

A Systematic Review on Peer Assessment: Intrapersonal and Interpersonal Factors

Ernesto Panadero^{1 & 2}, Maryam Alqassab³, Javier Fernández Ruiz⁴ & Jose Carlos G. Ocampo¹

¹ Facultad de Educación y Deporte, Universidad de Deusto, Bilbao, Spain.

² IKERBASQUE, Basque Foundation for Science, Bilbao, Spain.

³ Departamento de Educación, Universidad de Las Palmas de Gran Canarias, Spain.

⁴ Área de Psicología Evolutiva y de la Educación, Universidad de León, Spain.

Author note

Ernesto Panadero <https://orcid.org/0000-0003-0859-3616>

Maryam Alqassab <http://orcid.org/0000-0002-6574-5113>

Javier Fernández Ruiz <https://orcid.org/0000-0001-5419-7687>

Jose Carlos G. Ocampo <https://orcid.org/0000-0002-2180-7618>

Recommended citation: Panadero, E., Alqassab, M. Fernández-Ruiz, J., & Ocampo, J. C. G. (2023). A systematic review on peer assessment: intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2023.2164884>

We have no known conflict of interest to disclose.

Acknowledgements. The authors would like to thank Gavin Brown, Kit Double and Jan-Willem Strijbos for their efforts and valuable feedback on the instrument proposed.: Research funded by Spanish National R+D call from the Ministerio de Ciencia, Innovación y Universidades (Generación del conocimiento 2020), Reference number: PID2019-108982GB-I00. The second author's research is funded by the Spanish Ministry of Science and Innovation (Ministerio de Ciencia e Innovación) under the Juan de la Cierva Incorporación program (Reference number: IJC2020-043302-I).

Correspondence concerning this manuscript should be addressed to: Javier Fernández. Despacho 223, Área de Psicología Evolutiva y de la Educación Campus de Vegazana s/n. Universidad de León 24071, León (Spain). E-mail: javier@fernandezruiz.com

Abstract

Peer assessment is a popular research topic as it takes place in various educational settings around the globe. Due to its nature, a number of intrapersonal and interpersonal factors are involved in its implementation. Empirical research on these factors in peer assessment has been increasing in the last years and our aim is to systematically review them to extract conclusions about the relationships between peer assessment and intrapersonal and interpersonal factors. Using different search strategies, we reviewed 69 articles. We investigated: (1) the characteristics of the included studies, (2) the trends in the bidirectional relationships of intrapersonal \leftrightarrow peer assessment and interpersonal \leftrightarrow peer assessment, (3) the empirical findings in the relationships between intrapersonal factors and peer assessment, and (4) the empirical findings in the relationships between interpersonal factors and peer assessment. We have identified six intrapersonal factors: motivation, self-efficacy, emotions, trust in the self as assessor, fairness, and comfort; and five interpersonal factors: social connections, trust in the other as assessor, psychological safety, value diversity/congruence, and interdependence. The results showed clear directions for some of those factors and are preliminary in some of the others. This review offers directions to improve the quality of peer assessment research and explores the role of bidirectionality for future research, including an instrument to report the characteristics of future peer assessment studies to facilitate better reports and research designs.

Keywords: peer assessment, intrapersonal factors, interpersonal factors, human factors, systematic literature review

A Systematic Review on Peer Assessment: Intrapersonal and Interpersonal Factors

Peer assessment can be defined as an ‘arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status’ (Topping, 1998 p. 250). This implies two premises. First, peer assessment involves students in assessment with the promise of several gains, among them increasing students’ autonomy, self-regulated learning, and performance (Reinholz, 2016; Double et al., 2020; Ibarra-Sáiz et al., 2020). Such involvement entails that students’ individual characteristics (e.g., motivation to perform peer assessment, emotions experienced in relation to peer assessment) are likely to affect different peer assessment processes but also the outcomes of peer assessment itself (Topping et al., 2000; Strijbos et al., 2009; Panadero, 2016). We will call these factors *intrapersonal factors*, referring to variables that are within or related to the learner.

Second, peer assessment involves at least one student assessing another or receiving feedback, which makes it a social enterprise (van Gennip et al., 2009). Therefore, there are interpersonal factors (friendship, psychological safety) that are also expected to influence peer assessment processes and outcomes (Carvalho, 2013; Panadero, 2016; Rotsaert, 2017). Students often question the relevance of peer assessment because of social and interpersonal problems such as lack of trust in the peer as an assessor. To ameliorate negative peer assessment experiences, some have stressed the importance of creating a climate or an assessment culture which is perceived as safe and in which participants agree on the goals and values of the assessment practice (Dochy et al., 1999; Tillema et al., 2011).

To highlight the – long ignored – social nature of peer assessment processes, van Gennip, and colleagues (2009) conducted the first literature review on the interpersonal factors influencing peer assessment, recommending that their effects should be considered in research

and implementation. Following the same call, Panadero (2016) argued that ‘PA does not happen in a vacuum; rather it produces thoughts, actions, and emotions as a consequence of the interaction of assesses and assessors’ (p. 247). Panadero identified ten themes (motivation; trust in self and peer as assessors; friendship marking, etc.) and categorized them under three categories related to the intrapersonal (human) and interpersonal (social) factors of peer assessment implementation.

Since then, plenty of assessment scholars have supported the notion of going beyond the conventional summative study to a more interpersonally conscious approach (Strijbos & Wichmann, 2018). For example, some studies investigated the use of rubrics to avoid friendship bias (Panadero et al., 2013), or the effects of peer assessment on learning motivation (Chang et al., 2020), or the effects of anonymity on peer assessment (Vanderhoven et al., 2015; Rotsaert et al., 2018). Despite the sudden outburst of studies investigating interpersonal and intrapersonal factors in peer assessment in the last years, the relationship between peer assessment and these factors is still unclear. Our aim is to update the current status of knowledge and to clarify the empirical evidence on this topic.

A brief story about peer assessment

Peer assessment is known as a complex activity that involves many variables (Gielen et al., 2011). This complexity is represented by the multiple attempts to classify the variables needed to be considered when implementing peer assessment. Topping (1998) proposed a typology of 17 variables. Since then, several researchers have refined the typology by classifying the variables under themes or clusters and/or extending them to include up to 20 variables (van den Berg et al., 2006; van Gennip et al., 2009; Gielen et al., 2011; Adachi et al., 2018). These variables have been classified under clusters such as the use and management of peer

assessment, the learning environment, or the interactions between peers (Gielen et al., 2011; Adachi et al., 2018). This complexity highlights the importance of the roles that intrapersonal and interpersonal factors play in peer assessment.

The impact of peer assessment on performance has been the main focus of peer assessment reviews (Dochy et al., 1999; Topping, 2003; van Zundert et al., 2010) and meta-analyses (Li et al., 2020; Double et al., 2020). Double and colleagues (2020) found that peer assessment had a more positive effect on students' performance compared to teacher's assessment or no assessment, regardless of students' educational level and some feedback characteristics. Similarly, Li and colleagues (2020) found a positive effect of peer assessment on performance compared to no assessment or teacher assessment, that assessor training was the strongest factor influencing the impact of peer assessment on performance, and reported that computer-mediated peer assessment seems to be more beneficial than paper-based.

Previous reviews on intrapersonal and interpersonal factors in peer assessment

In their systematic literature review, van Gennip and colleagues (2009) investigated the effects of interpersonal variables and peer assessment design format (i.e. structural features) on learning (i.e. objective learning, learning as perceived by learners, and beliefs). They identified four interpersonal variables pertaining to relational interactions between team members, namely, psychological safety, value diversity, interdependence and trust. The review revealed that students mostly felt safe during peer assessment, and they trusted themselves and their peers as assessors. However, this conclusion was based on a limited number of studies as only four of the 15 empirical studies included in the review investigated interpersonal variables, and trust was not perceived positively in one of these four studies. The authors reported that none of the four studies investigated interpersonal variables in relation to learning outcomes of peer assessment.

Taking a broader perspective, Panadero (2016) conducted a narrative review of empirical peer assessment studies investigating interpersonal (social) factors that were identified by van Gennip and colleagues (2009), but also intrapersonal (human) factors influencing peer assessment and as outcomes of peer assessment. Panadero's review included 26 articles published between 1995 and 2014, identifying ten factors that had been explored. The main conclusion was that 'more intensive PA [peer assessment] implementations produce better human and social outcomes, especially as students gain a deeper understanding of the complexities of PA' (p. 261). Additionally, Panadero proposed a long list of potential research topics that could be conducted and emphasized the need for stronger research designs.

Both reviews accentuated the role of interpersonal or/and intrapersonal factors in shaping peer assessment processes and, ever since, interest has grown in this topic.

A new categorization of factors related to peer assessment

The two previous reviews used different categorizations. van Gennip and colleagues (2009) focused only on four interpersonal variables: (1) psychological safety, (2) value diversity, (3) interdependence and (4) trust in the self or the other as an assessor. One of the reasons for the limited number of factors might have been the low number of studies identified. On the other hand, Panadero (2016) found ten factors in the existing literature (see Figure 1) that he classified as three main self-explanatory categories - interpersonal, intrapersonal and cognitive - though he did not provide a strong rationale for the classification due to length limits.

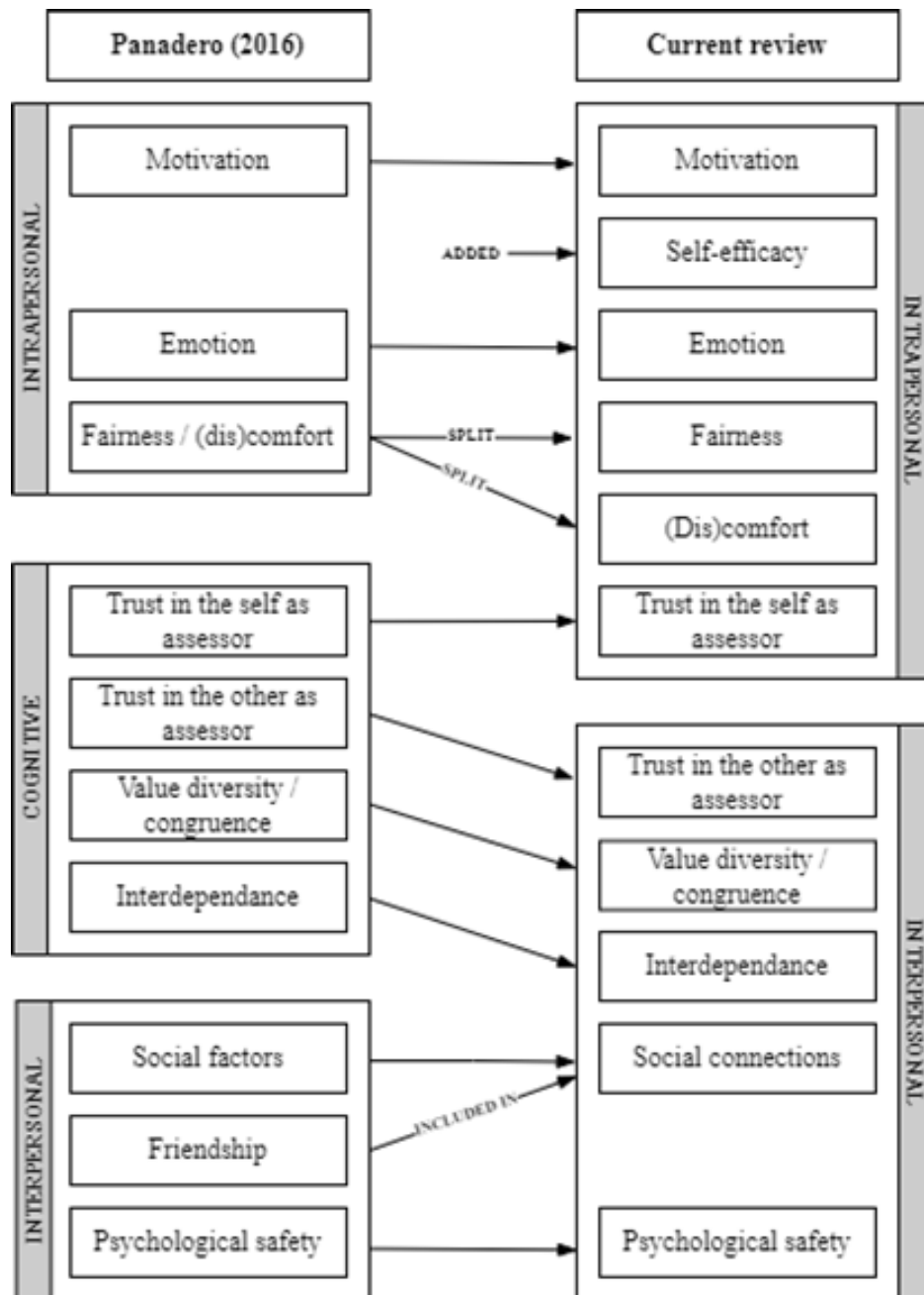


Figure 1. Illustration of the adaptation made to the factors and categories by Panadero (2016) for this review

In the conceptualization of the factors investigated in our review, we opted for a two-category approach as it is more parsimonious and represents a more accurate depiction of the factors' organization. We rearranged the ten original factors used by Panadero (2016) under either intrapersonal or interpersonal categories, split fairness and (dis)comfort into two factors,

and fused social factors and friendship in one factor. We added a new factor ‘self-efficacy’.

Figure 1 and Table 1 present the 11 factors and their corresponding definitions.

Table 1

Intrapersonal and interpersonal factors of peer assessment as defined by Panadero (2016) compared to the revised definitions used in the current study

Categories	Definition From Panadero (2016)	New definition & categories
Intrapersonal factors: refers to elements of students’ own processes or that influence these individual processes		
(1) Motivation	Having to do with the reasons or goals students have for performing PA and the effects of PA on motivation	Variables related to why students initiate, continue, or terminate a certain behavior at a particular time (e.g., goals, interest)
		Self-efficacy: individual's belief in their capacity to execute behaviors necessary to produce specific performance attainments (It refers to beliefs about academic ability or task specific ability; not to be confused with trust in the self as assessor)
(2) Emotion	Focused on the feelings connected to PA and its implementation	Variables related to the feelings and their regulation in the PA process (e.g., anxiety, happiness)
(5) Trust in the self as assessor	Focusing on belief that one can be a good assessor	Belief about the ability to perform peer assessment as an assessor
(7) Fairness/(dis)comfort	Referring to beliefs about the appropriateness of PA and the student’s level of comfort within assessor and/or assessee roles	Fairness: Students perceptions that the PA is free from bias, dishonesty, or injustice (Dis)comfort: a state of (un)ease and (dis)satisfaction regarding PA processes
Interpersonal factors: refers to elements of the peer interaction or that influence the interaction in the peer assessment process		
(3) Social connections	(Named social factors) related to aspects of social connections, such as enhancing peer scores to be more liked by classmates	Variables related to aspects of social connections and how they might interact with PA (e.g., not giving negative feedback to not affect social relationships) Examples: peer familiarity, friendship (marking), interpersonal author trait, peer pressure, fear of disapproval.
(6) Trust in the other as assessor	Related to having confidence in a peer who is capable of performing PA	Confidence in a peer’s capability to perform a fair and/or accurate PA
(8) Psychological safety	Focusing on the extent to which students feel safe to give sincere feedback as an assessor and do	Extent to which students feel safe to give sincere feedback as an assessor and do not fear receiving

Categories	Definition From Panadero (2016)	New definition & categories
	not fear inappropriate negative feedback as an assessee	inappropriate negative feedback or to respond to negative feedback
(9) Value diversity/congruence	Referring to the level of goal similarity between assessor and assessee	Variables referring to the similarity or differences of the assessor' and assessee' PA goals
(10) Interdependence	Examining the degree to which assessors and assesseees are mutually dependent on each other	Variables referring to the degree to which assessors and assesseees are mutually dependent on each other in the PA activity

Note. Categories are numbered following the original numbering in Panadero (2016) for ease of tracking

Bidirectionality of the relationship between intrapersonal factors \leftrightarrow peer assessment and interpersonal factors \leftrightarrow peer assessment

In their reviews van Gennip and colleagues (2009) and Panadero (2016) elucidated the importance of interpersonal and/or intrapersonal factors in peer assessment, yet both reviews overlooked a crucial aspect: these effects can be bidirectional. As much as intrapersonal and interpersonal factors can affect peer assessment, so can peer assessment affect these factors. For example, students' level of motivation can be expected to influence the level of participation in peer assessment, but also taking part in peer assessment can influence students' motivation. One study may explore how the students' goal orientation affects the type of peer assessment processes those students activate; while another study can explore whether non-anonymous versus anonymous peer assessment produce different effects on students' goal orientations. In other words, peer assessment can be investigated as both a dependent and independent variable, as much as intrapersonal and interpersonal factors can also be both. Investigating the directionality of the effects is important for at least three reasons. First, it specifies more clearly what is known from the current peer assessment research. Second, it gives direction to future studies, for example, by identifying that one of the factors has only been studied in one direction. And third, investigating bidirectionality will allow for more precise interpretation of results,

which is important for primary research but even more so for reviewing the field. To our knowledge, the bidirectional relationships between peer assessment and intrapersonal/interpersonal factors have not been investigated before in the peer assessment literature.

Aim and Research Questions

Our aim is to systematically review the evidence on the relationship between peer assessment and intrapersonal and interpersonal factors. We analyze two bidirectional relationships: (1) intrapersonal factors and peer assessment, and (2) interpersonal factors and peer assessment. We will explore the following research questions (RQ):

RQ1. What are the characteristics of the studies investigating intrapersonal and interpersonal factors in relation to peer assessment?

RQ2. What are the trends in the bidirectional relationships intrapersonal \leftrightarrow peer assessment, and interpersonal \leftrightarrow peer assessment?

RQ3. What are the relationships between intrapersonal factors and peer assessment?

RQ4. What are the relationships between interpersonal factors and peer assessment?

Method

Search Strategies

The literature search was carried out using four strategies. First, articles citing van Gennip and colleagues (2009) and Panadero (2016) indexed in Google Scholar were recorded. Second, literature searches in PsycINFO and ERIC were conducted using the following combination of key terms: ('peer assessment'; 'peer feedback'; 'peer review; 'peer evaluation'; 'peer grading') + ('motivation'; 'emotion'; 'interpersonal'; 'friendship'; 'anonym*'; 'blind'; 'confidential'). Those searches were conducted with year restriction (>2014), as earlier studies

were already covered in Van Gennip et al.'s and Panadero's reviews. Third, the first author screened his personal database to find possible relevant articles. Fourth, several peer assessment experts were consulted for references.

The four strategies were used over several iterations from September 2020 to September 2021. Studies from various disciplines (e.g. business, medicine and social science) and educational levels (e.g. elementary, secondary, tertiary and graduate) were included. In order to be included in this review, an article should: (1) contain empirical evidence on the effects of peer assessment on the interpersonal or intrapersonal factors or vice versa; (2) be published and peer-reviewed, including doctoral dissertations; (3) be conducted in educational contexts; and (4) be available in English.

A total of 511 articles were identified using all the search strategies. Of the total number, 499 articles were identified through the electronic databases search, and 12 articles were identified through the experts' database. After removing duplicates, the first and fourth authors independently screened a total of 267 abstracts. Of this number, 129 records (19 dissertations/theses, 1 technical report, 109 articles) were identified, 56 records (2 dissertations/theses, 54 articles) were identified in second phase, 76 records (3 dissertations/theses, 73 articles) were identified in the third phase, and 6 records (i.e., 6 articles) were identified in the fourth phase. The 267 abstracts comprised 220 journal articles, 24 dissertations/theses and 23 others (e.g., technical reports, conference proceedings, chapters). We excluded 202 articles from the list because they did not meet the inclusion criteria. Additional articles (26 articles from Panadero's review published in 2016 and 6 articles from expert suggestions) were identified.

A total of 97 full-text articles were accessed and further screened by the second, third and fourth authors for eligibility, and 28 articles were excluded. Finally, 69 qualified articles were included and analysed in this systematic review. Figure 2 shows a flow diagram of the literature search.

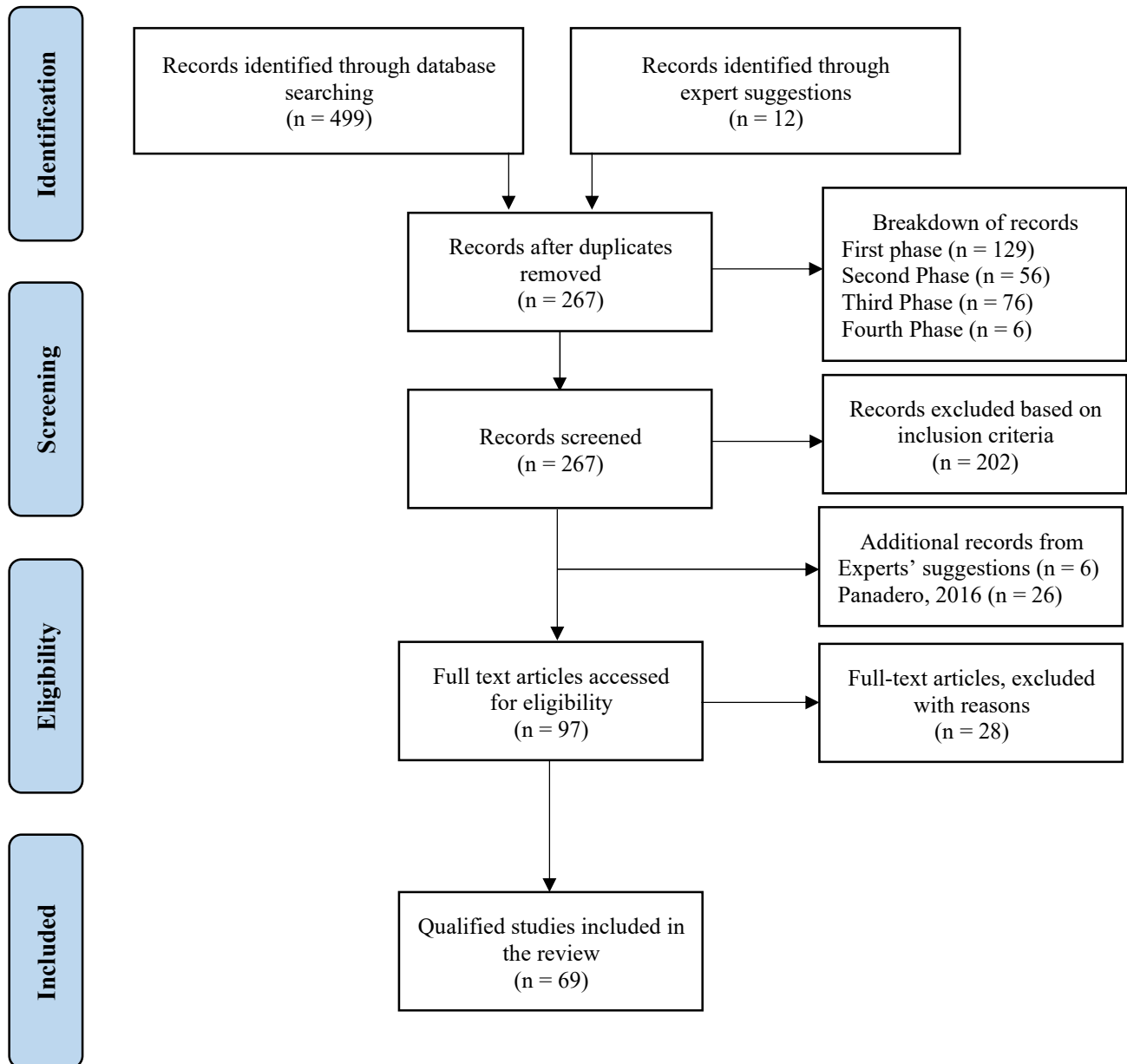


Figure 2. Literature search and inclusion process.

Coding procedure

Coding categories. Relevant data including study details (e.g., aim, demographic, design, main results) and peer assessment characteristics (e.g., terminology, type of activity, aids) were extracted from the 69 articles in a data extraction template prepared by the first author (see Supplementary material 1 for the detailed coding scheme of information extracted for each included study and Supplementary material 2 for the full dataset).

Quality of the study coding. The overall quality of the included studies was evaluated using a rubric developed by the authors assessing each study in terms of empirical design, quality of measurements, pedagogical quality, dependent variable/s quality and sample quality (see Supplementary Material 3). For each of these quality elements, the study was quantitatively rated as weak (0 points), medium (1 point) or robust (2 points). A total score was then computed for each paper representing the overall quality of the study. The score could range between 0 and 14 points.

Coding of study directionality. Each included study was categorised depending on the direction of the investigated relationships between peer assessment and inter/intrapersonal factors. We determined the directionality based on one or more of the following: (1) the theoretical assumptions guiding the study (e.g. peer assessment improves motivation versus higher motivation leads to a better peer assessment), (2) study hypotheses if specified, and (3) whether the inter/intrapersonal factors were reported as outcomes of the peer assessment or as factors influencing peer assessment outcomes in the results of the study. We identified four categories: Intrapersonal → PA, PA → Intrapersonal, Interpersonal → PA, and PA → Interpersonal. Depending on its focus, one study could be assigned to more than one of the four categories. The second, third, and fourth authors independently identified the directionality of

each included study. Disagreements were discussed between the coders until a 100% agreement was reached.

Coding reliability. Three of the authors worked together to establish an acceptable interrater agreement of 90%. During this process, 14 (14.4%) of the included papers were independently coded by at least two authors (first, third and fourth author). The first, third and fourth authors independently coded seven out of the 14 papers while the remaining 7 papers were coded by the third and fourth authors. This was done over several iterative rounds, each followed by a discussion of the disagreements until consensus was reached. The initial inter-rater percentage agreement between the first, third and fourth authors was 82.4%, which increased progressively until the three authors independently coded three consecutive papers with 100% agreement. The second author performed an independent coding of four out of the 14 papers selected to establish inter-rater agreement that were coded by the other three authors and reached an interrater agreement of 93.3%.

Data Analysis

The aim of this study was to review the empirical evidence of the relationship between intrapersonal and interpersonal factors and peer assessment, but the strength of this relationship was not the primary focus of the study. Therefore, we performed narrative content analysis (Dochy 2006). The characteristics of the study were descriptively analyzed to answer RQ1. For the rest of RQs, each study was categorized according to the directionality of its results based on the four categories.

Results

RQ1. What are the characteristics of the studies investigating intrapersonal and interpersonal factors, in relation to peer assessment?

We analysed the 69 included studies in terms of (a) research design characteristics (research design, sample size, gender distribution, educational level, subject domain and country) and (b) peer assessment intervention characteristics (terminology, grades, use of aids, length intervention, anonymity and constellations).

Research design characteristics.

This data can be found in Supplementary material 4.

Research design. Only 12% of the reviewed studies were experimental. The remaining studies had descriptive designs (48%) including qualitative, case studies or correlational studies, followed by quasi-experimental (40%) designs.

Sample size. Thirty-one of the studies (44.3%) had samples of $N = 80+$; twenty-six studies (37.1%) had samples of $N = 31$ to 79 ; and thirteen studies (18.6%) ranged from $N = 1$ to 30 participants. One study (Harris and Brown 2013) reported the participation of three teachers and their students but did not provide the number of students.

Gender distribution. Twenty-one (30%) of the studies did not report the gender of their participants. For the remaining studies, the majority of participants were females.

Educational level. The vast majority of studies were conducted in higher education (49; 70%) followed by secondary education (11; 15.7%) and elementary education (3; 4.3%). Some studies (7; 10%) had mixed educational levels.

Subject domain. Nineteen studies (27.1%) were in the area of science, technology, engineering and mathematics; sixteen (22.9%) in the area of teaching and instruction; fifteen (21.4%) in the area of languages and writing instruction; four (5.7%) in the domain of multimedia and arts. The domains of business and management, humanities and social science,

and mixed domains each had three (4.3%) included studies. Seven studies (10%) did not indicate the subject domain.

Country. The reviewed studies were conducted in different countries around the world with Taiwan and China dominating (18.6% & 10% respectively). Other countries including the Netherlands, United States, Germany, Turkey, Iran, Spain, Belgium, Brunei, the United Kingdom, Australia, Hong Kong, Korea, New Zealand, Portugal, Scotland, South Africa, and Sweden each had between five and one studies (7.1-1.4%). (1.4%), two (2.9%) studies were conducted in several countries, and ten (14.3%) did not indicate the country.

Peer Assessment Intervention Characteristics

This data can be found in Supplementary material 5.

Terminology. More than half of the studies used the term ‘peer assessment’ (41; 58.6%), followed by ‘web-based or online peer assessment’ (11; 15.7%). The term ‘peer feedback’ was used in ten (14.3%) studies, and five (7.1%) studies used a combination of ‘peer feedback + ‘peer assessment’. Other terms like ‘peer review’ or more context specific terms (e.g., ‘peer review of teaching/PRT’) appeared in two (2.9%) studies each.

Grades. Twenty-five (35.7%) studies did not have a grade component in their peer assessment activity, while in 45 (64.3%) studies the assessor gave some form of grade to the assessee.

Use of aids. Twenty-nine (41.4%) of the included studies used criteria, prompts or scales as the main assessment aid in their study. Nineteen (27.14%) studies utilised rubrics, and twelve (17.1%) studies combined different assessment aids. Eight (11.4%) studies did not use any peer assessment aids. The use of assessment aids was not applicable for two (2.9%) studies due to their design.

Length of intervention. In sixteen (22.9%) of the studies students only practiced peer assessment once before engaging in the actual peer assessment. There were fifteen (21.4%) studies in which students performed peer assessment for more than six times, and fourteen (20%) between two to five times, ten (14.3%) studies did not include practice, and in five (7.4%) studies the frequency of practice was unclear.

Anonymity. Almost a quarter of the studies (16; 22.9%) did not report the anonymity mode used, fourteen (20%) studies were conducted non-anonymously, thirteen (18.6%) studies were fully anonymous, eleven (15.7%) studies used a single-blind mode (identity of either the assessor or the assessee was known), nine (18.6%) studies combined different anonymity modes.

Constellations. Only three studies (Topping et al., 2000; Domínguez et al., 2016; Chen et al., 2020) described their peer assessment design based on a published peer assessment inventory: respectively, Topping (1998), Gielen et al. (2011), and Adachi et al. (2018).

RQ2. What are the trends in the bidirectional relationships intrapersonal \leftrightarrow peer assessment, and interpersonal \leftrightarrow peer assessment?

A key aspect of our review was to explore the directionality of effects between intrapersonal and interpersonal factors and peer assessment (See Table 2). Our results of studies on intrapersonal factors and peer assessment show that they mostly focus on the effects of peer assessment on intrapersonal factors (peer assessment \rightarrow intrapersonal), with a total of 52 out of 69 (75%) studies exploring this direction. This is by far the most studied direction, as the majority of scholars explored how being involved in peer assessment affects five intrapersonal factors: motivation (21 studies), self-efficacy (6 studies), emotion (9 studies), trust in the self as assessor (15 studies), fairness (15 studies), and discomfort (9 studies). Conversely, intrapersonal factors effects on peer assessment (intrapersonal \rightarrow peer assessment) has been explored only in

nine studies. These studies investigated the impact of motivation (3 studies), emotion (1 study), trust in the self as assessor (6 studies), and fairness (2 studies) on peer assessment, while discomfort and self-efficacy have not been studied. This finding shows that while there is interest in this area, it is the least developed research topic, and we still know little about how individual characteristics influence peer assessment processes.

Table 2

Frequency of studies in each factor arranged by directionality

Intrapersonal factors						
N	Variable			Variable	N	
9	Intrapersonal (general)	→		→	52	
3	Motivation	→	Peer Assessment	→	Motivation	21
0	Self-efficacy	→		→	Self-efficacy	6
1	Emotion	→		→	Emotion	9
6	Trust in the self as assessor	→		→	Trust in the self as assessor	15
2	Fairness	→		→	Fairness	15
0	Discomfort	→		→	Discomfort	9
Interpersonal factors						
N	Variable			Variable	N	
15	Interpersonal (general)	→		→	18	
12	Social connections	→	Peer Assessment	→	Social connections	5
4	Trust in the other as assessor	→		→	Trust in the other as assessor	11
2	Psychological safety	→		→	Psychological safety	5
2	Value diversity/congruence	→		→	Value diversity/congruence	3
2	Interdependence	→		→	Interdependence	2

Note: Some studies investigated multiple variables so they were included in different categories. This implies two things. First, the sum of the four total studies is higher than the number of studies included in this review. Second, the sum of the different factors in each category is higher than the total studies number.

Findings on the directionality of interpersonal factors and peer assessment relationships appeared to be more balanced than the intrapersonal factors. When it comes to the impact of interpersonal factors on peer assessment (interpersonal → peer assessment), this relationship has

been explored in 15 studies investigating the following factors: social connections (12 studies), trust in the other as assessor (4 studies), psychological safety (2 studies), value diversity/congruence (2 studies) and interdependence (2 studies). The impact of peer assessment on interpersonal factors (peer assessment → interpersonal), has been explored in 18 studies investigating the following factors: social connections (5 studies), trust in the other as assessor (11 studies), psychological safety (5 studies), value diversity/congruence (3 studies), and interdependence (2 studies).

While the four directions have been explored, the most studied aspect is how being involved in peer assessment influences intrapersonal factors (e.g., motivation, self-efficacy). Interestingly, although the directionality of the social factors of peer assessment has been studied in a more balanced way, studying interpersonal factors in relation to peer assessment is less predominant than studying intrapersonal factors.

Finally, we investigated how many studies had explored the bidirectional relationship empirically. We only detected one study by van Gennip et al. (2010) that examined both bidirectional relationships, but without acknowledging that feature of their investigation.

RQ3. What are the results on the relationships between intrapersonal factors and peer assessment?

The six intrapersonal factors identified were motivation, self-efficacy, emotions, trust in the self as assessor, fairness, and comfort. The results are detailed in Supplementary material 6 (two tables).

Motivation

The effect of motivation → peer assessment has been studied in three publications. Two studies showed that having higher motivation had a positive effect on peer assessment as

students used deeper approaches to learning (Filius et al., 2019), and committed less mistakes in peer assessment (Guo & Lei, 2020). Zou and colleagues (2018) found that higher levels of motivation led to reviewing less papers but spending more time on each paper instead of assessing them superficially.

The effects of peer assessment → motivation is the most investigated directionality (21 studies), and most studies have found positive effects (McMahon, 2010; Hwang et al., 2014; Lai & Hwang, 2015; Hsia et al., 2016; Kuo et al., 2017; Berndt et al., 2018; Ersöz & Şad, 2018; Huisman et al., 2018; Chien et al., 2020; Hassell & Lee, 2019; Li, 2019; Li et al., 2020; Chen et al., 2020; Lim et al., 2021). However, Chang and colleagues (2020) did not find a significant increase in learning motivation after peer assessment compared to teachers' feedback. Strijbos, and colleagues (2010) did not find a significant effect of the impact of peer feedback content and senders' competency levels on willingness to improve (i.e. motivation). In a recent study, Strijbos and colleagues (2021) found that fairness of peer assessment positively predicted motivation (willingness to improve). Other studies found negative effects on motivation. The study by Johnson and Winterbottom (2011) found a decrease in learning-goal orientation, although the qualitative data showed that some students moved towards learning-goal orientation. In one study students who assessed excellent-quality peer work experienced decreased motivation (Rogers & Feller, 2016), but this seems to be a logical result and does not overshadow the overall positive effect. Three studies suggested that motivation to peer assess can be negatively related to knowing your peer (Corgnet, 2012), and positively to using a web-based peer assessment (Kim & Ryu, 2013) and perceptions of fairness (Strijbos et al., 2021).

Self-efficacy

This factor has only been studied in one direction: peer assessment → self-efficacy, mostly because the other direction is operationalized most of the time as trust in self as assessor given the specificity of the self-efficacy construct. Three of the six studies investigating this factor suggest a positive relationship that peer assessment enhances students' self-efficacy (Kuo et al., 2017) even when compared to teachers' feedback (Feyli & Ayatollahi, 2016; Chang et al., 2020). Liu and colleagues (2016) did not find a negative effect of peer assessment on creative self-efficacy. Similarly, Ghahari and Farokhnia (2018) did not find a significant change in students' self-efficacy after being involved in peer assessment compared to teacher assessment. However, one study found a negative effect of peer assessment on self-efficacy when students had to assess an excellent-quality peer work compared to poor-quality peer work (Rogers & Feller, 2016).

Emotions

The effect of emotions → peer assessment has only been investigated in one study. Alqassab and colleagues (2019) found that anxiety was associated with producing less accurate peer feedback and that curiosity and confusion were not significantly related to accuracy. The effects of peer assessment → emotions have been investigated in nine studies. Cheng and colleagues (2014) found that receiving positive peer comments elicited more positive emotions; Strijbos and colleagues (2010) found that elaborate peer feedback from a competent peer led to negative affect; Berndt and colleagues (2018) found that receiving elaborate peer feedback triggered more positive affect; Omar and colleagues (2018) found that 30% of their participants did not enjoy providing peer feedback while 43.5% enjoyed it. These findings seem to suggest that receiving positive or elaborate peer feedback can lead to experiencing more positive emotions unless the feedback is not from a competent peer.

Trust in the self as assessor

Six studies have explored the effects of trust in the self as assessor → peer assessment. There is a clear pattern as most of these studies found trust in self as assessor had positive effects on different peer assessment outcomes including adoption of deep learning approach, feedback accuracy and conceptions of peer assessment (van Gennip et al., 2010; Cheng & Tsai, 2012; Harris & Brown, 2013; Rotsaert et al., 2017). In a study by Zou and colleagues (2018) students with more trust were found to review less peer papers because they spent longer time on each peers' work as they assessed in a deeper way. In one study trust in self as assessor did not predict peer feedback accuracy (Alqassab et al., 2019).

When it comes to the effects of peer assessment → trust in the self as assessor, eight studies found that taking part in peer assessment increased trust (Stanier, 1997; Prins et al., 2005; McMahon, 2010; Hsia et al., 2016; Ching & Hsu, 2016; Liu et al., 2018; Rotsaert et al., 2018; Seifer & Feliks, 2019). Four studies did not find an effect (Sluijsmans et al., 2002; van Gennip et al., 2010; Omar et al., 2018; Shen et al., 2020), while two studies found negative effects (Kilickaya, 2017; Alqassab et al., 2018). Hassel and Lee (2019) concluded that their participants did not believe peer assessment was a substitute for teacher's feedback. With such mixed results, it seems that the features of the intervention could be key to obtain either a positive or neutral result, but our coding of intervention did not identify a clear pattern in the direction of these effects.

Fairness

Only two studies have explored the effect of fairness → peer assessment finding that fairness was related to the engagement in and the quality of peer assessment (Carvalho, 2013;

Zou et al., 2018), so that higher perceptions of fairness would be preferable. Zou and colleagues (2018) found that more than one third of their participants doubted that peer assessment was fair.

A higher number of studies have investigated the effects of peer assessment → fairness. We identified two types of studies. First, studies focusing on whether students perceived peer assessment as fair after practicing it. Five of these studies have found positive results (Falchikov, 1995; Li & Steckelberg, 2004; Güler, 2017; Seifert & Feliks, 2019; Vander Schee & Birrittella, 2021) and three had negative results (McConlogue, 2012; Wilson et al., 2015; Ersöz & Şad, 2018). The second type of studies focused on comparing the differential effects of specific peer assessment design elements on fairness. These studies found that: (a) peer assessment was perceived as fair as teacher's feedback (Falchikov, 1995); (b) two studies found that students perceived anonymous and non-anonymous peer assessment as equally fair (Vanderhoven et al., 2015; Güler, 2017) while another found anonymous fairer (Lin, 2018); (c) traditional peer assessment was perceived as equally fair to online peer assessment (Wen & Tsai, 2006); (d) elaborated feedback from a competent peer was perceived as fairer (Strijbos et al., 2010; Berndt et al., 2018); and (e) a rubric did not increase fairness when compared to a group that performed peer assessment without it (Panadero et al., 2013).

(Dis)comfort

The effect of discomfort → peer assessment was not explored in any of the studies. Conversely, the impact of peer assessment → (dis)comfort has been investigated in nine studies, either by examining whether after practicing peer assessment students reported being more comfortable, or through comparing the differential effects of specific peer assessment design elements on students' comfort. In three of these studies students reported feeling more comfortable after participating in peer assessment (Prins et al., 2005; Ching & Hsu, 2016; Omar

et al., 2018), and one showed a negative result (Topping et al., 2000). In one study by Stanier (1997) 40% of the participants reported being uncomfortable, yet it is unclear from the report whether the remaining 60% were comfortable. The remaining four studies showed that: (a) peer assessment versus a control group experienced the same level of comfort (Liu et al., 2016); (b) anonymity had no significant effect on comfort in one study (Vanderhoven et al., 2015) and a positive effect in another (Raes et al., 2015); and (c) a rubric did not increase comfort when compared to a group that performed peer assessment without it (Panadero et al., 2013).

RQ4. What are the results on the relationships between interpersonal factors and peer assessment?

The five interpersonal factors identified were social connections, trust in the other as assessor, psychological safety, value diversity/congruence, and interdependence. The results are detailed in Supplementary material 7 (two tables).

Social connections

Twelve studies have explored the influence of social factors such as peer familiarity, friendship marking or peer pressure on peer assessment. The effect of friendship as a social factor on peer assessment has been mostly studied for its potential bias effects. The findings of these studies are clear as four studies (Harris & Brown, 2013; Domínguez et al., 2016; Kilickaya, 2017; Ersöz & Şad, 2018) have found friendship bias effects on peer assessment scores or students' perceptions of peer assessment, while only one study did not find an effect (Panadero et al., 2013). Two studies found that when students were more aware of the negative effect of friendship on peer assessment: (a) they perceived the educational value of peer assessment more positively (Rotsaert et al., 2017), and (b) they performed more back-evaluations (Zou et al., 2018). Another study found that friendship can facilitate team performance but had a negative

effect when the outcome of the peer evaluation determines the economic profit each member receives (Corgnet, 2012).

The second line of research has explored how other social factors beyond friendship bias influence perceptions of peer assessment. One study found that the perceived quality of peer solutions depended on how the assessor perceived the assessee's personality traits (Roscoe et al., 2020). While knowing the peer was found to make students more comfortable, they also recognized that not knowing the peer led to more objective peer assessment (van Heerden & Bharuthram, 2021). Another study reported a 60% acceptance of peer assessment if students were matched considering their social relationship (Anaya et al., 2019). Finally, one study found that the social relationship could influence the depth of the feedback exchange (To & Panadero, 2019).

Five studies have explored the effects of peer assessment → social connections. In these studies, the effects of anonymity was examined, with one study finding that anonymity did not decrease peer pressure (Raes et al., 2015), another study finding that anonymity decreased peer pressure and fear of disapproval (Vanderhoven et al., 2015), and a third study finding a decrease in perceptions of friendship marking and fear of disapproval when participants experienced an anonymous peer assessment phase followed by a non-anonymous phase (Rotsaert et al., 2018). Finally, one study found that students do not like marking friends (Falchikov, 1995), and another that a quarter of the participants reported that peers prioritize personal problems or friendship causing peer assessment to not be fair (Izgar & Akturk, 2018).

Trust in the other as assessor

Four studies have examined the effects of trust in the other as assessor → peer assessment finding a clear positive pattern: higher trust was associated with adopting deeper approaches to

learning (Cheng & Tsai, 2012), and with better perceptions on the educational value of peer assessment (Rotsaert et al., 2017; Carlsson Hauff & Nilsson, 2022). These findings are not surprising given that trust in the other as assessor has been found to be a strong predictor of perceptions of peer assessment (van Gennip et al., 2010).

Eleven studies have explored the effects of peer assessment practice → trust in the other as assessor. In these studies, various effects were reported: (1) that the practice of peer assessment increased trust in others (Prins et al., 2005; van Gennip et al., 2010); (2) but also that it did not have an effect on trust in others (Rotsaert et al., 2018); (3) that trust was high after peer assessment (Ching & Hsu, 2016; Seifert & Feliks, 2019); (4) that trust was ‘adequate’ (Li & Steckelberg, 2004; Hassel & Lee, 2019); (5) that trust was low (Lin et al., 2002; Wilson et al., 2015; Ghahari & Farokhnia, 2017; To & Panadero, 2019). Thus, it is difficult to reach a clear conclusion, and one of the reasons is the disparity in the quality of the reports.

Psychological safety

Two studies explored the effects of psychological safety → peer assessment finding that this factor had positive effects on learning approaches (Cheng & Tsai, 2012) and that it was a strong predictor of conceptions of peer assessment (van Gennip et al., 2010). Regarding the reversed direction, in two studies it has been found to be higher in peer assessment than in teacher assessment (van Gennip et al., 2010; Seifert & Feliks, 2019); two specific interventions (online role-play to deliver peer assessment and asking the assessee to leave the room) increased psychological safety (Ching & Hsu, 2016; Lim et al., 2021), and a faded-anonymity intervention found no effect (Rotsaert et al., 2018).

Value diversity/congruence

When it comes to the effects of value diversity/congruence → peer assessment, two studies have found that low value diversity was associated with adopting deep learning approach (Cheng & Tsai, 2012) and value diversity negatively predicted conceptions of peer assessment (van Gennip et al., 2010). Regarding the reversed direction, two studies have found higher value congruence after peer assessment when compared to teacher assessment (van Gennip et al., 2010; van Gennip, 2012) and another study found that anonymity increased value congruence (Rotsaert et al., 2018). The findings of these studies all support the positive effect of value congruence in peer assessment and suggest that it can be achieved with practice.

Interdependence

When it comes to the interdependence → peer assessment, one study found that it did not predict conceptions of peer assessment (van Gennip et al., 2010), while another descriptive study found that students who adopted cooperative non-competitive interdependence in peer assessment also adopted deep approaches to learning (Cheng & Tsai, 2012). Regarding the reversed direction, one study did not find a significant change in interdependence after peer assessment (compared to teacher assessment) (van Gennip et al., 2010), and another study did not find a significant effect in interdependence between anonymous and non-anonymous peer assessment (Liu et al., 2018). Research on this factor is still limited, and the current findings do not clarify the role of interdependence in peer assessment since this factor relies on the assumption that peer assessment is an interactive activity which is still not the case in the majority of peer assessment studies.

Discussion

Our aim was to review the evidence on the relationship between peer assessment and intrapersonal and interpersonal factors by exploring two bidirectional relationships: (1) intrapersonal factors \leftrightarrow peer assessment, and (2) interpersonal factors \leftrightarrow peer assessment.

Characteristics of the studies (RQ1)

We found that only a limited number of studies were experimental (12%), a similar finding has been reported in a recent review with only 24% of the analysed 449 studies using (quasi-) experimental approaches (Alqassab et al., 2023). The total sample sizes of the included studies varied along with the complexity of the research design and most were dominated by females. Most studies were conducted in higher education, a finding consistently reported by previous systematic reviews and meta-analyses (Panadero & Alqassab 2019; Double et al., 2020; Alqassab et al., 2023). The subject domains were mostly (a) science, technology, engineering, and mathematics, (b) teaching and instruction, and (c) language teaching and instruction. The studies were conducted in different countries around the globe, yet more studies seemed to be conducted in Taiwan and China.

The peer assessment intervention characteristics revealed that the most used terminology was ‘peer assessment’ followed by ‘web-based peer assessment’ and ‘peer feedback’. Around 60% of the studies had peer assessment activities that involved scoring, and more than 85% of the studies used at least one type of peer assessment aid, including criteria, prompts, rubrics or a combination of them. The length of intervention varied across different studies, yet in around 40% of the studies the students practiced peer assessment between two to six times. While a quarter of the included studies did not report the anonymity mode implemented, 50% used some sort of anonymity either double-blind, single blind or a combination.

Directionality of the studies (the bidirectionality hypothesis) (RQ2)

An innovative aspect of this review is the bidirectionality hypothesis: exploring the impact of intrapersonal and interpersonal factors on peer assessment, as well as the impact of peer assessment on these variables. Although the two prior literature reviews on intrapersonal and interpersonal factors (van Gennip et al., 2009; Panadero, 2016) have instigated a systematic consideration of these factors in relation to peer assessment, they did not explicitly address the bidirectionality of these effects. We believe that the directionality perspective adopted in this review helps us to better interpret the results of different studies because it clarifies where we are putting the focus. Our results showed that the most studied aspect (75%) was how being involved in peer assessment influences intrapersonal factors that lay within the assessor or assessee such as motivation, self-efficacy or trust in self. This finding shows that researchers are focusing more on how a social activity like peer assessment impacts individuals, and not focusing so much on the social and interactional processes themselves.

This finding should be a wakeup call for the field on how we conceptualize our research. Interestingly, the social aspects of peer assessment, encompassed in the interpersonal factors, have been studied in a more balanced way, yet it is a less predominant topic than intrapersonal factors. In their review, van Gennip and colleagues (2009) highlighted that none of the interpersonal factors were linked to peer assessment learning outcomes. Our findings show that despite the growing number of studies on interpersonal factors this observation is still applicable, especially for intrapersonal factors. One explanation for this gap is that the social factors in peer assessment received attention only a decade ago after van Gennip and colleagues (2009), and it might need more time to catch up with studies addressing the intrapersonal factors.

We also explored how many studies had examined the bidirectional relationship empirically. Strikingly, out of the 69 studies there was only a single study (van Gennip et al.,

2010) that explored the bidirectional relationships. In this study, the researchers investigated reactions to 360° feedback that implies analysing the self and the external sources, which explains the adoption of the bidirectionality perspective. Nevertheless, at no point in the article the exploration were such bidirectional relationships mentioned, even though the same author (van Gennip et al., 2009) produced the first review on interpersonal variables in peer assessment. This finding points out that the bidirectionality of the relationships between peer assessment and intrapersonal and interpersonal factors, despite their importance, has been overlooked by researchers.

Intrapersonal \leftrightarrow peer assessment (RQ3)

We identified some tentative patterns. It seems that higher levels of motivation are associated with more engagement in peer assessment activities (e.g., adopting deeper approaches, less error, providing more detailed feedback) and that motivation can also contribute to positive peer assessment experience. Being involved in peer assessment may contribute to higher levels of self-efficacy. Importantly, these relationships are probably shaped by students' individual characteristics, specifically the learner's ability, as both students' motivation and self-efficacy were found to be negatively influenced by peer assessing excellent-quality work (Rogers & Feller, 2016). Although this was only investigated in one study it can serve as preliminary evidence highlighting the importance of systematically taking learners' individual characteristics into account.

Despite the importance of emotions (Alqassab et al., 2019), this factor is still not studied systematically, especially their role in shaping the peer assessment experience. The very limited evidence (one study) showed that, as can be expected, anxiety is associated with negative peer assessment outcomes (Alqassab et al., 2019). More studies explored emotions as outcomes of

peer assessment and these studies seem to suggest that positive emotions can be experienced in response to receiving positive or elaborate peer feedback, but not when the elaborated feedback was from a highly competent peer (Cheng et al., 2014). Again, this shows the significant role of the learners' individual characteristics in moderating the relationship between peer assessment and relevant intrapersonal factors.

It is difficult to extract conclusions about emotions for at least two reasons. First, emotions are a multidimensional construct that can be measured in different ways (e.g., dimensional versus discrete) (Pekrun & Bühner, 2014), and the included studies did not use the same measurement. Second, peer assessment is a complex educational phenomenon that can trigger different types of emotions depending on the process investigated (e.g., assessing, being assessed, responding to the peer comments) (Panadero, 2016; Rotsaert et al., 2017). And third, different emotions can be expected to be experienced in varying degrees depending on students' characteristics, their perceptions of their peers and the characteristics of the learning environment.

Trust in the self as assessor has received abundant attention. Considering the results of the included studies, it seems safe to assume that learners can benefit from having trust in themselves as assessors. However, participating in peer assessment can have a negative impact on this (Kilickaya, 2017; Alqassab et al., 2018). It is worthwhile to explore under what conditions this negative impact occurs and whether training or/and instructional scaffolds can alleviate it. It will be crucial to investigate what is the interpretation of the decrease in trust after practicing peer assessment. One hypothesis is that students, as novices, might over-estimate their ability to assess their peers, but after practicing peer assessment their perceptions might become more realistic (Alqassab et al., 2018).

Our findings about fairness and comfort suggest that high levels of these factors can be beneficial for peer assessment. Most of the included studies seem to report positive effects. However, it is still unclear which peer assessment design elements can contribute to higher levels of perceptions of fairness and comfort. Like trust in self, learners can report lower fairness and comfort after being involved in peer assessment (McConlogue, 2012; Stanier, 1997). The challenge is then to explore the conditions under which students' perceptions of fairness and comfort would be negatively harmed, and what peer assessment design elements can alleviate this effect.

Interpersonal \leftrightarrow peer assessment (RQ4)

These studies have a strong focus on social connections, in particular friendship. Despite the evidence that friendship can lead to bias in peer assessment and that students are aware of this bias, students seem to prefer being assessed by peers they are familiar with for psychological reasons, and when students are aware of friendship bias they are more likely to value peer assessment (Rotsaert et al., 2017). On the other hand, students seem to have issues trusting their peers to assess them (Lin et al., 2002; Ghahari & Farokhnia, 2017), but trust was found to be associated with valuing peer assessment for learning (Rotsaert et al., 2017; Carlsson Hauff & Nilsson, 2022). Linking these two findings indicates that students want to be involved in peer assessment with familiar peers while acknowledging its drawbacks.

Psychological safety, value congruence and interdependence are potential contributors to a more positive interpersonal peer assessment experience, and the challenge is to determine how to design learning environments that facilitate this. Some peer assessment interventions such as role-play (Ching & Hsu, 2016) or asking the assessee to leave the room (Lim et al., 2021) might be efficient to support psychological safety. Our review shows that the number of studies

investigating interpersonal factors is increasing. For instance, value diversity and interdependence were not identified in the studies reviewed by van Gennip and colleagues (2009) while in our review they were investigated in five and four studies respectively. According to Van Gennip et al.'s review, only one study examined psychological safety and three studied trust. From the limited number of studies published back then, Van Gennip and colleagues concluded that students seemed to feel safe and to trust themselves and their peers in peer assessment. In our review, we have identified seven studies on psychological safety and a total of 36 studies investigating trust (in self and others) in relation to peer assessment which provided a clearer picture about these interpersonal factors. Despite some positive interpersonal experiences reported by students, they can also express negative experiences.

Next steps: realistic and ideal scenarios

The poor reporting of the peer assessment intervention characteristics is still persistent as only 3 out of the 69 reviewed studies used one of the existent inventories, which clearly specify what elements need to be considered and reported (see Alqassab et al., 2023). Thus, the first crucial step we can collectively take is to ensure we use the available inventories to systematically describe all peer assessment design elements. To facilitate the design and reporting of different peer assessment designs, we have design an instrument that researchers can use and include as a supplementary material to their articles (see Supplementary material 8 via: https://osf.io/5k42z/?view_only=c77740eca9ef44978e1ac47abcaef7c). We encourage authors, reviewers, and editors, to include this instrument in every upcoming peer assessment empirical paper. This will help to clarify how peer assessment was implemented in each study, facilitate replications, systematic literature reviews and meta-analyses, and set next directions for future research.

The second step to consider is the underconceptualization of crucial aspects such as the bidirectionality hypothesis and the lack of clear definitions of the variables / constructs under investigation. Obviously, when it comes to the bidirectionality, it makes sense that it has been underconceptualized as, to our knowledge, our review is the first publication presenting this idea. However, it has also happened with the variables/constructs that have been explored. Therefore, we encourage authors, reviewers, and editors, to make a collective effort to ensure a systematic reporting of all peer assessment design characteristics in every peer assessment empirical paper, tackling the underconceptualization issue by encouraging systematic use of clearly defined constructs. For this, we hope the use our instrument presented in the previous paragraph, could significantly help the field.

The third step to consider is the lack of replication studies in the field and the massive presence of weak research design. We urgently need stronger empirical research, as has been emphasized before (see previous reviews with specific recommendations e.g., Panadero, 2016; Panadero & Alqassab, 2019), as well as replication studies to have solid conclusions about the relationships between intrapersonal and interpersonal factors and peer assessment.

Limitations

We conducted a narrative review which is a qualitative approach that does not provide conclusions based on statistical analysis. However, a meta-analysis with the aims of our study is not feasible for two main reasons. First, the number of studies per each intrapersonal and interpersonal factors is small if we exclude qualitative non-experimental studies. Second, and more importantly, the diversity of studies would not be clearly captured as a meta-analysis requires more homogeneous studies.

Another limitation is the very small number of studies included in our review exploring group peer assessment. Obviously, group peer assessment can have different intrapersonal and interpersonal implications.

Lastly, due to the large number of descriptive and correlational studies we had to infer the directionality of the relationships between peer assessment and the intrapersonal and interpersonal factors of interest from the research questions or aims of these studies. Some studies also explored the relationships between different intra and interpersonal factors (Strijbos, Pat-El & Narciss, 2021), which adds to the challenge of conducting a systematic review of their effects. We hope that the instrument we are providing might help with this issue.

Conclusion

Peer assessment is a powerful instructional and learning activity but full of complexities. While the number of studies exploring the effects of and on intrapersonal and interpersonal factors has increased, there seems to be a heterogeneity in how these factors are investigated in relation to peer assessment. Our review presents four key outcomes that could contribute to a better understanding of such heterogeneity: (1) a new categorization of intrapersonal and interpersonal factors; (2) the idea of the bidirectional relationships between peer assessment and intrapersonal and interpersonal factors that is new to the field and could shed light on our research; (3) an overview of the empirical results of studies on interpersonal and intrapersonal factors of peer assessment; and (4) an instrument to report the characteristics of future peer assessment studies to facilitate better reports and research design. Seven years ago, one of us (Panadero, 2016) proposed a shift from exclusively exploring “peer scoring” to start paying more attention to the intrapersonal and interpersonal factors which were brought to the peer assessment literature by Van Gennip et al. (2009). We hope that in some years, when someone

reviews the literature again with the help of better reporting and the inclusion of the instrument we propose here, we will be able to extract even more specific guidelines for the implementation and investigation of peer assessment. Until then, we adhere to Panadero's view (2016) that formative implementations of peer assessment might produce better intrapersonal and interpersonal results as they offer more insights to the students about the why, what, and how of peer assessment.

References

(* the studies included in the review).

- Adachi, C., Tai, J., & Dawson, P. (2018). A framework for designing, implementing, communicating and researching peer assessment. *Higher Education Research & Development, 37*(3), 453–467. <https://doi.org/10.1080/07294360.2017.1405913>
- Alqassab, M., Strijbos, J., Panadero, E., Fernández Ruiz, J., Warren, M., & To, J. (In press). A systematic review of peer assessment design elements. *Educational Psychology Review*.
- *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>

- *Anaya, A. R., Luque, M., Letón, E., & Hernández-del-Olmo, F. (2019). Automatic assignment of reviewers in an online peer assessment task based on social interactions. *Expert Systems*, 36(4), e12405. <https://doi.org/10.1111/exsy.12405>
- *Berndt, M., Strijbos, J.-W., & Fischer, F. (2018). Effects of written peer-feedback content and sender's competence on perceptions, performance, and mindful cognitive processing. *European Journal of Psychology of Education*, 33(1), 31–49. <https://doi.org/10.1007/s10212-017-0343-z>
- *Carlsson Hauff, J., & Nilsson, J. (2021). Students' experience of making and receiving peer assessment: The effect of self-assessed knowledge and trust. *Assessment & Evaluation in Higher Education*, 0(0), 1–13. <https://doi.org/10.1080/02602938.2021.1970713>
- *Carvalho, A. (2013). Students' perceptions of fairness in peer assessment: Evidence from a problem-based learning course. *Teaching in Higher Education*, 18(5), 491–505. <https://doi.org/10.1080/13562517.2012.753051>
- *Chang, S.-C., Hsu, T.-C., & Jong, M. S.-Y. (2020). Integration of the peer assessment approach with a virtual reality design system for learning earth science. *Computers & Education*, 146, 103758. <https://doi.org/10.1016/j.compedu.2019.103758>
- *Chen, I.-C., Hwang, G.-J., Lai, C.-L., & Wang, W.-C. (2020). From design to reflection: Effects of peer-scoring and comments on students' behavioral patterns and learning outcomes in musical theater performance. *Computers & Education*, 150, 103856. <https://doi.org/10.1016/j.compedu.2020.103856>
- *Cheng, K.-H., Hou, H.-T., & Wu, S.-Y. (2014). Exploring students' emotional responses and participation in an online peer assessment activity: A case study. *Interactive Learning Environments*, 22(3), 271–287. <https://doi.org/10.1080/10494820.2011.649766>

- *Cheng, K.-H., & Tsai, C.-C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology*, 28(4), Article 4. <https://doi.org/10.14742/ajet.830>
- *Chien, S.-Y., Hwang, G.-J., & Jong, M. S.-Y. (2020). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-Speaking performance and learning perceptions. *Computers & Education*, 146, 103751. <https://doi.org/10.1016/j.compedu.2019.103751>
- *Ching, Y.-H., & Hsu, Y.-C. (2016). Learners' Interpersonal Beliefs and Generated Feedback in an Online Role-Playing Peer- Feedback Activity: An Exploratory Study. *The International Review of Research in Open and Distributed Learning*, 17(2). <https://doi.org/10.19173/irrodl.v17i2.2221>
- *Corgnet, B. (2012). Peer evaluations and team performance: When friends do worse than strangers. *Economic Inquiry*, 50(1), 171–181. <https://doi.org/10.1111/j.1465-7295.2010.00354.x>
- Dochy, F. (2006). *A guide for writing scholarly articles or reviews for the Educational Research Review*. <https://www.sciencedirect.com/journal/educational-research-review>
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331–350. <https://doi.org/10.1080/03075079912331379935>
- *Domínguez, C., Jaime, A., Sánchez, A., Blanco, J. M., & Heras, J. (2016). A comparative analysis of the consistency and difference among online self-, peer-, external- and instructor-assessments: The competitive effect. *Computers in Human Behavior*, 60, 112–120. <https://doi.org/10.1016/j.chb.2016.02.061>

- 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>
- *Ersöz, Y., & Şad, S. N. (2018). Facebook as a Peer-Assessment Platform: A Case Study in Art Teacher Education Context. *International Journal of Assessment Tools in Education*, 5(4), 740–753. <https://doi.org/10.21449/ijate.478277>
- *Falchikov, N. (1995). Peer Feedback Marking: Developing Peer Assessment. *Innovations in Education and Training International*, 32(2), 175–187. <https://doi.org/10.1080/1355800950320212>
- *Feyli, S., & Ayatollahi, M. A. (2016). The Effects of Classroom Peer-feedback on Advanced Students' Self- Confidence. *Journal of Modern Research in English Language Studies*, 3(1), 87–67.
- *Filius, R. M., Kleijn, R. A. M., Uijl, S. G., Prins, F. J., Rijen, H. V. M., & Grobbee, D. E. (2019). Audio peer feedback to promote deep learning in online education. *Journal of Computer Assisted Learning*, 35(5), 607–619. <https://doi.org/10.1111/jcal.12363>
- *Ghahari, S., & Farokhnia, F. (2017). Triangulation of language assessment modes: Learning benefits and socio-cognitive prospects. *Pedagogies: An International Journal*, 12(3), 275–294. <https://doi.org/10.1080/1554480X.2017.1342540>
- *Ghahari, S., & Farokhnia, F. (2018). Peer versus teacher assessment: Implications for CAF triad language ability and critical reflections. *International Journal of School & Educational Psychology*, 6(2), 124–137. <https://doi.org/10.1080/21683603.2016.1275991>

- Gielen, S., Dochy, F., Onghena, P., Struyven, K., & Smeets, S. (2011). Goals of peer assessment and their associated quality concepts. *Studies in Higher Education, 36*(6), 719–735.
<https://doi.org/10.1080/03075071003759037>
- *Güler, Ç. (2017). Use of WhatsApp in Higher Education: What's Up With Assessing Peers Anonymously? *Journal of Educational Computing Research, 55*(2), 272–289.
<https://doi.org/10.1177/0735633116667359>
- *Guo, X., & Lei, P.-W. (2020). Effect of Quality Characteristics of Peer Raters on Rating Errors in Peer Assessment. *International Journal of Testing, 20*(3), 206–230.
<https://doi.org/10.1080/15305058.2020.1720216>
- *Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*, 101–111.
<https://doi.org/10.1016/j.tate.2013.07.008>
- *Hassell, D., & Lee, K. Y. (2019). Evaluation of Self and Peer Assessments in a Second Year Engineering Module. *IAFOR Journal of Education, 7*(2), 105–130.
<https://doi.org/10.22492/ije.7.2.06>
- *Hou, H.-T., & Cheng, K.-H. (2012). Analyzing the latent emotional transfer pattern (LETP) of a learning community in an online peer-assessment activity. *British Journal of Educational Technology, 43*(4), E113–E116. <https://doi.org/10.1111/j.1467-8535.2012.01301.x>
- *Hsia, L.-H., Huang, I., & Hwang, G.-J. (2016). A web-based peer-assessment approach to improving junior high school students' performance, self-efficacy and motivation in performing arts courses. *British Journal of Educational Technology, 47*(4), 618–632.
<https://doi.org/10.1111/bjet.12248>

- *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>
- *Hwang, G.-J., Hung, C.-M., & Chen, N.-S. (2014). Improving learning achievements, motivations and problem-solving skills through a peer assessment-based game development approach. *Educational Technology Research and Development*, 62(2), 129–145. <https://doi.org/10.1007/s11423-013-9320-7>
- Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., Boud, D., Rotsaert, T., Brown, S., Salazar, M. L. S., & Gómez, H. M. R. (2020). El futuro de la evaluación en la educación superior. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 26(1), Article 1.
<https://doi.org/10.7203/relieve.26.1.17323>
- *Izgar, G., & Akturk, A. O. (2018). A Mixed Method Research on Peer Assessment. *International Journal of Evaluation and Research in Education (IJERE)*, 7(2), 118–126.
<https://doi.org/10.11591/ijere.v7i2.12770>
- *Johnson, N., & Winterbottom, M. (2011). Supporting girls' motivation in science: A study of peer- and self-assessment in a girls-only class. *Educational Studies*, 37(4), 391–403.
<https://doi.org/10.1080/03055698.2010.508605>
- *Kilickaya, F. (2017). Peer assessment of group members in tertiary contexts. In M. Sowa & J. Krajka (Eds.), *Innovations in languages for specific purposes—Present challenges and future promises* (pp. 329–343). Peter Lang.
- *Kim, M., & Ryu, J. (2013). The development and implementation of a web-based formative peer assessment system for enhancing students' metacognitive awareness and

- performance in ill-structured tasks. *Educational Technology Research and Development*, 61(4), 549–561. <https://doi.org/10.1007/s11423-012-9266-1>
- *Kuo, F.-C., Chen, J.-M., Chu, H.-C., Yang, K.-H., & Chen, Y.-H. (2017). A Peer-Assessment Mobile Kung Fu Education Approach to Improving Students' Affective Performances. *International Journal of Distance Education Technologies (IJDET)*, 15(1), 1–14. <https://doi.org/10.4018/IJDET.2017010101>
- *Lai, C.-L., & Hwang, G.-J. (2015). An interactive peer-assessment criteria development approach to improving students' art design performance using handheld devices. *Computers & Education*, 85, 149–159. <https://doi.org/10.1016/j.compedu.2015.02.011>
- Li, H., Xiong, Y., 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. (2019). Using game-based training to improve students' assessment skills and intrinsic motivation in peer assessment. *Innovations in Education and Teaching International*, 56(4), 423–433. <https://doi.org/10.1080/14703297.2018.1511444>
- *Li, L., Gao, F., & Guo, S. (2020). The effects of social messaging on students' learning and intrinsic motivation in peer assessment. *Journal of Computer Assisted Learning*, 36(4), 439–448. <https://doi.org/10.1111/jcal.12409>
- *Li, L., & Steckelberg, A. (2004). *Using Peer Feedback to Enhance Student Meaningful Learning*. Paper presented at the Conference of the Association for Educational Communications and Technology, Chicago, USA

- *Lim, W., Son, J.-W., & Kang, S.-H. (2021). How Reducing Discomfort Impacts Peer Assessments of Preservice Teachers. *Sustainability*, *13*(11), 6435.
<https://doi.org/10.3390/su13116435>
- *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>
- *Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2002). Student Attitudes toward Networked Peer Assessment: Case Studies of Undergraduate Students and Senior High School Students. *International Journal of Instructional Media*, *29*(2), 241–254.
- *Liu, C.-C., Lu, K.-H., Wu, L. Y., & Tsai, C.-C. (2016). The Impact of Peer Review on Creative Self-efficacy and Learning Performance in Web 2.0 Learning Activities. *Journal of Educational Technology & Society*, *19*(2), 286–297.
- *Liu, X., Li, L., & Zhang, Z. (2018). Small group discussion as a key component in online assessment training for enhanced student learning in web-based peer assessment. *Assessment & Evaluation in Higher Education*, *43*(2), 207–222.
<https://doi.org/10.1080/02602938.2017.1324018>
- *McConlogue, T. (2012). But is it fair? Developing students' understanding of grading complex written work through peer assessment. *Assessment & Evaluation in Higher Education*, *37*(1), 113–123. <https://doi.org/10.1080/02602938.2010.515010>
- *McMahon, T. (2010). Combining peer-assessment with negotiated learning activities on a day-release undergraduate-level certificate course (ECTS level 3). *Assessment & Evaluation in Higher Education*, *35*(2), 223–239. <https://doi.org/10.1080/02602930902795919>

- *Omar, S. N. P., Shahrill, M., & Sajali, M. Z. (2018). The Use of Peer Assessment to Improve Students' Learning of Geometry. *European Journal of Social Science Education and Research*, 5(2), 203. <https://doi.org/10.26417/ejsr.v5i2.p203-222>
- Panadero, E. (2016). Is It Safe? Social, Interpersonal, and Human Effects of Peer Assessment: A Review and Future Directions. In *Handbook of Human and Social Conditions in Assessment* (pp. 247–266). Routledge.
- *Panadero, E., Romero, M., & Strijbos, J.-W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, 39(4), 195–203. <https://doi.org/10.1016/j.stueduc.2013.10.005>
- 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>
- Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotions. In *International handbook of emotions in education* (pp. 571-589). Routledge.
- *Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. (2005). Formative peer assessment in a CSCL environment: A case study. *Assessment & Evaluation in Higher Education*, 30(4), 417–444. <https://doi.org/10.1080/02602930500099219>
- *Raes, A., Vanderhoven, E., & Schellens, T. (2015). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education*, 40(1), 178–193. <https://doi.org/10.1080/03075079.2013.823930>

Reinholz, D. (2016). The assessment cycle: A model for learning through peer assessment.

Assessment & Evaluation in Higher Education, 41(2), 301–315.

<https://doi.org/10.1080/02602938.2015.1008982>

*Rogers, T., & Feller, A. (2016). Discouraged by Peer Excellence: Exposure to Exemplary Peer Performance Causes Quitting. *Psychological Science*, 27(3), 365–374.

<https://doi.org/10.1177/0956797615623770>

*Roscoe, R., Wilson, J., Patchan, M., Chen, D., & Johnson, A. (2020). Modeling Student Evaluations of Writing and Authors as a Function of Writing Errors. *Journal of Language and Education*, 6(2), 147–164. <https://doi.org/10.17323/jle.2020.10316>

Rotsaert, T. (2017). *The social nature of peer assessment in secondary and higher education: Examining students' perceptions on interpersonal processes and peer feedback quality in anonymous face-to-face settings using Mobile Response Technology* [Dissertation]. University of Ghent.

*Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. *Studies in Educational Evaluation*, 53, 29–40.

<https://doi.org/10.1016/j.stueduc.2017.02.003>

*Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education*, 33(1), 75–99. <https://doi.org/10.1007/s10212-017-0339-8>

- *Seifert, T., & Feliks, O. (2019). Online self-assessment and peer-assessment as a tool to enhance student-teachers' assessment skills. *Assessment & Evaluation in Higher Education*, 44(2), 169–185. <https://doi.org/10.1080/02602938.2018.1487023>
- *Shen, B., Bai, B., & Xue, W. (2020). The effects of peer assessment on learner autonomy: An empirical study in a Chinese college English writing class. *Studies in Educational Evaluation*, 64, 100821. <https://doi.org/10.1016/j.stueduc.2019.100821>
- *Sluijsmans, D., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2002). Peer Assessment Training in Teacher Education: Effects on performance and perceptions. *Assessment & Evaluation in Higher Education*, 27(5), 443–454. <https://doi.org/10.1080/0260293022000009311>
- *Stanier, L. (1997). Peer assessment and group work as vehicles for student empowerment: A module evaluation. *Journal of Geography in Higher Education*, 21(1), 95–98. <https://doi.org/10.1080/03098269708725413>
- *Strijbos, J.-W., Narciss, S., & Dünnebieer, 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>
- Strijbos, J.-W., Ochoa, T. A., Sluijsmans, D. M. A., Segers, M. S. R., & Tillema, H. H. (2009). Fostering Interactivity through Formative Peer Assessment in (Web-Based) Collaborative Learning Environments: In C. Mourlas, N. Tsianos, & P. Germanakos (Eds.), *Cognitive and Emotional Processes in Web-Based Education* (pp. 375–395). IGI Global. <https://doi.org/10.4018/978-1-60566-392-0.ch018>

- *Strijbos, J.-W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the Feedback Perceptions Questionnaire. *Studies in Educational Evaluation*, 68, 100980. <https://doi.org/10.1016/j.stueduc.2021.100980>
- Strijbos, J.-W., & Wichmann, A. (2018). Promoting learning by leveraging the collaborative nature of formative peer assessment with instructional scaffolds. *European Journal of Psychology of Education*, 33(1), 1–9. <https://doi.org/10.1007/s10212-017-0353-x>
- Tillema, H. H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning – A review of research studies. *Studies in Educational Evaluation*, 37(1), 25-34. <https://doi.org/10.1016/j.stueduc.2011.03.004>
- *To, J., & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assessment & Evaluation in Higher Education*, 44(6), 920–932. <https://doi.org/10.1080/02602938.2018.1548559>
- Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, 68(3), 249–276. <https://doi.org/10.3102/00346543068003249>
- Topping, K. (2003). Self and Peer Assessment in School and University: Reliability, Validity and Utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising New Modes of Assessment: In Search of Qualities and Standards* (pp. 55–87). Springer Netherlands. https://doi.org/10.1007/0-306-48125-1_4
- *Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative Peer Assessment of Academic Writing Between Postgraduate Students. *Assessment & Evaluation in Higher Education*, 25(2), 149–169. <https://doi.org/10.1080/713611428>

- van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*(3), 341–356.
<https://doi.org/10.1080/03075070600680836>
- *van Gennip, N. A. E. (2012). *Assessing together: Peer assessment from an interpersonal perspective* [Leiden University]. <https://hdl.handle.net/1887/20012>
- 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>
- *van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*(4), 280–290.
<https://doi.org/10.1016/j.learninstruc.2009.08.010>
- *van Heerden, M., & Bharuthram, S. (2021). Knowing me, knowing you: The effects of peer familiarity on receiving peer feedback for undergraduate student writers. *Assessment & Evaluation in Higher Education, 1–11*. <https://doi.org/10.1080/02602938.2020.1863910>
- van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270–279. <https://doi.org/10.1016/j.learninstruc.2009.08.004>
- *Vander Schee, B. A., & Birrittella, T. D. (2021). Hybrid and Online Peer Group Grading: Adding Assessment Efficiency While Maintaining Perceived Fairness. *Marketing Education Review, 1–9*. <https://doi.org/10.1080/10528008.2021.1887746>

- *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>
- *Wen, M. L., & Tsai, C.-C. (2006). University Students' Perceptions of and Attitudes Toward (Online) Peer Assessment. *Higher Education, 51*(1), 27–44. <https://doi.org/10.1007/s10734-004-6375-8>
- *Wilson, M. J., Diao, M. M., & Huang, L. (2015). 'I'm not here to learn how to mark someone else's stuff': An investigation of an online peer-to-peer review workshop tool. *Assessment & Evaluation in Higher Education, 40*(1), 15–32. <https://doi.org/10.1080/02602938.2014.881980>
- *Yu, S. (2021). Giving genre-based peer feedback in academic writing: Sources of knowledge and skills, difficulties and challenges. *Assessment & Evaluation in Higher Education, 46*(1), 36–53. <https://doi.org/10.1080/02602938.2020.1742872>
- *Zhou, J., Zheng, Y., & Tai, J. H.-M. (2020). Grudges and gratitude: The social-affective impacts of peer assessment. *Assessment & Evaluation in Higher Education, 45*(3), 345–358. <https://doi.org/10.1080/02602938.2019.1643449>
- *Zou, Y., Schunn, C. D., Wang, Y., & Zhang, F. (2018). Student attitudes that predict participation in peer assessment. *Assessment & Evaluation in Higher Education, 43*(5), 800–811. <https://doi.org/10.1080/02602938.2017.1409872>

Supplementary material 1. Detailed coding scheme of the peer assessment design elements investigated in the current review

#	Information	Description
1	Author information	Name of authors
2	Year	Year of publication
3	Study aim/s	Aim/s of the study (if mentioned)
4	Research questions	RQs of the study (if mentioned)
5	Sample demographics	Participants' description and country/region of the study
6	Educational stage	Primary, Secondary, Higher Education, Other
7	Variables of interest	Variables investigated in the study which are of interest for this review (described in table 1).
8	PA terminology used	Term used for referring to Peer assessment (e.g., peer rating, peer feedback)
9	Type of PA activity	Description of how the PA activity was performed.
10	Grading	Did the task involve giving grades?
11	Reporting of PA typology elements	Did the study use a PA constellation (e.g., Topping's, 1998)
12	PA aids	Aids provided to support/scaffold peer assessment (e.g., rubrics/prompts)
13	Length of intervention/training	Total duration of the PA training and activity.
14	Anonymity	Double blind, single blind (for the assessor or assessee), or public
15	Task and subject	Description of the task/output assessed in the PA activity and the subject
16	Independent variable/s	All study's IVs (if mentioned) not limited to our variables of interest
17	Dependent variable/s	All study's DVs (if mentioned) not limited to our variables of interest
18	Type of data collection	Tools used in data collection (e.g., self-report questionnaires, interviews, etc.)
19	Research method	Research method used, based on Montero and Leon (2007): (a) Descriptive design or multiple case study (DSO), (b) quasi-experimental study (ERIG), (c) experimental study (ERG)
20	Gender differences	Was gender difference explored, yes or no.
21	Procedure	A summary of the procedure followed in the study
22	Results	A summary of the main findings about our variables of interest
23	Conclusions	Study conclusions about our variables of interest
24	Direction of effect	Synthesis of results (only for studies with experimental manipulation)
25	Directionality	Intrapersonal → PA, Interpersonal → PA, PA → Intrapersonal and PA → Interpersonal
26	Observations	Other remarks about the article

Note. PA = peer assessment; PF = peer feedback

Supplementary material 2. Database available upon request to the contact author

Excel file

Supplementary material 3. Study quality rubric

		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)
Quality of measurements		The study uses ad-hoc instruments without reporting reliability indexes	The study uses ad-hoc instruments reporting reliability indexes	The study uses validated instruments only
Pedagogical quality	Training	No training	Simple (e.g., just giving a short talk about PA)	Advanced (e.g., giving feedback on PA)
	Practice (Times they give or receive feedback)	1	2-5	6 or +
	PA aids	0	1	2 or +
Quality of DVs		Students' perceptions of...	SRL measured by self-report	"Objective" variables (academic performance, etc.)
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

Supplementary material 4. Research design characteristics

#	Author(s) (year)	Research design	Sample size	Gender distribution	Educational level	Subject domain	Country
1	Alqassab et al., 2018	ERIG	N = 43	34 females & 9 males	Higher education	Mathematics Education	Germany
2	Alqassab et al., 2019	ERG	N = 53; n _{control} = 27; n _{experimental} = 26	39 females & 14 males	Higher education	Geometry (Mathematics education)	Germany
3	Anaya et al., 2019	DSO	N = 100, task performed by 51	Not stated	Higher education	Teaching (Bachelor in computer engineering)	Not stated
4	Berndt et al., 2018	ERG	N = 45; n _{CGF-H} = 10; n _{CGF-L} = 12; n _{ESF-H} = 11; and n _{ESF-L} = 12.	35 females & 10 males	Higher education	Academic Writing	Germany
5	Carlsson Hauff & Nilsson, 2021	ERIG	N = 94	48% females	Higher education	Marketing	Sweden
6	Carvalho, 2013	DSO	N = 120	82 females & 38 males	Higher education	Business case (Management)	Portugal
7	Chang et al., 2020	ERIG	N = 51	Not stated	Elementary education	Creation of a VR object (Natural sciences)	Not stated
8	Chen et al., 2020	ERIG	N = 141 (n _{peer scoring} = 47, n _{peer comments} = 49, N _{PC+PS} = 45)	Not stated	Higher education	Musical theatre	Taiwan
9	Cheng & Tsai, 2012	DSO	N = 23	15 females & 8 males	Higher education	Communication technology	Taiwan
10	Cheng et al., 2014	DSO	N = 65	58 females & 7 males	Higher education	Communication technology	Taiwan
11	Chien et al., 2020	ERIG	N = 69; n _{control} = 36; n _{experimental} = 33	Not stated	Secondary education	English	Taiwan
12	Ching & Hsu, 2016	DSO	N = 16	6 females & 10 males	Higher education	Instructional design	USA
13	Corgnet, 2012	ERG	N = 66; n _{asymmetric peer evaluation} = 24; n _{symmetric peer evaluation} = 21; n _{equal split (no PA)} = 21	Not stated	Higher education	Mathematics	Spain
14	Domínguez et al., 2016	ERIG	N = 97	17 females & 80 males	Higher education	Computing project management	Spain
15	Ersoz & Sad, 2018	DSO	N = 10	2 females & 8 males	Higher education	Instructional technologies, material design	Turkey
16	Falchikov, 1995	DSO	N = 13	12 females & 1 male	Higher education	Human developmental psychology	Not stated
17	Feyli & Ayatollahi, 2016	ERIG	N = 60 (57 with filled questionnaires)	14 females & 46 males	Mixed (EFL learners)	English	Iran
18	Filius et al., 2019	DSO	N = 108; 14 participants interviewed	68 females & 38 males (2 not responded)	Mixed (College students enrolled in the Small Private Online Courses (SPOC) and other individuals enrolled at a Massive Online Open Courses (MOOC)).	Public health, human rights, clinical epidemiology, child development	Europe, 59.3%; Asia, 20.4%; Africa, 8.3%; North America, 6.5%; and South America, 5.6%

#	Author(s) (year)	Research design	Sample size	Gender distribution	Educational level	Subject domain	Country
19	Ghahari & Farokhnia, 2017	ERIG	N = 39; n _{peer assessment} = 12; n _{teacher assessment} = 13; n _{self-assessment} = 14	26 females & 13 males	Mixed (Secondary and Higher education)	English	Iran
20	Ghahari & Farokhnia, 2018	ERIG	N = 39; n _{peer assessment} = 12; n _{teacher assessment} = 13; n _{self-assessment} = 14	26 females & 13 males	Mixed (Secondary and Higher education)	English	Iran
21	Güler, 2017	ERIG	N = 84	34 females & 50 males	Higher education	Computer network and communication	Turkey
22	Guo & Lei, 2020	ERIG	N = 838	406 females & 432 males	Secondary education	Writing (Chinese)	China
23	Harris & Brown, 2013	DSO	Not stated (3 teachers and their students)	Not stated	Mixed (Primary, secondary, & intermediate education)	English	New Zealand
24	Hassel & Lee, 2019	DSO	N = 96; 59 participated in peer assessment and 65 filled questionnaires	Not stated	Higher education	Engineering (petroleum refining)	Brunei
25	Hou & Cheng, 2012	DSO	N = 65	Not stated	Higher education	Multimedia production	Taiwan
26	Hsia et al., 2016	ERIG	N = 163; n _{control} = 90; n _{experimental} = 73	82 females & 81 males	Secondary education	Performing arts	Taiwan
27	Huisman et al., 2018	ERIG/ ERG ¹	N = 83	77 females & 6 males	Higher education	Educational sciences/Family pedagogy	Netherlands
28	Hwang et al., 2014	ERIG	N = 167; n _{control} = 85; n _{experimental} = 82	Not stated	Elementary education	Game development (Science)	Taiwan
29	Izgar & Akturk, 2018	DSO	N = 27	Not stated	Higher education	Special teaching methods	Turkey
30	Johnson & Winterbottom, 2011	DSO	N = 28	28 females	Secondary education	Science	Not stated
31	Kilickaya, 2017	DSO	N = 35	25 females & 10 males	Higher education	Teaching Language skills	Turkey
32	Kim & Ryu, 2013	ERG	N = 122 n _{Web-based formative PA} = 42, n _{peer assessment} = 39, n _{self-assessment} = 41	Not stated	Higher education	Instructional design	Korea
33	Kuo et al., 2017	ERIG	N = 42; n _{control} = 22; n _{experimental} = 20	27 females & 15 males	Higher education	Physical education	Taiwan
34	Lai & Hwang, 2015	ERIG	N = 103; n _{control} = 52; n _{experimental} = 51	Not stated	Secondary education	Artwork design	Taiwan
35	Li, 2019	ERIG	N = 48	35 females & 13 males	Higher education	Not stated	Not stated
36	Li, Gao & Guo, 2020	ERIG	N = 79	71 females & 8 males	Higher education	English reading and writing	China
37	Li & Steckelberg, 2004	ERIG	N = 48; n _{control} = 21; n _{experimental} = 27	Not stated	Higher education	Instructional technology	USA

¹ Not stated how the assignment to experimental groups was made.

#	Author(s) (year)	Research design	Sample size	Gender distribution	Educational level	Subject domain	Country
38	Lim et al., 2021	DSO	N= 81 students (25 students (follow up interview))	52 females & 29 males	Higher education	Mathematics	USA
39	Lin, 2018	ERG	N = 32 students n experimental = 16; n control = 16	23 females & 9 males	Higher education	Teacher education	Taiwan
40	Lin et al., 2002	DSO	N Secondary= 57; N Higher Education = 41	Secondary: 88% males & 12% females; undergraduate: 85% males & 15% females	Mixed (Secondary and Higher education)	Electrical Engineering & Operating Systems	Taiwan
41	Liu et al., 2016	ERIG	N = 53; n Control = 26; n Experimental = 27	26 females & 27 males	Elementary education	Web 2.0 Storytelling Course.	Taiwan
42	Liu et al., 2017	ERIG	N = 81	71 females and 10 males	Higher education	Foreign Trade Practical Operation Process.	China
43	McConlogue, 2012	DSO	N = 82; n Masters = 26; n Mixed = 56	Not stated	Higher education	Engineering	United Kingdom
44	McMahon, 2010	DSO	N = 5	Not stated	Higher education	Training & Development Module	Not stated
45	Omar et al., 2018	DSO	N = 23	23 females	Secondary education	Mathematics	Brunei
46	Panadero et al., 2013	ERIG	N = 209; n Rubric = 104; n Non-rubric = 105	182 females & 27 males	Higher education	Learning & Development	Spain
47	Prins et al., 2005	DSO	N = 27	14 females & 13 males	Higher education	Agriculture, Water Management, Energy, Spatial Planning.	Belgium, Germany, Netherlands, Poland, and Spain
48	Raes et al., 2015	ERIG	N = 51; n Anonymous = 27; N Non-Anonymous = 24	92% females	Higher education	Educational Studies Course	Not stated
49	Rogers & Feller, 2016	ERG	N= 361	48% females	Mixed	Not stated	Not stated
50	Roscoe et al., 2020	DSO	N = 70	34.3% females	Higher education	Not stated	USA
51	Rotsaert et al., 2018	ERIG	N = 46	84.44% females	Higher education	Educational Studies Course	Not stated
52	Rotsaert et al., 2017	DSO	N = 3680; n Level 1 = 32.2%; n Level 2 = 34.4%; n Level 3 = 33.5%	1675 females & 1398 males	Secondary education	Not stated	Belgium.
53	Seifert & Felix, 2019	DSO	N = 300; n Bachelor's = 169; n Master's = 93	Not stated	Higher education	Teaching and Learning in Technological Environments	Not stated
54	Shen et al., 2020	ERG	N = 70	65 females & 5 males	Higher education	English (EFL)	China
55	Sluijsmans et al., 2002	ERIG	N = 93; n Experimental = 50; n Control = 43	74 females & 19 males	Higher education	Designing Creative Lessons Course (For the Domains Art, Dutch Language, & Music)	Netherlands
56	Stanier, 1997	DSO	N = 36	Not stated	Higher education	Interdisciplinary Studies in Environmental and Applied Sciences	United Kingdom

#	Author(s) (year)	Research design	Sample size	Gender distribution	Educational level	Subject domain	Country
57	Strijbos et al., 2010	ERG	N = 89; n _{CGF-H} = 18; n _{CGF-L} = 18; n _{ESF-H} = 18; and n _{ESF-L} = 17.	68 females & 21 males	Higher education	Academic Writing	Germany
58	Strijbos et al., 2021	DSO	N = 1486	817 females & 713 males	Secondary education	Writing	Netherlands
59	To & Panadero, 2019	DSO	N = 11; n Teachers = 7	6 females and 5 males	Higher education	University English, Educational Issues I And II and Human Development Theories, And Disaster Management and Society Development in Asia	Hong Kong
60	Topping et al., 2000	DSO	N = 12	10 females, 2 males	Higher education	Educational Psychology	Scotland
61	Van Gennip, 2012	ERIG	N = 106 students; n _{PA condition} = 25; n _{PA+ condition} = 45, n _{teacher-assessment condition} = 36	Not stated	Secondary education	Metal work & electronics	Netherlands
62	Van Gennip et al., 2010	ERIG	N = 62; n _{Experimental} = 45; n _{Control} = 17	62 males	Secondary education	Metal work & electronics	Netherlands
63	Van Heerden & Bharuthram, 2021	DSO	N = 64	Not stated	Higher education	Not stated	South Africa
64	Vander Schee & Birrittella, 2021	DSO	N = 827; n _{Pre COVID} = 697; n _{Post-COVID} = 130	52% females	Higher education	Introduction to Marketing	USA
65	Vanderhoven et al., 2015	ERIG	N = 69; N _{Experimental} = 33 N _{Control} = 36	72% females	Secondary education	Not stated	Belgium
66	Wen & Tsai, 2006	DSO	N = 280; n _{Undergraduates} = 85.1%; n _{Graduates} = 14.9%	58.1% males and 41.9% females	Higher education	Not stated	Taiwan
67	Wilson et al., 2015	DSO	N = 282	Not stated	Higher education	Science of Sex; Financial Techniques, Instruments and Markets; Finance and Finance Reporting, Intercultural Relations; International Television and Beyond; Transnational Communication; Key Competencies in Inclusive Education; Integrated Marketing Communications; And, History in The Secondary School II.	Australia

#	Author(s) (year)	Research design	Sample size	Gender distribution	Educational level	Subject domain	Country
68	Zhu et al., 2019	DSO	N = 149; n Course A = 70; n Course B = 79	Not stated	Higher education	English Writing	China
69	Zou et al., 2018	DSO	N = 105; n Engineering = 59%; n English = 41%	54% Females	Higher education	Engineering Students: English Reading and Writing Class; and, English Students: Advanced Thesis Writing for EFL Students Class.	China

Supplementary material 5. Peer assessment intervention characteristics

#	Author(s) (year)	PA terminology	Did PA contain grades?	Use of PA aids	Length of intervention	Anonymity ²	PA task & subject	Constellations
1	Alqassab et al., 2018	Peer feedback	No	Yes (criteria, prompts, & worked-example)	6 weeks (including 2 intensive training sessions); 4 practice sessions	Fictional assessee	Geometry construction task. Mathematics education.	No
2	Alqassab et al., 2019	Peer feedback	No	No	1 practice session with no training.	Fictional assessee	Geometry proof. Mathematics for pre-service teachers.	No
3	Anaya et al., 2019	OPA (online peer assessment)	Yes	Yes (Rubric)	No intervention	Not stated	Modular teaching mini-video. Computer engineering.	No
4	Berndt et al., 2018	peer feedback	No	Yes (Criteria)	1 practice session with no training.	Fictional assessor	Written text. Academic writing.	No
5	Carlsson Hauff & Nilsson, 2021	Peer assessment	No	No	None. But PA activity was done 6 times, no time element mentioned.	Not stated	Peer assessment on six reports discussing 2-3 articles related with consumer behavior. Master's course in marketing.	No
6	Carvalho, 2013	Peer assessment	Yes	Yes (Criteria)	No training	Single blind (assessee)	Contribution to teamwork (process); Business Case (Management course)	No
7	Chang et al., 2020	Peer assessment	No	Yes (Rubric)	They do not report training, and they only did two sessions of practice.	Not stated	Creation of an object in a VR environment. Natural sciences.	No
8	Chen et al., 2020	Peer assessment	Yes, for the scoring only and scoring + commenting conditions; No for the commenting only condition.	Yes (Rubric)	4 weeks intervention with 2 practice sessions and no training.	Single blind (assessee)	Video-recorded musical performance. Musical theatre course.	Adachi, Tai, & Dawson (2018)
9	Cheng & Tsai, 2012	(online) Peer assessment	Yes	Yes (assignment requirements)	5 practice sessions (each PA activity lasted for 1 week); No training.	Double blind	Five assignments (image editing, image composition, computer graphics outlining, computer symbol design, and computer self-portrait sketching). Communication technology course.	No
10	Cheng et al., 2014	(online) Peer assessment	Yes	No	PA activity lasted for 2 weeks (at least 5 practice sessions); No training.	Double blind	Student-created video project. Advanced communication technology course	No
11	Chien et al., 2020	Peer assessment	No	Yes (Rubric)	3 weeks span, with a total of 350 min.	Double blind	Learning film (watching SVVR materials and conducting English speaking training). English.	No

² **Single blind (assessor):** only the assessee knows the identity of the assessor; **Single blind (assessee):** only assessor knows the identity of the assessee; **Double blind:** both the assessor and the assessee do not know the identity of each other.

#	Author(s) (year)	PA terminology	Did PA contain grades?	Use of PA aids	Length of intervention	Anonymity ²	PA task & subject	Constellations
12	Ching & Hsu, 2016	Peer feedback	No	Yes (Feedback prompts)	4 weeks	Public.	Case analysis of an instructional design problem. Instructional design class.	No
13	Corgnet, 2012	Peer evaluation & peer assessment	No (they decide how much money each group member receives out of team's profit)	No	1 practice session with no training.	Single blind (assessee)	Real-life problem. Mathematics task.	No
14	Domínguez et al., 2016	Peer assessment	Yes	Yes (Rubric)	No training. No length given for the actual intervention.	Single blind (assessee)	Creation of websites with multimedia products. Computing project management.	Gielen et al., 2011
15	Ersoz & Sad, 2018	Peer assessment	No	None	Not applicable	Public	Commenting on artworks posted in a Facebook group. Instructional technologies and material design.	No
16	Falchikov, 1995	Peer assessment & peer feedback	Yes	Yes (student generated criteria)	Assessed more than one student; brief training.	Grading: Single blind (assessee) assessee. Oral feedback: Public	Oral presentation of a written summary of experimental study of interest. Human developmental psychology.	No
17	Feyli & Ayatollahi, 2016	Peer feedback	No	No	No training. Intervention: 14 sessions, bi-weekly.	Public	Oral discussion. English.	No
18	Filius et al., 2019	Peer feedback	No	Yes (Exemplars, and deep approach to learning questionnaire which worked like a rubric)	Feedback instruction given and exemplars were shown in both text and video modality. Length of training was not mentioned.	In small private online course (SPOC): Public In MOOC course: double blind	Providing peer feedback on how "deep" the peer feedback given by a peer is. Multiple subjects (i.e., "Public health, human rights, clinical epidemiology, and child development").	No
19	Ghahari & Farokhnia, 2017	Peer assessment	Yes	Yes (criteria)	Semester long (30 hours); 8 practice sessions; training (explanation of assessment criteria)	Not stated but probably public because they discussed the feedback with their peers	Writing task (1 paragraph describing previous class session). English course.	No
20	Ghahari & Farokhnia, 2018	Peer assessment	Yes	Yes (criteria)	8 practice session; training (explanation of assessment criteria)	Not stated but probably public because they discussed the feedback with their peers	Writing task (1 paragraph describing previous class session). English course.	No

21	Güler, 2017	Peer assessment	Yes	Yes (Rubric)	Training: 1 week Intervention: 5 weeks No specific time or number of hours given per session	Depending on the condition: double blind or public.	Peer scoring on group presentation via WhatsApp. Computer network and communication course.	No
22	Guo & Lei, 2020	Peer assessment	Yes	Yes (Rubric & scoring sheet)	Training: a session, peer grading 8 essays. Intervention: a session (or one full day, depending on the experimental condition) peer grading to 4 additional peer essays.	Not stated	Written essay. Chinese.	No
23	Harris & Brown, 2013	Peer assessment	Yes	Yes (Criteria & exemplars)	Length not mentioned; No training.	Not stated	Oral presentations, essays, projects. English.	No
24	Hassel & Lee, 2019	Peer assessment	Yes	Yes (Rubric)	1 practice session (lasted 2 hours); Training (1.5 hours).	Double blind	Essay of open-ended problem. Engineering (Petroleum refining).	No
25	Hou & Cheng, 2012	(online) Peer assessment	No	No	Length not specified but there were 21 groups and each group provided PF to other groups so multiple practice sessions were likely to take place; No training.	Not stated	Film (group project). Multimedia production course.	No
26	Hsia et al., 2016	Peer assessment (web-based)	Yes	Yes (Rubrics)	The teacher provided training in the 1st week. Students in the experimental group only had 1 round of PA. The whole intervention ran for 7 weeks, with 1 session a week (45 min/session).	Single blind (assessee)	Creation of a script and performance based on a selected theme (Chinese folk story). Junior High School Performing Arts course.	No
27	Huisman et al., 2018	Peer feedback	No	Yes (Assessment criteria)	One session for training (verbal instructions). PA was provided/received once.	Double blind	Written essay. Educational sciences/Family pedagogy.	No
28	Hwang et al., 2014	Peer assessment	Yes	Yes (criteria)	4 weeks intervention (1 practice session per week); no training.	Not stated	Game development; Science	No
29	Izgar & Akturk, 2018	Peer assessment	Yes	Yes (Rubric)	10 weeks, 4 hrs a week. Training in the form of co-creating the rubric with the students.	Assessor knows the identity of the assessee but unclear if assessors are anonymous.	Evaluating classmates after a presentation of a topic. Special Teaching Methods II course.	No

30	Johnson & Winterbottom, 2011	Peer assessment	Yes	Yes (criteria)	Study lasted 22 weeks but unclear how many time PA was practices; no training.	Not stated	Project work. Science.	No
31	Kilickaya, 2017	Peer assessment	Yes (30% from peers and 70% from instructor)	Yes (Peer evaluation form and team dynamics rubric)	Length not mentioned. Participants underwent a tutorial session about how to conduct PA and SA through hands-on experience on the rating scale.	Public	Students performed a micro teaching activity and evaluated their group mates after. Teaching Language Skills 1.	No
32	Kim & Ryu, 2013	(online) Peer assessment	Yes	Yes (Rubric)	1 hour PA practice; 20 minutes training.	Not stated	Ill-structured instructional design task. Instructional design course (preservice teachers)	No
33	Kuo et al., 2017	(online) Peer assessment	Yes	Yes (Criteria)	3 weeks; 2 practice sessions (assessing 2 peers in each session); no training.	Single blind (assessee)	Video recorded Kung Fu Tai-Chi movements. Physical education course.	No
34	Lai & Hwang, 2015	Peer assessment	Yes	Yes (Exemplars, general assessment criteria)	Length of intervention: 500 min (80 min dedicated to assessment criteria development and 60 min dedicated to peer assessment). Training: PA performed 10 times during training and more than 10 times during the task.	Double blind	Poster. Art.	No
35	Li, 2019	Peer assessment	Yes	Yes (Rubric)	Students peer assessed two drafts during training. Length of intervention not stated.	Not stated	WebQuest Project (inquiry-oriented activity). Subject not stated.	No
36	Li et al., 2020	Peer assessment	Yes	Yes (Rubric)	Length not reported, but participants were taught the dimensions of the rubric and practiced on 2 essays.	Public (experimental Group) Not stated for control group.	PA on argumentative essays. English reading and writing course.	No
37	Li & Steckelberg, 2004	(online) Peer feedback	No	Yes (criteria)	2 practice sessions (i.e., assessing 2 peers); brief training (discussing assessment criteria).	Double blind	Web-based project. Instructional technology course (preservice teachers).	No
38	Lim et al., 2021	Peer feedback/PRT (Peer Review of Teaching)	No	Yes (Rubric)	Length of PRT not mentioned, but the intervention ran from fall of 2018 to fall of 2020 (in different cohorts).	Single blind (assessee)	Preservice teachers peer reviewed their classmates after a microteaching activity. Middle grades mathematics methods course.	No

39	Lin, 2018	Peer assessment, peer feedback	No	Yes (Assessment Criteria)	Last 6 class sessions (total time not given). Explanation of the assessment dimensions, examples for each dimension, assessment criteria, and the use of the onPear system.	Public (control group). Single blind (assessee) (experimental group).	Assessment of peer microteaching video via the onPear system in Facebook. Teacher training course.	No
40	Lin et al., 2002	Peer assessment	Yes	Yes (Discussed criteria)	60 minutes; 4 practice attempts (i.e., assessing 4 peers); training =Discussed assessment criteria (for 30 minutes for the high school students & three sessions for the undergraduate students).	Not stated	Project. Electrical Engineering & Operating Systems courses.	No
41	Liu et al., 2016	Peer Review	Yes	Yes (Rubric)	Students were guided to evaluate others in the 1st and 2nd week of peer review exercise. Intervention ran for 8 weeks, 1 hour per week.	Assessor knows the identity of the assessee but unclear if assessors are anonymous.	Students had to create a story using an iPad, which their classmates could see. Web 2.0 storytelling course.	No
42	Liu et al., 2018	Web based peer assessment	Yes	Yes (Rubric)	"Two class sessions in a row." Total time not reported. Training conducted online.	Not stated	Writing of English counter-offer letters. Foreign Trade Practical Operation Process.	No
43	McConlogue, 2012	Peer assessment	Yes	Yes (discussed/constructed criteria)	4-5 practice sessions (i.e., assessing 4-5 peers); training (discussion of criteria + separate practice session).	Double blind	Written tasks (lab reports & Literature review). Engineering course.	No
44	McMahon, 2010	Peer assessment	No	Yes (Criteria)	15 weeks; practice sessions not reported (15?); no training.	Not mentioned	Action plans. Training & Development Module	No
45	Omar et al., 2018	Peer assessment	No	Yes (Assessment prompt)	Intervention one hour; training one session (length not stated).	Double blind	Problem solving. Mathematics	No
46	Panadero et al., 2013	Peer assessment	Yes	Yes (criteria; only for experimental condition)	1 practice session; no training.	Not stated	Concept map. Learning & development course.	No
47	Prins et al., 2005	Peer assessment	No	Yes (rubric, feedback rules, assessment template, examples of last year reports)	Mini-session of training (1/2 hr); 3 assessment activities.	Not stated	Written report; Agriculture, water management, energy, spatial planning.	No
48	Raes et al., 2015	Peer assessment	Yes	Yes (student generated rubric).	3 weeks of practice each student peer assessed 16 times; additionally 2 training sessions (introduction & discussion, development of criteria, practice PF).	Single blind (assessee) scoring followed by public oral feedback. Then either single blind assessee (condition A) or public (condition B)	Conducting Workshop (group assignment). Educational Studies course.	No
49	Rogers & Feller, 2016	Peer assessment	Yes	Yes (Assessment scale)	Not stated. Probably one session (it was outside class-time).	Single blind (assessor)	Essay. No subject specified.	No

50	Roscoe et al., 2020	Peer assessment	Yes	Yes (Grading scale)	Length of intervention: one session (60-90 minutes). No training.	Fictional assessee	Essay. No subject specified.	No
51	Rotsaert et al., 2018	Peer assessment/ peer feedback	Yes	Yes (student-generated assessment rubric + feedback criteria)	Length of intervention 4 sessions (weeks?); 7 practice sessions (i.e., each student assessed 7 times); training session (with practice).	Single blind (assessee) (during first 2 weeks of intervention) then public (during last 2 weeks)	Conducting Workshop (group assignment). Educational Studies course	No
52	Rotsaert et al., 2017	Peer assessment	NA (survey study)	NA (survey study)	NA (survey study)	NA (survey study)	Not stated.	No
53	Seifert & Feliks, 2019	Peer assessment	Yes	Yes (Rubric)	None.	Single blind (assessee) because some wrote their names on their work by mistake.	Self-assessment and peer-assessment for various online assignments submitted in various courses. (No specific assignment or task, or number of online assignments peer assessed was mentioned. But a respondent mentioned about a writing task in an interview transcript.) Teaching and learning in technological environments courses.	No
54	Shen et al., 2020	Peer assessment	No	Yes (Rubrics/guidance sheets)	Two hours session for training, one hour session for intervention.	Public	Written assignment. English (EFL)	No
55	Sluijsmans et al., 2002	Peer assessment	Yes	Yes (Criteria)	Training (four tasks of 1 hour each including introduction to PA, defining criteria, discussing guidelines for giving constructive PF, and judging performance of a peer) = This is the PA intervention; during PA activity each student had to assess three (3 practice attempts).	Assessor knows the identity of the assessee but unclear if assessors are anonymous	Videotapes of creative lessons. Designing creative lessons course (for the domains Art, Dutch language, & Music).	No
56	Stanier, 1997	Peer assessment	Yes	Yes (Negotiated criteria)	Training provided (Introduction session and discussion of criteria). Unclear for how long they practiced it and how many times they did PA.	Not stated	Brochure making. Interdisciplinary Studies in Environmental and Applied Sciences	No
57	Strijbos et al., 2010	Peer feedback	No	Yes (criteria)	1 practice session with no training.	Fictional assessor	Written text. Academic writing.	No
58	Strijbos et al., 2021	Peer feedback	No PA activity was done but students received a scenario	Yes (see observations)	NA	Fictional assessor	Writing formal business letter (see observations). All disciplines.	No

67	Wilson et al., 2015	(online) Peer assessment	Yes	Unclear	Unclear but appears to be multiple assessments (see procedure).	Depending on the practice model (see study)	Variety of tasks. Multiple courses: "Science of Sex (BIOL 260), Financial Techniques, Instruments and Markets (ACST 201), Finance and Finance Reporting (ACST 852), Intercultural Relations (ICOM 101), International Television and Beyond (ICOM 201), Transnational Communication (ICOM 817), Key Competencies in Inclusive Education (TEP 248), Integrated Marketing Communications (MKTG 204) and History in the Secondary School II (TEP 428). "	No
68	Zhou et al., 2019	Peer assessment	Yes	Rubrics	Not stated	Double blind	Written task (3). English writing.	No
69	Zou et al., 2018	Peer assessment	Yes	Yes (Rating scales for PA and back evaluation)	None.	Double blind	"The engineering students wrote one autobiography and one summary of a novel read in class. The English students wrote a literature review and methodology section for their senior thesis paper." Engineering students: English Reading and Writing class; and, English students: Advanced Thesis Writing for EFL Students class.	No

Supplementary material 6. Results of peer assessment ↔ intrapersonal variables*Table 1 of 2 for supplementary material 6.* Results of intrapersonal variables → peer assessment

Article #	Author & year	Aim(s)	Study design	Variables of interest	Study quality	Results
2	Alqassab et al., 2019	To investigate relationships between: (a) peer feedback accuracy and types (b) peer feedback accuracy, peer feedback provision beliefs, perceptions of feedback message, and emotions (c) Task performance (proof comprehension) beliefs about peer feedback provision and perceptions of peer feedback message	ERG	Emotions (curiosity, confusion, anxiety), trust in self as assessor	9	Emotions: Anxiety associated with less accurate peer feedback. No significant effects were found for feeling curious or confused on peer feedback accuracy. Trust in self as assessor: Did not predict peer feedback accuracy or task performance.
6	Carvalho, 2013	To investigate the relationships between students' perceptions of fairness of peer assessment and perceived conflict & friendship marking (problems with PA)	DSO	Fairness	4	Perceptions of fairness of peer assessment negatively associated with occurrence of problems and difficulties in peer assessment (conflict & friendship marking).
9	Cheng & Tsai, 2012	To explore the relationships between students' approaches to learning and their interpersonal perspectives (i.e., psychological safety, value diversity, trust and social interdependence) in online peer assessment.	DSO	Trust in self as assessor	3	Students who trusted themselves to peer assess were more likely to adopt cohesive conceptions and deep approaches to learning; students who had trust in their peers were inclined to use deep approaches.
18	Filius et al., 2019	To investigate the effect of asynchronous audio peer feedback on deep approach to learning in online education.	DSO	Motivation (measured by feeling personally committed in the mechanism scale)	9	Feeling personally committed was a significant predictor of deep approach to learning while providing and receiving peer feedback. Participants significantly felt more personally committed to provide peer feedback than to receive.
22	Guo & Lei, 2020	To investigate the effects of key variables related to peer raters' qualities: content knowledge, previous rating experience, training on rating tasks, and rating motivation.	ERIG	Motivation	13	Motivation-enhancing interventions (especially monetary incentives) reduce peer assessment errors. These effects are mediated by self-paced training.
23	Harris & Brown, 2013	To explore teachers' and students' perspective of peer and self-assessment in New Zealand.	DSO	Trust in self as assessor	2	Lack of trust seems to have a negative effect on peer assessment accuracy
52	Rotsaert et al., 2017	To investigate the relationship between students' perceptions of peer assessment and interpersonal factors.	DSO	Trust in self as assessor	4	Trust in self as assessor positively related to perceived value of peer assessment.
62	Van Gennip et al., 2010	To investigate: (a) how a peer assessment intervention can influence interpersonal variables (psychological safety, interdependence, value diversity, and trust) (b) the relationship between interpersonal variables, and students' conceptions of peer assessment and their learning	ERIG	Trust in self as assessor	8	Trust in self as assessor was a strong predictor of conceptions of peer assessment.

69	Zou et al., 2018	"This study explores the underlying factors that influence students' participation in peer assessment in writing instruction."	DSO	Motivation, fairness, and trust in self as assessor	6	<p>Motivation and trust in self: Students with positive attitude (including motivation and trust) towards peer assessment appeared to review less papers (i.e., less participation in PA) which had an explanation: "Positive attitude was associated with lower participation, in part because students worked hard on each review rather than doing many reviews superficially".</p> <p>Trust in self as assessor: More than 26% disagreed or strongly disagreed that they were competent enough to perform peer assessment (38% were undecided).</p> <p>Fairness: More than 34% of students agreed or strongly agreed that they doubt the fairness of peer review. The more students doubted the fairness of peer assessment the less likely they participated in the peer assessment (i.e., review less papers).</p>
----	------------------	--	-----	---	---	--

Table 2 of 2 for supplementary material 6. Results of peer assessment → intrapersonal variables

Article #	Author & year	Aim(s)	Study design	Variables of interest	Study quality	Results
1	Alqassab et al., 2018	To investigate the impact of a structured peer feedback training on assessment skills and perceptions of preservice mathematics teachers considering their domain knowledge.	ERIG	Trust in self as assessor	10	Students trust in self as assessors decreased significantly (with a large effect size, $\eta_p^2 = 0.20$) after the peer feedback training regardless of their domain knowledge level.
4	Berndt et al., 2018	"To investigate the impact of peer-feedback content (concise general feedback [CGF] vs. elaborated specific feedback [ESF]) and sender's competence level (low vs. high) on peer-feedback perceptions, text-revision performance, peer-feedback recall performance, and mindful cognitive processing."	ERG	Fairness (perceived adequacy), motivation (willingness to improve), emotions (affect)	9	Fairness and motivation: Receiving peer feedback from high competent peer led to more willingness to improve (i.e., motivation) and to more perceived adequacy of peer feedback (i.e., fairness) than low competent peer. Emotions: Elaborate specific peer feedback led to more positive affect than general peer feedback.
7	Chang et al., 2020	To investigate the effects of peer assessment compared to teacher feedback on students' performance and perceptions.	ERIG	Motivation (Students' learning, motivation), self-efficacy	8	Peer assessment leads to higher self-efficacy compared to teacher feedback. No significant differences were found in learning motivation.
8	Chen et al., 2020	The aim of this study was to investigate the effects of different modes of peer assessment (peer comments + peer scores vs. peer comments only vs. peer scores only) on students' peer assessment behaviors, theater performance, peer feedback quality, validity of peer assessment grades (compared to teacher's) and learning motivation.	ERIG	Motivation	8	Students in the peer commenting group reported higher levels of learning motivation than those in the peer scoring group.
10	Cheng et al., 2014	"The purpose of this study is to explore the relationships among the students' emotional responses, their participation in online peer assessment and their project performance."	DSO	Emotions	7	Students who received more positive comments reported more positive emotion and higher levels of participation in responding to their peers' comments. (Emotional response [valence] to receiving peer feedback and responding to peers' comments)
11	Chien et al., 2020	"To explore whether learning in the spherical video-based virtual reality (SVVR) environment with the PA approach could improve students' English-speaking performance, learning motivation and critical thinking skills, and reduce their English learning anxiety."	ERIG	Motivation, emotion (anxiety)	9	Motivation: Peer assessment based SVVR can enhance motivation. Anxiety: Peer assessment based on SVVR can reduce English learning anxiety.
12	Ching & Hsu, 2016	"This study explored learners' perceptions of the interpersonal factors critical in a role-playing peer- feedback activity, as well as the types of peer feedback learners generated."	DSO	Trust in self as assessor, comfort	3	Trust in self as assessor: Participants perceived the feedback they provided to their peers as satisfactory (M = 3.07, on a 5-point scale). Comfort: About 63% of participants reported feeling more comfortable critiquing peer's work using the role-play approach.
13	Corgnat, 2012	To investigate the impact of using peer assessment to share a jointly produced outcome on team performance.	ERG	Motivation	5	Negative effect of peer evaluation on performance and motivation when the members knew each other. When participants in the peer assessment were in groups with classmates only around 26% of them reported that their motivation increased as opposed to 42% of participants in groups without classmates.
15	Ersoz & Sad, 2018	"...to explore how a small group of visual art education students used Facebook as a peer-assessment platform...to better understand the advantages and disadvantages of this particular practice."	DSO	Motivation, fairness	1	Motivation: Facebook-based peer assessment improved students' motivation. Fairness: Some participants reported that criticism of their work on Facebook was unfair and disruptive.

16	Falchikov, 1995	"The present study attempted to capitalize on the benefits of peer assessment in terms of improving the learning process, sharpening critical abilities and increasing student autonomy, while at the same time addressing problems encountered in earlier studies."	DSO	Fairness	4	Best liked features of peer feedback marking as reported by participants was fairness of the system. Peer feedback was rated as fairer than traditional teacher assessment but as less enjoyable yet both types of assessment were reported as being equally not enjoyable.
17	Feyli & Ayatollahi, 2016	To explore the effects of peer and teacher feedback on students' self-confidence.	ERIG	Self-efficacy	5	Peer feedback and teacher feedback improved students' self-confidence. Mean self-confidence at posttest for the peer feedback condition was larger than teacher feedback condition.
19	Ghahari & Farokhnia, 2017	To investigate the effect of peer assessment compared to teacher assessment on language learners' writing skills and self-efficacy.	ERIG	Self-efficacy	9	Students' self-efficacy did not change after being involved in peer assessment or teacher assessment, and no differences in self-efficacy were found between participants in the peer and teacher assessment conditions. Focus group revealed that students in the peer assessment condition did not believe that peer assessment can enhance their motivation.
20	Ghahari & Farokhnia, 2018	To explore whether peer assessment is an appropriate alternative to teacher assessment.	ERIG	Emotions (embarrassment)	9	Qualitative data showed that some learners in the peer assessment condition reported feeling embarrassed to talk to other sex members (for sociocultural reasons).
21	Güler, 2017	" The aim of this study is to investigate the use of WhatsApp application in anonymous peer assessment in higher education."	ERIG	Fairness	10	Fairness: Students in both the anonymous and non-anonymous conditions perceived peer assessment via WhatsApp as fair. No significant differences in students' ratings of fairness were found between the anonymous and the non-anonymous conditions.
24	Hassel & Lee, 2019	To investigate if: (a) engineering students can perform accurate peer assessment when provided with appropriate training and rubric. (b) students valued self and peer assessment exercise and in what ways they found it effective/ineffective. (c) differences in performance and perception of peer assessment based on course of study.	DSO	Trust in self as assessor; motivation	4	Trust in self as assessor: Students did not consider providing peer feedback as an acceptable substitution to teacher feedback (M = 2.74; 7-point scale) Motivation: Students reported that the peer assessment activity motivated them to learn (M = 4.74; 7-point scale).
25	Hou & Cheng, 2012	To investigate students' emotions experienced during online peer assessment.	DSO	Emotion (emotional state of assessors and assessees)	6	In peer assessment, students appeared to experience positive emotions continuously and these emotions did not seem to shift to neutral or negative emotions. Also, neutral and negative emotions do not shift to positive emotions.
26	Hsia et al., 2016	To investigate: (a) the impact of web-based peer assessment on students' performance (b) the relationships between the students' self-efficacy in peer assessment (i.e., trust in self as assessor), learning motivation, and performance.	ERIG	Motivation and Trust in self as assessor.	9	Trust in self as assessor: Students in the experimental group reported high levels of trust in self as M = 4.10 on a 5-point scale; no comparisons were made with the control group). A positive significant correlation was found between performance and trust in self as assessor (r = 0.30). Motivation: Students in the experimental group reported high levels of learning motivation M = 4.13 (on a 5-point scale; no comparisons were made with the control group). A positive significant correlation was found between performance and motivation (r = 0.19).
27	Huisman et al., 2018	(a) To compare the impact of providing versus receiving peer feedback on students' performance in academic writing. (b) To investigate the relationship between the nature of received peer feedback, students' perceptions, and writing performance.	ERIG/ ERG ³	Motivation (willingness to improve)	7	Explanatory peer feedback leads to more willingness to improve. Students' willingness to improve was not significantly related to increase in writing performance.

³ Not stated how the assignment to experimental groups was made.

28	Hwang et al., 2014	To investigate the effect of peer assessment on students' learning achievement, motivation, and problem-solving skills in science game development.	ERIG	Motivation	8	Peer assessment had a positive effect on students' learning motivation as those in the experimental group reported significantly higher levels of learning motivation than the experimental group at the post-test which was supported by qualitative findings.
30	Johnson & Winterbottom, 2011	To explore the effects of self- and peer assessments on students' motivation in a girls-only biology class.	DSO	Motivation (goal-orientation)	4	Contradictory results: the quantitative data showed a decrease in learning-goal orientation (negative result), while the qualitative data showed that some students moved towards such an orientation (positive result).
31	Kilickaya, 2017	To investigate university language students' perceptions of peer assessment.	DSO	Trust in self as assessor	3	Most participants stated that they lacked confidence to assess their peers because the peer assessment training was not sufficient or because they did not have enough prior experience with peer assessment.
32	Kim & Ryu, 2013	To investigate the impact of a web-based formative peer assessment (WFPAS) on students' metacognitive awareness and learning achievement compared to traditional peer assessment and self-assessment.	ERG	Motivation	10	Motivation: Positive effect was found of WFPAS on motivation compared to traditional peer and self-assessment, with large effect sizes (Cohen's d) ranging between .55 and .70.
33	Kuo et al., 2017	To investigate the effect of incorporating a peer-assessment teaching model into mobile learning system on physical education students' outcomes.	ERIG	Motivation, self-efficacy	7	Positive effects of mobile-based peer assessment on learning motivation and physical self-efficacy, as the peer assessment group experienced improvement in learning motivation and self-efficacy at the post-test whereas no significant change was found for the mobile-based learning condition (without peer assessment).
34	Lai & Hwang, 2015	To investigate the effect of interactive peer assessment criteria development approach on students' learning outcomes.	ERIG	Motivation	13	Learning motivation: Interactive assessment criteria development approach in peer assessment can lead to higher levels of learning motivation, as students in this condition reported significantly higher motivation at the post-test than the control group (peer assessment without the interactive assessment-criteria development).
35	Li, 2019	"(...) to examine the effects of game based assessment training on students' assessment skills and intrinsic motivation."	ERIG	Motivation (intrinsic)	8	Game-based peer assessment training can lead to higher levels of intrinsic motivation, as students in this condition reported significantly higher levels of intrinsic motivation compared to the control (conventional peer assessment).
36	Li et al., 2020	To investigate the effects of peer assessment enabled by: (a) wiki, (b) small messaging group, and (c) big messaging group on students' performance and motivation.	ERIG	Motivation (intrinsic)	10	Peer assessment using small messaging group produces higher levels of intrinsic motivation compared to big messaging group, and wiki only settings.
37	Li & Steckelberg, 2004	"This study aims to gain a better understanding of how the newly arisen social messaging may impact the practice of peer assessment."	ERIG	Fairness	7	Students in the peer assessment (experimental) group considered peer assessment to be fair ("The peers' comments on my work were fair" M = 4.23, on a 5-point scale). Perceptions of the control group (no peer assessment) were not measured.
38	Lim et al., (2021)	To investigate the impact of a peer feedback model on preservice teachers' perceptions of peer feedback.	DSO	Motivation (engagement)	6	Motivation (engagement): An increased level of participation in peer feedback was observed in all courses.
39	Lin, 2018	To investigate the effect of anonymity in a Facebook-based peer assessment on students' peer feedback and their perceptions of learning and peer assessment.	ERG	Fairness	10	Participants in the anonymous condition rated peer feedback as less fair than those in the identified condition.

41	Liu et al., 2016	"This study aimed to explore how peer review may affect creative performance and self-efficacy in a Web 2.0 learning context."	ERIG	Self-efficacy, comfort ⁴ (reactions to potential feedback)	6	Self-efficacy: No evidence that peer assessment negatively influenced creative self-efficacy as no significant differences were found between the experimental (peer assessment) group and the control group (no peer assessment) in all creative self-efficacy measures, with small effect sizes (Cohen's d) ranging between .31 and .42. Comfort: Both groups showed high self-efficacy in the dimension of reaction to potential feedback, M = 3.88 and 3.65 for the experimental and the control groups respectively.
42	Liu et al., 2017	To investigate the effect of a web-based training involving group discussions on students' performance and perceptions of peer assessment and the effect of anonymity in this context.	ERIG	Trust in self as assessor	9	Overall, students had high trust in self as assessors during the web-based peer assessment task, M = 4.11 (on a 5-point scale) with no significant differences between the anonymous and non-anonymous conditions. However, students were less confident regarding giving helpful opinions or suggestions, telling whether peer has done his/her best, and ignoring unreasonable received feedback
43	McConlogue, 2012	To evaluate students' perceptions of the reliability of marking of peer assessment of divergent written tasks in two engineering courses.	DSO	Fairness	5	Many students in the study considered peer assessment unfair.
44	McMahon, 2010	To investigate the impact of peer assessment and negotiated learning activities on students' perceptions in an outcome-based curriculum.	DSO	Motivation, trust in self as assessor	4	Students reported that peer assessment improved their motivation and their trust in self as assessors.
45	Omar et al., 2018	"To examine the effect of the use of peer assessment in a Brunei Mathematics classroom in the learning of Geometry"	DSO	Trust in self as assessor, comfort, emotion (enjoyment)	4	Trust in self as assessor: Criteria discussion did not lead to trust in the self as an assessor. Comfort: Peer assessment created a feeling of comfort, especially when providing feedback instead of grading. Enjoyment: 30% of students reported that they did not enjoy giving peer feedback while 43.5% reported enjoy it.
46	Panadero et al., 2013	To investigate the effect of rubric and friendship on the reliability and validity of peer scoring, students' performance, perceived comfort and fairness of peer assessment.	ERIG	Comfort and Fairness	5	Comfort: No significant effect of rubric was found on perceived level of comfort. Fairness: No significant effect of rubric was found on perceived fairness of peer assessment.
47	Prins et al., 2005	To investigate students' perceptions of the design of instructional support for peer assessment within a computer supported collaborative learning environment.	DSO	Trust in self as assessor, comfort	7	Trust in self as assessor: Overall, students trusted themselves during peer assessment. Comfort: Five out of the seven students felt comfortable to perform peer assessment.
48	Raes et al., 2015	To investigate the effect of increasing anonymity using a Classroom Response Technology (CRT) to reduce undesirable social effects of peer assessment.	ERIG	Comfort	9	Anonymous peer assessment was perceived as more comfortable than non-anonymous peer assessment.
49	Rogers & Feller, 2016	To explore why students' success declined when they read and assessed their peers' excellent essays.	ERG	Motivation, self-efficacy	8	Assessing excellent-quality peer work leads to a decreased motivation (i.e., discouragement) in the assessor. Self-efficacy: Participants in the excellent-essay condition reported being significantly less able to write an essay as good as the essay they have assessed.
51	Rotsaert et al., 2018	To investigate the impact of fading anonymity on peer feedback quality and students' perceptions of peer assessment.	ERIG	Trust in self as assessor	10	A significant increase in trust in self was found after the initial anonymous peer assessment sessions (although trust in self was high to start with). Students' trust in self as assessor did not change when transitioned from the anonymous to the non-anonymous sessions.
53	Seifert & Feliks, 2019	To investigate students' and lecturers' perceptions of the use of self and peer assessment in teacher training.	DSO	Trust in self as assessor, fairness	4	Trust in self as assessor: Students reported being able to perform peer assessment accurately (M = 3.7 on a 5-point scale). Fairness: Most participants perceived peer assessment as fair (M = 4.40 on a 5-point scale).

⁴ The authors consider it as a sub-scale of (creative) self-efficacy, yet it clearly aligns with our definition of comfort.

54	Shen et al., 2020	"To evaluate the effects of using peer assessment in an English writing class to enhance Chinese college students' learner autonomy"	ERG	Trust in self as assessor (learner autonomy)	10	Peer assessment did not lead to improved trust in the self as assessor, as no significant differences were found between the experimental (i.e., peer assessment) and the control (traditional teacher assessment) groups.
55	Sluijsmans et al., 2002	To investigate the impact of peer assessment training (within a course) on students' assessment skills, task performance, and perceptions of the activity.	ERIG	Trust in self as assessor (assessment skill)	9	No effect of peer assessment training was found on increasing trust in self as assessor.
56	Stanier, 1997	To investigate students' perceptions of group work and peer assessment.	DSO	Trust in self as assessor, discomfort	2	Trust in self as assessor: At the beginning students raised concerns regarding their ability to assess their peers' work, but after receiving some support they showed an increased acceptance of the activity. Discomfort: Around 40% of the participants reported feeling uncomfortable. The methodological information is quite limited, but it seems logical to expect 60% did not.
57	Strijbos et al., 2010	To investigate the impact of peer feedback content (concise general vs. elaborated specific feedback) and sender's competency level (high vs. low) on students' perceptions of peer feedback and their academic writing performance.	ERG	Fairness (perceived adequacy), motivation (willingness to improve), emotions (affect)	9	Fairness: Elaborate specific feedback from a high competent peer was perceived as more adequate (i.e., fair) than from low competent peer. Motivation (willingness to improve): No significant effects of peer feedback content and senders' competency levels were found on willingness to improve. Emotions (affect): Elaborate specific feedback from a high competent peer led to more negative affect.
58	Strijbos et al., 2021	"The present study reports on the structural validity and measurement invariance of the Feedback Perceptions Questionnaire (FPQ) developed by Strijbos et al. (2010).	DSO	Fairness (perceived adequacy), motivation (willingness to improve), emotions (affect)	7	Fairness of peer feedback was found to be a strong positive predictor of willingness to improve (i.e., motivation), and affect.
60	Topping et al., 2000	To investigate the reliability and validity of qualitative peer assessment and students' perceptions of the activity.	DSO	Discomfort	4	All participants (12 students) reported experiencing discomfort during peer assessment.
62	Van Gennip et al., 2010	To investigate: (a) how a peer assessment intervention can influence interpersonal variables (psychological safety, interdependence, value diversity, and trust) (b) the relationship between interpersonal variables, and students' conceptions of peer assessment and their learning	ERIG	Trust in self as assessor	8	Students' trust in self as assessors in the peer assessment and the teacher assessment conditions did not differ significantly.
64	Vander Schee & Birrittella, 2021	To examine peer group grading's efficiency while maintaining perceived fairness.	DSO	Fairness	4	Students perceived that peer group grading was fair, regardless of whether it was conducted online or hybrid (online + face-to-face).
65	Vanderhoven et al., 2015	To investigate the effect of anonymity-using Classroom Response Technology (CRT)- on reducing undesirable social effects of peer assessment.	ERIG	Discomfort, fairness	9	Discomfort: No effect was found on anonymity in reducing assessors' levels of discomfort (which was already low to start with for both the anonymous and non-anonymous conditions). Fairness: No effect was found for anonymity on perceptions of peer assessment fairness as both conditions did not differ significantly and both forms of peer assessment were considered fair ($M = 3.50$ and 3.10 on a 5-point scale for the non-anonymous and anonymous conditions).
66	Wen & Tsai, 2006	To explore university students' perceptions and attitudes towards online peer assessment.	DSO	Fairness	2	No significant differences were found between perceptions of fairness of traditional peer assessment and online peer assessment, yet the mean scores for both items measuring fairness were low ($M_{\text{traditional}} = 2.81$, $SD = .87$, $M_{\text{online}} = 2.73$, $SD = .93$ on a 5-point scale).
67	Wilson et al., 2015	To investigate students' perceptions of fairness, trust, and temporality of online peer review Moodle tool.	DSO	Fairness	1	Students did not perceive peer assessment as fair.

68	Zhu et al., 2019	To investigate the social-affective impact of peer assessment.	DSO	Emotions	5	Several emotions were mentioned by students in relation to peer assessment including gratitude and being flattered, unsureness, nervousness, disappointment, and anger.
----	---------------------	---	-----	----------	---	---

Supplementary material 7. Results of peer assessment ↔ interpersonal variables*Table 1 of 2 for supplementary material 7. Results of interpersonal variables → peer assessment*

Article #	Author (year)	Aim(s)	Study design	Variables of interest	Study quality	Results
3	Anaya et al., 2019	To explore how social relations between assessors and assesseees can encourage learners to participate in the peer review process.	DSO	Social connections	3	Using social relations to assign students to peer assessment resulted in 60% acceptance rate for participation in peer assessment as reviewers.
5	Carlsson Hauff & Nilsson, 2021	To investigate the effects of trust in peer and self-assessed knowledge on students' perceptions of making and receiving peer assessment before and after the task.	ERIG	Trust in the other as assessor	7	Trust in the others as assessors positively associated with positive expectation of making and receiving peer assessment (i.e., feeling comfortable, making fair assessment, and learning from peer assessment).
9	Cheng & Tsai, 2012	To explore the relationships between students' approaches to learning and their interpersonal perspectives (i.e., psychological safety, value diversity, trust and social interdependence) in online peer assessment.	DSO	Psychological safety, value diversity, trust in the other as assessor, interdependence	3	Psychological safety: Students with high psychological safety were more likely to have cohesive conceptions and adopt deep learning approaches. Value diversity: Students with low goal value diversity were more likely to adopt cohesive conceptions and deep learning approach. Those with low value diversity for criteria possessed cohesive conceptions and deep approaches. Trust in the others as assessors: Those who had trust in their peers were inclined to use deep approaches. Social interdependence: Most students showed cooperative/non-competitive type of social interdependence (48%); these students mainly adopted cohesive conceptions and deep approaches to learning. Some students showed cooperative/competitive type of social interdependence (26%); most of them adopted cohesive conceptions and deep approach to learning.
13	Corgnet, 2012	To investigate the impact of using peer assessment to share a jointly produced outcome on team performance.	ERG	Social connections (friendship)	5	Friendship can facilitate team performance when outcomes are shared equally but seems to have a negative effect when peer evaluation determines the profit each member receives.
14	Domínguez et al., 2016	"...examining the consistency and differences among self-, peer-, external- and instructor-assessment." "...two additional, and secondary, goals: identify differences in the given assessments depending on both the quality of the assessed products and on the competency of the assessing student and analyse the impact of this activity on the learning process."	ERIG	Social connections (friendship)	7	Friendship (peers from the same university) leads to biased (less strict) peer assessment scores.
15	Ersoz & Sad, 2018	"...to explore how a small group of visual art education students used Facebook as a peer-assessment platform, as well as to better understand the advantages and disadvantages of this particular practice."	DSO	Social connections (friendship)	1	As a disadvantage of peer assessment, participants reported friendship-based favouritism in peer assessment that can lead to biased comments.

23	Harris & Brown, 2013	To explore teachers' and students' perspective of peer and self-assessment in New Zealand.	DSO	Social connections (friendship)	2	One teacher reported observing inaccuracy in peer assessment due to friendship, and students shared strong concerns regarding the negative impact of peer and self-assessment on social relationships.
31	Kilickaya, 2017	To investigate university language students' perceptions of peer assessment.	DSO	Social connections (friendship)	3	Most of the participants suffered from friendship bias, and they reported that their peer assessment was influenced by their relationships with the peer they assessed (a friend vs. someone I don't get along with).
46	Panadero et al., 2013	To investigate the effect of rubric and friendship on the reliability and validity of peer assessment, students' performance and their perceived comfort and fairness of peer assessment.	ERIG	Social connections (friendship)	5	No significant effect of rubric was found on reducing friendship bias.
50	Roscoe et al., 2020	To investigate the influence of students' perceptions of author characteristics and error in peer work on evaluation of quality of the assessed work.	DSO	Social connections (i.e., interpersonal author traits: generosity, kindness, and loyalty)	7	The perceived quality of the assessed work (written essays) was related to assessee's personality traits as perceived by assessors (e.g., generosity, kindness, loyalty).
52	Rotsaert et al., 2017	To investigate the relationship between students' perceptions of peer assessment and interpersonal factors.	DSO	Trust in the other as assessor, social connections (friendship marking, fear of disapproval)	4	Trust: positively associated with perceived educational value of peer assessment. Social factors (negative interpersonal processes): Awareness of negative interpersonal processes (i.e., fear of disapproval and friendship marking) positively predicted the perceived educational value of peer assessment.
59	To & Panadero, 2019	"This qualitative study explores peer assessment effects on the self-assessment process of 11 first-year undergraduates and the factors limiting peer influence."	DSO	Social connections (fear of disapproval ⁵)	5	Social factors (fear of disapproval): Data revealed that peer relationships could influence the depth of feedback exchange; six (out of the 11) students reported experiencing tension in feedback communication.
62	Van Gennip et al., 2010	To investigate: (a) how a peer assessment intervention can influence interpersonal variables (psychological safety, interdependence, value diversity, and trust) (b) the relationship between interpersonal variables, and students' conceptions of peer assessment and their learning	ERIG	Psychological safety, trust in the other as assessor, value diversity, interdependence	8	Trust in the other as assessor and psychological safety were strong positive predictors of conceptions of peer assessment. Value diversity: negatively predicted conceptions of peer assessment. Interdependence: Dependence of the self and dependence of the peer did not significantly predict conceptions of peer assessment. A mediation effect of conceptions of peer assessment on the relationships between interpersonal factors and perceived learning.
63	van Heerden & Bharuthram, 2021	To explore the effect that peer familiarity has on receiving feedback from peers.	DSO	Social connections	5	Knowing the peer leads to a positive peer assessment experience (i.e., more comfort in peer assessment). Some students indicated that not knowing the peer made peer assessment more objective.
69	Zou et al., 2018	"This study explores the underlying factors that influence students' participation in peer assessment in writing instruction."	DSO	Social connections (friendship)	6	The more students were concerned about the negative impact of peer assessment on interpersonal relationship the more they were likely to do back-evaluation (indicating the helpfulness of the peer assessment received on a five-point rating scale)

⁵ The authors call it tension in feedback communication

Table 2 of 2 for supplementary material 7. Results of peer assessment → interpersonal variables

Article #	Author & year	Aim(s)	Study design	Variables of interest	Study quality	Results
12	Ching & Hsu, 2016	"This study explored learners' perceptions of the interpersonal factors critical in a role-playing peer-feedback activity, as well as the types of peer feedback learners generated."	DSO	Psychological safety, trust in the other as assessor	3	Psychological Safety: Participants perceived the online role-play activity to deliver peer feedback to be beneficial for their psychological safety (means ranged between 2.5 and 3.57 on a 4-point Likert scale). Trust in the other as assessor: Overall, participants trusted their peers as assessors (means ranging from 3.07 to 3.36 on a 4-point Likert scale).
16	Falchikov, 1995	"The present study attempted to capitalize on the benefits of peer assessment in terms of improving the learning process, sharpening critical abilities and increasing student autonomy, while at the same time addressing problems encountered in earlier studies."	DSO	Social connections (friendship)	4	Students did not like marking friends.
20	Ghahari & Farokhni, 2018	To explore whether peer assessment is an appropriate alternative to teacher assessment.	ERIG	Trust in the other as assessor	9	In focus groups students raised trust issues regarding the expertise and knowledge of their peers and believed that the teacher should be responsible for providing feedback and not their peers.
24	Hassel & Lee, 2019	To investigate if: (a) engineering students can perform accurate peer assessment when provided with appropriate training and rubric. (b) students valued self and peer assessment exercise and in what ways they found it effective/ineffective. (c) differences in performance and perception of peer assessment based on course of study.	DSO	Trust in the other as assessor	4	Students felt that their peers had adequate knowledge to assess them (M = 4.39 on a 7-point Likert scale).
29	Izgar & Akturk, 2018	To investigate the relationship between peer and teacher assessment and students' perceptions of the process.	DSO	Social connections (friendship)	2	28% of the participants mentioned in the interviews that peers do not make fair assessment because they prioritise personal problems or friendship.
37	Li & Steckelberg, 2004	"This study was designed primarily to investigate the impact of anonymous peer feedback on student meaningful learning in higher education."	ERIG	Trust in the other as assessor	7	Peer assessment led to perceptions of trust in others as assessors: "Peers have adequate knowledge to evaluate my work" M = 3.73 on a 5-point Likert scale for the experimental group.
38	Lim et al., 2021	To investigate the impact of a peer feedback model on preservice teachers' perceptions of peer feedback.	DSO	Psychological safety ⁶	6	The employed feature of the peer feedback activity of making the presenter leave the room after a teaching demonstration helped the students to provide critical peer feedback with little discomfort.

⁶ The authors used comfort however their description aligns with our definition of psychological safety

40	Lin et al., 2002	(a) To assess the correlation between peers' and tutor's marking and compare the markings between undergraduate and high school students. (b) To assess student perceptions of peer assessment and compare the different perceptions of peer assessment between undergraduate and senior high school students	DSO	Trust in the other as assessor	5	In general, students reported low levels of trust in peer as assessor, with undergraduate students having significantly higher means ($M = 3.00$ $SD = .74$) for the item "I agree that peers have adequate knowledge to evaluate my work" than high school students ($M = 2.55$ $SD = .86$), $t = 2.78$, $p < .01$
42	Liu et al., 2017	To investigate the effect of a web-based training involving group discussions on students' performance and perceptions of peer assessment and the effect of anonymity in this context.	ERIG	Interdependence	9	Overall, students were satisfied with the group collaboration process ($M = 5.22$ on a 7-point scale). There were no significant differences between the anonymous and non-anonymous conditions.
47	Prins et al., 2005	To investigate students' perceptions of the design of instructional support for peer assessment within a CSCL environment.	DSO	Trust in the other as assessor	7	Students appeared to have more positive attitude towards being assessed by others at the post-test.
48	Raes et al., 2015	To investigate the effect of increasing anonymity in peer assessment using a Classroom Response Technology (CRT) to reduce undesirable social effects.	ERIG	Social connections (peer pressure)	9	Overall students experienced low levels of peer pressure (largest mean score = 2.59 on a 5-point scale). Students experienced significantly less peer pressure while providing non-anonymous written peer feedback rather than oral feedback.
51	Rotsaert et al., 2018	To investigate the impact of fading anonymity on peer feedback quality and students' perceptions of peer assessment.	ERIG	Psychological safety, trust in the other as assessor, value congruency, social connections (friendship marking & fear of disapproval)	10	Psychological safety: no significant change was observed over time. Those who presented their workshop in the non-anonymous phase reported significantly higher levels of psychological safety (M difference = .54, $r = .42$). Trust in peer as assessor: no significant change was found as a result of fading anonymity. Value congruency: improved significantly after the introduction of the anonymous phase (effect size $\eta^2 = .20$). Social factors: perceived friendship marking significantly increased from the anonymous to the non-anonymous phase but it was low when the non-anonymous sessions ended (Cohen's $d = .34$). Fear of disapproval: decreased significantly after the non-anonymous sessions in comparison to the anonymous sessions (Cohen's $d = .56$).
53	Seifert & Feliks, 2019	To investigate students' and lecturers' perceptions of the use of self and peer assessment in teacher training.	DSO	Trust in the other as assessor	4	Participants reported that their peers understood their work ($M=4.2$, on a 5-point scale).
59	To & Panadero, 2019	"This qualitative study explores peer assessment effects on the self-assessment process of 11 first-year undergraduates and the factors limiting peer influence."	DSO	Trust in the other as assessor	5	Trust in the other as assessor: Participants reported distrust in peers' evaluation competence.
61	Van Gennip, 2012	To investigate the influence of differently arranged peer assessment activities on psychological safety and value congruency.	ERIG	Psychological safety, value congruency	9	Psychological safety: was significantly higher for the peer assessment conditions compared to teacher assessment (i.e., control) condition. Value congruency: was significantly higher for the peer assessment+ condition (that involved reflection on the roles of assessors and assesses) than the control (teacher assessment) condition.

62	Van Gennip et al., 2010	To investigate: (a) how a peer assessment intervention can influence interpersonal variables (psychological safety, interdependence, value diversity, and trust) (b) the relationship between interpersonal variables, and students' conceptions of peer assessment and their learning	ERIG	Psychological safety, Trust in the other as assessor, value diversity, and interdependence	8	Psychological safety: Students felt safer in a peer assessment condition compared to teacher assessment condition. Trust in peer as assessor: Students reported more trust in peer as assessor after the peer assessment intervention. No significant differences in trust were found between the peer assessment and the teacher assessment conditions. Value diversity: Students reported a lower value diversity after the peer assessment intervention. Students perceived more value unanimity (i.e., low value diversity) in a peer assessment setting compared to teacher assessment settings. Interdependence: did not significantly change after the peer assessment intervention. No significant differences were found between the peer assessment and teacher assessment conditions in the perceptions of trust interdependence.
65	Vanderhoven et al., 2015	To investigate the effect of anonymity-using Classroom Response Technology (CRT)- on reducing undesirable social effects of peer assessment.	ERIG	Social connections (peer pressure, fear of disapproval)	9	Peer pressure: Participants experienced significantly less peer pressure in the anonymous CRT condition than the non-anonymous condition. Fear of disapproval: Participants experienced significantly less fear of disapproval in the anonymous CRT condition than the non-anonymous condition.
67	Wilson et al., 2015	To investigate students' perceptions of fairness, trust, and temporality of online peer review Moodle tool.	DSO	Trust in the other as assessor	1	In general, students' trust in peer as assessor was low.

Supplementary material 8.

Instrument to report the characteristics of peer assessment designs (version 1, January 2023)

Designed by 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*. Use that citation if you include the instrument in your publication.

Also accessible (check for updates to the instrument): https://osf.io/5k42z/?view_only=c77740eca9ef44978e1ac47abcaef7c

TITLE OF THE ARTICLE

Click here to add text

Authors

Click here to add text

Our study investigates:

- The impact of peer assessment on interpersonal/intrapersonal factors
- The impact of interpersonal/intrapersonal factors on peer assessment
- The impact of peer assessment on performance/behaviour
- The impact of moderator/mediator variables on peer assessment outcomes

Describe the characteristics of your peer assessment study in the table below.

Context		
Category	Description ⁱ	Our study
1 Subject domain	Subject domain the study was done in (e.g., mathematics, instructional sciences, accounting, etc.)	Click here to add text
2 Place/Time	Where was the PA conducted? (In class or out of class?)	<input type="checkbox"/> In class/class time <input type="checkbox"/> Out of class/free time
3 Setting	Formal or informal education setting?	<input type="checkbox"/> Formal <input type="checkbox"/> Informal
4 Requirement	Was PA compulsory or voluntary for assessor/assessee?	<input type="checkbox"/> Compulsory <input type="checkbox"/> Voluntary
5 Alignment	Was the PA activity aligned to curriculum, learning goals or teaching?	<input type="checkbox"/> Yes <input type="checkbox"/> No
Instructional design		
6 Purpose	What was the assessment purpose of the PA activity? (Formative, summative or both?)	<input type="checkbox"/> Formative <input type="checkbox"/> Summative <input type="checkbox"/> Both
7 Object	What was assessed? (e.g., written assignment, oral presentation, contribution to group work)	Click here to add text
8 Product/Output	What was the output of the PA? (e.g., score, written feedback, oral feedback, or a combination)	Click here to add text
9 Relation to staff assessment	Was PA done without staff assessment (substitutional) or in addition to staff assessment (supplementary)?	<input type="checkbox"/> Substitutional <input type="checkbox"/> Supplementary
10 Official weight	Did participation in the PA activity or the grade given by peer(s) contribute to learners' final grades?	<input type="checkbox"/> No <input type="checkbox"/> Yes-for participation in PA <input type="checkbox"/> Yes-for PA grade <input type="checkbox"/> Yes-both (PA & participation) <input type="checkbox"/> Other: Click here to add text
11 Reward	Was there a reward for participation in PA?	<input type="checkbox"/> No <input type="checkbox"/> Yes-course credit <input type="checkbox"/> Yes-incentives (e.g., free time, money, etc.) <input type="checkbox"/> Other: Click here to add text
12 Directionality	Was the learner assessing another without being assessed (unidirectional) or acting as both assessor and assessee (bidirectional)?	<input type="checkbox"/> Unidirectional <input type="checkbox"/> Bidirectional

13	Degree of interactivity	How did the assessee demonstrate engagement and response to PA?	<input type="checkbox"/> Reactive: assessee responds to assessor <input type="checkbox"/> Reciprocal: same learners assess each other on same task <input type="checkbox"/> Negotiated: PA was done more than once on the same task and both parties negotiated it
14	Frequency	How often was the PA of the same task done? (Once, twice, etc.)	Click here to add text
15	Group constellation	Did members of the same group assess each other (intragroup) or peers from another group (intergroup) or both?	<input type="checkbox"/> Intragroup <input type="checkbox"/> Intergroup <input type="checkbox"/> Both
16	Constellation assessor	The number of assessors assigned to each assessee	Click here to add text
17	Constellation assessee	The number of assessees per assessor	Click here to add text
18	Unit of assessment (assessor)	At what level did the assessor(s) perform PA? Individual, group, or both?	<input type="checkbox"/> Individual <input type="checkbox"/> Group <input type="checkbox"/> Both
19	Unit of assessment (assessee)	At what level did the assessee(s) experience the PA? Individual, group, or both?	<input type="checkbox"/> Individual <input type="checkbox"/> Group <input type="checkbox"/> Both
20	Privacy	Did assessor and assessee know the identity of each other? (Was PA public, single-blind, or double-blind?)	<input type="checkbox"/> Public <input type="checkbox"/> Single-blind (for assessor) <input type="checkbox"/> Single-blind (for assessee) <input type="checkbox"/> Double blind (anonymous) <input type="checkbox"/> Other: Click here to add text
21	Contact	Was PA done face-to-face or online? How was the contact between assessor and assessee?	<input type="checkbox"/> Face-to-face synchronous: same time same place (ST SP) <input type="checkbox"/> Online synchronous: same time different place (ST DP), e.g., skype, chat, etc. <input type="checkbox"/> Online asynchronous: different time different place (DT DP), e.g., e-mail, SWoRD, etc. <input type="checkbox"/> Other: Click here to add text
22	Matching	How were assessor and assessee matched for the PA activity? (e.g., random, skill, self-select, gender, friendship, performance, etc.)	Click here to add text
23	Format	How was the PA guided?	<input type="checkbox"/> Freestyle (no instruction provided to learners) <input type="checkbox"/> Guided instruction <input type="checkbox"/> Guided prompts <input type="checkbox"/> Guided criteria/rubrics <input type="checkbox"/> Guided criteria/rubrics and prompts <input type="checkbox"/> Other: Click here to add text
24	Training	Did the learners receive PA training at any time? If yes, describe the moment in which they received it and the type of training.	Click here to add text
25	Revision	Did learners revise their work after receiving or providing PA?	<input type="checkbox"/> No <input type="checkbox"/> Yes
26	Scope of involvement	How were learners involved in the PA activity?	<input type="checkbox"/> Development of assessment criteria <input type="checkbox"/> Gave PA only <input type="checkbox"/> Received PA only <input type="checkbox"/> Gave and received PA <input type="checkbox"/> Additional self-assessment <input type="checkbox"/> Other: Click here to add text
Outcomes			
27	PA Outcomes	These variables are directly measured as outcomes of the PA activity (i.e., why was the PA activity conducted?) Select all the options that apply to your study from the right column.	<input type="checkbox"/> Beliefs & perceptions: including perceptions of learning capacity to perform PA or any perceptions related to the PA processes (e.g., fairness, usefulness), metacognition and self-regulation, attitudes and beliefs (e.g., self-efficacy), teachers' perceptions/conceptions. <input type="checkbox"/> Emotions and motivation: emotions experienced by learners (e.g., achievement emotions, social emotions, etc.) & all motivational beliefs (e.g., learning motivation). <input type="checkbox"/> Performance: academic/domain specific performance, achievement, improved draft/work (i.e., revision).

	<ul style="list-style-type: none"> <input type="checkbox"/> Skills: quality of contribution to the group, professional behaviour, problem solving skills, work habits, interpersonal skills, metacognitive & self-regulatory skills. <input type="checkbox"/> Reliability of PA: consistency of PA scores compared to other peer assessors, or PA over several rounds. <input type="checkbox"/> Validity of PA: accuracy of PA compared to teachers/tutor/expert's assessment. <input type="checkbox"/> PF content: characteristics of the feedback messages including type, focus, quality, frequency of comments/posts (i.e., participation), etc. <input type="checkbox"/> PF processing: Includes implementation, reactions to PF, seeking PF, coping with PF, etc. <input type="checkbox"/> Other: Click here to add text
Moderators/mediators	
<p>28 PA Moderators/mediators</p> <p style="text-align: center;">Variables that are not usually manipulated but are taken into account when investigating PA. Select the variables that have been explored in your study from the right column.</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Gender: of assessor/assessee. <input type="checkbox"/> Ability & Skills: includes prior knowledge, prior performance, achievement level, GPA, finished high school, previous level of education, year of enrolment, etc. <input type="checkbox"/> Skills: reviewing ability, computer skills, etc. <input type="checkbox"/> Age: of assessor/assessee. <input type="checkbox"/> Culture, ethnicity, nationality or race <input type="checkbox"/> Other: Click here to add text
<p>This tool is based on: Alqassab, M., Strijbos, J., Panadero, E., Fernández Ruiz, J., Warren, M., & To, J. (2023). A systematic review of peer assessment design elements. <i>Educational Psychology Review</i>.</p>	

Note. PA = peer assessment; PF = peer feedback

ⁱ For more detailed descriptions refer to the Online Resource 3 by Alqassab et al. (2023) via: https://osf.io/4jbr3/?view_only=9f5b223115f244ac88ac5b054eb21149

If you have further questions you can contact us: erlaresearch@gmail.com