlated with post-feedback outcomes. The first four dimensions rely on the analyses of actual feedback messages, while perceptions are based on students' self-reports of perceived feedback quality. Noroozi et al. (2025) have shown that students' perceived learning from peer assessment activities does not necessarily correspond with their actual performance improvements. Inaccurate judgments of feedback quality, likely due to low competence in identifying the quality of peer feedback, may explain why perceived feedback quality did not significantly correlate with post-feedback outcomes.

Another key finding regarding the dimensions of peer feedback quality concerns feedback features. Feedback features as a moderator approached statistical significance ($p = .0504$). Moreover, measures that included both verification and elaboration components showed significantly stronger correlations with post-feedback outcomes than those including only verifications or elaborations. This may suggest that assessing peer feedback quality should account for both verification and elaboration to better predict performance and learning outcomes. This finding is consistent with Mason and Bruning's (2001) conclusion that effective feedback in computer-based instruction involves both components. Although elaboration provides more detailed information about

performance, and a higher proportion of elaboration is regarded as an indicator of higher feedback quality (Gielen & De Wever, 2015), verification can complement elaboration by offering concise and preliminary information.

Post-feedback outcomes were another primary variable in this meta-analysis, encompassing the categories of revision, task performance, learning, and course achievement. Peer feedback quality was significantly and positively correlated with all categories of post-feedback outcomes, but the strength of the relationship between peer feedback quality and post-feedback outcomes varied based on the type of outcome measures. Notably, the strongest correlation was found for the learning outcome. Additional analyses were conducted to examine the correlations with different post-feedback outcomes separately for provided and received feedback. The correlation coefficients shown for each outcome type in Fig. 4 implied a different pattern in the results for provided and received feedback. While the provided feedback was more strongly correlated with learning, the received feedback was more strongly correlated with revision. This may suggest that the benefits of peer feedback provision can become more evident in new transfer tasks, while peer feedback reception can contribute to taking action in the current tasks.

4.2. Effects of moderator variables

Regarding student characteristics, neither *gender* nor *educational level* significantly moderated the relationship between peer feedback quality and post-feedback outcomes in the present study. The finding that education level was no significant moderator aligns with previous meta-analyses on peer assessment (Li et al., 2021; Yan et al., 2022; Zhan et al., 2023). For gender, previous research (Almuadi & Barr, 2023; Noroozi et al., 2020, 2023; Ocampo et al., 2024) has favored female students in terms of peer feedback quality; however, there is conflicting evidence about the impact of gender on peer feedback-related performance improvements (Noroozi et al., 2020, 2023). In light of this, our meta-analysis contributes to the interpretation that gender does not necessarily significantly affect the correlation between peer feedback quality and post-feedback outcomes. However, further research on this issue is needed.

In addition to gender and educational level, we examined students' peer feedback role (i.e., feedback provision and reception) in our analyses. Peer feedback role significantly moderated the correlation between peer feedback quality and post-feedback outcomes. The correlation was significantly stronger for the provided feedback quality than for the received feedback quality. This suggests that producing high-quality peer feedback may be more effective in enhancing task performance and learning than receiving high-quality peer feedback. The stronger relationship between provided feedback quality and post-feedback outcomes aligns with empirical research comparing the learning benefits of peer feedback provision and reception (Chen et al., 2024; Cui & Schunn, 2024; Illana-Mahiques, 2021; Li et al., 2010). Moreover, it is consistent with our peer feedback model and discussions that frame feedback provision as a constructive learning activity (Wu & Schunn, 2023) and a problem-solving process that promotes higher-order cognitive engagement (Cho & MacArthur, 2011; van Popta et al., 2017).

For the received feedback, post-feedback outcomes likely depend on the peer feedback quality, student characteristics, and students' processing of peer feedback according to the feedback-student interaction model (Lipnevich & Smith, 2022) and our peer feedback model (Fig. 1). Student characteristics such as self-regulation, feedback literacy, and motivation can influence the extent to which feedback is acted upon (Winstone et al., 2017). Empirical studies (e.g., Bahr et al., 2025; Lipnevich et al., 2021) also indicated the significant effect of students' personality and cognitive abilities on the receptivity of instructional feedback. Furthermore, deep processing of peer feedback across different levels (i.e., cognitive, affective, and behavioral) may not be possible for all students. These complex interactions may also help explain the weaker relationship we found between the received peer feedback and post-feedback outcomes.

Among the variables concerning instructional support, *training* emerged as a significant moderator. In this meta-analysis, training referred to preparatory activities in which students were introduced to assessment criteria, example work or comments, or guidance on how to give or use peer feedback before conducting peer assessments. Studies that implemented training had a stronger correlation between peer feedback quality and post-feedback outcomes. Previous meta-analyses have similarly emphasized the importance of training as a moderator (Li et al., 2020; Zheng et al., 2020). These findings suggest that training strengthens the connection between feedback provision and reception and subsequent stages, such as revision and performance on future tasks.

In contrast to training, *on-task instructional* support *for feedback provision and reception* did not significantly moderate the relationship between peer feedback quality and post-feedback outcomes. In Hornstein et al.'s (2025) meta-analysis, instructional support for peer feedback provision that included both preparatory activities like training and other aids positively affected the quality of peer feedback provision and reception, but not subject-matter-related knowledge. On the other hand, in our meta-analysis, the moderating effect of training was analyzed separately from on-task instructional support, which enabled us to elucidate the role of instructional support introduced before and during peer assessment. We also involved different measures (e.g., revised task performance, performance in subsequent tasks or tests) as post-feedback outcomes in our meta-analysis. Finally, it is important to note that instructional support for peer feedback provision (e.g., with rubrics, prompts, and templates) was available in most studies reviewed in this meta-analysis (47 out of 53). The imbalance between categories of this variable may also have resulted in a nonsignificant moderating effect.

Concerning the nonsignificant moderating effect of instructional support for feedback reception, Hornstein et al. (2025) also found that this type of support did not influence feedback reception or subject-matter-related knowledge in their meta-analysis. The complexity of the feedback reception stage, as explained in the discussion part regarding the peer feedback role, may have led to a non-changing correlation between peer feedback quality and post-feedback outcome, even if this stage is supported.

Anonymity was also a nonsignificant moderator in this meta-analysis, consistent with prior meta-analyses examining the effects of peer assessment on various learning outcomes (e.g., Li et al., 2021; Zhan et al., 2023; Zheng et al., 2020). Although anonymity is recommended to decrease the psychosocial demands of peer assessment (e.g., Alemdag & Yildirim, 2022a; Vanderhoven et al., 2015),

the overall study results concerning its effect on peer assessment outcomes are conflicting (Panadero & Alqassab, 2019). Ensuring anonymity may not be sufficient to alter students' overall perceptions or beliefs about peer assessment (e.g., Li, 2017). The reduced accountability due to anonymity may also prevent learners from providing more effective feedback and from utilizing the feedback they receive (Panadero & Alqassab, 2019). Moreover, most reviewed studies included one kind of instructional support (e.g., training and on-task support for peer feedback provision), which may have influenced the moderating effect of anonymity.

4.3. Limitations and future directions

Although this meta-analysis clarifies the overall relationship between peer feedback quality and post-feedback outcomes and identifies important moderators, some limitations need to be considered. The first limitation is that we investigated the variables of interest at the person level rather than the comment level. Specifically, we included only studies that analyzed peer feedback at the provider or receiver level by examining the relationship between the quality of feedback a student provided or received and post-feedback outcomes. However, the unit of analysis could also be a single comment, allowing researchers to examine receptivity at the comment level. To provide more fine-grained analyses, future meta-analyses on peer feedback quality may examine comment-level data.

Second, there were imbalances in the availability of data on certain variables, which may have influenced moderator analyses and limited the scope for additional analyses. For example, most studies involved participants in higher education, focused on written assignments, and provided no support for feedback reception. Future studies could explore peer assessment for different tasks (e.g., problem-solving, presentations) in K-12 settings, while integrating support to enhance students' processing of peer feedback during the feedback reception stage.

Third, regarding feedback quality dimensions, only one study included in this meta-analysis examined overall peer feedback accuracy as a standalone factor. Assessing the accuracy of feedback content requires more expert knowledge of the learning task and evaluation criteria. In contrast, other dimensions of peer feedback quality (e.g., features or amount) can be coded by trained raters without subject domain knowledge or analyzed automatically. Moreover, assessing feedback accuracy often requires an additional level of analysis once feedback is segmented into more basic units of quality (e.g., suggestive or explanatory comments, as in feedback features), resulting in integrative measures that account for more than one quality dimension. Such methodological requirements may explain why few studies have examined this dimension. However, we should note that the accuracy of peer feedback may be more strongly correlated with performance improvement and learning because, when measuring other feedback dimensions (e.g., amount or features), inaccurate information that can mislead students may also be included. Hence, future research should focus more on measuring the overall content accuracy of qualitative feedback.

Fourth, in our additional analyses, we subdivided feedback features into verification, elaboration, and a combination of both. These categories could be expanded for more detailed classification. For example, Gielen and De Wever's (2015) coding scheme distinguishes verifications as positive, negative, or neutral, and elaborations as informative or suggestive. As more studies report correlations between these subcategories and post-feedback outcomes, more nuanced analyses of feedback features may be conducted in future meta-analyses.

Fifth, we analyzed two variables (i.e., gender and educational level) concerning student characteristics as potential moderators. We assume that other important variables may also moderate the correlation between feedback quality and post-feedback outcomes, as indicated in our peer feedback model (Fig. 1). Students' initial task knowledge and skills are one of them. For example, to decipher and apply complex peer feedback to revise their initial task performance, students need relevant task knowledge and skills. Consequently, the influence of a specific feedback quality indicator may depend on its adaptivity or how well it fits the receiver's readiness or needs. Because primary studies rarely report standardized baseline metrics, we were unable to model this interaction in our meta-analysis. Hence, it is recommended for future primary studies to explicitly consider and report the role of this student-level characteristic in their analyses. Future meta-analyses are also needed to examine whether initial task knowledge and skills, as well as other student characteristics, moderate the relationship between peer feedback and post-feedback outcomes.

Sixth, it is important to note that this meta-analysis is based on correlations; thus, no causal conclusions can be drawn. Future meta-analyses synthesizing intervention studies that manipulated peer feedback quality would be needed to investigate its effect on post-feedback outcomes.

Our final suggestion concerns the design of peer feedback activities in future empirical studies. In bidirectional peer assessment, students take the roles of both feedback providers and receivers, as outlined in our peer feedback model. While it is often implemented in educational settings, students' experiences in the feedback provider role (e.g., the cognitive demand of generating feedback for a peer) may influence learners' expectations and behavioral responses when transitioning to the receiver role (e.g., processing and implementing peer feedback regarding their own work; Alemdag & Yildirim, 2022b; McConlogue, 2015). Hence, future primary research studies need to experimentally isolate or statistically control for the interaction effects between the two roles and the potential effects of their temporal order. In such research designs, the effects of feedback provision or feedback reception across different outcomes (i.e., revision, performance, learning, and course achievement) can be investigated more rigorously.

4.4. Theoretical implications

This paper proposes a comprehensive yet parsimonious model of peer feedback that describes multiple dimensions of peer feedback quality and post-feedback outcomes, outlines the key processes involved in peer feedback activities, and considers student roles and characteristics and contextual factors (Fig. 1). Most importantly, it illustrates the interactions among different feedback agents,

including peers, the self, and the guiding educator. By serving multiple functions, our peer feedback model addresses the need for more interactional feedback models, as highlighted by Panadero and Lipnevich (2022).

The findings of this meta-analysis also provide evidence about when peer feedback may be more strongly linked to post-feedback outcomes. The feedback features measured as feedback quality indicators, the types of post-feedback outcomes, the role of the student as a feedback agent (i.e., feedback provider and receiver), and training as instructional support among contextual factors may change this relationship based on our synthesis of current empirical research. These results highlight the importance of considering not only the operationalization of the feedback message and feedback implementation but also the interplay of the feedback agent and contextual factors in shaping the relationship between them, although our meta-analysis remains limited in explaining the role of student characteristics.

4.5. Practical implications

Effectively implementing peer assessment in classrooms presents challenges for students and educators, given the multiple feedback components (e.g., feedback message and implementation) to consider (Panadero & Lipnevich, 2022). This meta-analysis highlights the positive relationship between peer feedback quality and post-feedback task performance and learning across different quality dimensions (e.g., amount, content, and features). Practitioners may therefore focus on ensuring that students both provide and receive high-quality feedback. Specifically, educators can take into account multiple dimensions of peer-feedback quality when assessing the quality of the feedback comments students produce and when explaining to students how to provide effective feedback. Students can also be encouraged to prepare feedback that includes both evaluative statements (e.g., positive and negative verifications) and elaborations (e.g., justifications and concrete suggestions for improvement), as feedback that contained both components appeared to be more strongly associated with better post-feedback outcomes in this meta-analysis.

Furthermore, given the stronger relationship between provided feedback quality and post-feedback outcomes, educators are encouraged to assign students the role of feedback providers during peer assessment activities. Additionally, due to the fact that training significantly moderates the relationship between peer feedback quality and outcomes, it is recommended that educators offer preparatory training to familiarize students with assessment criteria, examples of high-quality work, and effective strategies for providing and receiving feedback.

Finally, the significant and positive correlations between feedback quality dimensions and post-feedback outcomes identified in this meta-analysis can inform the development of online peer assessment systems powered by artificial intelligence (AI). Such systems could automatically analyze peer feedback based on established quality metrics and offer immediate feedback on the feedback itself (Topping et al., 2025).

5. Conclusion

Previous research offers multiple ways to define and measure peer feedback quality, leading to inconsistencies across studies. Additionally, mixed results about how peer feedback quality relates to post-feedback outcomes raise critical questions about what truly defines quality feedback and how it influences students' application of feedback in later tasks.

In this meta-analysis, we conceptualized peer feedback quality through five dimensions of amount, content, accuracy, features, and perceptions, drawing from recent conceptual frameworks. We also proposed a model of peer feedback that integrates agents, message, implementation, student characteristics, and context to account for the mechanisms through which peer feedback quality shapes post-feedback outcomes and to explain the potential reasons for the heterogeneity observed across studies. We synthesized data from 49 studies, encompassing 53 independent samples and 241 effect sizes, to examine the relationship between these dimensions of feedback quality and outcomes such as revision, performance, and learning. We identified a significant positive overall correlation ($r = .20$), with differences depending on specific feedback characteristics and outcome measures.

Notably, our moderator analyses showed that the quality of feedback students provided had a stronger association with improved post-feedback outcomes than the quality of feedback they received. Furthermore, the presence of training prior to peer assessment amplified this effect. These findings highlight the importance of carefully defining feedback quality, preparing students through training, and emphasizing their active role as feedback providers to maximize the benefits of peer assessment.

In sum, this study confirms that high-quality peer feedback positively correlates with students' subsequent performance and learning. The results have important implications for educators aiming to improve peer assessment practices. Specifically, educators may prioritize preparatory training before peer assessment activities and actively assign students the role of feedback providers, as these conditions most consistently strengthen the quality-outcome relationship. The findings also carry implications for the development of AI-driven systems capable of automatically analyzing peer feedback across multiple quality dimensions and offering real-time guidance, extending the reach of high-quality peer assessment beyond what is feasible through manual instructor review alone.

CRediT author statement

Ecenaz Alemdag: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing.

Anastasiya Lipnevich: Methodology, Writing – review and editing, Supervision.

Lisa Bardach: Methodology, Writing – review and editing, Supervision.

Marlit Annalena Lindner: Conceptualization, Methodology, Writing – review and editing, Supervision.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the first author used ChatGPT for language editing to improve the readability of the manuscript. After the use of AI, the authors reviewed the adapted text thoroughly and take full responsibility for the content of the published article.

Funding

This research was supported by departmental funding from the joint Chair of Digital Education at Europa-Universität Flensburg (EUF) and the Leibniz Institute for Science and Mathematics Education (IPN), Kiel, Germany. Lisa Bardach is supported by research funding No P-EDU-23-28, which is co-funded by the European Union (the project "Breakthrough in Educational Research" No 10-044-P-0001) under the 1st April 2025 Agreement with the Research Council of Lithuania (RCOL, LMTLT) and the 14th April 2025 Joint Activity Agreement with Kaunas University of Technology.

Declaration of competing interests

We have no conflicts of interest to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.edurev.2026.100848.

Data availability

Data will be made available on request.

References

Alemdag, E., & Yildirim, Z. (2022a). Design and development of an online formative peer assessment environment with instructional scaffolds. *Educational Technology Research and Development, 70*, 1359–1389. https://doi.org/10.1007/s11423-022-10115-x

Alemdag, E., & Yildirim, Z. (2022b). Effectiveness of online regulation scaffolds on peer feedback provision and uptake: A mixed methods study. *Computers & Education, 188*, Article 104574. https://doi.org/10.1016/j.compedu.2022.104574

Almuadi, M., & Barr, M. (2023). Exploring gender disparities in online peer feedback system: An analysis of feedback levels and quality. In Z. Kubincová, T. Hao, N. Capuano, M. Temperini, S. Ge, Y. Mu, P. Fantozzi, & J. Yang (Eds.), *International symposium on emerging technologies for education* (pp. 232–248). Singapore: Springer Nature Singapore.

Alqassab, M., Strijbos, J. W., Panadero, E., Ruiz, J. F., Warrens, M., & To, J. (2023). A systematic review of peer assessment design elements. *Educational Psychology Review, 35*(1). https://doi.org/10.1007/s10648-023-09723-7

Alqassab, M., Strijbos, J. W., & Ufer, S. (2018). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education, 33*(1), 11–30. https://doi.org/10.1007/s10212-017-0342-0

Alqassab, M., Strijbos, J. W., & Ufer, S. (2019). Preservice mathematics teachers' beliefs about peer feedback, perceptions of their peer feedback message, and emotions as predictors of peer feedback accuracy and comprehension of the learning task. *Assessment & Evaluation in Higher Education, 44*(1), 139–154. https://doi.org/10.1080/02602938.2018.1485012

Andrade, H. L. (2013). Classroom assessment in the context of learning theory and research. In J. H. McMillan (Ed.), *Sage handbook of research on classroom learning* (pp. 17–34). California: Sage Publications.

Bahr, J. L., Höft, L., Lipnevich, A., Meyer, J., & Jansen, T. (2025). Exploring students' receptivity to feedback: A latent profile analysis. *Assessment in Education: Principles, Policy & Practice*, 1–19. https://doi.org/10.1080/0969594X.2025.2467676

Boetje, J., & van de Schoot, R. (2024). The SAFE procedure: A practical stopping heuristic for active learning-based screening in systematic reviews and meta-analyses. *Systematic Reviews, 13*(1). https://doi.org/10.1186/s13643-024-02502-7

Borenstein, M. (2024). Avoiding common mistakes in meta-analysis: Understanding the distinct roles of Q, i-squared, tau-squared, and the prediction interval in reporting heterogeneity. *Research Synthesis Methods, 15*(2), 354–368. https://doi.org/10.1002/jrsm.1678

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

Bowman, N. A. (2012). Effect sizes and statistical methods for meta-analysis in higher education. *Research in Higher Education, 53*(3), 375–382. https://doi.org/10.1007/s11162-011-9232-5.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Chen, F., Zhang, S., Liu, Q., Yu, S., Li, X., & Zheng, X. (2024). Supporting learning performance improvement: Role of online group assessment. *Education and Information Technologies*, 1–26. https://doi.org/10.1007/s10639-024-12907-5

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology, 103*(1), 73–84. https://doi.org/10.1037/a0021950

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

Costley, J., Southam, A., Boitcov, M., Zhang, H., Shulgina, G., Baldwin, M., & Fanguy, M. (2025). Creative self-efficacy and its relationship to peer feedback. *Innovations in Education & Teaching International*, 1–14.

Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods and Applications, 19*, 497–515. https://doi.org/10.1007/s10260-010-0142-z

Cui, Y., & Schunn, C. D. (2024). Peer feedback that consistently supports learning to write and read: Providing comments on meaning-level issues. *Assessment & Evaluation in Higher Education*, 1–14. https://doi.org/10.1080/02602938.2024.2364025

Dmoshinskaia, N., Gijlers, H., & de Jong, T. (2022). Does learning from giving feedback depend on the product being reviewed: Concept maps or answers to test questions? *Journal of Science Education and Technology, 31*(2), 166–176.

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

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

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*(1), 70–120. https://doi.org/10.3102/0034654312474350

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology, 43*(6), 543–549. https://doi.org/10.1016/0895-4356(90)90158-L

Fisher, Z., Tipton, E., & Zhipeng, H. (2023). Robumeta: Robust variance meta-regression. *R package version 2.1*. https://CRAN.R-project.org/package=robumeta.

Fleckney, P., Thompson, J., & Vaz-Serra, P. (2025). Designing effective peer assessment processes in higher education: A systematic review. *Higher Education Research and Development, 44*(2), 1–16. https://doi.org/10.1080/07294360.2024.2407083

Gao, X., Noroozi, O., Gulikers, J., Biemans, H. J., & Banihashem, S. K. (2024). A systematic review of the key components of online peer feedback practices in higher education. *Educational Research Review, 42*. https://doi.org/10.1016/j.edurev.2023.100588, 100588.

Gielen, M., & De Wever, B. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior, 52*, 315–325. https://doi.org/10.1016/j.chb.2015.06.019

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*(4), 304–315. https://doi.org/10.1016/j.learninstruc.2009.08.007

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

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

He, W., & Gao, Y. (2023). Explicating peer feedback quality and its impact on feedback implementation in EFL writing. *Frontiers in Psychology, 14*, Article 1177094. https://doi.org/10.3389/fpsyg.2023.1177094

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ (Clinical Research ed.), 327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hornstein, J., Keller, M. V., Greisel, M., Dresel, M., & Kollar, I. (2025). Enhancing the peer-feedback process through instructional support: A meta-analysis. *Educational Psychology Review*. https://doi.org/10.1007/s10648-025-10017-3

Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2017). Peer feedback on college students' writing: exploring the relation between students' ability match, feedback quality and essay performance. *Higher Education Research and Development, 36*(7), 1433–1447. https://doi.org/10.1080/07294360.2017.1325854

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 & Evaluation in Higher Education, 43*(6), 955–968. https://doi.org/10.1080/02602938.2018.1424318

Illana-Mahiques, E. (2021). Re-thinking peer reviewing in the virtual context: The roles of giving and receiving online feedback in L2 Spanish classrooms. *Languages, 6*(3). https://doi.org/10.3390/languages6030151, 151.

Jin, X., Jiang, Q., Xiong, W., Feng, Y., & Zhao, W. (2024). Effects of student engagement in peer feedback on writing performance in higher education. *Interactive Learning Environments, 32*(1), 128–143. https://doi.org/10.1080/10494820.2022.2081209

Kerman, N. T., Banihashem, S. K., Karami, M., Er, E., Van Ginkel, S., & Noroozi, O. (2024). Online peer feedback in higher education: A synthesis of the literature. *Education and Information Technologies, 29*(1), 763–813. https://doi.org/10.1007/s10639-023-12273-8

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 133*(1), 159–174. https://doi.org/10.2307/2529310

Li, H., Bialo, J. A., Xiong, Y., Hunter, C. V., & Guo, X. (2021). The effect of peer assessment on non-cognitive outcomes: A meta-analysis. *Applied Measurement in Education, 34*(3), 179–203. https://doi.org/10.1080/08957347.2021.1933980

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193–211. https://doi.org/10.1080/02602938.2019.1620679

Li, L. (2017). The role of anonymity in peer assessment. *Assessment & Evaluation in Higher Education, 42*(4), 645–656. https://doi.org/10.1080/02602938.2016.1174766

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525–536. https://doi.org/10.1111/j.1467-8535.2009.00968.x. https://bera-journals.onlinelibrary.wiley.com/doi/10.1111/j.1467-8535.2009.00968.x. (Accessed 15 April 2010)

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 & Education, 116*, 81–92. https://doi.org/10.1016/j.compedu.2017.08.010

Lipnevich, A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education, 6*, Article 720195. https://doi.org/10.3389/feduc.2021.720195

Lipnevich, A. A., Gjicali, K., Asil, M., & Smith, J. K. (2021). Development of a measure of receptivity to instructional feedback and examination of its links to personality. *Personality and Individual Differences, 169*, Article 110086. https://doi.org/10.1016/j.paid.2020.110086

Lipnevich, A. A., & Smith, J. K. (2022). Student–feedback interaction model: Revised. *Studies In Educational Evaluation, 75*, Article 101208. https://doi.org/10.1016/j.stueduc.2022.101208

Liu, X., & Li, L. (2014). Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment. *Assessment & Evaluation in Higher Education, 39*(3), 275–292. https://doi.org/10.1080/02602938.2013.823540

Liu, Y., Groen, M. A., & Cain, K. (2024). The association between morphological awareness and reading comprehension in children: A systematic review and meta-analysis. *Educational Research Review, 42*, Article 100571. https://doi.org/10.1016/j.edurev.2023.100571

Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science, 40*, 257–275. https://doi.org/10.1007/s11251-011-9177-2

Lu, Q., Yao, Y., & Zhu, X. (2023). The relationship between peer feedback features and revision sources mediated by feedback acceptance: The effect on undergraduate students' writing performance. *Assessing Writing, 56*, Article 100725. https://doi.org/10.1016/j.asw.2023.100725

Lu, Q., Zhu, X., & Cheong, C. M. (2021). Understanding the difference between self-feedback and peer feedback: A comparative study of their effects on undergraduate students' writing improvement. *Frontiers in Psychology, 12*, Article 739962. https://doi.org/10.3389/fpsyg.2021.739962

Ludemann, P. M., & McMakin, D. (2014). Perceived Helpfulness of Peer Editing activities: First-year students' views and writing performance outcomes. *Psychology Learning and Teaching, 13*(2), 129–136. https://doi.org/10.2304/plat.2014.13.2.129

Mason, B. J., & Bruning, R. (2001). Providing feedback in computer-based instruction: What the research tells us. *Center for instructional innovation*. University of Nebraska–Lincoln. Retrieved March 5, 2025, from http://dwb.unl.edu/Edit/MB/MasonBruning.html.

McConlogue, T. (2015). Making judgements: Investigating the process of composing and receiving peer feedback. *Studies in Higher Education, 40*(9), 1495–1506. https://doi.org/10.1080/03075079.2013.868878

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Qureshi, R., Mattis, P., Lisy, K., & Mu, P.-F. (2017). Chapter 7: Systematic reviews of etiology and risk. In E. Aromataris, & Z. Munn (Eds.), *Joanna Briggs Institute reviewer's manual*. The Joanna Briggs Institute. https://doi.org/10.46658/JBIMES-24-06.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. Van Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–144). New York, NY: Lawrence Erlbaum Associates.

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning. *Digital Education Review, 23*(1), 7–26.
Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*, 375–401. https://doi.org/10.1007/s11251-008-9053-x
Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education, 46*(5), 756–778. https://doi.org/10.1080/02602938.2020.1823314
Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090
Noroozi, O., Alqassab, M., Taghizadeh Kerman, N., Banihashem, S. K., & Panadero, E. (2025). Does perception mean learning? Insights from an online peer feedback setting. *Assessment & Evaluation in Higher Education, 50*(1), 83–97. https://doi.org/10.1080/02602938.2024.2345669
Noroozi, O., Banihashem, S. K., Taghizadeh Kerman, N., Parvaneh Akhteh Khaneh, M., Babaee, M., Ashrafi, H., & Biemans, H. J. (2023). Gender differences in students' argumentative essay writing, peer review performance and uptake in online learning environments. *Interactive Learning Environments, 31*(10), 6302–6316. https://doi.org/10.1080/10494820.2022.2034887
Noroozi, O., Hatami, J., Bayat, A., Van Ginkel, S., Biemans, H. J., & Mulder, M. (2020). Students' online argumentative peer feedback, essay writing, and content learning: Does gender matter? *Interactive Learning Environments, 28*(6), 698–712. https://doi.org/10.1080/10494820.2018.1543200
Novakovich, J., & Long, E. C. (2013). Digital performance learning: Utilizing a course weblog for mediating communication. *Journal of Educational Technology & Society, 16*(4), 231–241.
Ocampo, J. C., Panadero, E., Zamorano, D., Sánchez-Iglesias, I., & Diez Ruiz, F. (2024). The effects of gender and training on peer feedback characteristics. *Assessment & Evaluation in Higher Education, 49*(4), 539–555. https://doi.org/10.1080/02602938.2023.2286432
Page, M. J., Moher, D., 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., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ (Clinical Research ed.), 372*, Article n160. https://doi.org/10.1136/bmj.n160
Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education, 44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1600186
Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education, 48*(8), 1053–1075. https://doi.org/10.1080/02602938.2023.2164884
Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, Article 100416. https://doi.org/10.1016/j.edurev.2021.100416
Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098–1120. https://doi.org/10.1037/edu0000103
Pearson, K. (1907). *On further methods of determining correlation*. London: Cambridge University Press.
Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175–181. https://doi.org/10.1037/0021-9010.90.1.175
Pustejovsky, J. (2024). _clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections_. *R package version 0.5.11*. https://CRAN.Rproject.org/package=clubSandwich.
Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing, 25*(4), 553–581. https://doi.org/10.1177/0265532208094276
Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*(1), 60–78. https://doi.org/10.1002/jrsm.1095
Strijbos, J. W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*(4), 291–303. https://doi.org/10.1016/j.learninstruc.2009.08.008
Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology, 2*, 85–112. https://doi.org/10.1007/s40865-016-0026-5
Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods, 5*(1), 13–30. https://doi.org/10.1002/jrsm.1091
Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics, 40*(6), 604–634. https://doi.org/10.3102/1076998615606099
Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*(1), 20–27. https://doi.org/10.1080/00405840802577569
Topping, K. J., Gehringer, E., Khosravi, H., Gudipati, S., Jadhav, K., & Susarla, S. (2025). Enhancing peer assessment with artificial intelligence. *International Journal of Educational Technology in Higher Education, 22*(1). https://doi.org/10.1186/s41239-024-00501-1, 3.
van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence, 3*(2), 125–133. https://doi.org/10.1038/s42256-020-00287-7
Van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review, 20*, 24–34. https://doi.org/10.1016/j.edurev.2016.10.003
Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education, 81*, 123–132. https://doi.org/10.1016/j.compedu.2014.10.001
Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*(2), 112–125. https://doi.org/10.1002/jrsm.11
Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538
Wu, H., Zhu, H., & Yang, X. (2022). The effects of providing peer feedback on learners' genre awareness in English as a foreign language business letter writing. *Frontiers in Psychology, 13*, Article 1059555. https://doi.org/10.3389/fpsyg.2022.1059555. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1059555/full. (Accessed 1 December 2022)
Wu, Y., & Schunn, C. D. (2020). When peers agree, do students listen? The central role of feedback quality and feedback frequency in determining uptake of feedback. *Contemporary Educational Psychology, 62*, Article 101897. https://doi.org/10.1016/j.cedpsych.2020.101897
Wu, Y., & Schunn, C. D. (2021a). From plans to actions: A process model for why feedback features influence feedback implementation. *Instructional Science, 49*(3), 365–394. https://doi.org/10.1007/s11251-021-09546-5
Wu, Y., & Schunn, C. D. (2021b). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal, 58*(3), 492–526. https://doi.org/10.3102/0002831220945266
Wu, Y., & Schunn, C. D. (2023). Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback. *Contemporary Educational Psychology, 73*, Article 102160. https://doi.org/10.1016/j.cedpsych.2023.102160
Yan, Z., Lao, H., Panadero, E., & Fernández-Castilla, B. (2026). A meta-analytical review of the effects of self-assessment and peer-assessment interventions on self-regulated learning strategies and affective outcomes. *Assessment in Education: Principles, Policy & Practice*, 1–39. https://doi.org/10.1080/0969594X.2026.2681630
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*, Article 100484. https://doi.org/10.1016/j.edurev.2022.100484
Yu, F. Y., & Wu, C. P. (2016). Predictive effects of the quality of online peer-feedback provided and received on primary school students' quality of question-generation. *Journal of Educational Technology & Society, 19*(3), 234–246.

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

Zhang, H., Costley, J., Courtney, M., Shulgina, G., & Fanguy, M. (2025). The impact of different peer feedback types on student academic writing performance from dyadic and individual analyses. *Education and Information Technologies, 30*, 6339–6366. https://doi.org/10.1007/s10639-024-13032-z

Zhang, Y., & Schunn, C. D. (2023). Self-regulation of peer feedback quality aspects through different dimensions of experience within prior peer feedback assignments. *Contemporary Educational Psychology, 74*, Article 102210. https://doi.org/10.1016/j.cedpsych.2023.102210

Zhang, Y., Schunn, C. D., & Wu, Y. (2024a). What does it mean to be good at peer reviewing? A multidimensional scaling and cluster analysis study of behavioral indicators of peer feedback literacy. *International Journal of Educational Technology in Higher Education, 21*(1). https://doi.org/10.1186/s41239-024-00458-1, 26.

Zhang, Y., Schunn, C. D., & Wu, Y. (2024b). Interconnecting peer feedback literacy: Exploring the relationship between providing and acting on peer feedback. *Studies In Educational Evaluation, 83*, Article 101411. https://doi.org/10.1016/j.stueduc.2024.101411

Zheng, L., Zhang, X., & Cui, P. (2020). The role of technology-facilitated peer assessment and supporting strategies: A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(3), 372–386. https://doi.org/10.1080/02602938.2019.1644603

Zong, Z., Schunn, C. D., & Wang, Y. (2021). What aspects of online peer feedback robustly predict growth in students' task performance? *Computers in Human Behavior, 124*, Article 106924. https://doi.org/10.1016/j.chb.2021.106924