European Journal of Psychology of Education (2026) 41:90
https://doi.org/10.1007/s10212-026-01148-4

Assessor versus assessee: a scoping review of the impact of peer assessment roles on cognitive, intrapersonal, and interpersonal outcomes

David Zamorano[1] · **Maryam Alqassab**[2,3] · **Ernesto Panadero**[1,4,5]

Received: 26 May 2025 / Revised: 7 May 2026 / Accepted: 27 May 2026 /
Published online: 23 June 2026
© The Author(s) 2026

Abstract
Peer assessment can shape students' learning across cognitive, intrapersonal, and interpersonal outcomes. When students participate in peer assessment, they can take on two distinct roles, as assessors or as assessees, each of which may shape their learning differently. However, only recently, there has been interest in comparing how each role contributes to students' learning. This comparison comes with a challenge: students usually play both roles reciprocally, making it difficult to disentangle the effect of each. We conducted a scoping review to examine the methodological approaches, and the evidence generated from peer assessment studies comparing the effects of the assessor and assessee roles on cognitive, intrapersonal, and interpersonal learning outcomes. Our review identified 19 studies, most conducted in higher education contexts. They compared the differential impact on cognitive ($n=9$), intrapersonal ($n=9$), and interpersonal ($n=1$) outcomes. Among the studies focusing on cognitive outcomes, a subset used designs that specifically isolated each role, yet even these produced inconsistent results: in some studies students acting as assessors outperformed assessees, while in others both roles led to similar gains. In contrast, research focusing on intrapersonal and interpersonal outcomes is largely exploratory, though it offers a useful foundation for developing more focused designs aimed at specific variables. We propose a set of guidelines for future research aimed at systematically disentangling the impact of the assessor and assessee roles on the three aforementioned outcomes.

Keywords Peer assessment · Roles · Assessor · Assessee · Cognitive outcomes · Intrapersonal outcomes · Interpersonal outcomes · Scoping review

Introduction

During peer assessment, students can evaluate peers' work and receive evaluations in return (Black & Wiliam, 1998a, 1998b; Topping, 2017). This process has been shown to influence a variety of learning outcomes including cognitive, intrapersonal, and interpersonal. For example, students who engage in peer assessment show a significant increase in academic performance, writing quality, and problem-solving — cognitive outcomes (Double

Extended author information available on the last page of the article

et al., 2020; Li et al., 2020; Yan et al., 2022). But learning is not only about performance. Peer assessment also changes how students think and feel about themselves, their perceived learning, their emotions and comfort — intrapersonal outcomes (Panadero, 2016; Panadero et al., 2023; van Gennip et al., 2009). And because peer assessment is inherently social, it shapes how students see and treat each other, the trust they place in their peers or even their friendship — interpersonal outcomes (Panadero et al., 2023). However, the peer assessment process involves multiple phases during which students can play two distinct roles as assessors or assessees, each potentially resulting in different cognitive and affective mechanisms (Bauer et al., 2023; Kollar & Fischer, 2010).

Previous studies suggested that acting as an assessor or assessee can lead to different learning outcomes. For example, Lundstrom and Baker (2009) showed that engaging in the assessor role may contribute more to improvements in students' task performance or writing compared to when the students engage as an assessee. Some evidence also suggests a differential role impact on intrapersonal and interpersonal outcomes. For example, Zhan (2024) found that when students played the role of assessees, they experience more discomfort compared to acting as assessors. Nevertheless, disentangling the impact of each peer assessment role is not straightforward in practice. In most classrooms, students naturally assess and are assessed by peers, which makes it difficult to determine which experience drove learning. To address this gap, the present scoping review systematically examined the empirical studies comparing the effects of peer assessment roles on cognitive, intrapersonal, and interpersonal outcomes. Specifically, we synthesised both the methodological characteristics and the findings of studies that have sought to disentangle the impact of each role in order to evaluate the nature of current empirical evidence on this topic.

The two roles of peer assessment: assessor and assessee

During peer assessment, students evaluate the work of their peers by providing grades, qualitative comments, or both, and in turn may receive evaluations of their own work from peers (Alqassab & Panadero, 2020). Since the assessor and assessee roles involve distinct processes with different cognitive and emotional demands, several theoretical frameworks and models of peer assessment have explicitly distinguished between them. For example, Kollar and Fischer (2010) and Bauer et al. (2023) emphasized that each role activates unique cognitive mechanisms. When students are assessors, they are expected to interpret and evaluate peers' work, often identifying strengths, weaknesses, or errors and articulating this in the form of feedback or scores (Bauer et al., 2023). This process involves active cognitive engagement, including error detection, application of criteria, and potentially metacognitive reflection. In contrast, when they are assessees, they interpret the feedback they receive, judging its relevance, and deciding how (or whether) to incorporate it into their own work (Aben et al., 2024; Bauer et al., 2023). Additionally, Aben et al., (2019, 2024) addressed how intrapersonal and interpersonal processes differ during peer assessment when students act as assessors or assessees. Different processes, then, may contribute differently to learning outcomes.

Cognitive, intrapersonal and interpersonal outcomes

Part of students' learning from peer assessment is understood through its effect on students' cognitive, intrapersonal and interpersonal outcomes. By considering these three dimensions, we can better understand the complexity of peer assessment: a practice that

can shape students' academic performance while also shaping what they think and feel about themselves and their social experiences (Panadero et al., 2023). Cognitive outcomes refer to gains in knowledge, skills, or task performance from peer assessment, often measured through writing quality, problem-solving accuracy, or exam scores. These outcomes have received the most attention in the experimental research and are the focus of several meta-analyses, which consistently show moderate positive effects of peer assessment on academic performance (Double et al., 2020; Li et al., 2020; Yan et al., 2022). Some studies have also identified moderators, such as online implementation or assessor training, that may enhance these effects (Li et al., 2020).

Intrapersonal outcomes capture how students think and feel about themselves and the task, as a result of peer assessment. There is a wide variety of these outcomes, some examples include perceived learning, self-efficacy, confidence when giving or receiving feedback, and emotional responses such as anxiety or enjoyment. There is a growing interest in studying these outcomes as they reflect students' engagement with the process and their self-perceptions as learners (Panadero et al., 2023; van Gennip et al., 2009).

Meanwhile, interpersonal outcomes are concerned with how peer assessment shapes relationships among students, including the trust they place in their peers, sense of fairness, psychological safety, and social connectedness. Although interpersonal outcomes are the least frequently studied in peer assessment research (Panadero et al., 2023), they are crucial to comprehend how peer assessment shapes social interactions and influence students' future willingness to give honest evaluations or to act on the evaluation received (Panadero, 2016; van Gennip et al., 2009).

Despite the growing interest in studying the impact of peer assessment across all three outcomes: what students learn, how they feel, and how they relate to peers (Double et al., 2020; Li et al., 2020; Panadero et al., 2023; Yan et al., 2022), how each peer assessment role differently contributes to those effects remains unclear. A student who evaluates a peer's writing may develop different skills than when they receive the evaluation from their peer (Lundstrom & Baker, 2009). Li et al.'s (2020) meta-analysis revealed that, students acting as assessors compared students acting as assessees had similar results on cognitive outcomes. However, these results need careful interpretation. To reach this conclusion, Li and colleagues compared imbalanced effect sizes (reciprocal peer assessment $k = 108$; assessor-only $k = 20$ and assessee-only $k = 6$). Moreover, assessor-only and assessee-only students came from different studies, so any difference between them could reflect differences in study context rather than peer assessment role. Beyond the cognitive outcomes, a systematic comparison of the effects of peer assessment on intrapersonal and interpersonal outcomes is limited. Although some studies have suggested that students tend to feel more comfortable as assessors than as assessees, perceive greater learning benefits from the assessor role, and experienced more distrust toward peers when on the receiving end of evaluation (Ion et al., 2019; Su, 2023; Zhan, 2024) no integrative synthesis has yet examined intrapersonal or interpersonal outcomes across peer assessment roles. One of the limitations of previous literature that compared the impact of roles on different outcomes —including those in Li et al.'s (2020) meta-analysis, is that the studies' samples acted as both assessors and assessees within the same peer assessment activity. Accordingly, it remains unclear what the existing studies truly reveal about the impact of each peer assessment role on different learning outcomes. There is a need for a systematic examination of the nature of the existing studies that examined peer assessment role effects across cognitive, intrapersonal, and interpersonal outcomes.

Aim and research questions

The aim of this review is to analyze the empirical research that compared the effects of the assessor and assessee roles on cognitive, intrapersonal, and interpersonal outcomes. In doing so we aim to highlight methodological strengths and gaps, clarify the current evidence of peer assessment roles effects, and provide suggestions to guide future work on how peer assessment roles can support learning. We investigated two research questions (RQ):

RQ1 What are the methodological characteristics of studies comparing the effects of the assessor and assessee roles on cognitive, intrapersonal and interpersonal outcomes?

RQ2 What are the effects of the assessor and assessee roles on cognitive, intrapersonal and interpersonal outcomes?

Method

Search strategies

The search was conducted over two phases (in June 2022 and in September 2024), following the PRISMA model (Page et al., 2021) and using two databases: PsycINFO and ERIC. The following keyword combinations were used: TI ("peer-assessment" OR "peer-feedback") AND AB ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* OR assessee*"). Additional keyword combinations were made with synonyms, found in the reproducible searches for all databases in Appendix A. No time limit was set. The second literature search conducted in phase 2 was done to update the search done in phase 1. In phase 1, following Page et al. (2021), we conducted additional search strategies to ensure a thorough exploration of the existing literature. These strategies were: manual screening of the references of the included articles, including recommendations from the authors of the selected articles, expert recommendations, and screening of personal library records. The inclusion criteria used in this review were: a) the study was conducted in a formal educational context (primary, secondary, higher education or academia), b) the study is empirical (experiments, quasi-experiments, pre-experimental), c) the study was published in English, d) the study compared differences between assessors and assessees in cognitive or intrapersonal or interpersonal peer assessment outcomes. Accordingly, studies involving reciprocal peer assessment were included, but only where the outcomes for each role were separately discussed. The search from both phases resulted in 230 studies; the literature search process is shown in Fig. 1.

Selection

After removing duplicates, the first and the second author independently screened the title and abstract of 10% ($n = 21$) of the articles, reaching an acceptable inter-rater agreement ($\kappa = 0.772$) in the first round. Discrepancies were discussed, and the process was repeated in a second round, resulting in full agreement (100%); then, the first author screened the remaining studies ($n = 166$). After excluding the studies that did not meet the inclusion

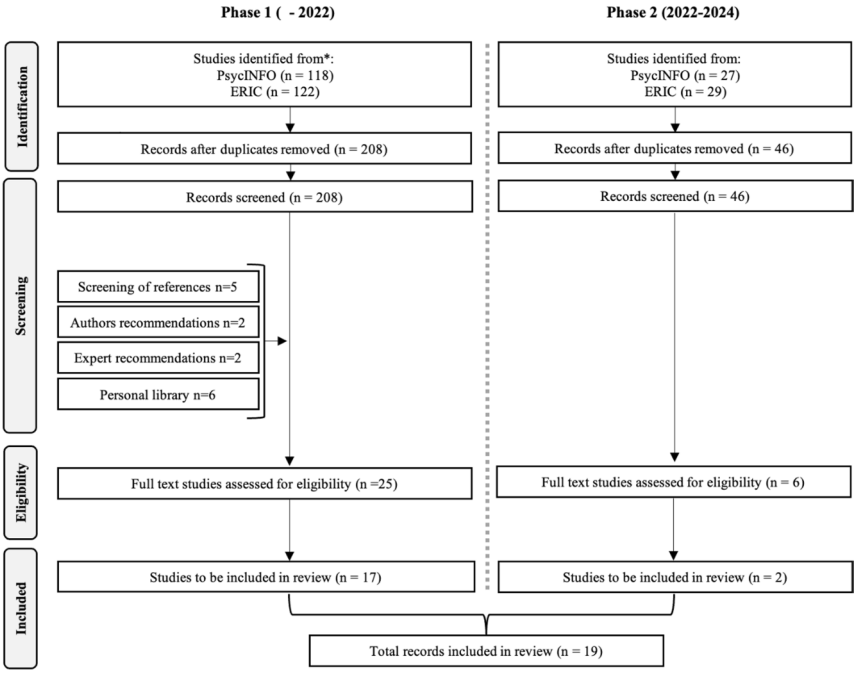

Fig. 1 PRISMA flow diagram. Literature search and inclusion process

criteria, a total of 25 studies were included. The full texts of the 25 studies were read and assessed for inclusion by the first author, 17 were included. In phase 2, 56 records were identified through the databases search, and after removing duplicates, 46 records remained. After screening, six were selected to have a full-text read to be assessed for inclusion, 2 were included. In total, 19 studies from both Phase 1 and 2 were included in the final review.

Coding procedure

Coding the characteristics of studies, treatment of roles and the quality of study

For all the included 19 studies, we used part of the coding scheme of peer assessment design by Alqassab et al. (2023) only incorporating data from study details (e.g., aim, demographic, design, main results). Additionally, we coded for directionality and quality of outcome measures. Directionality was included to better understand how they investigated and analyzed them (e.g., whether students engaged in both roles or were assigned a single role or engaged in both assessing and being assessed). Quality of outcome measures were analyzed in detail: the instruments used, their reported reliability, the methods for analyzing qualitative data, and the theoretical frameworks guiding their interpretation, depending on the student's approach. This provides a specific detail on the constructs measured, which is especially important considering the review's emphasis on outcomes. The detailed coding scheme can be found in Supplementary Material A, and the information extracted for the full dataset can be found in Supplementary Material B.

Additionally, we evaluated the overall quality of the included studies using an adapted version of a rubric previously applied in Panadero et al. (2023) that assesses each study quality in terms of empirical design, pedagogical quality, dependent variable/s and quality of the sample (see Appendix B). For each of these quality criteria, the study was quantitatively rated as weak (0 points), medium (1 point), or robust (2 points). A total score (ranging between 0–12) was computed by summing all points of each criterion representing the overall quality of the study.

Coding reliability

To establish a reliable extraction of the study details and the quality of the included studies, the first and the second authors independently coded 15% of the included articles ($n=3$). When computing an inter-rater reliability index Krippendorff's α has been shown to produce inflated results with skewed distributions, for example when two raters have consistent agreement in a category (Feinstein & Cicchetti, 1990). This was the case in our study, specifically for some categories of the coding scheme on which raters agreed repeatedly "Type of study" or "Aims". Therefore, following Alqassab et al. (2023), we report both the percentage agreement and Krippendorff's α to offer transparent measures. Due to the reduced number of studies scanned through the two inter-rater rounds (three articles per round), we established a threshold of 100% agreement for each category. As indicated in Appendix C, on the second round, we achieved full agreement on all study categories. Then, the first author coded the remaining articles ($n=16$).

Data analysis

We employed a scoping review methodology to thoroughly explore how peer assessment roles, whether as assessor or assessee, have been examined in relation to cognitive, intrapersonal, and interpersonal outcomes. This approach was chosen to help identify central concepts, types of evidence, and existing research gaps (Levac et al., 2010; Tricco et al., 2016). While our search process was systematic and transparent, the inherent flexibility of the scoping review allowed for the inclusion of diverse study designs without necessitating a formal quality appraisal. Instead, we evaluated and reported the quality of the included studies as part of our findings to inform future research (e.g., Misiejuk & Wasson, 2021; Van der Kleij & Lipnevich, 2021). To evaluate how studies isolate assessor-versus-assessee effects, we examined five methodological indicators, identified via our coding framework and quality rubric, that we considered essential for establishing clear role-specific effects: the study's overall quality score; the research design (e.g., experimental, quasi-experimental, pre-experimental); methodological approach (e.g., quantitative, qualitative, mixed-methods); directionality (e.g., whether students served in both roles or were assigned to distinct conditions); and quality of outcome measures (e.g., instrument used, reliability reports, previously used). We organized the results based on each outcome type, cognitive and intrapersonal and interpersonal outcomes for RQ1 and RQ2.

Results

General characteristics of the included studies

The sample sizes of the included studies ranged from 15 to 188 participants (See Table 1). The studies encompass students from primary (5.3%), secondary (15.8%), higher (52.6%),

vocational (21.1%) and postgraduate (5.3%) education. The studies were conducted in various subject domains; the most common being education (21.1%) and L2 writing (15.8%), the remaining varied: humanities, education, engineering, writing and technology (see Table 1). Regarding the type of study, the most common were quantitative (52.6%), followed by mixed methods (26.3%), and lastly purely qualitative (21.1%).

RQ1 What are the methodological characteristics of studies exploring the assessor and assessee roles effects on cognitive, intrapersonal and interpersonal outcomes?

Starting with an overview of the outcomes variables explored, nine (47.4%) out of the included studies explored cognitive peer assessment outcomes. All of them used task performance as their outcome of interest (see Table 2), this includes traditional outcomes such as test performance, and/or essay writing but also practical skills that were rated with a score (Double et al., 2020). Nine (47.4%) out of the included studies explored the differential impact of peer assessment roles on intrapersonal outcomes. The intrapersonal outcomes studied by the reviewed studies varied: six studies explored perceived learning, one emotion (enjoyment) and two comfort. Only one study (5.2%) explored interpersonal outcomes (see Table 3) and focused on trust.

Regarding cognitive outcomes, the mean study quality score rubric for these studies was low: 5.33 out of 12 ($SD = 1.50$). All of the studies had a quantitative approach. From these studies, 55.6% (five studies) had a quasi-experimental design (Çevik, 2015; Crinon & Marin, 2010; Huisman et al., 2018; Lee, 2015; Lundstrom & Baker, 2009). In these

Table 1 Description of sample size, educational level, subject domain, and type of study

Study	Sample	Ed. Level	Subject domain	Type of study
Lee (2015)	128	Higher Education	Entrepreneur management	Quantitative
Crinon & Marin (2010)	68	Primary education	Scientific writing	Quantitative
Lundstrom & Baker (2009)	91	Vocational Education	English writing	Quantitative
Huisman et al. (2018)	83	Higher Education	Education	Quantitative
Çevik. (2015)	70	Higher Education	Education	Quantitative
Wu & Schunn (2020)	185	Secondary Education	Language and Composition	Quantitative
Lu & Zhang (2012)	181	Secondary Education	Humanities	Quantitative
Lu & Law (2012)	181	Secondary Education	Humanities	Quantitative
Zhang et al. (2021)	27	Higher Education	Educational technology	Quantitative
Cao et al. (2019)	15	Higher Education	L2 writing	Qualitative
Martin & Sippel (2021)	120	Vocational Education	L2 speaking	Mixed methods
Li & Grion (2019)	40	Higher Education	Education	Qualitative
Su & Huang (2022)	32	Vocational Education	L2 writing	Mixed methods
Gukas et al. (2008)	42	Higher Education	Medicine	Quantitative
Ertmer et al. (2007)	15	Vocational Education	Technology	Mixed methods
Nicol et al. (2014)	82	Higher Education	Engineering	Mixed methods
Ion et al. (2019)	188	Higher Education	Education	Quantitative
Zhan (2024)	70	Higher Education	General Education	Qualitative
Su (2023)	40	Postgraduate Education	L2 writing	Mixed methods

Table 2 Studies on the role effect on cognitive outcomes

Study	Design	Variable of interest	Directionality	Results
Lee (2015)	Quasi-experimental	Task performance (Quality of business plan)	Condition 1: Unidirectional: only assessor. Condition 2: Unidirectional only assessee	On the second round of revision, students in the assessor's group significantly improved their tasks compared to the assessee group
Crinon & Marin (2010)	Quasi-experimental	Task performance (Quality written draft)	Condition 1: Unidirectional: only assessor. Condition 2: Unidirectional: only assessee	The assessor's condition task contained significantly more relevant information than the one of the assessee group
Lu & Zhangl. (2012)	Pre-experimental [Cross-sectional]	Task performance (Final project scores)	Bidirectional	Both assessing and being assessed predicted the final project scores. More specifically, providing cognitive comments and receiving praise
Lu & Law (2012)	Pre-experimental [Cross-sectional]	Task performance (Liberal studies project performance)	Bidirectional	Providing suggestions to peers and providing feedback on identifying problems (assessor) and positive affective feedback from peers (assessee) predicted the final project performance
Çevik (2015)	Experimental	Problem solving performance	Condition 1: Unidirectional: only assessor. Condition 2: Unidirectional: only assessee	Both assessors and assessee improved their problem-solving skills. However, assessors significantly outperformed assessee in justifying their problem-solving with reasoning
Lundstrom & Baker (2009)	Quasi-experimental; Pretest–post-test	Task performance (Quality of writing)	Condition 1: Unidirectional: only assessor. Condition 2: only assessee	The assessor showed significantly greater improvements in specific writing areas than the assessee, particularly the lower (beginning) assessor group

Table 2 (continued)

Study	Design	Variable of interest	Directionality	Results
Zhang et al. (2021)	Pre-experimental [Longitudinal]	Creativity performance on task	Bidirectional	The assessors' conditions appear to be more beneficial than being assessed by a peer
Huisman et al. (2018)	Experimental	Task performance (Writing performance)	Condition 1: Unidirectional: only assessor. Condition 2: only assessee	Students acting as assessors improved their essays to a similar degree as students in the assessee role
Wu & Schunn (2020)	Pre-experimental [Cross-sectional]	Writing performance	Bidirectional	Both assessor and assessee roles positively influenced document improvements. The role of assessee especially influenced "low-level" improvements

studies, each student was assigned to one condition (assessor or assessee), measures were gathered before and/or after their assignment to the role condition and were later compared, which allowed the isolation of the effect of the role. The remaining four studies (44.4%) were pre-experimental and followed a quantitative approach (Lu & Law, 2012; Lu & Zhang, 2012; Wu & Schunn, 2020; Zhang et al., 2021). In these studies, all students performed both roles. The effects of roles on cognitive outcomes were therefore inferred from assessment-related indicators associated with acting as an assessor or an assessee. For example, cognitive outcomes were predicted based on the number and characteristics of comments provided (for the assessor) or received (for the assessee). Under this logic, if the number of comments received predicted task performance, this was interpreted as evidence that the assessee role was more beneficial. Compared with designs that isolate each role in separate groups, this approach requires greater inferential assumptions and limits the identification of role-specific effects.

Regarding quality of outcome measures, four of the nine studies (44.4%) used a rubric to score a task, followed an inter-rater agreement process and reported reliability indices: Cohen's kappa (Çevik, 2015), intra-class correlation (Lu & Zhang, 2012), FACET (Lundstrom & Baker, 2009) and Cronbach's alpha (Zhang et al., 2021). Lu and Law (2012) employed a rubric and Lee (2015) a list with the main criteria, but did not follow an inter-rater agreement process. Huisman et al. (2018) specified the percentage of each main criteria (e.g., content (30%) and structure (20%) (…)) but did not specify the criteria or the use of a rubric to score the task, however they performed an inter-rater agreement. Wu and Schunn (2020) did not score a task but counted the number of implementations relating the comments received with the second draft of the task, in this process they computed an inter-rater agreement. Similarly, Crinon & Marin (2010) counted the document improvements with the inclusion of relevant information but did not report an inter-rater agreement.

With respect to intrapersonal outcomes, the mean study quality score of the nine studies exploring intrapersonal outcomes studies was 4.20 out of 12 ($SD = 1.83$). From the studies that examined intrapersonal outcomes, two (22.2%) used a quasi-experimental design but operationalized peer assessment roles differently. While Martin and Sippel (2021) used a between-subject design that assigned students to two distinct conditions, assessor or assessee, Su and Huang (2022) implemented a within-subjects design alternating the order of the roles, one group assessed before being assessed and vice versa. Both designs allowed for the isolation of role effects while maintaining participation in both roles. These two studies focused on perceived learning. The remaining seven studies (77.7%) studying intrapersonal outcomes were pre-experimental studies with a quantitative (Gukas et al., 2008; Ion et al., 2019), **qualitative** (Cao et al., 2019; Crinon & Marin, 2010; Su & Huang, 2022) or mixed-method approach (Ertmer et al., 2007; Nicol et al., 2014). These studies explored perceived learning (four studies), emotions (enjoyment) (one study) and comfort (2 studies). All of these studies researched both roles similarly: participants engaged in both roles, and their perceptions were collected only after the entire process of peer assessment.

Regarding quality of outcome measures, compared to cognitive outcomes, studies drew on multiple sources of information. The instruments employed include closed and open questionnaires, interviews or semi-structured interviews and written reflection of feedback process. From the studies that used closed questionnaires, seven used rating scales, specifically Likert scales. Among them, five (Ertmer et al., 2007; Gukas et al., 2008; Ion et al., 2019; Nicol et al., 2014; Su, 2023), used ad-hoc rating scales but did not report reliability indices. In contrast, two studies also used ad-hoc scales but did report reliability, specifically, Cronbach's α indexes that indicated high reliability $\alpha = 0.84$ (Martin & Sippel, 2021), $\alpha = 0.87$ and $\alpha = 0.84$ for two scales used by Su and Huang (2022). For qualitative data

Table 3 Studies on the role effect on intrapersonal and interpersonal outcomes

Study	Design	Variable of interest	Directionality	Results
Cao et al. (2019)	Pre-experimental [Case study]	Perceptions of learning (assessing and being assessed)	Bidirectional	Of the 15 interviewees, seven perceived benefits in both roles, four preferred receiving feedback, and three providing feedback
Martin & Sippel (2021)	Quasi-experimental	Beliefs about peer corrective feedback; Learners' experiences on peer assessment roles	Condition 1: Unidirectional: only assessor. Condition 2: only assessee	Quantitative results showed a significant positive shift in beliefs about peer (corrective) feedback, but only in the assessor condition. Qualitative analysis shows no noteworthy changes over time between groups
Li & Grion (2019)	Pre-experimental [Cross-sectional]	Perceptions of learning (assessing and being assessed)	Bidirectional	Qualitative analysis showed that students reported that both roles promoted their learning in the project. However, more students emphasized giving feedback as a process for learning
Su & Huang (2022)	Experimental	Perceptions of enjoyment (assessing and being assessed)	Bidirectional	Quantitative analysis showed that students enrolled in the assessor role were significantly favored in terms of enjoyment. The qualitative analysis aligns with the results
Gukas et al. (2008)	Pre-experimental [Cross-sectional]	Perceptions of comfort (assessing and being assessed)	Bidirectional	Descriptive analysis of a questionnaire showed that most students felt more comfortable being assessed than when assessing
Ertmer et al. (2007)	Pre-experimental [Longitudinal]	Perceptions of learning (assessing and being assessed)	Bidirectional	Quantitative analysis showed that students perceive both roles contribute to their learning. Qualitative analysis shows students reference more how providing improves their learning

Table 3 (continued)

Study	Design	Variable of interest	Directionality	Results
Nicol et al. (2014)	Pre-experimental [Cross-sectional]	Perceptions of learning and benefits from assessing and being assessed by a peer	Bidirectional	Quantitative descriptive analysis revealed mixed perceptions, though qualitative data suggested reviewing was more beneficial than receiving
Ion et al. (2019)	Pre-experimental [Cross-sectional]	Perceptions of learning (assessing and being assessed)	Bidirectional	Quantitative analysis showed students perceived greater learning benefits as assessors than as assessees
Su (2023)	Pre-experimental [Longitudinal]	Perceptions of comfort (assessing and being assessed)	Bidirectional	Quantitative analysis showed that students felt significantly more comfortable giving feedback than receiving it
*Zhan (2024)	Pre-experimental [Cross-sectional]	Perceptions of peer assessment	Bidirectional	Participants encountered cognitive challenges when acting as assessors and socio-affective challenges as assessees, notably: distrust of peers

*The only study to examine interpersonal outcomes; Martin and Sippel (2021) has more conditions one with teacher feedback and a control

(analyzed qualitatively), such as open questionnaires, semi-structured interviews, and retrospective journals, all studies used a deductive thematic analysis approach (Gukas et al., 2008; Li & Grion, 2019; Nicol et al., 2014; Su, 2023). Three studies integrated theoretical models/framework into their analysis (Cao et al., 2019; Ertmer et al., 2007; Martin & Sippel, 2021).

The only study focusing on interpersonal outcomes, which focused on trust was a pre-experimental, qualitative investigation. This study had a quality score of 5 out of 12. In this study, students participated in both roles and their perceptions were collected only after the entire process (Zhan, 2024). This study employed a qualitative approach by analyzing retrospective journals using deductive thematic analysis to identify emergent themes (Zhan, 2024).

RQ2 What are the effects of the assessor and assessee roles on cognitive, intrapersonal and interpersonal outcomes?

Our second goal was to present the main insights gained from examining the 19 studies on the effects of peer assessment roles on outcomes. As in the previous section, we organize our results by distinguishing between the three main outcomes, cognitive and intrapersonal and interpersonal outcomes. The latter two are grouped together because only one study explored interpersonal outcomes.

All of the nine studies that focused on peer assessment cognitive outcomes, used task performance as their outcome of interest (see Table 2), this includes traditional outcomes such as test performance, and/or essay writing but also practical skills that were rated with a score (Double et al., 2020). Four studies out of these studies, found the assessor role had greater impact on task performance, while five studies found that both roles led to similar task performance. Considering the study quality score, studies that found the assessor role had greater effect on task performance had a mean quality score of 5.25 out of 12 ($SD = 1.70$) and similarly studies that found no role effect had a mean score of 5.40 out of 12 ($SD = 1.73$).

From the studies that explored the impact of peer assessment roles on intrapersonal factors, specifically the six that explored perceived learning, three found that students perceived greater learning when engaging in the role of assessor. These studies had a mean study quality score of 3 out of 12 ($SD = 1$). The other three studies on perceived learning, found students perceived similar learning from both roles. These studies had a mean study quality score of 4.33 out of 12 ($SD = 0.57$).

Among the remaining studies that explored the effect of peer assessment roles on intrapersonal factors, one studied one type of emotions (enjoyment) and two studied comfort. Regarding enjoyment, Su and Huang (2022) found the assessor role was generally enjoyed by more students than the assessee role. The quality score of this study was 8 out of 12. As for comfort, Su (2023) found that students reported higher levels of comfort when assessing compared to being assessed by peers. This study had a quality score of 5 out of 12. Gukas et al. (2008), with a quality score of 2 out of 12, compared each role using descriptive analysis and found similar levels of comfort while receiving 81% and 72% when giving.

Lastly, the only study on the impact of peer assessment roles on interpersonal factors, Zhan (2024) found that when acting as assessees, students experienced more socio-affective challenges, one particularly prominent being their distrust of peers. This study had a quality score of 5 out of 12.

Discussion

In this scoping review, we examined the existing body of literature about how the roles of assessor and assessee differently contribute to students' learning from peer assessment, focusing on cognitive, intrapersonal, and interpersonal outcomes. Following an analysis of the methodological quality of the existing studies we assess the strength of research evidence on the differential impact of peer assessment roles. Our synthesis identified methodological inconsistencies and mixed findings making it difficult to draw clear conclusions about the differential effects of peer assessment roles. In what follows, we reflect on these patterns and outline directions for future research.

RQ1 Methodological characteristics of studies comparing peer assessment roles

Our review revealed that cognitive and intrapersonal outcomes received considerable attention, each addressed in 47.4% of the reviewed studies. This reflects broader trends in the peer assessment literature, where cognitive outcomes, typically measured by task performance, have been consistently prioritized (Double et al., 2020; Li et al., 2020; Yan et al., 2022). Conversely, only one study covered interpersonal outcomes (Zhan, 2024). Research interest in intrapersonal and interpersonal dimensions is growing, however, in the studies that compare roles, they still remain underrepresented, especially interpersonal outcomes (Panadero et al., 2023).

Despite the apparent attention to intrapersonal outcomes, among the nine studies addressing them, six focused specifically on students' perceptions of learning. Only two of them addressed comfort and one examined emotional response (enjoyment). The strong emphasis on perceived learning aligns with a trend to treat such perceptions as proxies for actual learning (Panadero et al., 2018). This assumption is questionable and evidence from peer assessment research shows that perceived learning does not necessarily align with actual performance (Noroozi et al., 2024). Consequently, with researchers keeping a narrow focus on perceived learning, the field risks overlooking other intrapersonal and interpersonal outcomes that are equally relevant to peer assessment and are far more underexplored (e.g., emotion regulation, trust, or psychological safety).

The overall quality of the studies was clearly low. Studies examining cognitive outcome studies scored highest with a mean score close to 5 (out of 12), followed by studies on interpersonal outcomes (5.0) and intrapersonal outcomes, with a score of 4. These scores highlight persistent weaknesses in study design and quality of outcome measures. Design-wise, a substantial proportion of studies employed non-experimental approaches: 44.4% in the cognitive outcome group (4 out of 9), 77.8% in the intrapersonal group (7 out of 9), and 100% in the single interpersonal study. Most pre-experimental designs involved students participating in both roles and collecting outcome data only after the full peer assessment cycle. Such designs are limited in their ability to isolate role-specific effects. This is particularly problematic given that disentangling these effects was a main goal of the studies reviewed. That said, a subset of studies adopted more rigorous designs that allowed for role isolation. For cognitive outcomes, five studies (55.6%) used quasi-experimental or experimental designs where participants were assigned to either assessor or assessee conditions (e.g., Huisman et al., 2018; Lundstrom & Baker, 2009). For intrapersonal outcomes, two studies (22.2%) implemented either a quasi-experimental between-subjects design (Martin & Sippel, 2021) or an experimental within-subjects design with counterbalancing (Su &

Huang, 2022). These designs represent meaningful efforts to improve internal validity and offer more robust insights into peer assessment role effects.

Regarding the quality of outcome measures, we identified three critical patterns. First, for studies that examined cognitive outcomes, four of the nine studies (44.4%) used rubrics and reported an inter-rater reliability to score task performance, typically with robust indices such as Cohen's kappa (e.g., Çevik, 2015; Lundstrom & Baker, 2009). Although these practices were used by less than half of the studies, they reflect a positive increasing attention to construct validity, an essential concern in performance-based assessments (Messick, 1995; Perlman, 2003). Second, for intrapersonal outcomes, seven out of nine studies used closed-ended questionnaires, all based on ad-hoc instruments. However, only two of these reported reliability indices (e.g., Cronbach's α in Martin & Sippel, 2021; Su & Huang, 2022), raising concerns about psychometric robustness. The absence of validated instruments undermines the interpretability of findings and limits comparability across studies. Third, for both intrapersonal and interpersonal outcomes, all qualitative studies employed deductive thematic analysis. Yet only three of these studies explicitly anchored their coding schemes in theoretical frameworks (e.g., Cao et al., 2019; Ertmer et al., 2007; Martin & Sippel, 2021). This lack of theoretical grounding has been noted in previous reviews as a key limitation of the field (Panadero & Alqassab, 2019; Panadero et al., 2023). Without a coherent conceptual base, qualitative findings risk becoming descriptive rather than explanatory, and their contribution to theory-building remains limited.

The current evidence comparing the effects of peer assessment roles seems to be methodologically inconsistent. While some studies — particularly those using well-constructed experimental designs and validated tools — provide promising efforts, the field requires more rigorous standards in both research designs and outcomes measurement. This is especially urgent for studies investigating intrapersonal and interpersonal outcomes of peer assessment, where the variety of constructs and instruments demands greater precision and consistency.

RQ2 What are the effects of peer assessment roles on

Cognitive outcomes

A subset of studies compared the effects of the assessor and assessee roles with sufficient rigor (e.g., Çevik, 2015; Huisman et al., 2018; Lundstrom & Baker, 2009). Although one study, Lundstrom and Baker (2009), reported significant advantages for students acting as assessors, the other two studies found no significant differences between peer assessment roles (e.g., Çevik, 2015; Huisman et al., 2018), aligning with the findings of Li et al. (2020), whose meta-analysis reported no overall effect. Taken together, this body of evidence suggests that neither role can be favored over the other, and consequently reciprocal peer assessment should not be dismissed. In contrast, one possible explanation for the inconsistent results might be related to how peer assessment activities were designed in these studies. Specifically, the peer assessment design decisions researchers and teachers make including anonymity or formative or summative approach of peer assessment can moderate how each peer assessment role plays out, (Bauer et al., 2023). For example, it has been suggested that non-anonymous peer assessment may intensify social pressure and promote more superficial engagement in the assessor role, potentially resulting in lower-quality and less critical feedback (Panadero & Alqassab, 2019). Additionally, assessors on formative non-anonymous dialogic peer assessment did not perceive it as socially

challenging, possibly maintaining feedback quality (e.g., Fraile et al., 2024). The limited number of studies identified in this review does not allow us to explore how these design decisions moderate peer assessment role effects, warranting further investigation.

The small number of rigorous role-comparison studies provides a starting point for future hypothesis testing (e.g., Çevik, 2015; Huisman et al., 2018; Lundstrom & Baker, 2009). To support such research, we reported sample sizes, means, standard deviations, and calculated effect sizes when significant effects on cognitive outcomes were found (see Supplementary Material B). This information can be used to justify sample sizes in future studies and to interpret results based on previously observed effect sizes from significant findings (Lakens, 2022).

A separate group of studies were descriptive studies. Rather than assigning students to roles, they quantified the peer feedback students provided and received, its amount, type, and quality, and used these as predictors of subsequent performance (e.g., Wu & Schunn, 2020). When a characteristic associated with the assessor role predicted outcomes (e.g., amount of feedback provided), they interpreted this as evidence that the assessor role was more beneficial, and vice versa. While these studies can provide a useful exploration of peer assessment roles, they fail to establish causal relationships.

Intrapersonal and interpersonal outcomes

Among the nine studies addressing intrapersonal outcomes, six focused on perceived learning. While the majority found some perceived benefit in both roles, two studies warrant particular attention due to their more rigorous methodology. Martin and Sippel (2021) found that students reported greater perceived learning after acting as assessors. In contrast, Cao et al. (2019) found no clear preference or advantage for either role in their qualitative case study. These diverging results highlight the need for more systematic research on perceived learning and other intrapersonal outcomes of peer assessment. Beyond perceived learning, only one study examined emotional response (i.e., enjoyment) and two studies focused on comfort. Su and Huang (2022) found that students generally enjoyed the assessor role more than the assessee role. Similarly, Su (2023) reported that students felt significantly more comfortable when assessing the work of their peers than when being assessed. These findings suggest that the assessor role may offer a more emotionally secure or enjoyable experience. Among the studies addressing intrapersonal outcomes, Su's (2023) stands out for its methodological care in differentiating roles and measuring role-specific outcomes. Lastly, only one study examined interpersonal outcomes (Zhan, 2024) in which socio-affective challenges were explored revealing that students acting as assessees experienced more interpersonal discomfort, which came from skepticism toward their peers' competence.

While still limited and largely exploratory, these studies offer a foundation for hypothesis generation on interpersonal and intrapersonal peer assessment outcomes. Descriptive studies of peer assessment processes are crucial for building theory and informing future experimental work (Strijbos & Sluijsmans, 2010). Constructs such as perceived learning, emotional response, or trust -all influenced by the social dynamics of assessment- deserve deeper empirical attention using stronger designs.

Limitations

This review presents several methodological limitations inherent to the current state of the literature. First, the included studies displayed substantial heterogeneity in terms of

research design, outcome variables, and measurement instruments, which limits the possibility of cross-study comparisons and the generalizability of findings. Second, as is typical in scoping reviews, we did not conduct a formal risk of bias assessment or statistical synthesis, meaning that no conclusions could be drawn regarding the magnitude or consistency of the effects observed. Third, many of the studies, particularly those focusing on intrapersonal and interpersonal outcomes, relied on pre-experimental designs or non-validated instruments, which undermines the strength of the evidence on these outcomes and raises concerns about construct validity. Fourth, although a systematic and transparent search strategy was employed, the literature search was limited to two databases. These databases, PsycINFO and ERIC, were selected due to their comprehensive coverage of educational psychology and peer assessment research. To minimize the risk of missing relevant studies, multiple supplementary search strategies were applied in line with PRISMA recommendations. Fifth, most included studies were conducted in higher education contexts, limiting generalizability to younger learners. Future research should examine whether the patterns identified in these studies hold at K-12 educational levels. Finally, while this review applied a systematic and comprehensive search strategy, the diversity of terminology in peer assessment research may have led to the non-identification of relevant studies that explored role-related effects adopting different conceptual labels.

Recommendations for future research

Future peer assessment studies comparing how assessors and assessees impact cognitive outcomes should pay attention to both design and instruments. Studies would benefit from research designs that isolate each role (e.g., (quasi-)experimental designs). Particularly, given the lack of controlled designs identified in this review where comparing two groups — assessor and assessee — is the main aim. Naturally, non-experimental studies like descriptive or qualitative studies are also necessary for theory building and identifying processes that need attention, as shown in this review (Cao et al., 2019; Wu & Schunn, 2020). Regarding instruments, given that cognitive outcomes are typically operationalized as task performance, a minimum methodological requirement should be the use of scoring rubrics combined with inter-rater reliability procedures and clearly reported indices (see Jonsson and Svingby (2007) for practical guidance).

When it comes to comparing peer assessment roles, studying intrapersonal and interpersonal outcomes might require different design decisions than when studying cognitive outcomes. While we have addressed (quasi-)experimental designs as the gold standard for cognitive outcomes, this might not be realistic for intrapersonal and interpersonal outcomes, particularly in face-to-face naturalistic settings where the social nature of peer assessment is more evident. In those cases, it would be important to prioritize the timing of data collection. Capturing perceptions immediately after each role engagement, possibly through repeated measures designs, can reduce retrospective bias and yield more accurate insights into students' experiences (e.g., Martin & Sippel, 2021).

The low quality of outcome measures is another critical issue specific to intrapersonal and interpersonal outcomes. These outcomes are often complex and multidimensional, particularly emotions, making them difficult to operationalize and compare across studies (Panadero et al., 2023). For future studies, the priority should be to align recent taxonomies of intrapersonal and interpersonal outcomes with previously used tools to measure those constructs (e.g., Alqassab et al., 2019; Huisman et al., 2020; Pekrun et al., 2017; Prins et al., 2005). Additionally, when using closed-ended questionnaires, it is essential

that instruments be tailored to each role. The ideal solution involves mirrored scales, one for assessors and one for assessees, as demonstrated in Su's (2023) study on comfort.

There are studies that can serve as reference on how to apply these recommendations. We have mentioned some along the review (e.g., Çevik, 2015; Huisman et al., 2018; Lundstrom & Baker, 2009; Su, 2023). More recently, Zamorano et al., (2025, 2026) also provide illustrative examples. These studies managed to isolate peer assessment roles while incorporating robust cognitive, intrapersonal, and interpersonal outcomes measures.

Conclusions

Students who assess their peers and students who are assessed by their peers may be experiencing fundamentally different learning processes. However, the current body of research has not yet been rigorous enough to tell us how both peer assessment roles differently contribute to learning outcomes. While a handful of studies on cognitive outcomes are methodologically sound, they showed mixed results. Studies on intrapersonal and interpersonal outcomes remain exploratory and limited by construct validity and weak research designs. In class, students mostly engage in both roles during peer assessment activities. Knowing how differently each role shapes students' learning can help teachers support their students more effectively. We encourage researchers to keep exploring the differential impact of peer assessment roles in different learning outcomes as research evidence in this area appears to be very limited.

Appendix A Full Search Strategies for all electronic databases used in the systematic review

The search was conducted independently for each search terms (e.g., peer assessment, peer feedback, peer grading…etc) because of difficulties when including all terms simultaneously on the search engines.

ERIC (via EBSCOhost)TI ("peer assessment" OR "peer-assessment") AND AB ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

TI ("peer feedback" OR "peer-feedback") AND AB ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

TI ("peer review" OR "peer-review") AND AB ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

TI ("peer evaluation" OR "peer-evaluation") AND AB ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

TI ("peer grading" OR "peer-grading") AND AB ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

TI ("peer rating" OR "peer-rating") AND AB ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*"

TI ("peer scoring" OR "peer-scoring") AND AB ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

TI ("peer marking" OR "peer-marking") AND AB ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

Limiters applied: None

PsycINFO (ProQuest)

ti ("peer assessment" OR "peer-assessment") AND ab ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

ti ("peer feedback" OR "peer-feedback") AND ab ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

ti ("peer review" OR "peer-review") AND ab ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

ti ("peer evaluation" OR "peer-evaluation") AND ab ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

ti ("peer grading" OR "peer-grading") AND ab ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

ti ("peer rating" OR "peer-rating") AND ab ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

ti ("peer scoring" OR "peer-scoring") AND ab ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*").

ti ("peer marking" OR "peer-marking") AND ab ("role*" OR "assessor* or assessee*" OR "receiver* or giver*" OR "assessor* or assessee*")

Appendix B Detailed student quality rubric of the study details and peer feedback design elements

		Weak (0 point)	Medium (1 point)	Robust (2 points)
Empirical design		Descriptive Design with Structured Observation or multiple case study	Quasi-experimental Experimental design Intact Group **ERIG**	Experimental Experimental Design Random Group **ERG**
Pedagogical quality	**Training**	No training	Simple (e.g., just giving a short talk about peer assessment	Advanced: 1) the students get the chance to practice PA before the actual activity and/or 2) have the opportunity of giving feedback on peer assessment.
	Practice (Number of times they give or receive FB)	1	2 to 5	6 or +
	PA aids	0	1	2 or +
Quality of DVs		Students' perceptions of…	SRL measured by self-report	"Objective" variables (academic performance, etc.)

	Weak (0 point)	Medium (1 point)	Robust (2 points)
Quality of the sample*	Characteristics and selection process of the sample are not clearly described, or the sample is too small to extract conclusions.	The sample is well-sized, but the participants were chosen using convenience-criterion only.*	The sample is well-sized, and the participants were chosen randomly or based in specific criteria other than convenience.

Sample: For a ERG 10-15 participants per condition is usually ok.
For a ERIG let's aim for at less 20 per condition.
For descriptive studies collecting surveys at least 25, for interviews at least 10, for in depth case analysis 5.

*For studies utilizing questionnaires or non-experimental/quasi-experimental designs, selecting students from a single class constitutes convenience sampling

This rubric is an adapted version of the one proposed by Panadero et al. (2023). We removed two categories that were deemed redundant: (1) Length of intervention, as most interventions were short and relatively uniform in duration, and (2) Quality *of measures*, which focused primarily on quantitative metrics and did not account for qualitative measures. Although the latter were often exploratory in nature, we considered them crucial to provide a greater picture of the studies conducted on this review and thus included them as a distinct category in our main coding scheme, rather than evaluating them as a quality criterion

Appendix C Inter-rater reliability in percentage agreements and Krippendorff's α for study details, peer assessment instructional design and PA context

Coding category	First round		Second round	
	% Agreement	Krippendorff's α	% agreement	Krippendorff's α
Study details				
Type of Study	100%	1	100%	1
Aim(s)	100%	1	100%	1
Design	66.6%	0	100%	1
Independent variable(s)	33.3%	–0.25	100%	1
Dependent variable(s)	66,6%	0	100%	1
Type of variable(s) of interest	100%	1	100%	1
Sample size	100%	1	100%	1
Educational level	66.6%	0	100%	1
Sample characteristics	100%	1	100%	1
Length of Intervention	66.6%	0	100%	1
Procedure	100%	1	100%	1
Statistical analysis	100%	1	100%	1
Results	0%	–0.66	100%	1
Main conclusions	100%	1	100%	1
Accounting of roles				
Directionality	33.3%	–0.25	100%	1

Coding category	First round		Second round	
	% Agreement	Krippendorff's α	% agreement	Krippendorff's α
Outcomes				
Quality of outcome measures	100%	1	100%	1
Quality of the study				
Empirical design	100%	1	100%	1
Training	66.6%	0	100%	1
Practice	33.3%	–0.25	100%	1
Aids	0%	–0.66	100%	1
Quality of DVs	33.3%	–0.25	100%	1
Quality of the sample	33.3%	–0.25	100%	1

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10212-026-01148-4.

Acknowledgements The authors would like to thank Dr. Belén Fernández-Castilla for her guidance and tutoring during the meta-analysis initially presented in early versions of this study. We also gratefully acknowledge the reviewers of a previous version of the manuscript for their constructive feedback, which helped improve the quality of the work and better reflect its value to the field.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This project was funded by: Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad) National I+D Call (Convocatoria Excelencia) project reference PID2019-108982GB-I00; personal grant (Formación de Personal Investigador) reference PRE2020-092280; and Basque Government Call for Grants to support the activities of research groups of the Basque University System (2022-2025) project reference IT1624-22.

Data Availability The data supporting the findings of this study are available in Supplementary Material 2.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aben, J., Dingyloudi, F., Timmermans, A., Strijbos, J.-W., & Henderson, M. A. (2019). The impact of feedback in higher education: Improving assessment outcomes for learners. *The impact of feedback in higher education. Improving assessment outcomes for learners* (pp. 107–125). Springer International Publishing. https://doi.org/10.1007/978-3-030-25112-3

Aben, J., Mascareño, M., Anneke, L., Dingyloudi, F., & Strijbos, J.-W. (2024). The impact of interpersonal perceptions on the process of dealing with errors while providing and processing peer-feedback on writing. *Instructional Science*. https://doi.org/10.1007/s11251-024-09660-0

Alqassab, M., & Panadero, E. (2020). Peer assessment. In S. Brookhart (Ed.), *Routledge encyclopedia of education* (pp. 1–9). Routledge.

Alqassab, M., Strijbos, J. W., Panadero, E., Ruiz, J. F., Warrens, M., & To, J. (2023). A Systematic Review of Peer Assessment Design Elements. In *Educational Psychology Review* (Vol. 35, Issue 1). Springer US. https://doi.org/10.1007/s10648-023-09723-7

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

Bauer, E., Greisel, M., Kuznetsov, I., Berndt, M., Kollar, I., Dresel, M., Fischer, M. R., & Fischer, F. (2023). Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda. *British Journal of Educational Technology, 54*(5), 1222–1245. https://doi.org/10.1111/bjet.13336

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice,* 5:1, 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*(1), 81–90. https://doi.org/10.1177/003172171009200119

Cao, Z., Yu, S., & Huang, J. (2019). A qualitative inquiry into undergraduates' learning from giving and receiving peer feedback in L2 writing: Insights from a case study. *Studies in Educational Evaluation, 63*, 102–112. https://doi.org/10.1016/j.stueduc.2019.08.001

Çevik, Y. (2015). Assessor or assessee? Investigating the differential effects of online peer assessment roles in the development of students' problem-solving skills. *Computers in Human Behavior, 52*, 250–258. https://doi.org/10.1016/j.chb.2015.05.056

Crinon, J., & Marin, B. (2010). The role of peer feedback in learning to write explanatory texts: Why the tutors learn the most. *Language Awareness, 19*(2), 111. https://doi.org/10.1080/09658411003746604

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

Ertmer, P. A., Richardson, J. C., Belland, B., Camin, D., Connolly, P., Coulthard, G., Lei, K., & Mong, C. (2007). Using peer feedback to enhance the quality of student online postings: An exploratory study. *Journal of Computer-Mediated Communication, 12*(2), 412–433. https://doi.org/10.1111/j.1083-6101.2007.00331.x

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

Fraile, J., Monguillot, M., González-Arévalo, C., Lehane, P., & Panadero, E. (2024). Investigating anonymity in formative and summative peer assessment: Effects on university students' social-affective factors, perceptions and preference. *Studies in Educational Evaluation*. https://doi.org/10.1016/j.stueduc.2024.101410

Gukas, I. D., Miles, S., Heylings, D. J., & Leinster, S. J. (2008). Medical students' perceptions of peer feedback on an anatomy student-selected study module. *Medical Teacher, 30*(8), 812. https://doi.org/10.1080/01421590802262070

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

Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2020). A questionnaire to assess students' beliefs about peer-feedback. *Innovations in Education and Teaching International, 57*(3), 328–338. https://doi.org/10.1080/14703297.2019.1630294

Ion, G., Sánchez Martí, A., & Agud Morell, I. (2019). Giving or receiving feedback: Which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education, 44*(1), 124–138. https://doi.org/10.1080/02602938.2018.1484881

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*(4), 344–348. https://doi.org/10.1016/j.learninstruc.2009.08.005

Lakens, D. (2022). Sample size justification. *Collabra: Psychology, 8*(1), 1–32. https://doi.org/10.1525/collabra.33267

Lee, C.-Y. (2015). The effects of online peer assessment and family entrepreneurial experience on students' business planning performance. *Turkish Online Journal of Educational Technology - TOJET, 14*(1), 123–132.

Levac, D., Colquhoun, H., & Kelly, K. O. (2010). Scoping studies: advancing the methodology. *Implementation Science 2010, 5:69*, 1–18. https://doi.org/10.1017/cbo9780511814563.003

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., & Grion, V. (2019). The power of giving feedback and receiving feedback in peer assessment. *All Ireland Journal of Higher Education, 11*(2), 1–17.

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

Lu, J., & Zhang, Z. (2012). Understanding the effectiveness of online peer assessment: A path model. *Journal of Educational Computing Research, 46*(3), 313–333.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002

Martin, I. A., & Sippel, L. (2021). Providing vs. receiving peer feedback: Learners' beliefs and experiences. *Language Teaching Research*. https://doi.org/10.1177/13621688211024365

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.

Misiejuk, K., & Wasson, B. (2021). Backward evaluation in peer assessment: A scoping review. *Computers & Education*. https://doi.org/10.1016/j.compedu.2021.104319

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Noroozi, O., Alqassab, M., Taghizadeh Kerman, N., Banihashem, S. K., & Panadero, E. (2024). Does perception mean learning? Insights from an online peer feedback setting. *Assessment & Evaluation in Higher Education, 0*(0), 1–15. https://doi.org/10.1080/02602938.2024.2345669

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*. https://doi.org/10.1136/bmj.n160

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment. A Review and Future Directions. *Handbook of Human and Social Conditions in Assessment*, *July 2016*, 247–266. https://doi.org/10.4324/9781315749136

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 and Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2023.2164884

Panadero, E., Jonsson, A., & Alqassab, M. (2018). Providing Formative Peer Feedback: What Do We Know? *In the Cambridge Handbook of Instructional Feedback (Issue October)*. https://doi.org/10.1017/9781316832134.020

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The epistemically-related emotion scales. *Cognition and Emotion, 31*(6), 1268–1276. https://doi.org/10.1080/02699931.2016.1204989

Perlman, C. (2003). Performance Assessment: Designing Appropriate Performance Tasks and Scoring Rubrics. *Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators*, 497–506.

Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. W. (2005). Formative peer assessment in a CSCL environment: A case study. *Assessment and Evaluation in Higher Education, 30*(4), 417–444. https://doi.org/10.1080/02602930500099219

Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*(4), 265–269. https://doi.org/10.1016/j.learninstruc.2009.08.002

Su, W. (2023). Masked ball for all: How anonymity affects students' perceived comfort levels in peer feedback. *Assessment and Evaluation in Higher Education, 48*(4), 502–512. https://doi.org/10.1080/02602938.2022.2089348

Su, W., & Huang, A. (2022). More enjoyable to give or to receive? Exploring students' emotional status in their peer feedback of academic writing. *Assessment and Evaluation in Higher Education, 47*(7), 1005–1015. https://doi.org/10.1080/02602938.2021.2004389

Topping, K. (2017). Peer Assessment: Learning by Judging and Discussing the Work of Other Learners. *Interdisciplinary Education and Psychology*, *1*(1). https://doi.org/10.31532/interdiscipeducpsychol.1.1.007

Tricco, A. C., Lillie, E., Zarin, W., Brien, O. ', Colquhoun, H., Kastner, M., Levac, D., Ng, C., Sharpe, J. P., Wilson, K., Kenny, M., Warren, R., Wilson, C., Stelfox, H. T., & Straus, S. E. (2016). A scoping review on the conduct and reporting of scoping reviews. *BMC Medical Research Methodology, 16*(1), 1–10. https://doi.org/10.1186/s12874-016-0116-4

Van der Kleij, F. M., & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability, 33*(2), 345–373. https://doi.org/10.1007/s11092-020-09331-x

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*(1), 41–54. https://doi.org/10.1016/j.edurev.2008.11.002

Wu, Y., & Schunn, C. D. (2020). 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

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*(September), 1–15. https://doi.org/10.1016/j.edurev.2022.100484

Zamorano, D., Panadero, E., Alqassab, M., & Amillano, A. (2025). Effect of peer feedback role: Task performance, feedback implementation and perceptions on learning and trust. *Journal of Experimental Education, 0*(0), 1–25. https://doi.org/10.1080/00220973.2025.2466146

Zamorano, D., Panadero, E., Alqassab, M., & Gómez, J. F. (2026). Provider versus recipient : Peer feedback roles effect on task performance , perceptions of learning , trust , and epistemic emotions. *Studies in Educational Evaluation*, *88*(January). https://doi.org/10.1016/j.stueduc.2026.101560

Zhan, Y. (2024). Are they ready? An investigation of university students' difficulties in peer assessment from dual perspectives. *Teaching in Higher Education, 29*(4), 823–840. https://doi.org/10.1080/13562517.2021.2021393

Zhang, Y., Pi, Z., Chen, L., Zhang, X., & Yang, J. (2021). Online peer assessment improves learners' creativity: Not only learners' roles as an assessor or assessee, but also their behavioral sequence matter. *Thinking Skills and Creativity, 42*(June), 1–11. https://doi.org/10.1016/j.tsc.2021.100950

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Current themes of research and most relevant publications in the field of Psychology of Education:

David Zamorano. Is a research associate at the ERLA (Education, Regulated Learning & Assessment group) at Deusto University (Spain). His research focuses on educational assessment, specifically peer assessment and formative assessment tools.

Most relevant publications:

Zamorano, D., Panadero, E., Alqassab, M., & Amillano, A. (2025). Effect of Peer Feedback Role: Task Performance, Feedback Implementation and Perceptions on Learning and Trust. Journal of Experimental Education, 0(0), 1–25. https://doi.org/10.1080/00220973.2025.2466146.

Zamorano, D., Panadero, E., Alqassab, M., & Gómez-Varela, J. F. (2026). Provider versus recipient: Peer feedback roles effect on task performance, perceptions of learning, trust, and epistemic emotions. *Studies in Educational Evaluation, 88*, 101560. https://doi.org/10.1016/j.stueduc.2026.101560.

Panadero, E., Delgado, P., Zamorano, D., Pinedo, L., Fernández-Ortube, A., & Barrenetxea-Mínguez, L. (2025). Putting excellence first: How rubric performance level order and feedback type influence students' reading patterns and task performance. Learning and Instruction, 99(August 2024). https://doi.org/10.1016/j.learninstruc.2025.102168.

Maryam Alqassab. Is an assistant professor at Leiden University Graduate School of Teaching (ICLON). Her research focuses on didactics of research supervision, formative assessment, and (technology-enhanced) (peer) assessment and learning design

Most relevant publications:

Alqassab, M., Strijbos, J. W., Panadero, E., Ruiz, J. F., Warrens, M.,& To, J. (2023). A Systematic Review of Peer Assessment Design Elements. In Educational Psychology Review (Vol. 35, Issue 1). Springer US. https://doi.org/10.1007/s10648-023-09723-7.

Alqassab, M., Strijbos, J. W., & Ufer, S. (2018). The impact of peer solution quality on peer- feedback provision on geometry proofs: Evidence from eye-movement analysis. Learning and Instruction, 58(October 2017), 182–192. https://doi.org/10.1016/j.learninstruc.2018.07.003.

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. In Assessment& Evaluation in Higher Education (Vol. 44, Issue 1, pp. 139–154). https://doi.org/10.14742/ajet.830.

Ernesto Panadero. Professor and Director of CARPE (Centre for Assessment Research, Policy and Practice in Education) at Dublin City University (Ireland) and as an Ikerbasque Professor and international leader of ERLA (Education, Regulated Learning & Assessment group) at Deusto University (Spain). His research focuses on self-regulated learning and educational assessment

Most relevant publications:

Panadero, E. (2023). Towards a paradigm shift in feedback research: Five further steps influenced by self regulated learning theory. Educational Psychologist, 58(3), 193-204. https://doi.org/10.1080/00461520.2023.22236 42.

Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of Rubrics on Academic Performance, Self-Regulated Learning, and self-Efficacy: a Meta-analytic Review. Educational Psychology Review, 35(4), 113. https://doi.org/10.1007/s10648-023- 09823-4.

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.

Authors and Affiliations

David Zamorano[1] · Maryam Alqassab[2,3] · Ernesto Panadero[1,4,5]

✉ David Zamorano
david.zamorano@deusto.es

[1] Education Regulated Learning and Assessment Group (ERLA Group), Faculty of Education and Sport, Universidad de Deusto, Bilbao, Spain

[2] Graduate School of Teaching (ICLON), Leiden University, Leiden, The Netherlands

[3] Department of Online Learning and Instruction, Faculty of Educational Sciences, Open University, Heerlen, Netherlands

[4] Centre for Assessment Research Policy and Practice in Education (CARPE), School of Policy and Practice, Institute of Education, St. Patrick's Campus, Dublin City University, Dublin, Ireland

[5] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain