

The impact of peer assessment design on interpersonal processes: A systematic review

Abstract

Peer assessment can have a very positive impact on students' learning. However, it also poses some challenges such as trust issues, concerns about its fairness, feelings of discomfort, etc. It is known that the interpersonal processes of peer assessment can affect its outcomes. However, it is less clear how peer assessment should be implemented to prevent this. In this article, we reviewed 27 studies that have investigated how the instructional design elements of peer assessment influence interpersonal processes. A literature search was carried out on Scopus, PsycINFO, and Eric with a search string composed of three parts with keywords linked to (1) peer assessment, (2) educational contexts, and (3) interpersonal processes. The 14552 retrieved records were screened in terms of the inclusion criteria, which resulted in 27 studies being included and analysed through a narrative synthesis. The results revealed that the three most commonly-investigated design elements were privacy, format, and training. Privacy seems to have a positive impact on some interpersonal processes, but a detrimental effect on others. Concerning format, studies suggest that increased amounts of guidance are associated with a positive impact on interpersonal processes, but this result must be interpreted with caution due to the limitations of the studies' research designs. The use of training seems to be beneficial in minimising the negative effects of interpersonal processes. Finally, no conclusion could be drawn on three other design elements that were only investigated in one study each, and with weak research designs. These results allowed us to draw important recommendations for practitioners implementing peer assessment and to identify gaps and needs that could inform future research.

Keywords: peer evaluation, literature review, interpersonal processes, instructional design elements

1. Introduction

1.1 Peer assessment and its challenges

Formative assessment has the power to foster students' learning (Wisniewski et al., 2020). A key component of formative assessment – feedback – can help students close the gap between their actual and desired performances. Consequently, feedback is considered one of the most critical influences on student learning (Hattie & Timperley, 2007). Involving students in formative assessment by asking them to provide feedback to their peers can be particularly profitable for students' learning, given that peers are often conceptualised as potential mediators of students' learning (Panadero et al., 2018).

The term “peer assessment” is used to define situations in which students provide feedback to their peers. More precisely, peer assessment can be defined as “an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners, then learn further by giving elaborated feedback and discussing their judgment

with peers to achieve a negotiated agreed outcome” (Topping, 2017, p. 2). Other definitions of peer assessment have been proposed (e.g., Falchikov, 1995; Topping, 1998), but Topping’s (2017) definition has the advantage of clearly delimiting the concept of peer assessment, which makes it easier to determine if studies implemented peer assessment and, therefore, should be included in the review.

Recent meta-analyses have found that students can learn more from peer assessment than from teacher assessment (Double et al., 2020; Wisniewski et al., 2020). There are multiple reasons why peer assessment is an effective way to enhance students’ learning. Students are often able to explain what they have just learned in a more accessible and encouraging way than teachers (Nicol & Macfarlane-Dick, 2006). Moreover, with peer assessment, students perform both the role of assessee and assessor. Some studies suggest that it is the latter role that allows them to learn the most (e.g., Çevik, 2015; Lundstrom & Baker, 2009).

However, peer assessment does not automatically increase students’ learning. Indeed, even though meta-analyses found that peer feedback has a medium (Double et al., 2020) to strong effect (Wisniewski et al., 2020) on students’ learning, the effect sizes of each study are heterogeneous (Wisniewski et al., 2020). This variability is not specific to peer feedback. The effect of teacher feedback is also highly variable (Wisniewski et al., 2020), which suggests that feedback is a complex process with many elements involved that can increase or hinder its efficacy. The complexity of feedback was recently captured by the MISCA meta-model developed by Panadero and Lipnevich (2022). According to this model, five elements are important to take into consideration when implementing or researching feedback. The first element, student characteristics (i.e., gender, prior knowledge, self-efficacy, etc.), is an overlapping element situated in the centre of the model as the learner belongs at the very centre of the feedback process. The four other elements are the message (i.e., the information contained in the feedback), the implementation (i.e., the purpose of the feedback and the processes it triggers in the learner), the context (i.e., the instructional context in which the feedback takes place and the presentation of the feedback), and the agents (i.e., the agents involved in the feedback process, their role, and their interactions). These five elements are interrelated, and need to be aligned to allow the learner to use the feedback in such a way as to enhance their learning (Panadero & Lipnevich, 2022).

The “agents” element of the MISCA model is central to peer assessment (Panadero & Lipnevich, 2022). In terms of this element, the source of feedback plays an important role. Indeed, the specific characteristics of peer assessment pose specific challenges. Students tend to be reluctant to participate in peer assessment and to ignore the feedback they receive (e.g., Wichmann, 2018; Winstone et al., 2017), which risks hindering its efficacy. This can be explained by two fundamental aspects of peer assessment: students do not have the same status and expertise as their teachers (Gielen et al., 2010) and they can have various social connections with their peers (Harris & Brown, 2013). A series of challenges can arise from this, such as trust issues (e.g., Hanrahan & Isaacs, 2001), concerns about the fairness of peer assessment (e.g., Zou et al., 2018), feelings of discomfort during the peer assessment process (e.g., Stanier, 1997), the importance of “face-saving” (e.g., S. Yu et al., 2016), or bias due to the relationships

between peers (e.g., Domínguez et al., 2016). These challenges have in common that they emerge from the interpersonal processes of peer assessment.

Different terms are used in the literature to label the interpersonal processes of peer assessment, such as “social aspects” (Panadero, 2016, p. 247), “social factors” and “social process” (Topping, 2017, p. 13), “interpersonal variables” (van Gennip, 2010, p. 281). None of these terms have been clearly defined, and it is not clear which individual variables are associated with each of these terms. In their seminal paper, van Gennip and colleagues (2009) identified four “interpersonal variables”: trust, psychological safety, value congruency, and interdependence. In subsequent research, other researchers have studied other variables related to the interpersonal processes of peer assessment, such as perceived fairness, fear of disapproval, and feeling of discomfort (e.g., Raes et al., 2015; Vanderhoven et al., 2015). More recently, Panadero et al. (2023) have proposed a conceptualization in two categories: *intrapersonal factors* and *interpersonal factors*. The latter refers to “elements of the peer interaction or that influence the interaction in the peer assessment process” (Panadero et al., 2023, p. 1057) and only includes five variables (social connections, trust in the other as assessor, psychological safety, value diversity, interdependence). This conceptualization has the advantage of setting clear boundaries to the notion of interpersonal factors, but is quite restrictive. For example, notions that are more relevant in an Asian context than in a European context such as face-saving or group harmony (e.g., Fithriani, 2018; S. Yu et al., 2016), are not included in this conceptualization. In this review, we will adopt a broader conception of interpersonal processes and include all variables that have been described as “social” or “interpersonal” by the researchers studying them. As a result, some processes that are not included in Panadero et al.’s (2023) conceptualization are included in this review (e.g., face-saving, peer pressure). Moreover, we will not strictly follow Panadero et al.’s (2023) distinction between intra and interpersonal factors, as we consider any process that results from, or is strongly influenced by, the peer interactions that take place in peer assessment to be interpersonal. Therefore, four processes that would be considered intrapersonal factors in Panadero et al.’s (2023) conceptualisation (i.e., emotions due to the interpersonal processes, comfort, trust in the self as an assessor, and perceived fairness) are included in this review. These processes are also usually considered as interpersonal processes by the researchers investigating them (e.g., van Gennip et al., 2009; van Heerden & Bharuthram, 2021; Vanderhoven et al., 2015).

Given the challenges of peer assessment, it is important to implement peer assessment activities in a way that reduces the negative effects of interpersonal processes (Panadero et al., 2023). However, to be able to do this, it is necessary (1) to understand what the effects of these interpersonal processes are and (2) to understand how the negative effects of interpersonal processes can be minimized through peer assessment design. Previous literature reviews on the interpersonal processes of peer assessment provide a partial answer to these two questions.

1.2 Previous literature reviews: What do we already know?

Over the last two decades, the impact of the interpersonal processes of peer assessment has attracted a great deal of attention. Research on the subject has been synthesized in no less than

seven different literature reviews (Panadero, 2016; Panadero et al., 2018; Panadero & Alqassab, 2019, Panadero et al., 2023; Strijbos et al., 2009; Topping, 2017; van Gennip et al., 2009). These literature reviews aim to provide us with a better understanding of the role that interpersonal processes play in peer assessment, and the impact they have on its outcomes. Most of them only focus on the “agents” element of the MISCA model from Panadero and Lipnevich (2022) without considering the other elements. In the first systematic literature review on the subject, van Gennip and her colleagues (2009) aimed to understand the extent to which the outcomes of peer assessment are related to interpersonal processes. They only found four studies measuring interpersonal processes, none of which linked interpersonal processes to the learning outcomes of peer assessment. The review concluded that more research on the subject was needed. A similar conclusion was drawn in a thematic review published the same year, which deplored the lack of attention to interpersonal processes in peer assessment practices (Strijbos et al., 2009).

Following these calls, more researchers have investigated the interpersonal process of peer assessment, which has allowed subsequent reviews to be based on a more substantial body of evidence and to draw some conclusions. The systematic review conducted by Panadero (2016) revealed that the interaction between peer assessment and interpersonal processes should not be underestimated, and that the use of a formative approach to peer assessment could mitigate the negative effects of interpersonal processes. The same conclusions were highlighted in Panadero et al.’s (2018) thematic review of the main dissertation in the field. In a thematic review aimed at teachers, Topping (2017) warns that interpersonal processes of peer assessment can hinder its reliability and validity, and urges teachers to consider interpersonal processes. In addition, a review on the effect of anonymity in peer assessment (Panadero & Alqassab, 2019) showed the mixed impact of anonymity on interpersonal processes. Although empirical evidence is limited, this review suggests that anonymity has beneficial effects on some aspects (e.g., reducing peer pressure), but can have detrimental effects on others (e.g., reducing perceptions of fairness).

More recently, Panadero et al. (2023) conducted an update of the review published in 2016 (Panadero, 2016) aimed at synthesizing the results of research on the relationships between intra- and inter-personal factors of peer assessment. Even though the number of studies investigating these factors has increased, a major conclusion of this systematic review is that some of the limitations of these studies make it difficult to synthesize the results. These limitations were (1) a poor reporting of peer assessment intervention characteristics, (2) a lack of clear and agreed-upon definitions of the factors investigated, and (3) a lack of strong research design and a lack of replication studies. They proposed different ways to overcome these limitations in future research. First, to improve the reporting of peer assessment intervention, they developed an instrument that researchers can use to systematically describe the design elements of the peer assessment activities implemented in studies. Second, to address the under-conceptualization of the field, they proposed definitions of the main intra- and inter-personal factors of peer assessment, and presented the idea of bi-directionality of the relationship between peer assessment and intra- and interpersonal processes. This idea emphasizes the fact that intra- and inter-personal factors of peer assessment can both affect and be affected by peer

assessment, and that it is important to take into consideration the directionality of the effects. Third, they urged researchers in the field to conduct studies with stronger research designs and to conduct more replication studies.

While all the published reviews highlight the limitations of existing studies on the subject, they also allow us to better understand the impact that interpersonal processes can have on peer assessment processes and outcomes. For example, studies included in these reviews found that the social connections between students influence the scores they provide to their peers (e.g., Domínguez et al., 2016), that low perception of fairness is associated with less engagement in peer assessment (e.g., Zou et al., 2018), or that low levels of trust in the other as an assessor is associated with more effective approaches to learning during peer assessment (Cheng & Tsai, 2012). These results suggest that interpersonal processes of peer assessment are one of the factors that explain the variability in the efficacy of peer assessment in fostering learning, and that they have to be considered when designing peer assessment activities.

1.3 What is missing? Purpose of the present review study

The existing reviews have emphasised the importance of considering the interpersonal processes of peer assessment, but give little indication of how such assessment should be designed to take into account these interpersonal processes. The MISCA model (Panadero & Lipnevich, 2022) emphasises the importance of considering the interconnections between the different elements of feedback. To address the question of how to consider interpersonal processes such as to ensure they do not harm peer assessment outcomes, the interconnection between the “agents” element and the “context” element seems particularly important. The only review that incorporates the context element of feedback is one on the impact of anonymity (a contextual element) conducted by Panadero and Alqassab (2019), which revealed the mixed impact of anonymity. However, even though anonymity is an important aspect of the context of peer assessment, it is far from being the only one.

Indeed, when implementing peer assessment, the teaching team has a lot of decisions to make, such as matching the assessor and the assessee (e.g., randomly, by competence level...), the intended output of peer assessment (e.g., comments, scores...), and the guidance that will be offered to students (e.g., rubrics, prompts...). Alqassab et al. (2023, p. 18) label these decisions as relating to instructional design (i.e. “instructional design refers to what designers can manipulate”). These authors developed a typology including twenty-two elements (e.g., matching, training, directionality, frequency, privacy...). The present review is the first to examine the literature using this typology of instructional design elements as we strive to understand how peer assessment can be designed to minimize the negative effect of interpersonal processes.

In the present article, we set out to review studies that have investigated how the instructional design elements of peer assessment influence interpersonal processes. Echoing Panadero et al.’s (2023) idea of bi-directionality, we are interested here in investigating how peer assessment influences interpersonal processes, not how interpersonal processes influence peer assessment. Indeed, our aim is to understand how peer assessment can be designed to minimise the negative effects of interpersonal processes. The other direction (i.e. the influence of interpersonal

processes on peer assessment) has already been synthesised in several literature reviews (e.g., Panadero, 2016; van Gennip et al., 2009). Therefore, this study will review studies that have investigated the impact of peer assessment design elements (dependent variable) on interpersonal processes of peer assessment (independent variable).

RQ1. What are the characteristics of the studies investigating the impact of instructional elements on the interpersonal processes of peer assessment?

RQ2. Which instructional design elements have been manipulated and which interpersonal processes have been measured in these studies?

RQ3. What are the impacts of the manipulated instructional design elements on interpersonal processes?

This review will be of interest to both researchers and practitioners. For researchers, the results will give an overview of which instructional design elements have been manipulated to influence which interpersonal processes, and which relationships are still to be investigated. For practitioners, the results can provide guidance on how to design peer assessment activities.

2. Methodology

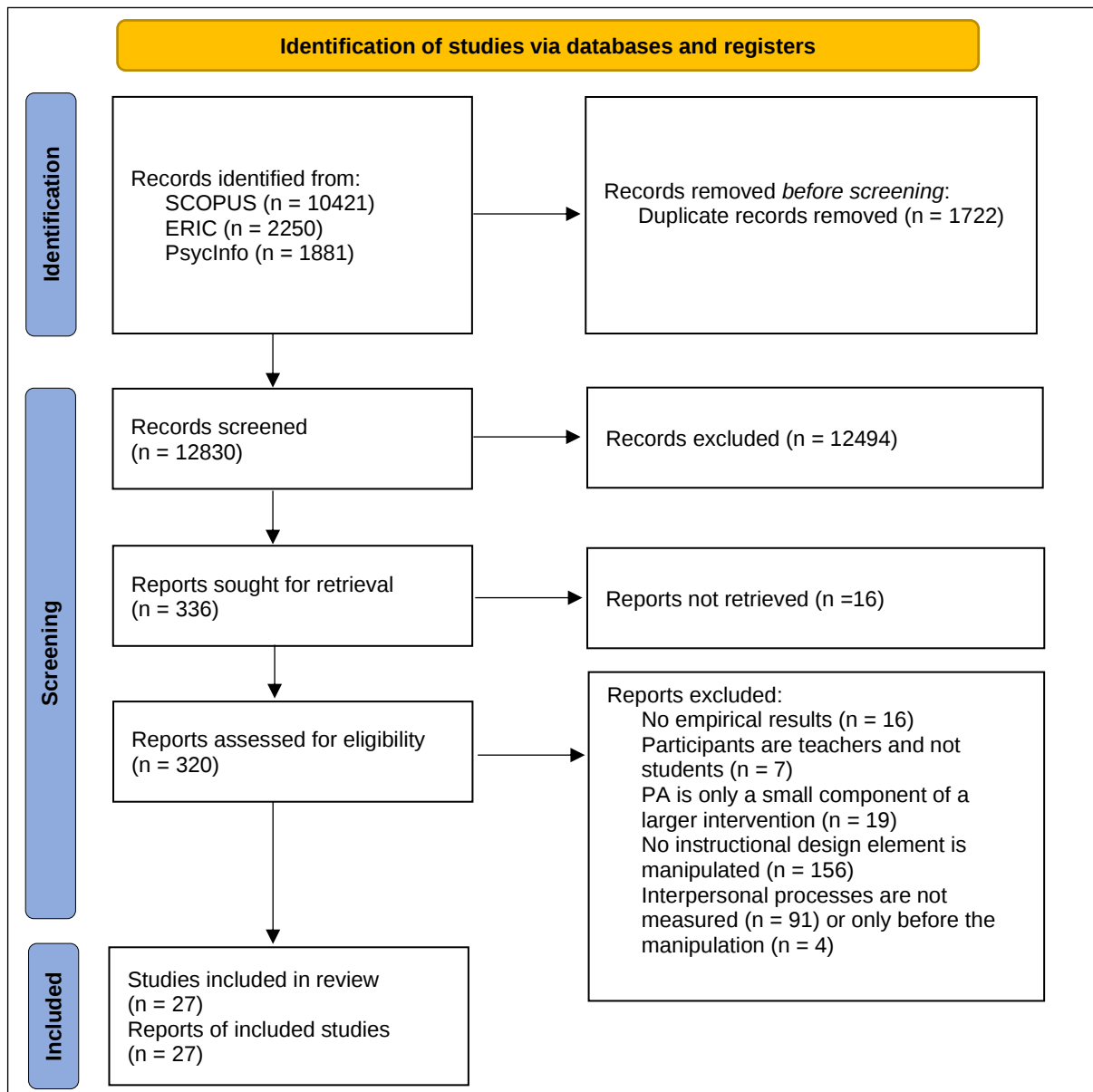
2.1 Search strategy

As it is recommended that at least two databases be used (Bramer et al., 2017; Ewald et al., 2022), the literature search was carried out using three complementary databases: Scopus (a general database), PsycINFO (a database specialised in the field of psychology), and ERIC (a database specialised in the field of education). The search strings consisted of three connected parts. The first was composed of peer assessment and its most common synonyms (“peer assessment”, “peer evaluation”, “peer feedback”, “peer grading”, “peer marking” and “peer review”). These search terms were included with and without hyphenation. The second part was composed of keywords linked to educational contexts (e.g., “student”, “school”, “college”, “elementary education”, etc.). We set out to include studies conducted in primary, secondary, and higher education. The keywords related to the first two parts were searched for in the titles, abstracts, and keywords of the articles. For the third part, related to interpersonal processes, a broader approach was adopted. To capture all relevant articles, a wide selection of search terms was used (e.g., “interpersonal”, “social”, “relation”, “psychological safety”, “trust”, “fear of disapproval”, “friendship”, etc.), and these terms were searched for in the full text. When available (i.e., for PsycINFO and Eric), the thesaurus was also used for each part of the search string. The complete search string is available in the Appendix 1.

A first search was conducted in July 2020, and an update was conducted in September 2023. The search was limited to peer-reviewed journal articles published in English. No time restrictions were placed; the results thus included all studies from these three databases until September 2023. In total, 14552 records were identified and exported to a reference management software that identified 1722 duplicates. These duplicates were manually removed, which left 12830 unique records (see Figure 1).

In addition, a snowballing method was used to search for additional publications. The references cited in the included articles were checked, as well as the other literature reviews on the subject (Panadero, 2016; Panadero et al., 2018; Panadero & Alqassab, 2019; Panadero et al., 2023; Strijbos et al., 2009; Topping, 2017; van Gennip et al., 2009). No additional records were identified through this snowballing procedure.

Figure 1. Flow diagram of selection of relevant publications for the review.



Note From: Page et al. (2021)

2.2 Selection process

Titles and abstracts were scanned using the following criteria: (1) The study was published in English in a scientific peer-reviewed journal, (2) the study involved empirical research, (3) the study was conducted in an educational context and involved students (4) the study investigated the interpersonal processes of peer assessment, (5) the peer assessment was performed on a learning task and with students from the same educational level and (6) the study manipulated at least one instructional design element.

The first criterion aimed to exclude studies that were of insufficient scientific quality to be accepted through a peer-review process. With regard to the second criterion, reviews, opinion

pieces, and studies presenting peer assessment tools were excluded. Quantitative, qualitative, and mixed-methods studies were included. The third criterion restricted the selection to studies conducted in an educational context (from primary school to higher education) and excluded studies done in a professional context (e.g., studies on peer reviews between researchers). The fourth criterion allowed us to only select studies that investigated at least one interpersonal process of peer assessment. As explained in the introduction, we adopted a broad definition of interpersonal processes of peer assessment and included all factors that have been considered as “social” or “interpersonal” by the researchers investigating them. The fifth criterion was used to remain in line with Topping’s (2017) definition of peer assessment. It excluded studies in which the peer assessment was conducted between students from different levels (e.g., master’s students assessing bachelor’s students’ work), and studies in which peer assessment was not used as a learning tool (e.g., peer assessment only used to weight teacher’s scores of group work). Finally, in line with our research question, studies were included if they manipulated at least one of the twenty-two elements of instructional design described in the typology of Alqassab et al. (2023). When the title and abstract did not provide sufficient information to determine if the study fully satisfied the inclusion criteria, the publication remained in the selection. If it became clear in the next step that a publication did not match the inclusion criteria, it was removed.

The screening of titles and abstracts was conducted by the first author. To verify its reliability, a second coder screened 10% of the references (1300 references). As most references did not meet the inclusion criteria, resulting in unbalanced marginals, Gwet’s AC₁ is the most reliable indicator (Black et al., 2016). According to the cut-offs discerned by Landis & Koch (1977), the Gwet’s AC₁ indicated an almost perfect agreement [Gwet’s AC₁ = 0.98 (95% CI, 0.969–0.985), $p < .001$] (Gwet, 2014).

Based on the title and abstract scan, 12494 records were excluded. Full-text versions of the remaining publications were obtained, with the exception of sixteen publications that could not be retrieved, even after a personal request to the authors. The full texts of the remaining 320 references were screened, and 293 publications were excluded based on the eligibility assessment (the reasons for exclusion are detailed in Figure 1). This resulted in 27 articles that were included in the review and subsequently analysed.

2.3 Data extraction and data analysis

As a variety of types of studies are included in the review, they were analysed through a narrative synthesis, following the guidelines of Petticrew and Roberts (2006). This narrative analysis consisted of three steps. First, the relevant information for each article was recorded using a data extraction form. This data form was inspired by other literature reviews (Heitink et al., 2016; Lindvall & Ryve, 2019) and consisted of the following sections:

- General information (authors’ names, publication date, publication title)
- Context (country, level of education, domain of education)
- Study aim (research question, interpersonal processes investigated as named in the study, instructional design elements manipulated as categorized by Alqassab et al., 2023)

- Research design (study type, study design, instruments)
- Sample (sample type, sample size, sample characteristics, sampling method)
- Results and conclusion

The characteristics of the study were descriptively analysed to answer our first research question (i.e., What are the characteristics of the studies investigating the impact of instructional elements on the interpersonal processes of peer assessment?). In the second step, each study was carefully read, and its results were explored in depth. In the third step, the results were analysed across the individual studies. The studies were organised based on (1) the instructional design elements they manipulated and (2) the interpersonal processes they measured (which allowed us to answer RQ2: Which instructional design elements have been manipulated and which interpersonal processes have been measured in these studies?). For each set of studies, the results were examined to judge whether they were consistent or contradictory across the studies. This allowed us to answer the third research question (What are the impacts of manipulated instructional design elements on interpersonal processes?) through a cross-study synthesis.

During the third step, the quality of the studies was also considered. The overall quality of the included studies was assessed, using three checklists proposed by Efron & Ravid (2019), one for quantitative studies, one for qualitative studies, and one for mixed-method studies. A second coder coded a subset of the quantitative, qualitative, and mixed-method studies. Gwet's AC₂ (using linear weights) was run to determine the agreement between the two raters on the quality appraisal with regard to 9 articles. The `agree.coeff2.r` script (Gwet, 2014) on R was used. There was a moderate to substantial agreement between the two raters, AC₂ = .573 (95% CI, .481 to .666), $p < .001$.

The aim of the quality appraisal is not to exclude any study, but to identify the weaknesses and strengths of each study to (1) consider them when synthesising the studies' results and (2) have an idea of the quality of the studies in the field to identify the limitations and perspectives. Indeed, even studies with important limitations can provide valuable information (Booth et al., 2016; Walsh & Downe, 2005), such as examples of innovative ways to design peer assessment activities, but it is important to consider the studies' quality. The critical appraisal of all 27 quantitative, qualitative, and mixed-method studies can be found in Appendices D, E, and F, respectively.

3. Results

3.1 RQ1: Description of included studies

The characteristics of the 27 included studies are quite homogeneous, as can be seen in Table 1. First, the studies have principally been conducted in three regions of the world: East and South Asia (China, Taiwan, Japan, Vietnam, and the Philippines), Europe (Belgium, Turkey, and Spain), and the United States of America. Second, most studies have been conducted in higher education, with only two studies conducted in secondary education, and one in primary education. Third, the majority of studies have been conducted in the context of education or English as a foreign language. Fourth, all the studies were published in the last fifteen years.

Concerning study design, most studies used a mixed method (14 studies), followed by a quantitative method (9 studies), and a qualitative method (4 studies). There is an overrepresentation of correlational studies (16 studies), with only eight quasi-experimental, and three experimental studies. The sample size ranges from 12 to 58 students for qualitative studies and from 16 to 209 students for mixed-method or quantitative studies.

3.2 RQ2: Investigated instructional design elements and interpersonal processes

To present their results, we organized studies based on the instructional design elements they manipulated (see Tables 2 and 3). The most commonly-investigated design element was privacy (15 studies), followed by the peer assessment format (6 studies) and the presence of training (4 studies). The matching of assessors and assessees, the product of assessment, and the unit of assessment (assessors) are each investigated in one study. The interpersonal processes of peer assessment investigated in the studies are, in order of frequency, *(dis)comfort* (11 studies), *trust* (8 studies), *emotions due to interpersonal processes* (7 studies), *perceived fairness* (5 studies), *peer pressure* (4 studies), *psychological safety* (3 studies), *friendship marking* (2 studies), *perceptions towards assessors* (2 studies), *fear of disapproval* (2 studies), *face-saving* (2 studies), and *value congruency* (1 study). Most studies investigated at least two interpersonal processes.

Table 1. Main characteristics of the included studies

Authors and date	Context		Research design			Sample size
	Country	Education setting	Subject	Study type	Study design	
Akçay et al. (2021)	Turkey	Higher education	Education	Qualitative	Correlational	N = 58
Cheng et al. (2023)	China	Higher education	Education	Mixed method	Experimental	N = 34
Ching & Hsu (2016)	USA	Higher education	Education	Mixed method	Correlational	N = 16
Güler (2017)	Turkey	Higher education	Education	Mixed method	Experimental	N = 84
Kaya (2021)	Turkey	Higher education	EFL	Qualitative	Correlational	N = 12
Kim (2019)	Japan	Higher education	EFL	Mixed method	Correlational	N = 64
Kim & Lan (2021)	China	Higher education	EFL	Mixed method	Correlational	N = 57
Li (2017)	USA	Higher education	Education	Quantitative	Quasi-experimental	N = 77
Lim et al. (2021)	USA	Higher education	Education	Mixed method	Correlational	N = 81
Lin (2018)	Taiwan	Higher education	Education	Quantitative	Experimental	N = 32
Ocampo et al. (2023)	Philippines	Higher education	Psychology	Quantitative	Quasi-experimental	N = 145
Panadero et al. (2013)	Spain	Higher education	Education	Quantitative	Quasi-experimental	N = 209
Parikh & Sheehan (2016)	USA	Higher education	Pharmacy	Quantitative	Correlational	N = 144
Praver et al. (2011)	Japan	Higher education	EFL	Quantitative	Correlational	N = 86
Raes et al. (2015)	Belgium	Higher education	Education	Mixed method	Quasi-experimental	N = 51
Rotsaert et al. (2018)	Belgium	Higher education	Education	Mixed method	Correlational	N = 46
Senden et al. (2023)	Belgium	Higher education	Education	Quantitative	Quasi-experimental	N = 89
Su (2023)	China	Higher education	EFL	Mixed method	Correlational	N = 40

Thanh & Gillies (2010)	Vietnam	Higher education	/	Mixed method	Correlational	N = 145
Tzeng et al. (2021)	USA	Higher education	Medicine	Mixed method	Correlational	N = 32
van Heerden & Bharuthram (2021)	South Africa	Higher education	Academic writing	Mixed method	Correlational	N = 64
Vanderhoven et al. (2015)	Belgium	Secondary education	Oral skills	Mixed method	Quasi-experimental	N = 69
Xue et al. (2023)	China	Higher education	Academic writing	Qualitative	Correlational	N = 12
Yu (2012)	Taiwan	Secondary School	Science	Quantitative	Quasi-experimental	N = 101
Yu & Liu (2009)	Taiwan	Higher education	Engineering	Mixed method	Correlational	N = 40
Yu & Wu (2011)	Taiwan	Primary education	Science	Quantitative	Quasi-experimental	N = 243
Zhang et al. (2022)	China	Higher education	Academic writing	Qualitative	Correlational	N = 12

3.3 RQ3: Impact of privacy on interpersonal processes

Peer assessment can be organized in terms of various levels of privacy (Alqassab et al., 2023): it can be entirely public (i.e., all names are disclosed to all participants), confidential (i.e., names of the assessee and the assessor are only disclosed to peers with whom they interact), single-blind (anonymity of the assessor or the assessee), or double-blind (anonymity of both assessor and assessee). Five studies have investigated the impact of double-blind anonymity (Kaya, 2021; Su, 2023; F. Yu, 2012; F. Yu & Liu, 2009; F. Yu & Wu, 2011), while seven studies have investigated the impact of single-blind anonymity, with only the identity of the assessor being concealed (Güler, 2017; L. Li, 2017; Lim et al., 2021; Lin, 2018; Raes et al., 2015; Rotsaert et al., 2018; Vanderhoven et al., 2015). Finally, three studies have combined the effects of anonymity and face-to-face contact by comparing online anonymous peer feedback with face-to-face non-anonymous peer feedback (Akçay et al., 2021; Kim, 2019; Kim & Lan, 2021). The results of these studies are summarized in Table 2.

Double-blind anonymity

In the three studies conducted by Yu and colleagues (F. Yu & Liu, 2009; F. Yu & Wu, 2011; F. Yu, 2012), the authors obtained mixed results on the impact of disclosing students' names or anonymity (either with a nickname or no name). In F. Yu and Liu's (2009) study on emotions relating to the interpersonal processes, higher education students could choose the identity revelation mode they wanted to use as assessor and as assessee. They expressed a strong preference for the anonymous mode, both as an assessor and as an assessee, because, as they reported, having their identity concealed helps them avoid feelings of unease, tensions, shameful feelings, and humiliating situations. The two other studies both investigated the impact of the different identity revelation modes on students' perceptions toward their assessors. In F. Yu & Wu's (2011) study, eight classes of primary school pupils were assigned to one of four conditions: real names, no name, nicknames, or self-choice (i.e., students could choose any of the previous modes). The perceptions toward their assessors were more positive in the real names and self-choice conditions which suggest that anonymity, at least when imposed, impairs assessee's perceptions of the assessors. However, these results were not replicated in F. Yu's (2012) study, conducted with secondary school pupils: no significant differences were found between the conditions regarding assessee's perceptions of the assessors.

The impact of double-blind anonymity on interpersonal processes was also investigated in two other studies: Kaya's (2021) study showed a mixed impact of privacy on emotions due to the interpersonal processes, while Su's (2023) study revealed a positive impact on comfort. In both studies, higher education students experienced non-anonymous peer feedback, followed by anonymous peer feedback to compare their perceptions. In Kaya's (2021) study, students felt tense when they were identifiable, while anonymity alleviated this tension by making them feel free to provide objective feedback. However, anonymous peer feedback also conveyed negative emotions (anger, upset, and disappointment) because, in the anonymous condition, students attributed the poor performance of their peers to a lack of effort, which was not the case in the non-anonymous condition. In a study by Su (2023), students felt significantly more comfortable

when providing and receiving anonymous peer feedback than was the case with non-anonymous peer feedback. They reported feeling more at ease and relaxed under conditions of anonymity.

Single-blind anonymity

The results from the three studies that examined the effects of the provision of anonymous feedback on peer pressure are also mixed. Vanderhoven et al. (2015) compared the perceptions of students who anonymously or non-anonymously provided feedback to their peers in class. Students felt significantly less peer pressure when they provided anonymous feedback (through a classroom-response technology) than when they provided non-anonymous feedback (through card raising). However, this impact of anonymity on peer pressure was not found in L. Li's (2017) and Raes et al.'s (2015) studies. In L. Li's (2017) studies, students were also assigned to provide anonymous or non-anonymous peer feedback, but the perceptions of peer pressure did not differ significantly between the groups. In Raes et al.'s (2015) study, students gave scores to their peers using classroom-response technology, participated in an oral feedback exchange moderated by the teacher, and wrote comments they uploaded online. In the first group, the uploaded feedback was anonymous, whereas it was non-anonymous in the second group. There was no significant difference in students' perceptions of peer pressure in the two groups, even though their answers to open questions suggested that an increase in anonymity is associated with a decrease in peer pressure.

Single-blind anonymity, like double-blind anonymity, seems to positively affect students' perceptions of comfort. Indeed, two studies (Lim et al., 2021; Raes et al., 2015) found a positive effect, while the third did not find any effect (Vanderhoven et al., 2015). In Lim et al.'s (2021) study, the assessee was asked to leave the room while their peers provided feedback on their performance (feedback that will be communicated afterwards by the teacher). Students reported that, compared to previous experiences of peer assessment, they found this type of peer assessment more comfortable because, as an assessor, they were able to express themselves without being afraid or hurting someone's feelings, and, as an assessee, they know the feedback they receive is authentic. In Raes et al.'s (2015) study, anonymity had a significant effect on comfort: students felt more comfortable when they provided anonymous feedback through classroom-response technology than when they provided non-anonymous oral or written feedback. In Vanderhoven et al.'s (2015) study, students' perceptions of discomfort were not significantly different whether they provided anonymous or non-anonymous peer feedback.

Güler (2017), Lin (2018), and Vanderhoven et al. (2015) all compared the impact of the anonymity of the assessor on students' perceptions of fairness, but found different results. There was no significant difference between students in the anonymous and non-anonymous conditions in the studies of Güler (2017) and Vanderhoven et al. (2015), while students' perceptions of fairness were significantly higher for students in the non-anonymous condition in the study of Lin (2018).

Finally, Rotsaert et al. (2018) manipulated privacy in yet another way and found mixed results. They investigated the effect of a transitional approach to anonymity. Each week, students provided written feedback on their peers' presentations through mobile-response technology

followed by an oral discussion moderated by the teacher. In the first two weeks, the written feedback was anonymous, while in the final two weeks, it was non-anonymous. Students' perceptions of friendship marking, trust, psychological safety, fear of disapproval, and value congruency were already quite positive prior to the intervention. Rotsaert et al.'s (2018) findings indicated a decrease in fear of disapproval, which was also found in Vanderhoven et al.'s (2015) study. In addition, Rotsaert and colleagues (2018) found an increase in friendship marking. None of the other students' perceptions evolved with the transition to anonymity, which suggests that, according to the authors, the transitory approach revealing students' identities had no deleterious effects.

Anonymity and contact

The last three studies found different effects depending on the interpersonal process investigated (comfort, trust, face-saving, peer pressure, or perceived fairness) when comparing online anonymous peer feedback with face-to-face non-anonymous peer feedback (Akçay et al., 2021; Kim, 2019; Kim & Lan, 2021). In the study of Kim (2019) and Kim and Lan (2021), one group of randomly assigned students experienced first non-anonymous face-to-face peer assessment and then anonymous peer assessment with no interaction, while another group experienced the activities in reverse order. In the study by Kim (2019), conducted on Japanese students, assessors felt more comfortable providing negative feedback in the anonymous condition. However, they also felt comfortable providing negative feedback face-to-face because, through the interactions, hedging was easier, and they could work collaboratively with their peers to make them realise their mistakes. As assessees, they trusted their peers more in the non-anonymous condition, because the interactions helped the assessors to formulate accurate feedback. Kim and Lan (2021) conducted their study with Chinese students. They hypothesised that students' concerns about face-saving would hamper their abilities to provide constructive feedback when non-anonymous. The results confirm that face-saving matters to students, yet, students preferred face-to-face feedback: they were keen to provide high-quality feedback, to save face, and they found it was easier to do so when they could interact with their peers. In Akçay et al.'s (2021) study, all the students first experienced face-to-face peer feedback, and then single-blind peer feedback concealing the identity of the assessor. Some students felt pressure to provide positive feedback and higher grades when it was non-anonymous. Anonymity was seen as a solution to this: the students felt that they could provide the deserved scores and comments when anonymous. However, other students viewed anonymous peer feedback as being less fair, precisely because they thought that the pressure felt during face-to-face feedback positively contributed to the objectivity of the peer assessment, even though the authors did not clarify why this might be.

Table 2. Privacy and interpersonal processes investigated in the included studies

Privacy	Impact	Emotions due to the interpersonal processes	Perceptions towards assessors	Comfort	Peer pressure	Perceived fairness	Fear of disapproval	Friendship marking	Trust	Psychological safety	Value congruency	Face-saving
Double-blind anonymity	Positive	Kaya (2021) ; Yu & Liu (2009)		Su (2023)								
	Negative	Kaya (2021)	Yu & Wu (2011)									
	Absent		Yu (2012)									
Single-blind anonymity	Positive			Lim et al. (2021); Raes et al. (2015)	Vanderhoven et al. (2015)		Rotsaert et al. (2018); Vanderhoven et al. (2015)					
	Negative					Lin (2018)		Rotsaert et al. (2018)				
	Absent			Vanderhoven et al. (2015)	Li (2017); Raes et al. (2015)	Güler (2017); Vanderhoven et al. (2015)			Rotsaert et al. (2018)	Rotsaert et al. (2018)	Rotsaert et al. (2018)	
Anonymity and absence of contact	Positive			Kim (2019)	Akçay et al. (2021)							
	Negative					Akçay et al. (2021)			Kim (2019)			Kim & Lan (2021)

Note: Studies with an experimental or quasi-experimental design are in bold. Studies are highlighted in green if the researchers found that the design element had a desirable effect on interpersonal processes and in red if they found an undesirable effect.

3.4 RQ3: Impact of peer assessment format on interpersonal processes

The format of peer assessment was manipulated in six studies. Format refers to the degree of guidance of the peer assessment activity, i.e., did the students receive a rubric, some prompts, etc. to guide them? (Alqassab et al., 2023).

The studies of Xue et al. (2023) and Zhang et al. (2022) found a positive impact of guidance on face-saving and a mixed impact on emotions due to the interpersonal processes. These two publications relate to the same data collection, which aimed to compare the impact of three assessment tools offering different types of guidance (Moodle, rain Classroom, and WeChat) on face-saving and emotions due to interpersonal processes. Moodle is a learning management system that pushes students to follow the instructions thoughtfully when providing feedback, Rain Classroom is a tool created to project presenters' slides on the audience's mobile application while allowing them to provide quick real-time feedback, and WeChat is a social media platform that attends to the participants' affective factors. All students had the opportunity to use the three tools on multiple occasions to provide peer feedback. Students evoked concerns about face-saving (Zhang et al., 2022) and communication anxiety (Xue et al., 2023). They thought the most appropriate tool to deal with these issues was WeChat because it allowed them to insert emojis and memes in their feedback to soften harsh feedback, which the teaching teams encouraged students to do. However, according to the students involved, WeChat also created an emotional burden because students needed to be connected to their peers' social media accounts to be able to use it.

Two studies (Ching & Hsu, 2016; Tzeng et al., 2021) found a positive impact of a peer assessment format with more guidance on students' perceptions of comfort, while a third study did not find any effect (Panadero et al., 2013). It is worth noting that in the two studies identifying a positive impact, there was no real control group. Tzeng et al. (2021) introduced a method called "Observation-Reaction-Feedback" that aimed to guide students when they provided feedback on their peers' clinical skills. Students could choose to use the method or not. Students who chose to use it were those least comfortable with peer feedback at the pre-test. At the post-test, their level of comfort caught up with the levels of other students, and, according to their answers to open questions, this was thanks to the method. In Ching and Hsu's (2016) study, students played a self-select stakeholder role while providing feedback to their peers (i.e., they provided feedback from a particular perspective). Students reported feeling comfortable while participating in the peer feedback activity. In their quasi-experimental study, Panadero et al. (2013) compared students' perceptions of comfort when using a rubric or not. In two classrooms students received a rubric to provide feedback to their peers, while in the two other classrooms, they did not. There was no difference in students' perceptions of comfort between the two groups.

In their study on the use of rubrics, Panadero et al. (2013) also compared students' perceptions of fairness and their tendency to overmark their friends. Even though peer ratings were globally more valid when students used a rubric, friendship marking was only present in these classrooms: when students did not use a rubric, they tended to overmark all their peers, not just

their friends. There was no impact in terms of the use of the rubric on students' perceptions of fairness.

Finally, the presence of increased amounts of guidance seem to positively impact students' perceptions of trust and psychological safety. In Cheng et al.'s (2023) experimental study, students were randomly assigned to a group that was guided by a general script during the peer feedback, or to a group that received a more detailed regulation script for dialogic feedback with prompts. Students' attitudes toward the climate of trust were significantly more positive in the group with the more detailed regulation script for dialogic peer assessment. In the interviews, students explained that the prompts allowed them to ensure they shared a common understanding of the assessment criteria, which raised their feelings of trust. This positive impact on trust was also present in Ching and Hsu's (2016) study, alongside a positive impact on psychological safety. Students reported feeling psychologically safe and confident in themselves and others when participating in a peer feedback activity where they had to provide feedback from the perspective of a chosen stakeholder.

3.5 RQ3: Impact of training on interpersonal processes

The impact of training on the interpersonal processes of peer assessment was investigated in four studies, two of them investigating the impact of a general training session (Ocampo et al., 2023; Parikh & Sheehan, 2016), while the other two investigated the impact of a training session specifically designed to target interpersonal processes (L. Li, 2017; Senden et al., 2023).

The studies investigating the impact of a general training session found it had a positive impact on comfort (Ocampo et al., 2023) and on trust (Ocampo et al., 2023; Parikh and Sheehan, 2016). In Parikh and Sheehan's (2016) study, students received a 30-minute instructional lecture explaining how to provide feedback to their peers. Most students reported trusting their abilities and that of their peers to conduct peer assessments after receiving the training. However, there was no control group or use of a pre-test to compare the results. Ocampo et al. (2023) also investigated the effect of general training on students' perceptions, but used a quasi-experimental design and a longer training session. Students in the control group received a 30-minute training on how to use the peer feedback tool, while students in the experimental group received a 180-minute training that also allowed them to practice peer feedback. All students assessed the same three writing samples created by the research team, even though they were led to believe they were real samples from their peers. The researchers manipulated the gender that was indicated on the samples. Students' perceptions were significantly different between trained and untrained assessors, but only for male assessors providing feedback to women: trained men felt more comfortable and confident than untrained men when providing feedback to women.

The studies investigating the impact of a specific training session found a positive effect on peer pressure (L. Li, 2017), but not on trust and psychological safety (Senden et al., 2023). In L. Li's (2017) study, already mentioned in the section on privacy, students in the training group received a training session specifically conceived to address the peer pressure students can feel when their identities are revealed. Students who received the training felt significantly less peer pressure than students who did not receive it, regardless of whether their identities were

revealed or not. Training seems therefore more effective than privacy when it comes to addressing feelings of peer pressure. Senden et al. (2023) also created a training session focusing on the interpersonal processes of peer assessment, but targeting trust and psychological safety. Students received either a 1-hour general training or a 2-hour training with a part designed to address their concerns regarding trust and psychological safety. Students' perceptions of psychological safety, trust in the self, and trust in the other were already quite high at the pre-test and remained so in both groups.

3.6 RQ3: Impact of other instructional design elements on interpersonal processes

The three remaining studies each manipulated a different instructional design element: the unit of assessment (Thanh & Gillies, 2010), the output of assessment (Praver et al., 2011), and the matching of assessor and assessees (van Heerden & Bharuthram, 2021). These three studies were correlational in nature.

Thanh and Gillies (2010) found a positive impact when it came to conducting peer assessment in a group on students' emotions due to the interpersonal processes and feelings of comfort. Students assessed the contributions of their group's members in two ways. First, they assessed them individually, and second, they assessed them collectively, through an open discussion involving all group members. They also collectively assessed other groups' performances. Students' feelings of discomfort and worry were significantly lower when doing the collective assessment than when doing the individual one, which was also confirmed by the qualitative results.

In their study, Praver et al. (2011) found that having to write comments in addition to providing scores led to negative emotions on the part of students. The students had to facilitate a reading circle and were assessed by their peers, either with scores only (in the first three rounds) or with scores and written comments (in the last three rounds). At the end of the course, students had to indicate which peer assessment output they found the more useful, and which caused more nervousness and embarrassment. The vast majority of students found it more useful to provide and receive a combination of scores and written comments. However, most students also indicated that providing and receiving comments in addition to scores made them more nervous, and that providing comments in addition to scores caused more embarrassment, while there was no difference in embarrassment for the receiving part.

Finally, van Heerden and Bharuthram (2021) found a mixed impact in terms of familiarity on interpersonal processes. In their study, after exchanging feedback with a peer, students indicated how well they knew this peer and how they felt during the peer feedback activity. Based on descriptive analyses, it seems that students who knew their peers well found the experience more comfortable, enjoyable, and less challenging than those who did not have such a close relationship. The students explained that knowing their peers made it easier to take feedback as constructive criticism instead of viewing it as being harsh and critical and that it was associated with an implied level of honesty and trust. However, some students thought that not knowing their peers was beneficial because it made the process more objective.

Table 3. Instructional design elements and interpersonal processes investigated in the included studies

Element	Impact	Emotions due to the interpersonal processes	Face-saving	Comfort	Perceived fairness	Friendship marking	Psychological safety	Trust	Peer pressure
Format (more guidance)	Positive	Xue et al. (2023); Zhang et al. (2022)	Xue et al. (2023) ; Zhang et al. (2022)	Ching & Hsu (2016); Tzeng et al. (2021)			Ching & Hsu (2016)	Cheng et al. (2023); Ching & Hsu (2016)	
	Negative	Xue et al. (2023); Zhang et al. (2022)				Panadero et al. (2013)			
	Absent			Panadero et al. (2013)	Panadero et al. (2013)				
Presence of a general training	Positive			Ocampo et al. (2023)				Ocampo et al. (2023), Parikh & Sheehan (2016)	
Presence of a specific training	Positive								Li (2017)
	Absent						Senden et al. (2023)	Senden et al. (2023)	
Unit of assessment (group)	Positive	Thanh & Gillies (2010)		Thanh & Gillies (2010)					
Product (presence of comments)	Negative	Praver et al. (2011)							
Matching (familiarity)	Positive	van Heerden & Bharuthram (2021)		van Heerden & Bharuthram (2021)				van Heerden & Bharuthram (2021)	
	Negative				van Heerden & Bharuthram (2021)				

Note: Studies with an experimental or quasi-experimental design are in bold. Studies are highlighted in green if the researchers found that the design element had a desirable effect on interpersonal processes and in red if they found an undesirable effect.

4. Discussion

Previous literature reviews have revealed that interpersonal processes can influence peer assessment outcomes and processes (e.g., Panadero et al., 2023; Topping, 2017). Therefore, when designing peer assessment activities, the impact of design elements on interpersonal processes should be considered. Otherwise, these activities may not achieve their goal of promoting student learning. This systematic literature review innovates by linking the “agents” and “context” elements of the MISCA model (Panadero & Lipnevich, 2022) to explore how peer assessment can be designed to avoid the negative impacts of interpersonal processes and maximise its effectiveness.

Twenty-seven studies that investigated the impact of the instructional design elements of peer assessment on interpersonal processes were reviewed. The included studies were limited to peer-reviewed papers. Even though we did not restrict the included papers to a specific year of publication, none of them were published prior to 2009. Moreover, almost half of the papers were published in the last three years, which makes this literature review particularly timely. It seems that calls for more attention to the interpersonal processes of peer assessment (Strijbos et al., 2009; van Gennip et al., 2009) have been heard, and that the research on the subject is rapidly developing. If this research attention persists, an update of the literature review presented in this paper will be necessary in a few years.

Our review has uncovered that privacy was the instructional design element whose impact on interpersonal processes has been investigated the most, with more than half of the included studies focusing on it. It seems that anonymity is seen as the easiest solution to minimising the negative impacts of interpersonal processes. There is consistent evidence from several studies that privacy increases feelings of comfort (Kim, 2019; Lim et al., 2021; Raes et al., 2015; Su, 2023), decreases fear of disapproval (Rotsaert et al., 2018; Vanderhoven et al., 2015), and alleviates the tensions students can feel (Kaya, 2021; F. Yu & Liu, 2009). It also seems that it decreases peer pressure (Açkay et al., 2021; Vanderhoven et al., 2015), even though some studies did not find any significant effect (L. Li, 2017; Raes et al., 2015). Yet, as already suggested by Panadero and Alqassab (2019), the results confirmed that, even if anonymity can be beneficial, it is not a magical solution to erase student’s concerns. On the contrary, it could even have detrimental effects on some processes: evidence suggests that privacy can convey negative emotions (Kaya, 2021) and has a detrimental impact on perceived fairness (Açkay et al., 2021; Lin, 2018), perceptions toward the assessors (F. Yu & Wu, 2011), and value congruency (Rotsaert et al., 2018). However, some studies did not find that privacy had any significant effect on these processes (Güler, 2017; Vanderhoven et al., 2015; F. Yu, 2012). Finally, some studies aimed to examine the effect of privacy, but this was confounded with the effect of face-to-face contact. The detrimental effect they found in the case of trust (Kim, 2019) and face-saving (Kim & Lan, 2021) seems to come from the absence of contact, not from anonymity itself.

The second most common instructional design element investigated in the included studies was the format of peer assessment. The five studies focusing on the impact of the format of peer assessment used different interventions, all of which aimed to provide increased amounts of

guidance to the students. Not surprisingly, it seems that offering more guidance has a positive impact: it could help students “save face” (Xue et al., 2023; Zhang et al., 2022) and increase students’ perceptions of comfort (Ching & Hsu, 2016; Tzeng et al., 2021), trust (Cheng et al., 2023; Ching & Hsu, 2016), and psychological safety (Ching & Hsu, 2016). However, these results must be interpreted with caution due to the limitations in the studies’ research designs, especially as the study with the stronger research design (Panadero et al., 2013) did not find any significant impact in terms of format on comfort and perceived fairness. Panadero et al. (2013) also identified a negative impact of a rubric on friendship marking. However, this negative impact should be put into perspective: it did not imply a less valid peer assessment for friends in the rubric condition, as students using a rubric only overscored their friends, while students without a rubric overscored all their peers.

The impact of the presence of training on interpersonal processes has been investigated in four studies and seems to have only a positive effect. Two studies have found that general training increased students’ feelings of comfort (Ocampo et al., 2023; Parikh & Sheehan, 2016), while one study has found that a training session specifically created to decrease students’ feelings of peer pressure achieved its target (L. Li, 2017). The absence of a significant effect related to training in the last study (Senden et al., 2023) could be explained by the fact that students had high levels of trust and psychological safety prior to the training. This would be consistent with the results obtained by Ocampo et al. (2023) that training only increased trust in the case of students who started with low levels of trust. In sum, while previous literature reviews have shown that providing training prior to peer assessment has a positive impact on the efficacy of peer assessment (Li et al., 2020), students’ peer assessment skills (van Zundert et al., 2010), and students’ attitudes toward peer assessment (van Zundert et al., 2010), the present review adds that training is also beneficial when it comes to minimising the negative effects of interpersonal processes.

Finally, it is impossible to arrive at any conclusion with regard to the other three design elements investigated in the included studies: the unit of assessment (Thanh & Gillies, 2010), the product of assessment (Praver et al., 2011), and the matching between assessors and assessees (van Heerden & Bharuthram, 2021). Indeed, each of these elements are only investigated in one study, and these studies have weak research designs. More research is needed to understand the impact that these three design elements can have on the interpersonal processes of peer assessment.

4.1 Gaps unveiled by the literature review and perspectives for future research

Our review found that the impact of only seven of the 21 instructional design elements identified by Alqassab et al. (2023) had been investigated. The impact of other elements on interpersonal processes should also be investigated. In the review conducted by Alqassab et al. (2023) it was found that peer assessment in research is not as diverse as it could theoretically be. With important similarities in the design of these activities, it may not be necessary to investigate the impact of each of the 21 instructional design elements. In fact, these authors found that – in the research published up to date – some instructional design elements with regard to peer assessment often occur together, and found five main design patterns of peer

assessment. For example, when peer assessment takes place face-to-face, it tends to be synchronous, formative, non-anonymous, bi-directional, and individual, with a matching of assessors and assessees based on self-selection, and with oral and written comments as an output (Alqassab et al., 2023). Rather than examining the impact of each element individually, it would be interesting to examine the impact of these patterns on the interpersonal processes of peer assessment. Moreover, these patterns are determined by 10 main instructional design elements (Alqassab et al., 2023), which could therefore be investigated as a priority. The present review found that the impact of only four of these elements (product, matching, contact, and privacy) on interpersonal processes has been investigated. The other elements are the purpose (formative or summative), the scope of involvement (e.g., giving and/or receiving feedback), the possibility of revision after the peer assessment, the unit of assessment assessor (is the peer assessment conducted individually or in a group?), the unit of assessment assessee (is the peer assessment conducted by an individual peer or by a group?), and the group constellation (is the peer assessment conducted between or within groups?). These elements are likely to influence interpersonal processes. For example, the question of perceived fairness should be more prominent in the case of summative assessment than in that of formative assessment.

The present review is innovative in terms of considering the interconnections between the “agents” and “context” elements of the MISCA model (Panadero & Lipnevich, 2022) when examining the impact of instructional design elements (context) on interpersonal processes (agents). However, instructional design elements are only one aspect of the “context” element in the MISCA model, other important aspects being the instructional context (e.g., does the peer assessment activity take place at the beginning, middle, or end of the course?), the classroom climate, and the general context (e.g., educational level). Investigating the influence of classroom climate might be particularly relevant as it seems to have an important impact on students’ perceptions of peer assessment (Rotsaert, 2017). Another perspective is to investigate the three other elements of the MISCA model (Panadero & Lipnevich, 2022). As student characteristics are the central element of the model, it might be relevant to focus first on this element and explore how these characteristics interact with interpersonal processes. Student characteristics include the intra-personal elements of the conceptualisation developed by Panadero et al. (2022), which were not included in the review presented in this paper. As these elements (e.g., self-efficacy) also play an important role in peer assessment (Panadero et al., 2023), it would be useful to conduct a literature review to investigate how they are affected by the instructional design elements of peer assessment.

Another important student characteristic to investigate is culture. The studies included in this review suggest that student culture is important as some interpersonal processes may be more relevant in an Asian context than in a European one (Kim & Lan, 2021; Xue et al., 2023; Zhang et al., 2022). These studies were conducted in homogeneous cultural contexts, but their findings raise questions about peer assessment in intercultural contexts such as which interpersonal factors are salient and how peer assessment activities should be designed in these contexts. Developing knowledge on this topic is important because the higher education population is increasingly multicultural. Six per cent of all higher education students in OECD countries are international students, with a much higher proportion in some countries such as 22 per cent in

Australia (OECD, 2023). This multicultural context offers opportunities to prepare students to work effectively in an increasingly intercultural workplace, but it is not without challenges (Popov et al., 2012). Research is lagging behind in providing educators with ideas on how to design inclusive and culturally sensitive peer assessment activities.

4.2 Limitations and directions for further research

The first limitation we encountered during the process was linked to the subject of the review itself. As already noted by Panadero et al. (2023), the different interpersonal processes of peer assessment are ill-defined. In the included studies, the processes investigated are rarely defined, and it was hard to tell if they covered the same reality from one study to another. On the one hand, some processes were given the same name but measured with regard to entirely different items (e.g., psychological safety in Ching & Hsu's 2016 study and Rotsaert et al.'s 2018 study). On the other hand, some processes were given different names but appeared to measure close processes by using very similar items (e.g., perception toward the assessor in F. Yu's 2012 study, and trust in peers' evaluative abilities in Rotsaert et al.'s 2018 study). This lack of clear definitions also influences the way these processes are measured. Most studies relied on non-validated scales which affects the results obtained and hampers comparisons between studies. In their literature review on the intra and interpersonal factors of peer assessment, Panadero et al. (2023) proposed definitions of the main processes, which can serve as a basis for developing research on the subject. We argue that a second important step to advance the research is to create and validate scales for each interpersonal process. In the meantime, we recommend researchers provide all the items used to measure interpersonal processes. Indeed, while conducting the review, we noticed that the scales used were not always available, which prevented us from verifying what researchers meant by the different interpersonal processes. Therefore, in this review, we found ourselves constrained to using the terms used by the authors of the studies.

The second limitation is the lack of variety in the contexts in which the studies were conducted. The vast majority of studies were conducted in higher education, most of them in the domain of education or EFL. This focus on higher education was also noted in other reviews on peer assessment (e.g., Alqassab et al., 2023; H. Li et al., 2020). It is problematic because nothing guarantees that results found with higher education students apply to students in primary and secondary education. For example, the only study conducted in primary education (F. Yu & Wu, 2011) showed that privacy led students to write impulsive comments evidencing irrational emotions (e.g., "POOR! POOR! POOR!", F. Yu & Wu, 2011, p. 2175). We can imagine that privacy will not have the same consequences for more mature students and therefore that privacy will not have the same impact on interpersonal processes either. For the studies conducted in higher education, the vast majority were conducted within the context of "soft disciplines", which present some important differences from "hard disciplines" regarding teaching and learning (Neumann et al., 2002). Hence, more studies conducted in primary and secondary education are needed, alongside more studies in "hard disciplines" of higher education.

Moreover, the included studies were conducted in a limited number of countries. It may be because we restricted our search to papers published in English, which may have led us to miss out on relevant studies published in other languages. Search in other languages should be conducted to verify if studies have been conducted in other parts of the world (e.g., studies conducted in South America and published in Spanish). If the absence of studies is confirmed, it would be necessary to conduct studies in a larger variety of countries and other parts of the world. Indeed, it has been argued that culture plays a role in explaining the relationships between instructional design elements and interpersonal processes (e.g., Fithriani, 2018). For example, students coming from collectivist cultures, which are more common in Asian countries, tend to work toward maintaining group harmony and mutual face-saving, which can affect their willingness to criticize their peers (Fithriani, 2018).

The third limitation is linked to the quality of the included studies. While performing the quality appraisal, we identified important limitations with regard to the included studies, such as a lack of information on the reliability and validity of the instruments used for quantitative and mixed-method studies, or a lack of detail on how the data were triangulated in the case of qualitative studies. There is a dearth of studies using strong research designs, which hampers the research on the subject. For example, concerning the study of the impact of privacy, studies with similar procedures have drawn different conclusions. Rotsaert et al. (2018) have investigated the impact of a transitional approach to anonymity by asking the same students to first conduct an anonymous peer assessment, and then a non-anonymous peer assessment. The same procedure was used by Kim (2019) and Kim and Lan (2021), but this time to compare the impact of anonymous and non-anonymous peer assessment, without taking into consideration the fact that the experience of the first activity could influence the second one. This illustrates the importance of using control groups composed of different students (Creswell, 2012). The predominance of studies with weak research designs is a problem that has been pointed out in the literature on peer assessment for some time now, but it seems that despite calls to take action (e.g., Strijbos and Sluijsmans, 2010; Van Zundert et al., 2010), this observation still holds for recently published studies. In addition to a lack of experimental and quasi-experimental designs, a frequent problem is that researchers do not determine if their samples are large enough to detect the potential effects. A power analysis was conducted in only one study (Ocampo et al., 2023), which was one of the studies that included the most participants. An explanation for these methodological limitations could be that studies on peer assessment are almost always conducted in real educational contexts and that the emphasis is put on ecological validity. It is however possible to combine ecological and internal validity as was showcased by a number of the included studies (e.g., Panadero et al., 2013; F. Yu, 2012).

4.3 Implications for practice

Based on the present review, we can offer three recommendations for practitioners implementing peer assessment. Moreover, we provide specific guidelines to follow when designing peer assessment activities in Appendix G. The first recommendation is that a training session prior to peer assessment is preferable. Previous literature reviews have already shown that the most critical factor in ensuring peer assessment effectiveness is the provision of training (e.g., H. Li et al., 2020). Our review evidenced that training also has a positive impact on

interpersonal processes such as students' feeling of comfort (Ocampo et al., 2023; Parikh & Sheehan, 2016). Combined, these findings suggest that training should be considered an indispensable part of a peer assessment exercise.

The second recommendation is to carefully consider the use of anonymity. Indeed, evidence suggests that anonymity can help make students more comfortable, yet it could also have some detrimental effects such as conveying negative emotions (Kaya, 2021) or creating negative perceptions of the peer as an assessor (F. Yu & Wu, 2011). Linked to this, we recommend gauging students' perceptions before deciding how to design peer assessment activities. Indeed, it seems from the data that the different interpersonal processes are not all impacted in the same ways by the instructional design elements. Privacy, for example, could be a good idea for students who are very uncomfortable with the idea of conducting a peer assessment (Kim, 2019; Lim et al., 2021; Raes et al., 2015; Su, 2023), but not for students who fear that peer assessment will be unfair (Açkay et al. 2021; Lin, 2018). It is therefore important to understand students' primary concerns, before deciding on designing the peer assessment activity.

Finally, it is important to bear in mind that interpersonal processes are not the only factor to consider when designing peer assessment. For example, it would be problematic to conclude from the study of Praver et al. (2011) that we should ask students to provide only scores, and not written comments, to prevent negative emotions from arising. Indeed it has been established that written comments are more beneficial for learning than scores (Lipnevich & Smith, 2009; Shute, 2008), so a wiser conclusion would be to ask students to provide comments while manipulating other design elements to mitigate the possible increase in negative emotions.

4.4 Conclusion

The present study reviewed 27 empirical studies on the impact of the instructional design elements of peer assessment on interpersonal processes. Privacy is the most commonly-investigated design element. It has a positive impact on some interpersonal processes and a negative impact on others. The second most commonly-investigated design element was the format of peer assessment: it seems that increased amounts of guidance have a positive impact on interpersonal processes, but studies with stronger research designs are needed to confirm this. The presence of training, the third most commonly-investigated design element, seems to also have a positive impact on interpersonal processes, which confirms that training is an important step with regard to peer assessment. The review revealed that the impact of other design elements has not been sufficiently studied to conclude anything, and that further research is needed. It also revealed some limitations in the field, such as a lack of clear definitions and adapted instruments to measure interpersonal processes or a dearth of studies using strong research designs. These gaps and needs in the research should inform further research. Nevertheless, the present review still provides some implications for practice, such as the importance of training students prior to peer assessment, and the importance of gauging students' perceptions to decide how best to design peer assessment activities.

References

- *Akçay, A. O., Güven, U., & Karahan, E. (2021). The Views of Pre-Service Elementary Teachers About Online and Traditional Peer Assessment. *International Journal of Assessment Tools in Education*, 8(2), 409-422. <https://doi.org/10.21449/ijate.762104>
- Alqassab, M., Strijbos, J.W., Panadero, E., Fernández Ruiz, J., & Warrens, M. (2023). A Systematic Review of Peer Assessment Design Elements. *Educ Psychol Rev* 35, 18, 3-36. <https://doi.org/10.1007/s10648-023-09723-7>
- Black, N., Mullan, B., & Sharpe, L. (2016). Computer-delivered interventions for reducing alcohol consumption: Meta-analysis and meta-regression using behaviour change techniques and theory. *Health Psychology Review*, 10(3), 341-357. <https://doi.org/10.1080/17437199.2016.1168268>
- Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Systematic Reviews*, 6(1), 245-245. <https://doi.org/10.1186/s13643-017-0644-y>
- Booth, A., Sutton, A., & Clowes, M. (2016). *Systematic approaches to a successful literature review* (Second ed.). SAGE.
- Çevik, Y. D. (2015). Assessor or assessee? Investigating the differential effects of online peer assessment roles in the development of students' problem-solving skills. *Computers in Human Behavior*, 52, 250-258. <https://doi.org/10.1016/j.chb.2015.05.056>
- *Cheng, L., Li, Y., Su, Y., & Gao, L. (2023). Effect of regulation scripts for dialogic peer assessment on feedback quality, critical thinking and climate of trust. *Assessment and Evaluation in Higher Education*, 48(4), 451-463. <https://doi.org/10.1080/02602938.2022.2092068>
- 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), 599-618. <https://doi.org/10.14742/ajet.830>
- *Ching, Y., & Hsu, Y. (2016). Learners' interpersonal beliefs and generated feedback in an online role-playing peer- feedback activity: An exploratory study. *International Review of Research in Open and Distance Learning*, 17(2), 105-122. <https://doi.org/10.19173/irrodl.v17i2.2221>
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson education.
- 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>
- Efron, S. E., & Ravid, R. (2019). *Writing the literature review: A practical guide*. Guilford Publications.
- Ewald, H., Klerings, I., Wagner, G., Heise, T. L., Stratil, J. M., Lhachimi, S. K., Hemkens, L. G., Gartlehner, G., Armijo-Olivo, S., & Nussbaumer-Streit, B. (2022). Searching two or more

- databases decreased the risk of missing relevant studies: A meta-research study. *Journal of Clinical Epidemiology*, 149, 154-164. <https://doi.org/10.1016/j.jclinepi.2022.05.022>
- Falchikov, N. (1995). Peer Feedback Marking: Developing Peer Assessment. *Innovations in Education and Training International*, 32, 175-187. <https://doi.org/10.1080/1355800950320212>
- Fithriani, R. (2018). Cultural influences on students' perceptions of written feedback in L2 writing. *Journal of Foreign Language Teaching & Learning*, 3(1), 1-13. <https://doi.org/10.18196/ftl.3124>
- Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment and Evaluation in Higher Education*, 36(2), 137-155. <https://doi.org/10.1080/02602930903221444>
- *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>
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters* 4). Gaithersburg, MD: Advanced Analytics, LLC.
- Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research and Development*, 20(1), 53-70. <https://doi.org/10.1080/07294360123776>
- 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>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50-62. <https://doi.org/10.1016/j.edurev.2015.12.002>
- *Kaya, F. (2021). Emotions related to identifiable/anonymous peer feedback: A case study with turkish pre-service english teachers. *Issues in Educational Research*, 31(4), 1088-1100.
- *Kim, S. (2019). Japanese student writers' perspectives on anonymous peer review, *ELT Journal*, 73(3), 296-305. <https://doi.org/10.1093/elt/ccy061>
- *Kim, S., & Lan, Y. (2021). L2 writers' perspectives on face-to-face and anonymous peer review: Voices from China. *Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras*, 35, 149-164. <https://doi.org/10.30827/portalin.vi35.13887>
- Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment and Evaluation in Higher Education*, 45(2), 193-211. <https://doi.org/10.1080/02602938.2019.1620679>
- *Li, L. (2017). The role of anonymity in peer assessment. *Assessment and Evaluation in Higher Education*, 42(4), 645-656. <https://doi.org/10.1080/02602938.2016.1174766>
- *Lim, W., Son, J., & Kang, S. (2021). How reducing discomfort impacts peer assessments of preservice teachers. *Sustainability (Basel, Switzerland)*, 13(11), 6435. <https://doi.org/10.3390/su13116435>

- *Lin, G. (2018). Anonymous versus identified peer assessment via a facebook-based learning application: Effects on quality of peer feedback, perceived learning, perceived fairness, and attitude toward the system. *Computers and Education*, 116, 81-92. <https://doi.org/10.1016/j.compedu.2017.08.010>
- Lindvall, J., & Ryve, A. (2019). Coherence and the positioning of teachers in professional development programs. A systematic review. *Educational Research Review*, 27, 140-154. <https://doi.org/10.1016/j.edurev.2019.03.005>
- Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4), 319-333. <https://doi.org/10.1037/a0017841>
- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18(1), 30-43. <https://doi.org/10.1016/j.jslw.2008.06.002>
- Neumann, R., Parry, S., & Becher, T. (2002). Teaching and learning in their disciplinary contexts: A conceptual analysis. *Studies in Higher Education*, 27(4), 405-417. <https://doi.org/10.1080/0307507022000011525>
- Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- *Ocampo, J. C. G., Panadero, E., & Díez, F. (2023). Are men and women really different? the effects of gender and training on peer scoring and perceptions of peer assessment. *Assessment and Evaluation in Higher Education*, 48(6), 760-776. <https://doi.org/10.1080/02602938.2022.2130167>
- OECD (2023), *Education at a Glance 2023: OECD Indicators*, OECD Publishing, Paris, <https://doi.org/10.1787/e13bef63-en>.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Journal of Clinical Epidemiology*, 10(1), 178-89. <https://doi.org/10.1186/s13643-021-01626-4>
- Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment. In G. Brown & L. Harris (Eds), *Handbook of human and Social Conditions in Assessment* (pp. 247-265). Routledge.
- Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment and Evaluation in Higher Education*, 44(8), 1253-1278. <https://doi.org/10.1080/02602938.2019.1600186>
- Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assessment and Evaluation in Higher Education*, ahead-of-print(ahead-of-print), 1-23. <https://doi.org/10.1080/02602938.2023.2164884>

- Panadero, E., Jonsson, A., & Alqassab, M. (2018). Providing formative peer feedback: What do we know? In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 409–431). Cambridge University Press. <https://doi.org/10.1017/9781316832134.020>
- *Panadero, E., Romero, M., & Strijbos, J. (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>
- *Parikh, B. T., & Sheehan, A. H. (2016). Student pharmacists' perceptions of the implementation of peer assessment instruction and activities within a didactic course. *Currents in Pharmacy Teaching and Learning*, 8(6), 821-826. <https://doi.org/10.1016/j.cptl.2016.08.012>
- Popov, V., Brinkman, D., Biemans, H. J. A., Mulder, M., Kuznetsov, A., & Noroozi, O. (2012). Multicultural student group work in higher education. an explorative case study on challenges as perceived by students. *International Journal of Intercultural Relations*, 36(2), 302-317. <https://doi.org/10.1016/j.ijintrel.2011.09.004>
- *Praver, M., Rouault, G., & Eidswick, J. (2011). Attitudes and affect toward peer evaluation in EFL reading circles. *The Reading Matrix*, 11(2), 89-101.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Wiley. <https://doi.org/10.1002/9780470754887>
- *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>
- *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. *Eur J Psychol Educ* 33, 75–99. <https://doi.org/10.1007/s10212-017-0339-8>
- * Senden, M., De Jaeger, D., & Coertjens, L. (2023). Safe and sound: Examining the effect of a training targeting psychological safety and trust in peer assessment. *Frontiers in Education*, 8, 1-13. <https://doi.org/10.3389/feduc.2023.1198011>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189. <https://doi.org/10.3102/0034654307313795>
- Stanier, L. (1997). Peer assessment and group work as vehicles for student empowerment: A module evaluation. *Journal of Geography in Higher Education*. <https://doi.org/10.1080/03098269708725413>
- 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 webbased education: Integrating human factors and personalization* (pp. 375–395). IGI Global.
- Strijbos, J., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, 20(4), 265-269. <https://doi.org/10.1016/j.learninstruc.2009.08.002>

- *Su, W. (2023). Masked ball for all: How anonymity affects students' perceived comfort levels in peer feedback. *Assessment and Evaluation in Higher Education*, 48(4), 502-512. <https://doi.org/10.1080/02602938.2022.2089348>
- *Thanh, P. T. H., & Gillies, R. (2010). Designing a Culturally Appropriate Format of Formative Peer Assessment for Asian Students: The Case of Vietnamese Students. *International Journal of Educational Reform*, 19(2), 72-85. <https://doi.org/10.1177/105678791001900201>
- 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. (2017). Peer assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology*, 1(1), 7. <https://doi.org/10.31532/InterdiscipEducPsychol.1.1.007>
- *Tzeng, A., Bruno, B., Cooperrider, J., Dinardo, P. B., Baird, R., Swetlik, C., Goldstein, B. N., Rastogi, R., Roth, A. J., Gilligan, T. D., & Rish, J. M. (2021). A structured peer assessment method with regular reinforcement promotes longitudinal self-perceived development of medical students' feedback skills. *Medical Science Educator*, 31(2), 655-663. <https://doi.org/10.1007/s40670-021-01242-w>
- 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 and Evaluation in Higher Education*, 46(8), 1191-1201. <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>
- * Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers and Education*, 81, 123-132. <https://doi.org/10.1016/j.compedu.2014.10.001>
- Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: A literature review. *Journal of Advanced Nursing*, 50(2), 204-211. <https://doi.org/10.1111/j.1365-2648.2005.03380.x>
- Wichmann, A., Funk, A., & Rummel, N. (2018). Leveraging the potential of peer feedback in an academic writing activity through sense-making support. *European Journal of Psychology of Education*, 33(1), 165. <https://doi.org/10.1007/s10212-017-0348-7>
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17-37. <https://doi.org/10.1080/00461520.2016.1207538>
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087-3087. <https://doi.org/10.3389/fpsyg.2019.03087>

- * Xue, S., Yang, Y., Du, J., & Liu, F. (2023). Multi-layered e-feedback anxiety: An action research study among chinese learners using peer feedback activities in an academic writing course. *Frontiers in Psychology*, 14, 1062517-1062517. <https://doi.org/10.3389/fpsyg.2023.1062517>
- *Yu, F. Y. (2012). Any effects of different levels of online user identity revelation? *Educational Technology & Society*, 15(1), 64-77.
- *Yu, F., & Liu, Y. (2009). Creating a psychologically safe online space for a student-generated questions learning activity via different identity revelation modes. *British Journal of Educational Technology*, 40(6), 1109-1123. <https://doi.org/10.1111/j.1467-8535.2008.00905.x>
- *Yu, F., & Wu, C. (2011). Different identity revelation modes in an online peer-assessment learning environment: Effects on perceptions toward assessors, classroom climate and learning activities. *Computers and Education*, 57(3), 2167-2177. <https://doi.org/10.1016/j.compedu.2011.05.012>
- Yu, S., Lee, I., & Mak, P. (2016). Revisiting chinese cultural issues in peer feedback in EFL writing: Insights from a multiple case study. *The Asia-Pacific Education Researcher*, 25(2), 295-304. <https://doi.org/10.1007/s40299-015-0262-1>
- *Zhang, M., He, Q., Du, J., Liu, F., & Huang, B. (2022). Learners' perceived advantages and social-affective dispositions toward online peer feedback in academic writing. *Frontiers in Psychology*, 13, 973478-973478. <https://doi.org/10.3389/fpsyg.2022.973478>
- Zou, Y., Schunn, C. D., Wang, Y., & Zhang, F. (2018). Student attitudes that predict participation in peer assessment. *Assessment and Evaluation in Higher Education*, 43(5), 800-811. <https://doi.org/10.1080/02602938.2017.1409872>

Appendix

Appendix A. Search string for ERIC

((MAINSUBJECT.EXACT("Friendship") OR MAINSUBJECT.EXACT("Interpersonal Relationship") OR MAINSUBJECT.EXACT("Trust (Psychology)") OR MAINSUBJECT.EXACT("Motivation") OR MAINSUBJECT.EXACT("Peer Relationship")) OR (motivation* OR emotion* OR interpersonal* OR "psychological safety" OR friend* OR "value diversit*" OR "value congruen*" OR interdependen* OR trust* OR anonymity OR relation* OR social OR "fear of disapproval" OR anxiety OR sham*)) AND ((MAINSUBJECT.EXACT("Higher Education") OR MAINSUBJECT.EXACT("College Students") OR MAINSUBJECT.EXACT("Elementary Schools") OR MAINSUBJECT.EXACT("Elementary Education") OR MAINSUBJECT.EXACT("Junior High Schools") OR MAINSUBJECT.EXACT("Secondary Education") OR MAINSUBJECT.EXACT("Elementary School Students") OR MAINSUBJECT.EXACT("High Schools") OR MAINSUBJECT.EXACT("Middle Schools") OR MAINSUBJECT.EXACT("Junior High School Students") OR MAINSUBJECT.EXACT("Middle School Students") OR MAINSUBJECT.EXACT("Compulsory Education")) OR ti(student* OR school* OR "elementary education" OR "secondary education" OR "elementary school*" OR "middle school*" OR learn OR learner* OR learning OR "higher education" OR college* OR "high school*" OR "highschool*" OR "grade 1" OR "grade 2" OR "grade 3" OR "grade 4" OR "grade 5" OR "grade 6" OR "grade 7" OR "grade 8" OR "grade 9" OR "grade 10" OR "grade 11" OR "grade 12" OR "primary education" OR "public school*" OR "secondary school*" OR "junior high" OR "K-12*" OR k12 OR preuniversit* OR "pre-universit*" OR universit*) OR ab(student* OR school* OR "elementary education" OR "secondary education" OR "elementary school*" OR "middle school*" OR learn OR learner* OR learning OR "higher education" OR college* OR "high school*" OR "highschool*" OR "grade 1" OR "grade 2" OR "grade 3" OR "grade 4" OR "grade 5" OR "grade 6" OR "grade 7" OR "grade 8" OR "grade 9" OR "grade 10" OR "grade 11" OR "grade 12" OR "primary education" OR "public school*" OR "secondary school*" OR "junior high" OR "K-12*" OR k12 OR preuniversit* OR "pre-universit*" OR universit*)) AND (MAINSUBJECT.EXACT("Peer Evaluation") OR ti("peer evaluation*" OR "peer assessment*" OR "peer feedback" OR "peer grading" OR "peer marking" OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR "peer-feedback" OR "peer-grading" OR "peer-marking" OR "peer-review*")) OR ab("peer evaluation*" OR "peer assessment*" OR "peer feedback" OR "peer grading" OR "peer marking" OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR "peer-feedback" OR "peer-grading" OR "peer-marking" OR "peer-review*"))

Appendix B. Search string for PsycInfo

((MAINSUBJECT.EXACT("Elementary Education") OR
MAINSUBJECT.EXACT("Elementary Schools") OR MAINSUBJECT.EXACT("High
Schools") OR MAINSUBJECT.EXACT("College Students") OR
MAINSUBJECT.EXACT("Higher Education") OR MAINSUBJECT.EXACT("High School
Education") OR MAINSUBJECT.EXACT("Middle Schools") OR
MAINSUBJECT.EXACT("Secondary Education") OR MAINSUBJECT.EXACT("Middle
School Education") OR MAINSUBJECT.EXACT("Junior High School Students") OR
MAINSUBJECT.EXACT("Middle School Students") OR
MAINSUBJECT.EXACT("Elementary School Students")) OR ab(student* OR school* OR
"elementary education" OR "secondary education" OR "elementary school*" OR "middle
school*" OR learn OR learner* OR learning OR "higher education" OR "college*" OR "high
school*" OR "highschool" OR "grade 1" OR "grade 2" OR "grade 3" OR "grade 4" OR
"grade 5" OR "grade 6" OR "grade 7" OR "grade 8" OR "grade 9" OR "grade 10" OR "grade
11" OR "grade 12" OR "primary education" OR "public school*" OR "secondary school*" OR
"junior high" OR "K-12*" OR k12 OR preuniversit* OR "pre-universit*" OR universit*)
OR ti(student* OR school* OR "elementary education" OR "secondary education" OR
"elementary school*" OR "middle school*" OR learn OR learner* OR learning OR "higher
education" OR "college*" OR "high school*" OR "highschool" OR "grade 1" OR "grade 2"
OR "grade 3" OR "grade 4" OR "grade 5" OR "grade 6" OR "grade 7" OR "grade 8" OR
"grade 9" OR "grade 10" OR "grade 11" OR "grade 12" OR "primary education" OR "public
school*" OR "secondary school*" OR "junior high" OR "K-12*" OR k12 OR preuniversit*
OR "pre-universit*" OR universit*)) AND (MAINSUBJECT.EXACT("Peer Evaluation") OR
ab("peer evaluation*" OR "peer assessment*" OR "peer feedback" OR "peer grading" OR
"peer marking" OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR "peer-
feedback" OR "peer-grading" OR "peer-marking" OR "peer-review*") OR ti("peer
evaluation*" OR "peer assessment*" OR "peer feedback" OR "peer grading" OR "peer
marking" OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR "peer-
feedback" OR "peer-grading" OR "peer-marking" OR "peer-review*")) AND
((MAINSUBJECT.EXACT("Trust (Social Behavior)") OR MAINSUBJECT.EXACT("Peer
Relations") OR MAINSUBJECT.EXACT("Motivation") OR
MAINSUBJECT.EXACT("Interpersonal Relationships") OR
MAINSUBJECT.EXACT("Friendship") OR MAINSUBJECT.EXACT("Anonymity")) OR
(motivation* OR emotion* OR interpersonal* OR "psychological safety" OR friend* OR
"value diversit*" OR "value congruen*" OR interdependen* OR trust* OR anonymity OR
relation* OR social OR "fear of disapproval" OR anxiety OR sham*))

Appendix C. Search string for Scopus

(TITLE-ABS-KEY ("peer evaluation*" OR "peer assessment*" OR {peer feedback} OR {peer grading} OR {peer marking} OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR {peer-feedback} OR {peer-grading} OR {peer-marking} OR "peer-review*")) AND (TITLE-ABS-KEY (student* OR school* OR {elementary education} OR {secondary education} OR "elementary school*" OR "middle school*" OR learn OR learner* OR learning OR {higher education} OR college* OR "high school*" OR highschool* OR {grade 1} OR {grade 2} OR {grade 3} OR {grade 4}) OR TITLE-ABS-KEY ({grade 5} OR {grade 6} OR {grade 7} OR {grade 8} OR {grade 9} OR {grade 10} OR {grade 11} OR {grade 12} OR {primary education} OR "public school*" OR "secondary school*" OR "junior high*" OR "K-12*" OR k12 OR preuniversit* OR "pre universit*" OR universit*)) AND (ALL (motivation* OR emotion* OR interpersonal OR {psychological safety} OR friend* OR "value diversit*" OR "value congruen*" OR interdependen* OR trust* OR anonymity OR relation* OR social OR {fear of disapproval} OR anxiety OR sham*))

Appendix D. Critical appraisal for quantitative studies

		Li (2017)	Lin (2018)	Ocampo et al. (2023)	Panadero et al. (2013)	Parikh & Sheehan	Prayer et al. (2011)	Senden et al. (2023)	Yu (2012)	Yu & Wu (2011)
General	Description should be specific and detailed	2	2	2	2	2	2	2	2	2
	Whether the study was experimental or nonexperimental is noted	2	2	2	2	2	2	2	2	2
Participants	There is a clear and detailed description of the participants and their characteristics	2	2	2	2	1	1	2	2	1
	There is an explanation of how the sample was selected	2	2	2	2	2	2	2	2	2
Data collection tools	Data collection tools are clearly described	2	2	0	0	2	0	2	2	2
	Information is provided as to whether existing, modified, or newly developed data collection tools are used	2	2	2	2	1	2	2	2	2
	There is information about the reliability and validity of the data collection tools	2	0	0	1	0	0	1	2	1
	When appropriate, sample items or questions are included	2	2	2	2	2	2	2	0	2
Procedures	The procedure for conducting the study are clearly and adequately described	2	2	2	2	2	2	2	2	2
	There is a description of the process, duration, and schedule of the study	2	2	2	2	2	2	2	2	2
	The data collection process is described	2	2	2	2	2	2	2	2	2
Data analysis	There is an explanation of the study's design and data analysis	2	2	2	2	2	2	2	2	2
	Statistical procedures used to test the study's hypotheses are listed when appropriate	2	2	2	2	2	2	2	2	2
Ethics	Ethical guidelines are followed	/	/	/	/	/	/	2	/	/
Results	The main findings are reported	2	2	2	2	2	2	2	2	2
	The results are clearly presented and easy to follow	2	2	2	2	2	2	2	2	2
	There are tables, graphs, or charts that help explain the findings	2	2	2	0	0	2	2	2	2
Discussion	The results are examined, explained and evaluated	2	2	2	2	2	2	2	2	2
	There is a discussion of the findings in relation to the original questions and/or hypotheses	2	2	2	2	2	2	2	2	2
	Findings are discussed in relation to prior research on the topic	1	2	2	2	2	0	2	2	2
	The study's conclusions are justified based on the results of the study	2	2	2	2	2	1	2	2	2
	The authors provided possible explanations for unexpected results	2	2	2	1	/	1	2	2	1
	Weaknesses and limitations of the study are discussed	2	2	2	2	2	2	2	1	1
	Implications of the findings for practitioners are presented	1	0	0	2	2	1	0	1	2
	There are suggestions for further research	2	2	2	2	1	1	2	1	2
Validity and reliability	There is evidence of the study's internal validity, especially in experimental studies	2	2	2	2	1	1	2	2	2
	There is evidence of external validity	2	2	2	2	2	2	2	2	2
	The validity of the data collection instrument(s) is presented	2	0	0	1	0	0	0	1	0
	The reliability of the data collection instrument(s) is presented	2	0	0	1	0	0	2	2	2
	The level of reliability of the measure is appropriate for its use in the study	2	/	/	1	/	/	1	2	2
	The reliability levels of the data collection instrument(s) are reported when validity is discussed	2	0	0	1	0	0	2	2	2

0 = not respected, 1 = partly respected, 2 = respected, / = no information

Appendix E. Critical appraisal for qualitative studies

		Akay et al. (2021)	Kaya et al. (2023)	Xue et al. (2023)	Zhang et al. (2022)
General	Description should be specific and detailed	1	1	1	1
Researcher's role	Researcher's subjectivity and preconceived notions about the topic are discussed	0	0	0	0
Setting	Detailed information about the setting and rationale for selecting it is provided	1	1	1	1
Participants	Background information about the participants is offered	1	1	1	1
	Participant selection is based on their knowledge of the topic and their ability and willingness to share it	2	2	2	2
Data collection tools	The process of data collection is clearly and comprehensively described	1	2	2	2
	Methods used to generate data are clearly described	1	1	2	2
	Multiple sources of data collection were used to enable triangulation	2	1	2	2
	Data collection strategies are appropriate for the research question(s)	2	2	2	2
Procedures	The procedure for conducting the study are clearly and adequately described	1	1	1	1
	A record of the process, including interviews, observations, protocols, field notes, and transcripts, is available in the appendix	0	0	0	0
Data analysis	The method of data analysis is logically developed and clearly presented	2	1	2	2
	The process for coding the data and for identifying categories, themes, and pattern is outlined	1	1	1	1
	The relationship among the pattern are supported by the data	2	2	2	2
Ethics	The identity of the participants and the settings is kept confidential	2	2	2	2
	Formal ethical permissions to conduct the study were obtained	/	/	2	2
	The participants were informed about the nature and research and their confidentiality was assured	/	/	2	2
Results	There is a clear explanation of how the findings are drawn from the data analysis	2	1	2	2
	The results are reported in a "thick" narrative	2	1	2	2
	The findings are supported by illustrative quotations and specific instances	2	1	2	2
	The findings clearly and adequately address the research question(s)	2	2	2	2
	Contradictory findings and outlier data are included in the research findings	2	1	1	1
Discussion	The meaning and implications of the study are discussed in relation to the research questions	2	2	2	2
	The discussion relates the study's findings to existing literature on the topic	2	2	2	2
	The findings are discussed within the broader social, cultural, and/or political contexts	0	0	1	1
	The researcher reflects on the experiences of conducting the study	0	0	0	0
Validity and trustworthiness	The credibility, transferability, dependability, and conformability of the study are discussed	0	0	0	0
	The research process and data collection phases entailed a prolonged engagement	/	/	/	/
	The researcher discussed personal biases and preconceived assumptions	0	0	0	0
	The research process, procedures, and documentation are recorded	2	/	2	2
	There is a rich "tick" description of the participants, the setting(s), and the events	1	1	1	1
	There is an explanation of how the data results were triangulated	1	0	1	1
	Processes for alternative possible interpretation and discrepant information are presented	0	0	0	0
	Procedures for checking the research accuracy from the perspective of participants and peer researchers are used	0	0	0	0

Appendix F. Critical appraisal for mixed method studies

		Cheng et al. (2023)	Ching & Hsu (2016)	Güler (2017)	Kim (2019)	Kim & Lan (2021)	Lim et al. (2021)	Raes et al. (2015)	Rotsaert et al. (2018)	Su (2023)	Thanh & Gillies (2010)	Tzeng et al. (2022)	van Heerden & Bhaunthram (2021)	Vanderhoven et al. (2015)	Yu & Liu (2009)
General	The reasons for choosing a mixed-methods approach are clearly explained	0	0	1	0	0	1	1	0	1	0	0	0	1	0
	The methods used match the research questions, and each method is designed to answer at least one research question	2	2	2	1	1	2	2	2	2	1	2	1	2	2
Researcher's role	The researcher role – an objective or subjective stance – is assumed to match the research question being investigated	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Site and participants	The site and participants are fully described and the reasons for selecting them are presented	1	1	1	1	1	1	1	1	1	0	1	0	0	0
Data collection tools	Data collection tools, both quantitative and qualitative, are described in detail	1	2	2	1	2	2	2	2	2	2	2	1	1	2
	Sample items from data collection tools are included	2	2	0	1	2	2	2	2	2	2	2	0	2	2
Procedures	The steps used to collect both quantitative and qualitative data are explicitly described	2	2	2	2	2	2	2	2	2	2	1	1	2	2
	The procedures for data analysis reflect the type of mixed data designs (embedded, two-phased or triangulation)	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Data analysis	It is indicated whether the analysis occurred sequentially or concurrently	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	It is indicated when the integration of quantitative and qualitative data occurred	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	It is clear whether the quantitative or qualitative approach is given priority	1	1	1	1	1	1	1	1	1	1	1	0	1	1
Ethics	The ethical considerations listed for quantitative and qualitative studies are followed	/	/	/	2	/	2	/	/	/	/	/	/	/	/
Results	Findings are presented in different formats based on the method and data source used to generate the data	2	1	1	1	1	1	2	2	1	1	2	2	2	/
	An explanation is offered of how findings from the quantitative and qualitative methods are integrated	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Discussion	The major findings are reiterated in relation to the study's questions	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	Conflicting findings are examined and properly explained	2	/	/	1	2	/	/	2	/	/	2	1	1	/
	The study's findings are evaluated in relation to prior research in the field	2	2	2	2	2	2	2	2	2	0	2	2	2	2

Validity, reliability, and trustworthiness	The procedure for ensuring the validity, reliability, and trustworthiness of mixed-methods research follow the specific requirements of each approach	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
--	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Appendix G. Guidelines for organising peer assessment activities

Design element	Guidelines	Description and references
Privacy	Organise peer assessment anonymously if students have significant feelings of discomfort and unease or significant fears of disapproval or peer pressure. Organise peer assessment non-anonymously if students perceive peer assessment as unfair or have negative perceptions of the assessors. Organise peer assessment in such a way that anonymity does not prevent interaction between students.	There is consistent evidence from several studies that privacy increases feelings of comfort (Kim, 2019; Lim et al., 2021; Raes et al., 2015; Su, 2023), decreases fear of disapproval (Rotsaert et al., 2018; Vanderhoven et al., 2015), and reduces tension that students may feel (Kaya, 2021; F. Yu & Liu, 2009). It also appears to reduce peer pressure (Açkay et al., 2021; Vanderhoven et al., 2015), although some studies did not find any significant effect (L. Li, 2017; Raes et al., 2015). Negative effects of anonymity have also been found. Evidence suggests that privacy can convey negative emotions (Kaya, 2021) and has a detrimental effect on perceived fairness (Açkay et al., 2021; Lin, 2018), perceptions towards assessors (F. Yu & Wu, 2011), and value congruency (Rotsaert et al., 2018). However, some studies did not find that privacy had a significant effect on these processes (Güler, 2017; Vanderhoven et al., 2015; F. Yu, 2012). Some studies have aimed to examine the effect of privacy, but this has been confounded with the effect of face-to-face contact. The detrimental effects on trust (Kim, 2019) and face-saving (Kim & Lan, 2021) found in these studies appear to be due to the absence of contact rather than anonymity per se.
Format of peer assessment	Offer guidance to students. The impact of social media (rather than learning management systems) is unclear. More research is needed.	The provision of either a specific method for guiding students (Tzeng et al., 2021), a detailed regulation script for dialogic peer feedback (Cheng et al., 2023), or the opportunity to adopt stakeholder perspective when providing feedback (Ching & Hsu, 2016) was associated with high levels of comfort (Ching & Hsu, 2016; Tzeng et al., 2021), trust (Cheng et al., 2023; Ching & Hsu, 2016), and psychological safety (Ching & Hsu, 2016). Two studies compared the effects of three peer assessment tools offering different types of guidance. These studies found that the use of a social media platform (WeChat) was associated with less worry about “saving face” and fewer negative emotions than learning management systems, but also created emotional distress (Xue et al., 2023; Zhang et al., 2022).

	The effect of using a rubric on interpersonal processes is unclear. Further research is needed.	A study on the effect of using a rubric (Panadero et al., 2013) found no positive effect on students' feelings of comfort or their perceptions of fairness and a negative effect on friendship marking. However, this negative effect should be put into perspective: it did not imply a less valid peer assessment of friends in the rubric condition, as students using a rubric only overrated their friends, whereas students without a rubric overrated all their peers.
Training	Train students on how to provide feedback prior to peer assessment. If students have specific concerns, it may be beneficial to also provide a specific training on these aspects as well.	Two studies have found that general training was associated with positive effects: high levels of trust (Ocampo et al., 2023), and high levels of comfort (Ocampo et al., 2023; Parikh & Sheehan, 2016), while one study has found that a training session specifically designed to reduce students' feelings of peer pressure achieved its goal (L. Li, 2017). A last study (Senden et al., 2023) did not find any significant effect of specific training on students' levels of trust and psychological safety, but the lack of effect could be explained by methodological limitations.
Matching of assessors and assessees	The effect of different strategies for matching assessors and assesses is unclear. Further research is needed.	One study found that students reported the experience of peer assessment as more comfortable, enjoyable, and less challenging when they had to provide feedback to someone they knew well than to someone they did not know well, but it also had a negative effect on perceived fairness. However, this study is the only one on the subject and has some important methodological limitations.
Product of peer assessment	The effect of the presence of comments on interpersonal processes is unclear. Further research is needed.	One study found that having to provide comments rather than just grades was seen as more useful, but also made students more nervous and embarrassed (Praver et al., 2011). However, this study is the only one on the subject and has some important methodological limitations.
Unit of assessment (assessors)	The effect of collective peer assessment (unit of assessors = group) is unclear. Further research is needed.	One study found that conducting peer assessment collectively as a group was associated with lower feelings of discomfort and worry than conducting it individually (Thanh & Gillies, 2010). However, this study is the only one on the subject and has some important methodological limitations.