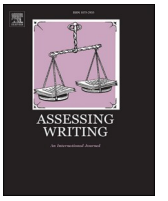
Assessing Writing 66 (2025) 100987

Contents lists available at ScienceDirect

Assessing Writing

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



Which gender provides more specific peer feedback? Gender and assessment training's effects on peer feedback specificity and intrapersonal factors

José Carlos G. Ocampo [a,b,c,*,1], Ernesto Panadero [a,d,e], David Zamorano [a], Iván Sánchez-Iglesias [f]

[a] Faculty of Education and Sports, University of Deusto, Bilbao, Spain
[b] Institute of Educational Science, University of Tübingen, Germany
[c] Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Germany
[d] Centre for Assessment Research Policy and Practice, School of Policy and Practice, Dublin City University, Ireland
[e] IKERBASQUE Basque Foundation for Science, Bilbao, Spain
[f] Psicobiología y Metodología de las Ciencias del Comportamiento, Universidad Complutense de Madrid, Madrid, Spain

ARTICLE INFO

Keywords:
Gender
Assessment training
Peer feedback content
Intrapersonal factors
Randomised-controlled design

ABSTRACT

This study investigated the effects of assessor gender (male vs. female), fictitious assessee gender (male vs. female), and assessment training (with vs. without) on peer feedback specificity (i.e. localisation and focus) and intrapersonal factors (i.e. trust in the self as an assessor and discomfort). This study involved 240 undergraduate psychology students (n_{Men}=120, n_{Women}=120), with half receiving assessment training and the other half receiving the task instructions. Participants were divided into eight subgroups based on training condition and their self-reported gender to provide peer feedback to three writing samples (poor, average, excellent quality) by fictitious male or female peer assessees in *Eduflow*. A total of 3017 peer feedback segments were analysed, revealing that trained or untrained male and female assessors were comparable in most peer feedback specificity categories when assessing fictitious male or female assessees. Nonetheless, we also found that female assessors excelled in certain categories of peer feedback specificity, while male assessors also demonstrated competencies in other categories. Results also showed that assessors who received assessment training provided localised peer feedback in all the writing samples. Finally, gender and training did not affect participants' trust in their abilities and (dis)comfort when providing peer feedback.

Studies have noted that peer feedback can positively impact cognitive, affective, and social outcomes for students (Kerman et al., 2023; Panadero et al., 2023). Nonetheess, evidence suggests that student characteristics can hinder the peer feedback process. For example, studies have found that male and female assessors generate feedback of varying characteristics, which may affect how assessees implement and benefit from peer feedback (e.g., Ocampo et al., 2023, 2025). Moreover, students' internal reactions towards

* Corresponding author at: Institute of Educational Science, University of Tübingen, Germany.
E-mail addresses: carlos.ocampo@uni-tuebingen.de (J.C.G. Ocampo), ernesto.panadero@dcu.ie, ernesto.panadero@deusto.es (E. Panadero), david.zamorano@deusto.es (D. Zamorano), i.sanchez@psi.ucm.es (I. Sánchez-Iglesias).

[1] The project was conceptualised at the University of Deusto as part of the first author's PhD dissertation. The writing and revision of the manuscript were carried out at the University of Tübingen.

https://doi.org/10.1016/j.asw.2025.100987
Received 24 April 2025; Received in revised form 23 September 2025; Accepted 28 September 2025
Available online 7 October 2025

peer feedback can undermine its effectiveness. Some students do not trust their peers' ability to provide quality feedback (e.g., To & Panadero, 2019), further emphasising the impact of individual characteristics on the peer feedback process.

The importance of investigating student characteristics in the feedback process has been highlighted in the seminal work of Lipnevich and Panadero (2021) and Panadero and Lipnevich, (2022), where they forwarded that students' characteristics (e.g. gender, emotions, abilities) are central to the feedback process because these individual characteristics may influence and moderate the other elements in the feedback process. Given this, exploring how students may be supported to provide sound assessments in written works is crucial. Specifically, it is important to investigate how issues related to gender and intrapersonal factors may be lessened by assessment training to enable students to develop peer feedback with better characteristics. Thus, this study aims to investigate the effects of gender and assessment training on peer feedback specificity and intrapersonal factors.

1. Applying the MISCA model in peer feedback research

Lipnevich and Panadero (2021) synthesised fourteen feedback models and typologies spanning four decades of feedback research. It mentioned how individual student characteristics influence how feedback is described, how students internally process feedback, how students interact with feedback agents, and how feedback is presented to impact learning. These elements served as the groundwork for the meta-model of feedback, the Message, Implementation, Student Characteristics, Context, and Agents, or the MISCA model (Panadero & Lipnevich, 2022). Since the MISCA model highlights the integrated interactions between these elements, we will investigate the interactions between the Student Characteristics, Message, and Context in this study. We chose these three elements because of their growing importance in the peer assessment literature (see Alqassab et al., 2023). These elements encompass certain variables (e.g., Student Characteristics: Intrapersonal Factors, Gender; Message: Peer Feedback Characteristics; Context: Scaffolds) that are gaining attention in peer assessment research because of their instrumental influence on effective peer assessment implementation (Kerman et al., 2023). Though the MISCA model was not explicitly built for peer assessment research, its generalisability transcends varied areas in feedback research, given the array of feedback typologies and models they systematically integrated. For instance, the MISCA model is now used in peer assessment research (e.g., Byl & Topping, 2023).

2. Student characteristics: what's the role of intrapersonal factors and gender?

The MISCA model highlights the central role of students in the feedback process, proposing that their complex individual characteristics may influence how they provide and process feedback (Panadero & Lipnevich, 2022). Evidence supporting this proposition is increasing in peer assessment literature. The review of Panadero et al. (2023) found that students generate various intrapersonal and interpersonal reactions due to the social nature of peer assessment. Interpersonal factors are social variables (e.g., trust in the peer as an assessor), and intrapersonal factors (e.g., trust in the self as an assessor) are variables internal to the learner that influence peer assessment processes and outcomes. This review also suggests that peer assessment influences intrapersonal and interpersonal factors, just as these factors impact peer assessment (i.e., bidirectional hypothesis). Since we focus on the internal reactions of the assessors when giving peer feedback in this article, we will review intrapersonal variables in the context of gender in the subsequent sections of this article.

While some found the absence of gender effects on intrapersonal issues during peer assessment (e.g., Gatfield, 1999), some studies found differences for men and women. Seroussi et al. (2023) found that men and women did not differ in comfort and trust in themselves when providing peer feedback, but women were more willing to provide peer feedback. On the contrary, Ocampo et al. (2022) found that men felt less trust in themselves as assessors and discomfort when assessing women's writing, especially if they did not receive assessment training. Moreover, studies also found that women experience more discomfort than men because of the social nature of peer assessment (e.g., Rotsaert et al., 2017; Zou et al., 2018). These studies purport that men and women differ in intrapersonal reactions due to peer assessment (or vice-versa). These findings underscore the importance of prioritising student characteristics as it may influence peer assessment processes and outcomes (Panadero et al., 2023). Importantly, it strengthens the interaction of Student Characteristics and Message elements proposed in the MISCA model discussed above (Panadero & Lipnevich, 2022). Given this, it is important to unpack the characteristics of students' peer feedback messages, as the content embedded in the 'Message' element can be one of the roots of students' adverse interpersonal reactions towards peer assessment.

3. Message: how is peer feedback message affected by student characteristics?

The characteristics of the feedback that students receive must be quality if we want them to use it actively. The MISCA model proposes that the feedback message has content and function categories (Panadero & Lipnevich, 2022) . Content refers to typologies integrated in the feedback (i.e., verification, valence, load, and type of information), while function pertains to the intended meaning (i.e., learning/performance, motivation/affect, or self-regulatory) of the feedback provided by the assessor (e.g., peer). In peer assessment literature, studies have focused on analysing the content characteristics that should be embedded in a peer feedback message, especially in writing instruction (e.g., Gielen & De Wever, 2015; Nelson & Schunn, 2009; Patchan et al., 2016). Among these, Patchan et al.'s (2016) work is acknowledged for its thoroughness and relevance in writing instruction research. In line with the MISCA model, Patchan et al.'s coding scheme captures the verification, valence, and load classifications while emphasising how information types (e.g., specificity) may affect implementation.

Patchan et al. (2016) investigated how peer feedback characteristics affect feedback implementation and revision quality. They built their framework on the earlier work of Nelson and Schunn (2009) by expanding the features of peer feedback that promote

implementation. Although subsequent revisions of Patchan et al.'s model have been proposed (e.g., Wu & Schunn, 2023), we believe that using Patchan and colleagues' model adds more value to this study as it fleshes out the specific content in peer feedback in the context of writing instruction. In their model, Nelson and Schunn (2009) and Patchan et al. (2016) mentioned that peer feedback specificity pertains to the specific details that are included in the peer feedback message. They framed that the specificity of a peer feedback message varies on a continuum from outcome feedback (e.g., right or wrong) to highly specific feedback (e.g., localisation, and focus).

Peer feedback localisation is a component of peer feedback specificity that refers to the extent to which feedback identifies the part of the work containing an issue (Nelson & Schunn, 2009). Localisation can be classified as general, text, quote, or topic, depending on how precisely the feedback points to the problematic section. On the other hand, peer feedback focus refers to the type of issue in the text that the assessor highlights in their feedback. This can be categorised into low prose, high prose, or substance, depending on whether the comments concern surface-level writing features, more complex stylistic aspects, or the content and argumentation itself (Patchan et al., 2016). On the receiving end, some studies show that the specificity of peer feedback did not influence performance (Gielen et al., 2010), while others found that receiving praise and localised peer feedback increased the likelihood of implementation (Patchan et al., 2016). However, while localised peer feedback was more likely to be implemented, it did not improve revision quality, whereas focused feedback, though less likely to be implemented, led to improved revisions (Patchan et al., 2016). In the context of gender, we found that men had better essay revisions after receiving specific peer feedback, while women had better essay revisions after receiving concise peer feedback (Ocampo et al., 2025).

On the providing end, some studies note that gender does not affect peer feedback specificity in writing contexts. For instance, gender did not influence the specific peer feedback provided in literature reviews (Ocampo et al., 2023). On the other hand, men were found to provide specific suggestions in their peer feedback of writing reports, but women's peer feedback was more localised and focused (Leung et al., 2010). Females' peer feedback was seen to be more justified and detailed in argumentative essays (Noroozi et al., 2022). However, given the mixed findings in the literature, it remains unclear which gender consistently provides specific peer feedback. It is crucial to examine how gender influences the specificity of peer feedback, as assessees may disregard the feedback from peers of a particular gender (Cao et al., 2019). Additionally, though studies highlight differences between males and females in the peer feedback they provide, no research has explored whether they differ in peer feedback specificity, except Leung et al. (2010), which utilised a descriptive approach. This is an important gap, given the evidence that localised and focused feedback can increase the likelihood of peer feedback uptake (Nelson & Schunn, 2009; Patchan et al., 2016).

As summarised by Lipnevich and Panadero (2021), the effectiveness of feedback in improving performance and achievement depends on the specific characteristics of the feedback provided. One would think that highly specific feedback would lead to more learning improvements, but the reviewed literature above showed mixed effects, especially in relation to gender. This proves the variability and the complexity of the elements within the (peer) feedback process. As a result, we must provide consistent and organised scaffolds to ensure that students provide quality peer feedback to minimise potential deterrents in the peer feedback process.

4. Context: can assessment training improve peer feedback and intrapersonal factors?

The MISCA model highlights instructional design as crucial for effective feedback activities (Panadero & Lipnevich, 2022). Providing scaffolding to students to compose specific, localised, and focused feedback has been shown to improve the content of peer feedback (Baker, 2016; Hoogeveen & van Gelderen, 2013). Literature notes that scaffolding through assessment training is a well-established strategy for enhancing peer feedback content. To illustrate, Min (2016) examined the instructional design of assessment training programs and identified that peer assessment discussion, feedback provision demonstration, and students' application of peer feedback provision are common components of assessment training programs. Studies implementing these components showed significant improvements in students' peer feedback, and their interpersonal and intrapersonal reactions during peer assessment (Alqassab et al., 2018; Lam, 2010; Min, 2016; Rotsaert et al., 2018; Sluijsmans et al., 2004).

Studies directly investigating assessment training's influence on minimising gender issues when providing peer feedback are scant and mixed. Trained male and female assessors did not differ in the number of specific peer feedback provided, but female assessors provided more general peer feedback than men (Hamer et al., 2015). Untrained male assessors were less reliable in peer assessments of literature reviews compared to trained male and female assessors (Ocampo et al., 2022). On the other hand, both male and female assessors showed gender bias towards opposite-sex peer assessees, even after receiving advanced assessment training (Aryadoust, 2016). Regarding assessment training's effects on intrapersonal factors, the recent review of Senden and Coertjens (2024) showed positive benefits favouring assessment training. For instance, trained male and female assessors reported comfort and trust in their abilities (Ocampo et al., 2022).

Overall, these findings show the importance of assessment training in enhancing peer feedback and intrapersonal factors. Nonetheless, studies that examine the combined effects of assessment training in minimising gender issues, improving peer feedback content, and intrapersonal factors are limited (see Ocampo et al., 2023, 2022). This is crucial since prior studies have found evidence of gender stereotypes perpetuating peer assessments (Cao et al., 2019; Torres-Guijarro & Bengoechea, 2017), which has the potential to influence the other elements in the feedback process (see Panadero & Lipnevich, 2022). Thus, it is necessary to explore how the student characteristics (i.e., gender and intrapersonal factors), message (i.e. peer feedback specificity), and context (i.e., assessment training) elements influence the peer feedback process.

5. Aim and research questions

This study is part of a larger research project examining the influence of gender on peer assessment. In our previous study (Ocampo et al., 2023), we focused on peer feedback characteristics that are related to problem identification and providing solutions. We investigated how verification (i.e., right or wrong) and elaboration (i.e., informative or suggestive) peer feedback is affected by gender and training. In this study, we extend our previous work by focusing on the specific characteristics of peer feedback messages. Specifically, we investigate how the localisation and focus of peer feedback are influenced by gender and training. We sought to investigate this because receiving specific feedback can help students improve their performance (Lipnevich & Panadero, 2021). However, students' individual characteristics may deter quality peer feedback provision. Nonetheless, assessment training can be a powerful scaffold to help students provide specific feedback (Baker, 2016; Hoogeveen & van Gelderen, 2013). In this study, we examine the interplay between the Message (i.e., peer feedback message specificity), Student Characteristics (i.e., gender and intrapersonal factors), and Context (i.e., Assessment Training) elements of the MISCA model (Panadero & Lipnevich, 2022). Specifically, we aim to investigate the effects of gender and assessment training on peer feedback specificity and intrapersonal factors.

We examine the following independent variations on peer feedback specificity (peer feedback localisation and focus) and intrapersonal factors (trust in the self as an assessor and dis/comfort) using a randomised controlled trial ($2 \times 2 \times 2$): (1) assessor gender (male vs. female), (2) fictitious assessee gender (male vs. female), and (3) assessment training (with vs. without). Our research questions are as follows:

RQ1. Does assessor gender, assessee gender and assessment training affect peer feedback localisation (i.e., general, text, quote, and topic)?

RQ2. Does assessor gender, assessee gender and training affect peer feedback focus (i.e., low prose, high prose, and substance)?

RQ3. Does assessor gender, assessee gender and training affect intrapersonal factors (i.e., trust in the self as an assessor and dis/comfort)?

6. Method

6.1. Participants

A total of 240 undergraduate psychology students participated in this study. Participants were enrolled in research writing courses at a private university in the Philippines. Data were collected during the 2021–2022 and 2022–2023 academic years. Based on their self-reported gender, there were 120 men and 120 women whose mean age was 20.8 (SD = 1.16). There were 121 participants (60 men and 61 women) who underwent assessment training and 119 participants who did not receive peer assessment training. Participants were further divided into eight subgroups (i.e., four per condition), where they provided peer feedback to fictitious male or female assessees (see Fig. 2 for distribution). Across all conditions, 61.7 % (i.e. 51.4 % in the assessment training and 48.6 % in the no assessment training condition) of participants reported prior experience giving peer feedback.

6.2. Materials

6.2.1. Writing samples

Participants provided peer feedback on three writing samples of poor, average, and excellent quality, as rated by three independent experts (i.e. psychology instructor, educational assessment expert, and writing instructor) using the same rubric as the participants, with high inter-rater reliability (ICC =.984). The writing samples were literature review drafts from former students to enhance ecological validity. The writing samples cover topics on e-learning (poor), passion (average), and burnout (excellent). To establish the assessee gender manipulation, assessors provided feedback on the same three writing samples, with only the gender of the assessee being altered based on their respective subgroup (see Fig. 2).

6.2.2. Writing Rubric

Participants used a department recommended rubric for research projects. Those in the assessment training condition were instructed to use it as a guide for composing their peer feedback. In contrast, those in the no assessment training condition only had the rubric in their Eduflow account for reference and no further instructions on using it. The writing rubric is in Supplementary Material 1.

6.2.3. Eduflow platform

Students gave peer feedback using Eduflow, a platform designed for assessment. Each student was assigned an individual account for the task (see Fig. 1). All participants were trained to submit their peer feedback in Eduflow.

6.3. Measures

6.3.1. Peer Feedback Specificity

Patchan et al.'s, (2016) definitions of peer feedback localisation and peer feedback focus were used to code peer feedback specificity in this study, as it is the most intricate one in the literature. Peer feedback localisation was further categorised into general, text,

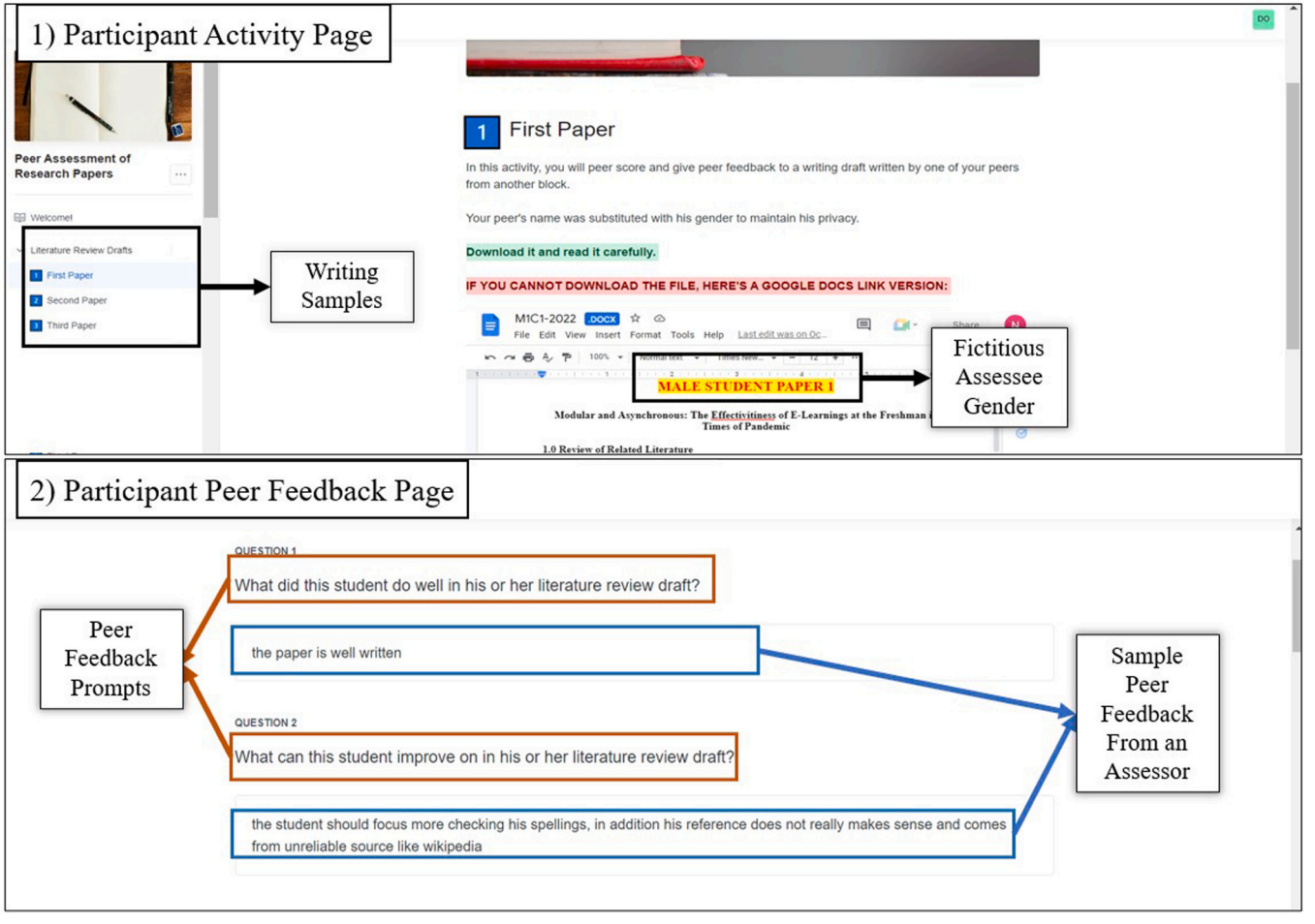


Fig. 1. Participant Eduflow Page.

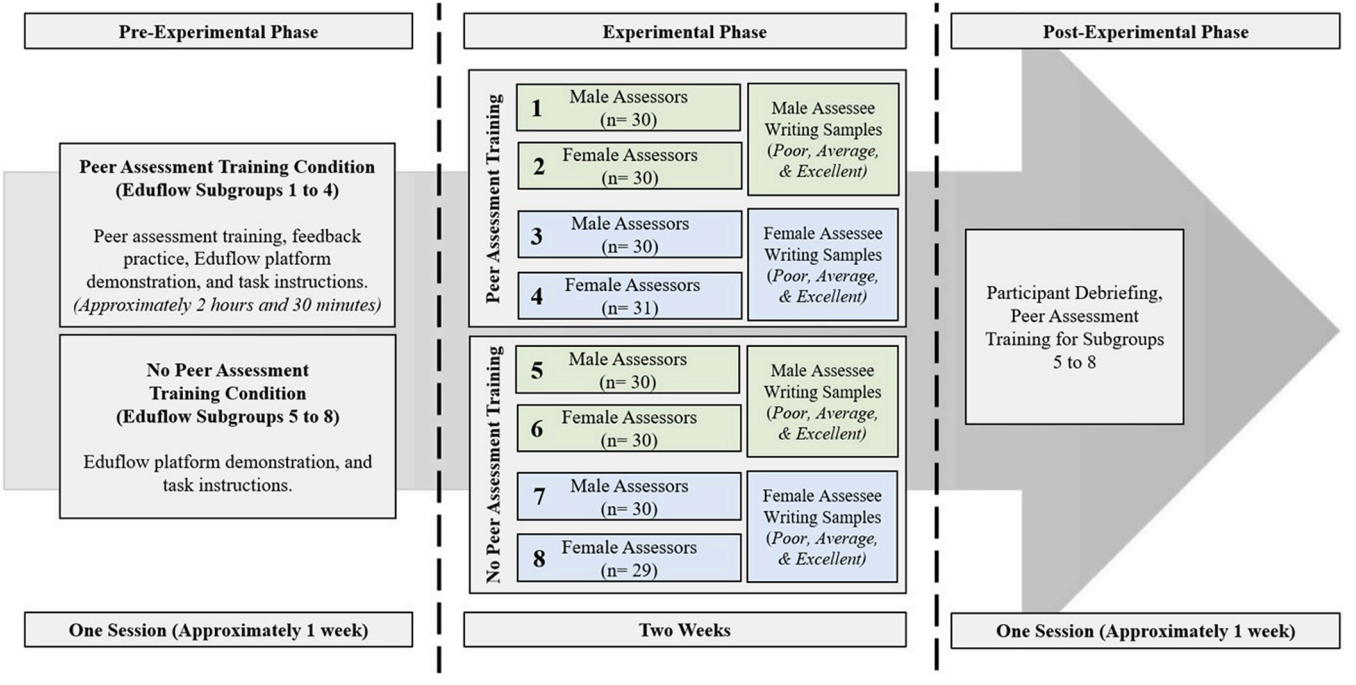


Fig. 2. Study Procedure Flowchart.

quote, and topic. On the other hand, the focus of peer feedback was further categorised and coded into low prose, high prose, and substance. Originally, this coding scheme was used only to code criticism peer feedback (i.e., feedback explaining problems or solutions) and not for positively phased peer feedback. However, we have also decided to use it to code positively phrased peer feedback because there was positively phrased peer feedback that showed specificity in our dataset.

Following Strijbos et al. (2006) peer feedback data were divided into syntactical units for content analysis. Random peer feedback messages were segmented by the first and third authors with high agreement (93 %), subsequent individual segmentation resulted in 3017 segments that were content analysed. The authors collaboratively discussed the coding scheme for content analysis before independently coding the segments for the presence (1) or absence (0) of peer feedback localisation and focus categories. To ensure a common understanding of the peer feedback coding scheme, 455 random segments were double-coded for interrater agreement. All Kappa values were above the conventional threshold of.80, indicating agreement between the two coders (Landis & Koch, 1977). The descriptions, examples, and interrater agreement are in the Appendix.

6.3.2. *Intrapersonal factors*

Participants responded to two single-item questions for intrapersonal factors based on Ocampo et al. (2022). The first question about trust in themselves as an assessor was, *"How effective did you feel in giving peer feedback to your peer's review of related literature drafts?"* (1 = Extremely Ineffective to 5 = Extremely Effective). The last question about discomfort was, "*How comfortable did you feel in giving peer feedback to your peer's review of related literature draft?"* (1 = Extremely Uncomfortable to 5 = Extremely Comfortable).

Table 1
Peer feedback localisation and focus for the poor writing sample.

		Poor Writing Sample							
	Groups	M	SD	Min	Max	*U*	*z*	*p*	*r*
General Localisation	Male Assessor	.629	.338	.00	1.00	5017.5	−4.22	< .001***	−0.27
	Female Assessor	.810	.239	.00	1.00				
	Male Assessee	.685	.333	.00	1.00	6527.5	−1.30	.194	−0.08
	Female Assessee	.754	.272	.00	1.00				
	Training	.713	.290	.00	1.00	6763.5	−.843	.399	−0.05
	No Training	.727	.322	.00	1.00				
Text Localisation	Male Assessor	.183	.237	.00	1.00	5830.5	−2.90	.004**	−0.19
	Female Assessor	.092	.160	.00	.800				
	Male Assessee	.138	.215	.00	1.00	7185.5	−.030	.976	0.00
	Female Assessee	.136	.200	.00	.667				
	Training	.161	.219	.00	1.00	6206.0	−2.08	.037*	−0.13
	No Training	.113	.192	.00	.800				
Quote Localisation	Male Assessor	.014	.069	.00	.500	6732.0	−1.74	.082	−0.11
	Female Assessor	.022	.071	.00	.500				
	Male Assessee	.011	.053	.00	.500	6934.5	−.986	.324	−0.06
	Female Assessee	.025	.084	.00	.500				
	Training	.020	.072	.00	.500	6979.0	−.819	.413	−0.05
	No Training	.017	.069	.00	.500				
Topic Localisation	Male Assessor	.092	.231	.00	1.00	6875.5	−.930	.352	−0.06
	Female Assessor	.045	.123	.00	.500				
	Male Assessee	.087	.192	.00	1.00	6384.0	−2.34	.019*	−0.15
	Female Assessee	.051	.179	.00	1.00				
	Training	.055	.166	.00	1.00	6796.5	−1.15	.248	−0.07
	No Training	.083	.204	.00	1.00				
Low Prose	Male Assessor	.263	.288	.00	1.00	5227.5	−3.74	< .001***	−0.24
	Female Assessor	.382	.243	.00	1.00				
	Male Assessee	.315	.251	.00	1.00	7060.0	−.266	.791	−0.02
	Female Assessee	.331	.293	.00	1.00				
	Training	.314	.244	.00	1.00	7051.5	−.281	.779	−0.02
	No Training	.332	.300	.00	1.00				
High Prose	Male Assessor	.256	.281	.00	1.00	6589.0	−1.18	.237	−0.08
	Female Assessor	.193	.200	.00	.667				
	Male Assessee	.216	.239	.00	1.00	6971.5	−.442	.658	−0.03
	Female Assessee	.232	.252	.00	1.00				
	Training	.245	.246	.00	1.00	6417.0	−1.51	.130	−0.10
	No Training	.203	.244	.00	1.00				
Substance	Male Assessor	.370	.333	.00	1.00	6849.0	−.661	.508	−0.04
	Female Assessor	.386	.259	.00	1.00				
	Male Assessee	.365	.305	.00	1.00	6859.5	−.642	.521	−0.04
	Female Assessee	.391	.292	.00	1.00				
	Training	.361	.284	.00	1.00	6727.0	−.890	.373	−0.06
	No Training	.396	.312	.00	1.00				

* $p < .05$
** $p < .01$
*** $p < .001$

6.4. Procedure

All procedures in this study were conducted in accordance with ethical guidelines and were approved by the respective University Ethics Committees: the authors' institution (ETK-33/20–21) and the participants' university (REC_22_028). This study was conducted in psychology research writing courses where peer feedback was used as an assessment strategy. The trained participants received peer assessment discussion, feedback provision demonstration, and practice peer feedback provision on Eduflow based on Lam (2010), which lasted for 2.5 h. The untrained participants received only task instructions and a brief demo of Eduflow, which took around 30 min. All participants had access to the writing rubric and answered these prompts to provide peer feedback:

1. What did this student do well in his/her literature review draft?
2. What can this student improve on in his/her literature review draft?

All participants were instructed that they would be providing peer feedback to literature review drafts written by students from other classes. Participants were told that the assessee's name had been replaced by their gender to ensure privacy and facilitate instructor identification. Anonymity was maintained throughout, and students were asked not to share details about the writing samples to preserve confidentiality. The study lasted four weeks of a 13-week semester, and participants were given two weeks to complete the feedback task. After submission, a debriefing was provided, and those in the no assessment training group were later offered the same assessment training. Assessment training lesson guides and peer assessment design elements in this study are found in

Table 2
Peer feedback localisation and focus for the average writing sample.

		Average Writing Sample							
	Groups	M	SD	Min	Max	*U*	*z*	*p*	*r*
General Localisation	Male Assessor	.632	.356	.00	1.00	5687.0	−2.93	.003**	−0.19
	Female Assessor	.771	.286	.00	1.00				
	Male Assessee	.694	.336	.00	1.00	7041.5	−.307	.759	−0.02
	Female Assessee	.710	.324	.00	1.00				
	Training	.723	.305	.00	1.00	6919.5	−.543	.587	−0.04
	No Training	.681	.353	.00	1.00				
Text Localisation	Male Assessor	.152	.246	.00	1.00	6702.5	−1.09	.274	−0.07
	Female Assessor	.111	.202	.00	.890				
	Male Assessee	.133	.231	.00	1.00	7148.5	−.113	.910	−0.01
	Female Assessee	.130	.221	.00	1.00				
	Training	.150	.214	.00	.800	6094.5	−2.43	.015*	−0.16
	No Training	.113	.236	.00	1.00				
Quote Localisation	Male Assessor	.010	.076	.00	.750	6726.0	−2.17	.030*	−0.14
	Female Assessor	.019	.068	.00	.400				
	Male Assessee	.020	.090	.00	.750	7073.5	−.579	.563	−0.04
	Female Assessee	.009	.048	.00	.400				
	Training	.021	.095	.00	.750	7072.0	−.584	.559	−0.04
	No Training	.007	.035	.00	.200				
Topic Localisation	Male Assessor	.105	.233	.00	1.00	6585.0	−1.71	.087	−0.11
	Female Assessor	.050	.166	.00	1.00				
	Male Assessee	.091	.227	.00	1.00	6880.5	−.889	.374	−0.06
	Female Assessee	.064	.178	.00	1.00				
	Training	.052	.169	.00	1.00	6389.5	−2.25	.024*	−0.15
	No Training	.103	.232	.00	1.00				
Low Prose	Male Assessor	.287	.323	.00	1.00	6691.5	−.975	.329	−0.06
	Female Assessor	.300	.251	.00	1.00				
	Male Assessee	.287	.290	.00	1.00	7041.5	−.304	.761	−0.02
	Female Assessee	.300	.289	.00	1.00				
	Training	.315	.320	.00	1.00	6863.5	−.645	.519	−0.04
	No Training	.271	.253	.00	1.00				
High Prose	Male Assessor	.280	.312	.00	1.00	6955.0	−.472	.637	−0.03
	Female Assessor	.275	.252	.00	1.00				
	Male Assessee	.316	.291	.00	1.00	6050.5	−2.21	.027*	−0.14
	Female Assessee	.238	.271	.00	1.00				
	Training	.306	.281	.00	1.00	6253.5	−1.82	.068	−0.12
	No Training	.248	.284	.00	1.00				
Substance	Male Assessor	.333	.306	.00	1.00	6445.0	−1.43	.152	−0.09
	Female Assessor	.377	.277	.00	1.00				
	Male Assessee	.335	.280	.00	1.00	6686.0	−.976	.329	−0.06
	Female Assessee	.376	.303	.00	1.00				
	Training	.325	.272	.00	1.00	6434.0	−1.45	.146	−0.09
	No Training	.386	.310	.00	1.00				

* $p < .05$
** $p < .01$

Supplementary Materials 1 and 2, respectively.

6.5. Research design and data analysis

This study used a 2 × 2 × 2 randomised controlled design with three independent variables: assessor gender (men vs women), fictitious assessee gender (man vs woman) and training condition (with vs without). Our dependent variables were the proportion of peer feedback specificity that is categorised in peer feedback localisation (i.e., general, text, quote, and topic) and peer feedback focus (i.e., low prose, high prose, substance), as well as self-report ratings for intrapersonal factors (i.e., trust in the self as an assessor and discomfort).

Following our previous approach to compare the effects of assessor gender, assessee gender, and assessment training on peer feedback specificity, we computed the proportion of peer feedback content for each assessor and per writing sample. Proportions were calculated by dividing the raw count of each peer feedback specificity category by the total number of segments per writing sample. The raw counts were used to provide descriptive information on the content of participants' peer feedback for each dimension. Additionally, participants' ratings for the two intrapersonal variable questions were used in the analysis. We used the Mann-Whitney U non-parametric test because normality assumptions were not achieved for the segments, raw counts, and proportions of all peer feedback specificity categories. On the other hand, a factorial ANOVA with Bonferroni correction was used for the intrapersonal variable questions. All analyses were conducted using SPSS version 28.

Table 3
Peer feedback localisation and focus for the excellent writing sample.

	Excellent Writing Sample								
	Groups	M	SD	Min	Max	*U*	*z*	*p*	*r*
General Localisation	Male Assessor	.581	.342	.00	1.00	4544.5	−5.09	< .001***	−0.33
	Female Assessor	.798	.235	.00	1.00				
	Male Assessee	.691	.309	.00	1.00	7184.0	−.031	.976	0.00
	Female Assessee	.688	.317	.00	1.00				
	Training	.676	.332	.00	1.00	7035.5	−.314	.753	−0.02
	No Training	.703	.291	.00	1.00				
Text Localisation	Male Assessor	.140	.225	.00	1.00	7011.5	−.413	.679	−0.03
	Female Assessor	.111	.175	.00	.670				
	Male Assessee	.128	.204	.00	.670	7174.5	−.056	.955	0.00
	Female Assessee	.123	.200	.00	.100				
	Training	.150	.205	.00	.670	6184.0	−2.23	.026*	−0.14
	No Training	.101	.196	.00	1.00				
Quote Localisation	Male Assessor	.009	.062	.00	.500	7196.0	−.028	.978	0.00
	Female Assessor	.004	.031	.00	.250				
	Male Assessee	.011	.066	.00	.500	7076.0	−.853	.394	−0.06
	Female Assessee	.002	.022	.00	.182				
	Training	.005	.042	.00	.400	7078.0	−.835	.403	−0.05
	No Training	.009	.055	.00	.500				
Topic Localisation	Male Assessor	.131	.233	.00	1.00	5841.5	−3.71	< .001***	−0.24
	Female Assessor	.032	.107	.00	1.00				
	Male Assessee	.082	.185	.00	1.00	7158.5	−.113	.910	−0.01
	Female Assessee	.081	.191	.00	1.00				
	Training	.091	.205	.00	1.00	6929.0	−.739	.460	−0.05
	No Training	.072	.168	.00	.667				
Low Prose	Male Assessor	.140	.235	.00	1.00	5198.5	−4.03	< .001***	−0.26
	Female Assessor	.249	.250	.00	1.00				
	Male Assessee	.184	.257	.00	1.00	6696.5	−1.01	.310	−0.07
	Female Assessee	.205	.239	.00	1.00				
	Training	.201	.243	.00	1.00	6859.0	−.686	.493	−0.04
	No Training	.188	.253	.00	1.00				
High Prose	Male Assessor	.215	.256	.00	1.00	6221.0	−1.91	.056	−0.12
	Female Assessor	.281	.277	.00	1.00				
	Male Assessee	.293	.276	.00	1.00	5851.0	−2.63	.008**	−0.17
	Female Assessee	.203	.253	.00	1.00				
	Training	.251	.255	.00	1.00	6967.0	−.454	.650	−0.03
	No Training	.246	.282	.00	1.00				
Substance	Male Assessor	.526	.343	.00	1.00	5704.0	−2.81	.005**	−0.18
	Female Assessor	.407	.289	.00	1.00				
	Male Assessee	.433	.322	.00	1.00	6320.5	−1.65	.098	−0.11
	Female Assessee	.501	.320	.00	1.00				
	Training	.464	.300	.00	1.00	7128.0	−.134	.893	−0.01
	No Training	.470	.345	.00	1.00				

* $p < .05$
** $p < .01$
*** $p < .001$

7. Results

7.1. RQ1. Does assessor gender, assessee gender, and assessment training affect peer feedback localisation?

Results for peer feedback localisation are shown in Table 1 for the poor sample, Table 2 for the average sample and Table 3 for the excellent sample. Regarding assessor gender, there were no significant differences in quote ($p = .082$) and topic ($p = .352$) localisation for the poor sample, text ($p = .274$) and topic ($p = .087$) localisation for the average sample, or text ($p = .679$) and quote ($p = .978$) localisation for the excellent sample. However, female assessors are likely to provide general localisations in poor ($U = 5017.5$, $z = -4.22$, $p < .001$, $r = -0.27$), average ($U = 5687.0$, z = -2.93, $p = .003$, $r = -0.19$), and excellent ($U = 4544.5$, $z = -5.09$, $p < .001$, $r = -0.33$) samples, as well as quote localisation in the average sample ($U = 6726.0$, $z = -2.17$, $p = .030$, $r = -0.14$). Male assessors also tend to provide text localisations in the poor sample ($U = 5830.5$, $z = -2.90$, $p = .004$, $r = -0.19$) and topic localisation in the excellent sample ($U = 5198.5$, $z = -4.03$, p = <.001, $r = -0.26$).

On the assessee's side, we found no differences in the general (*ps* =.194–.976), text (*ps* =.910–.976), and quote (*ps*=.324–394) localisations across the three writing samples, and topic localisation for the average ($p = .374$) and excellent ($p = .910$) samples. The only significant finding was for fictitious male who received topic localisations in the poor sample ($U = 6384.0$, $z = -2.34$, $p = .019$, $r = -0.15$).

Similarly, we did not find significant differences in the general (*ps* =.753–399), quote (*ps* =.559–.403) localisations in all the writing samples, and topic localisation for the poor ($p = .248$) and excellent ($p = .460$) samples. However, we found that trained assessors are likely to provide text localisations in the poor ($U = 6206.0$, $z = -2.08$, $p = .037$, $r = -0.13$), average ($U = 6094.5$, $z = -2.43$, $p = .015$, $r = -0.16$), and excellent ($U = 6184.0$, $z = -2.23$, $p = .026$, $r = -0.14$) samples. Trained assessors also provided significantly more topic localisations for the average sample ($U = 6389.5$, $z = -2.25$, $p = .024$, $r = -0.15$).

7.2. RQ2. Does assessor gender, assessee gender, and assessment training affect peer feedback focus?

Results for peer feedback focus are shown in Table 1 for the poor sample, Table 2 for the average sample and Table 3 for the excellent sample. We found that both male and female assessors provided similar levels of high-prose peer feedback for the poor ($p = .237$) and excellent ($p = .056$) samples, as well as substance-focused feedback for the poor ($p = .508$) and average ($p = .146$) samples. Both genders also provided comparable low-prose feedback for the average sample ($p = .329$). However, female assessors were more likely to provide low-prose feedback for poor ($U = 5227.5$, $z = -3.74$, $p < .001$, $r = -0.24$) and excellent ($U = 5198.5$, $z = -4.03$, $p < .001$, r = -0.26) samples. Finally, male assessors were more inclined to provide substance-focused feedback for excellent sample ($U = 5704.0$, $z = -2.81$, $p = .005$, $r = -0.18$). Regarding assessees, male and female assessees received similar types of focused peer feedback across all writing samples. There were no significant differences across the three writing samples for low-prose (*ps*

Table 4
Descriptive statistics for trust in the self as an assessor and (dis)comfort.

			Trust in the Self as an Assessor		(Dis)Comfort	
Condition	Assessor Gender	Assessee Gender	*M*	*SD*	*M*	*SD*
Training	Males	Males	4.03	0.72	3.63	0.89
		Females	3.87	0.78	3.63	0.72
		Total	3.95	0.75	3.63	0.80
	Females	Males	4.06	0.73	3.52	0.93
		Females	4.07	1.05	3.93	0.74
		Total	4.07	0.89	3.72	0.86
	Total	Males	4.05	0.72	3.57	0.90
		Females	3.97	0.92	3.78	0.74
		Total	4.01	0.82	3.68	0.83
No Training	Males	Males	3.90	0.80	3.50	1.23
		Females	3.77	1.14	3.30	1.09
		Total	3.83	0.98	3.40	1.15
	Females	Males	4.10	0.77	3.69	0.97
		Females	3.80	0.61	3.80	0.85
		Total	3.95	0.71	3.75	0.90
	Total	Males	4.00	0.79	3.59	1.10
		Females	3.78	0.90	3.55	1.00
		Total	3.89	0.85	3.57	1.05
Total	Males	Males	3.97	0.76	3.57	1.06
		Females	3.82	0.97	3.47	0.93
		Total	3.89	0.87	3.52	1.00
	Females	Males	4.08	0.74	3.60	0.94
		Females	3.93	0.86	3.87	0.79
		Total	4.01	0.80	3.73	0.88
	Total	Males	4.02	0.75	3.58	1.00
		Females	3.87	0.91	3.67	0.88
		Total	3.95	0.84	3.63	0.94

=.310–.791) and substance-focused (*ps*=.*098 to*.521) peer feedback. Both genders also received similar levels of high prose peer feedback in the poor sample ($p = .658$). However, fictitious male assessees were more likely to receive high-prose peer feedback for the average ($U = 6050.5$, $z = -2.21$, $p = .027$, $r = -0.14$) and excellent ($U = 5851.0$, $z = -2.63$, $p = .008$, $r = -0.17$) samples. Finally, trained and untrained assessors were comparable in terms of the low-prose (*ps*=.493–.779), high-prose (*ps* = .068–. 650), and substance-focused ($p = .146$–.893) peer feedback they provided across the three writing samples

7.3. RQ3. Does assessor gender, assessee gender, and assessment training affect interpersonal factors?

Descriptive statistics for intrapersonal factors are shown in Table 4, while the results for the intrapersonal variables trust in the self as an assessor and (dis)comfort are shown in Table 5. The results of the factorial ANOVA for trust in self as an assessor revealed a non-significant triple interaction, $F(1, 232) = 0.61$, $p = 0.44$. Additionally, no other main or interaction effects were significant. Similarly, results for (dis)comfort showed a non-significant triple interaction, $F(1, 232) = 0.05$, $p = 0.83$, with no other significant main or interaction effects. These findings indicate that trained and untrained male and female assessors had comparable levels of trust in their abilities and comfort when assessing male or female assessees.

8. Discussion

Despite the growing evidence about gender's influence on peer assessment, few studies have investigated ways to minimise potential gender issues in the peer feedback process and students' intrapersonal reactions. At the same time, while several studies found no differences between men and women in terms of peer feedback specificity and interpersonal factors (e.g., Gatfield, 1999; Hamer et al., 2015), some findings mention differences between men and women in these aspects (e.g., Noroozi et al., 2022; Seroussi et al., 2023). Therefore, this study aimed to investigate the effects of gender and assessment training on peer feedback specificity and intrapersonal factors. This study extended our earlier work that focused on verification and elaboration peer feedback, by looking at the intricacies of peer feedback messages as shown in how they are localised and focused (Ocampo et al., 2023). From the findings of the previous study and this study, we were able to further unpack aspects of Nelson and Schunn's (2009) peer feedback specificity continuum, disentangling outcome-related feedback (e.g., right or wrong/correct or incorrect, as in verification feedback in the first study) from highly specific feedback characteristics (e.g., localisation and focus, as examined in this study). From a theoretical standpoint, we utilised the MISCA model to systematically explore the interplay between feedback messages, student characteristics, and context elements in the peer feedback process. Specifically, with a gender-balanced sample, we investigated the effects of gender and assessment training on peer feedback specificity and intrapersonal factors using a randomised controlled trial in this study.

8.1. RQ1. Effects on peer feedback localisation

Our findings indicate that male and female assessors were comparable across several peer feedback localisation categories in all three writing samples. Specifically, both genders provided similar amounts of quote and topic localisation in the poor sample, text and topic localisation in the average sample, and text and quote localisation in the excellent sample. This suggests that, regardless of gender, assessors are equally capable of identifying problematic or commendable areas in the writing sample and in directing the assessees' attention to these areas through their feedback, supporting the earlier findings about the absence of gender bias when providing peer feedback (Hamer et al., 2015; Ocampo et al., 2023). Likewise, on the assessees' side, our findings indicate no significant differences in the localised peer feedback received across the writing samples. Specifically, there were no differences in general, text, or quote localisation. Additionally, topic localisation was consistent for the average and excellent samples. These results suggest that, regardless of gender, assessees received localised peer feedback, which may indicate the absence of gender differences, supporting previous findings in other peer assessment contexts (e.g., Falchikov & Magin, 1997). Overall, this suggests that assessor or assessee gender has no discernible impact on peer feedback localisation, indicating that assessors consistently provide localised feedback and assessees are equally likely to receive it, regardless of gender.

While many of our findings were non-significant, we also found significant differences in peer feedback localisation. First, we can say that women were consistent and were generally better assessors than men since their peer feedback used general localisation in all writing samples and quote localisation for the average sample. Previous studies have also forwarded this contention (Noroozi et al., 2022; Ocampo et al., 2023). Women may have allotted more effort to thoroughly review the writing samples and describe the general

Table 5
Factorial ANOVA results for trust in the self and perceptions of (dis)comfort.

	Trust in the Self			(Dis)Comfort		
Effect	*F* (1232)	*p*	η_p^2	*F* (1232)	*p*	η_p^2
Assessor Gender	1.16	0.28	0.01	3.24	0.07	0.01
Assessee Gender	1.92	0.17	0.01	0.46	0.50	0.00
Assessment Training	1.13	0.29	0.01	0.77	0.38	0.00
Assessor Gender × Assessor Gender	0.00	1.00	0.00	2.25	0.14	0.01
Assessment Training × Assessor Gender	0.00	0.99	0.00	1.09	0.30	0.01
Assessment Training × Assessee Gender	0.39	0.53	0.00	1.09	0.30	0.01
Assessor Gender × Assessor Gender × Assessment Training	0.61	0.44	0.00	0.05	0.83	0.00

location of an issue in the writing samples since they excel in linguistic tasks (e.g., writing tasks) and are more self-regulated than men (Voyer & Voyer, 2014). Second, the web-based nature of the peer feedback activity may have helped women to provide more localised peer feedback. Since the writing samples were easily accessible on the peer feedback page (see Fig. 1), female assessors might have capitalised on this feature by copy-pasting and quoting the specific areas in the average sample they had concerns about. This supports studies saying that women are more attuned and active in performing tasks in web-based learning environments (see Prinsen et al., 2007 for a review).

On the other hand, we found that male assessors could provide topic localisations in the poor sample and text localisations in the excellent sample. Our study adds to the literature that men can craft specific peer feedback by pointing to the topic in the essay where an issue might have occurred, contradicting students' stereotype that men provide superficial peer feedback (Cao et al., 2019). While we cannot entirely attribute this to training because of the non-parametric nature of the analysis, we believe men may have been able to localise their peer feedback because of the assessment scaffolding (e.g., prompts and rubric) available in the feedback pages, lending support once again to the strength of assessment scaffolds in improving feedback (Panadero & Lipnevich, 2022).

On the receiving side of the peer feedback, fictitious male assessees received topic localisations in the poor sample even if the writing samples were identical and only the gender of the fictitious assessee changed. This was similar to our previous finding, where fictitious males received more suggestions even though the writing samples were identical (Ocampo et al., 2023). We think this is because of the gender stereotype that men will not adhere to feedback if it is not detailed enough (Gray & Leith, 2004; Marshall & Smith, 1987). Future studies may investigate how implementation can be affected by peer feedback detail and gender.

Regarding training, both trained and untrained assessors were comparable in providing general and quote localisations in all writing samples and topic localisation for the poor writing sample, which aligns with some findings (Hamer et al., 2015). Since more than half of the participants have had peer feedback experience in the past (i.e., 61.7 %) and all participants received assessment scaffolds in their account regardless of their training condition, the experience (e.g., on using a rubric and answering prompts) of untrained assessors in their previous courses may have guided them to write localised peer feedback. However, we consistently found that trained assessors provided significantly more text localisations in the three writing samples and topic localisations in the poor sample. This supports prior studies' that assessment training improves peer feedback characteristics (Alqassab et al., 2018; Rotsaert et al., 2018; Sluijsmans et al., 2004). Apart from the assessment training implemented, we think the thorough explanation of assessment scaffolds may have contributed to this positive finding. This is encouraging because our assessment training was only carried out for a single session of 2.5 h. This strengthens a recent meta-analysis's call for the implemention of scaffolding (e.g., assessment training) in the peer assessment process (Hornstein et al., 2025).

8.2. RQ2. Effects on peer feedback focus

We found that male and female assessors provided comparable peer feedback across various focus categories, with both genders offering high-prose peer feedback in poor and excellent samples, substance-focused peer feedback in poor and average samples, and low-prose feedback in the average sample. Likewise, assessees received low-prose and substance-focused feedback across all samples and high-prose feedback in poor samples, regardless of their gender. This suggests that assessees will receive peer feedback that identifies minor to content-related issues in their essays, no matter what their gender is. Similarly, these findings indicate that gender-related differences are minimal in both providing and receiving peer feedback, reinforcing the notion that peer assessment can be an equitable process (Hamer et al., 2015; Leung et al., 2010). Nonetheless, our results suggest that female assessors provided low prose-focused peer feedback in the poor and excellent samples, while male assessees provided substance-focused peer feedback in the excellent sample. On the other hand, fictitious male assessees received high prose-focused peer feedback in the average and excellent writing samples

As low prose-focus peer feedback addresses surface-level issues (e.g., formatting, spelling, and word choice), we believe that women's tendency to be more thorough may have contributed to their greater attention to enumerating such issues (Prinsen et al., 2007; Voyer & Voyer, 2014). On the other hand, male assessors provided more substance-focused peer feedback than female assessors. Substance peer feedback deals with content-related (e.g., contradiction, missing content) and support-related (e.g., robustness of references) issues in the paper, which may require holistic cognitive processing. It appears that our female assessors were more geared to provide solution-oriented peer feedback to fix issues in the writing samples, while males were more concerned about the holistic ideas and insights in the writing sample, lending support to prior findings in this area (e.g., He & Wong, 2021; Kuhn & Holling, 2009). This may also provide a glimpse into gender's influence when participating in web-based learning environments, where women were observed to be more collaborative (e.g., providing solutions to writing issues) and men were more authoritative (e.g., showing concerns about facts) (Prinsen et al., 2007). This phenomenon was also seen in Cao et al.'s (2019) study. Future studies may want to investigate gender differences in cognitive processes in peer feedback.

On the other hand, we found that fictitious male assessees received high prose-focused peer feedback in the average and excellent writing samples. Several explanations can be drawn from this finding. First, unlike the earlier work of Patchan et al. (2016), where only criticism peer feedback was coded for feedback specificity, we have decided to code both criticism and positively phrased peer feedback for feedback specificity since it was present in our data. This means that assessors may have praised or criticised the substance of the male assessee's writing sample. Interestingly, although the writing samples of men and women were identical, assessors only did this to male assessees' work. We believe cultural stereotypes may have played a part in this result in a double-edged manner. Like the results for topic localisations, assessors may have given substance-focused peer feedback to men because they may not take notice of the feedback if it is not detailed enough. On the other hand, male assessees may have received substance-focused peer feedback because men are valued highly in society, especially in the Philippines, where the data collection took place (Hofstede et al., 2010).

Assessors may have seen men in a higher light due to their strong societal influence. Future studies may investigate cultural dynamics in peer feedback.

Why did assessment training not affect positively peer feedback focus? Several explanations can be made from this. First, it may be due to participants' prior peer feedback experience in their other courses. Additionally, the writing rubric they used in their courses was the same in this study. Therefore, combining these two may have equalised the effects of training for the untrained assessors. Untrained assessors may have capitalised on their experience and knowledge of the rubric to give focused feedback to the writing samples. Second, unlike peer feedback localisation, where the assessor had to pinpoint the specific part of the work with an issue, providing focused peer feedback that identifies content-related issues may be more challenging due to the complex cognitive demands involved in coming up with such peer feedback. For example, the assessor must possess adequate knowledge of rhetorics and writing to provide low-prose and high-prose focused peer feedback, as well as sufficient domain knowledge to provide substance-focused peer feedback (Patchan et al., 2016). At the same time, providing peer feedback to three writing samples of varying quality is a complex task that even high-ability assessors find challenging to make (Patchan & Schunn, 2015). Although half of our participants received training in providing peer feedback and all participants were instructed on academic writing in their research courses, they may not have been familiar with the specific topics they were tasked to assess. This aligns with what Patchan and Schunn (2015) found that only assessors with high domain knowledge were able to provide high prose and substance-focused peer feedback. In other words, training students to provide focused peer feedback may be challenging to achieve in such a short training duration, and it is suggested that future studies investigate the effects of assessment training on content-specific skills (see Hornstein et al., 2025).

8.3. RQ3. Effects on intrapersonal factors

We found that regardless of whether they received training or not and whether they assessed a male or female assessee, both male and female assessors reported the same levels of trust in their abilities as an assessor and comfort. These findings are in line with prior studies (Seroussi et al., 2023) and contradict findings where men (e.g., Ocampo et al., 2022) or women (e.g. Zou et al., 2018) experience adverse intrapersonal reactions during peer assessment activities. Several reasons can explain these findings.

First, regardless of whether they received training or not, most of our participants reported having prior experience with peer feedback. This may have balanced the effect of training on managing perceptions of trust and discomfort. Students may have been realistic in rating their levels of trust and discomfort, given their previous experience with peer feedback, such as in previous studies (Alqassab et al., 2018). Second, the web-based nature of the activity may have mitigated adverse intrapersonal reactions. Because the assessors were informed that their feedback would remain anonymous and the assessee would not know who provided it, this sense of anonymity may have helped them trust their abilities and feel more comfortable giving feedback. The presence of this feature has increased in many web-based peer assessment platforms (Ocampo & Panadero, 2023), and studies have noted its advantages on socio-affective outcomes (e.g., trust in the self as an assessor and discomfort) (Panadero & Alqassab, 2019). Third, though assessment training positively affected feelings of trust and comfort in our previous study with a similar design in peer scoring contexts (Ocampo et al., 2022), we believe the flexible nature of providing peer feedback may have allayed adverse reactions towards peer assessment. Assessors may have felt constricted by the need to assign fixed scores based on the pre-set rubric, while they may have felt independent to compose any peer feedback related to the writing sample, supporting previous findings (e.g., Fraile et al., 2023). Although we did not test for participants' affect when using a rubric vs. prompts in this study, it is still an interesting finding given that responding to prompts is a cognitively demanding activity that might generate adverse internal reactions due to the effort required (Panadero et al., 2014).

8.4. Limitations and future lines of research

First, our participants were limited to university students in the Philippines. As mentioned earlier, gender-related cultural perceptions may have played a part in the peer feedback provided and their intrapersonal reactions towards peer feedback. Subsequent studies should consider replicating the study in other cultures to see its generalisability. Second, though the training we implemented was well-organised and we found positive results for peer feedback localisation, the assessment training we implemented was brief. Future studies should consider implementing assessment training with a longer duration. Third, because our analyses were non-parametric, we were restricted in exploring possible interactions between factors. Future studies that achieve normality in their variables may apply parametric methods to test these interactions more comprehensively. Fourth, the assessees in this study were fictional. We did this because we did not want to influence the peer feedback participants would provide to the writing samples. Though this approach has been used in prior peer feedback studies (e.g., Strijbos et al., 2010), future studies may want to use real-life male and female assessees to increase the ecological validity of the study. Fifth, our measurement of intrapersonal factors was limited to single-item self-report questions. Although it has been a practice in prior peer assessment studies (Ocampo et al., 2022), studies should consider using more comprehensive measures to assess intrapersonal factors.

8.5. Pedagogical recommendations

Since the majority of the effect sizes are small based on conventional thresholds (Cohen, 1992), it can be assumed that differences between men and women in peer feedback specificity may not be drastic in assessing writing-related tasks. Thus, with circumstances similar to our study, instructors do not need to worry about it. Nevertheless, we advocate for the implementation of a well-organised assessment training program when assessing writing outputs, as it may influence peer feedback content. Students should be oriented

about rubric use and prompts (Baker, 2016; Hoogeveen & van Gelderen, 2013; Panadero & Lipnevich, 2022), and context-specific skills (Hornstein et al., 2025) when crafting peer feedback. As gender-related stereotypes may influence peer assessments, such as the notion that men are superficial assessors (Cao et al., 2019) or that women are strict assessors (Aryadoust, 2016), we recommend gender-awareness workshops prior to engaging in peer assessment given the gender-stereotypical tendencies in providing peer feedback found in this study.

9. Conclusions

In this study, we zeroed in on the interplay of student characteristics, context, and message elements of the MISCA model. Specifically, we investigated the effects of gender and assessment training on peer feedback specificity and intrapersonal factors using a randomised controlled trial with a gender-balanced sample. Our results generally suggest that gender had minimal observable effects on peer feedback and no effects at all on intrapersonal factors. These slight variations between men and women in peer feedback are what Hamer et al. (2015) would call "minor stylistic differences" (p. 162). Nonetheless, even if these differences are "minor", it is still necessary to be mindful of gender's effects as it may influence the peer assessment process (Kerman et al., 2023).

Funding

First author was funded during the conceptualisation and data collection stages of this paper (December 2020 to February 2024) by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 847624, and co-funded by several institutions. All authors funded via the Spanish National R+D call from the Ministerio de Ciencia, Innovación y Universidades (Generación del conocimiento 2019), Reference number: PID2019-108982GB-I00. Basque Government call for Ayudas para apoyar las actividades de grupos de investigación del sistema universitario vasco para los años 2022–2025. Reference number: IT1624-22. Third author is funded by the Formación de Personal Investigador, reference PRE2020-092280.

CRediT authorship contribution statement

Ocampo Jose Carlos: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Panadero Ernesto:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Zamorano David:** Validation, Methodology, Investigation, Data curation. **Sánchez-Iglesias Iván:** Validation, Methodology, Formal analysis, Data curation.

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

During the preparation of this work the author(s) used ChatGPT4 to shorten and simplify long sentences. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgements

Our sincerest gratitude to David Kofoed Wind for the Eduflow access, and Patricia Mae Taba, Rita Aringo, and Ian Llenares for the assistance in data collection.

Appendix. Peer Feedback Specificity Coding Scheme (Based on Nelson & Schunn, 2009 and Patchan et al., 2016)

Peer Feedback Specificity Category	Subcategory	Description	Example from the dataset
Peer Feedback Localisation	General	Mentioned a general location that could be found in any paper.	■ "The sentence structure is also good, almost no typographical errors." ■ "Also, there are few spelling errors and the student managed to minimize plagiarism by paraphrasing (but it can still be improved)."
	Text	Mentioned a specific location to identify issues or positive aspects of the paper.	■ "In 1.3, the spacing is not right, the gap from 1.1 and 1.2 is far from the gap of 1.2 and 1.3." ■ "There are no given DOIs or links at the end of each reference."
	Quote	Used a quote to highlight issues or positive aspects of the paper.	■ "I would like to know why did the student put this "There is a need to study passion in other context such as teaching in order to test the generalizablitiy [sic]…" in the paragraph?" ■ "Consider also rechecking in case of broken sentence fragments, such as this one "…""

(*continued on next page*)

(*continued*)

Peer Feedback Specificity Category	Subcategory	Description	Example from the dataset
	Topic	Mentioned the topic to identify issues or strengths in the paper occurred.	■ "Also, it is commendable how(…)(3) the discussion provides relevant examples (e.g., differentiation of harmonious passion and obsessive passion)" ■ "In *1.4 Modular Learning*, I think that the student failed to deliver what modular actually means."
Peer Feedback Focus	Low Prose	Surface-level feedback dealing with word choice (e.g., spelling) or paper format (e.g., indention, spacing)	■ "The student did well on the format of his paper." ■ "This misspelled word is "Effectivitiness", it should be "Effectiveness"" (Quote Localization) ■ "Lastly, the references were kind of messy and confusing due to the format that was not arranged properly"
	High Prose	High-level writing statements (e.g., clarity, use of transitions, strength of arguments, provision of support and counterarguments, insight).	■ "The organization of thoughts and the way she wrote it concisely and interconnected it with the progressive paragraphs, makes it more interesting to continue reading." ■ "He has also put too much information that he can summarize to shorten and make his RRL more direct and straight to the point."
	Substance	A statement dealing with content (i.e., correctness, comprehensiveness) and the level of robustness of support or reference used.	■ "The work introduced the topic that is needed, the ideas were well presented and organized, well supported with examples and theories." ■ "Examine and assess the articles you've chosen and most importantly, make sure that your sources are credible and reliable (…)"

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.asw.2025.100987.

Data Availability

The data that has been used is confidential.

References

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

Alqassab, M., Strijbos, J.-W., Panadero, E., Ruiz, J. F., Warrens, M., & To, J. (2023). A Systematic Review of Peer Assessment Design Elements. *Educational Psychology Review, 35*(1), 18. https://doi.org/10.1007/s10648-023-09723-7

Aryadoust, V. (2016). Gender and academic major bias in peer assessment of oral presentations. *Language Assessment Quarterly, 13*(1), 1–24. https://doi.org/10.1080/15434303.2015.1133626

Baker, K. M. (2016). Peer review as a strategy for improving students' writing process. *Active Learning in Higher Education, 17*(3), 179–192. https://doi.org/10.1177/1469787416654794

Byl, E., & Topping, K. J. (2023). Student perceptions of feedback in reciprocal or nonreciprocal peer tutoring or mentoring. *Studies in Educational Evaluation, 79*, Article 101304. https://doi.org/10.1016/j.stueduc.2023.101304

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Falchikov, N., & Magin, D. (1997). Detecting gender bias in peer marking of Students' group process work. *Assessment Evaluation in Higher Education, 22*(4), 385–396. https://doi.org/10.1080/0260293970220403

Fraile, J., Gil-Izquierdo, M., & Medina-Moral, E. (2023). The impact of rubrics and scripts on self-regulation, self-efficacy and performance in collaborative problem-solving tasks. *Assessment Evaluation in Higher Education, 48*(8), 1223–1239. https://doi.org/10.1080/02602938.2023.2236335

Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment Evaluation in Higher Education, 24*(4), 365–377. https://doi.org/10.1080/0260293990240401

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

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

Gray, C., & Leith, H. (2004). Perpetuating gender stereotypes in the classroom: a teacher perspective. *Educational Studies, 30*(1), 3–17. https://doi.org/10.1080/0305569032000159705

Hamer, J., Purchase, H., Luxton-Reilly, A., & Denny, P. (2015). A comparison of peer and tutor feedback. *Assessment Evaluation in Higher Education, 40*(1), 151–164. https://doi.org/10.1080/02602938.2014.893418

He, W., & Wong, W. (2021). Gender differences in the distribution of creativity scores: Domain-Specific patterns in divergent thinking and creative problem solving. *Frontiers in Psychology, 12*, Article 626911. https://doi.org/10.3389/fpsyg.2021.626911

Hofstede, G., Hofstede, G.J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind: intercultural cooperation and its importance for survival* (Revised and expanded third edition). McGraw-Hill. ⟨http://catdir.loc.gov/catdir/enhancements/fy1009/2010010437-b.html⟩.

Hoogeveen, M., & van Gelderen, A. (2013). What works in writing with peer response? A review of intervention studies with children and adolescents. *Educational Psychology Review, 25*(4), 473–502. https://doi.org/10.1007/s10648-013-9229-z

Hornstein, J., Keller, M. V., Greisel, M., Dresel, M., & Kollar, I. (2025). Enhancing the Peer-Feedback process through instructional support: a Meta-Analysis. *Educational Psychology Review, 37*(2), 42. https://doi.org/10.1007/s10648-025-10017-3
Kerman, N. T., Banihashem, S. K., Karami, M., Er, E., van Ginkel, S., & Noroozi, O. (2023). Online peer feedback in higher education: a synthesis of the literature. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12273-8
Kuhn, J.-T., & Holling, H. (2009). Gender, reasoning ability, and scholastic achievement: a multilevel mediation analysis. *Learning and Individual Differences, 19*(2), 229–233. https://doi.org/10.1016/j.lindif.2008.11.007
Lam, R. (2010). A peer review training workshop: coaching students to give and evaluate peer feedback, 114 *TESL Canada Journal*, 114. https://doi.org/10.18806/tesl.v27i2.1052.
Landis, J. R., & Koch, G. G. (1977). A One-Way components of variance model for categorical data. *Biometrics, 33*(4), 671–679. https://doi.org/10.2307/2529465
Leung, K., Chan, M., Maxwell, G., & Poon, T. (2010). A qualitative analysis of Sub-degree students commentary styles and patterns in the context of gender and peer e-Feedback. In P. Tsang, S. K. S. Cheung, V. S. K. Lee, & R. Huang (Eds.), *Hybrid Learning, 6248* pp. 149–159). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14657-2_15.
Lipnevich, A. A., & Panadero, E. (2021). A Review of Feedback Models and Theories: Descriptions, Definitions, and Conclusions. *Frontiers in Education, 6*. https://www.frontiersin.org/article/10.3389/feduc.2021.720195.
Marshall, S. P., & Smith, J. D. (1987). Sex differences in learning mathematics: a longitudinal study with item and error analyses. *Journal of Educational Psychology, 79*(4), 372–383. https://doi.org/10.1037/0022-0663.79.4.372
Min, H.-T. (2016). Effect of teacher modeling and feedback on EFL students' peer review skills in peer review training. *Journal of Second Language Writing, 31*, 43–57. https://doi.org/10.1016/j.jslw.2016.01.004
Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: how different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x
Noroozi, O., Banihashem, S. K., Taghizadeh Kerman, N., Parvaneh Akhteh Khaneh, M., Babayi, M., Ashrafi, H., & Biemans, H. J. A. (2022). Gender differences in students' argumentative essay writing, peer review performance and uptake in online learning environments. *Interactive Learning Environments*, 1–15. https://doi.org/10.1080/10494820.2022.2034887
Ocampo, J. C. G., Panadero, E., & Díez, F. (2022). Are men and women really different? The effects of gender and training on peer scoring and perceptions of peer assessment. *Assessment & Evaluation in Higher. Education, 48*(6), 760–776. https://doi.org/10.1080/02602938.2022.2130167
Ocampo, J. C., & Panadero, E. (2023). Web-Based Peer Assessment Platforms: What Educational Features Influence Learning, Feedback and Social Interaction? In O. Noroozi, & B. De Wever (Eds.), *The Power of Peer Learning*. Springer Open. https://doi.org/10.1007/978-3-031-29411-2_8.
Ocampo, J. C., Panadero, E., Sánchez-Iglesias, I., & Ruiz, F. D. (2025). "Due to the composition of the feedback, I think it's a girl": The effects of gender and peer feedback content on essay revisions and perceptions of Peer Feedback. *Studies in Educational Evaluation, 85*, 101455. https://doi.org/10.1016/j.stueduc.2025.101455
Ocampo, J. C., Panadero, E., Zamorano, D., Sánchez-Iglesias, I., & Diez Ruiz, F. (2023). The effects of gender and training on peer feedback characteristics. *Assessment & Evaluation in Higher. Education, 49*(4), 539–555. https://doi.org/10.1080/02602938.2023.2286432
Panadero, E., Alonso-Tapia, J., & Huertas, J.-A. (2014). Rubrics vs. self-assessment scripts: Effects on first year university students' self-regulation and performance / Rúbricas y guiones de autoevaluación: efectos sobre la autorregulación y el rendimiento de estudiantes universitarios de primer año. *Journal for the Study of Education and Development, 37*(1), 149–183. https://doi.org/10.1080/02103702.2014.881655
Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher. Education, 44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1600186
Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: intrapersonal and interpersonal factors. Assessment & Evaluation in Higher. *Education, 48*(8), 1053–1075. https://doi.org/10.1080/02602938.2023.2164884
Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416
Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: how peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098–1120. https://doi.org/10.1037/edu0000103
Patchan, M. M., & Schunn, C. D. (2015). Understanding the benefits of providing peer feedback: how students respond to peers' texts of varying quality. *Instructional Science, 43*(5), 591–614. https://doi.org/10.1007/s11251-015-9353-x
Prinsen, F. R., Volman, M. L. L., & Terwel, J. (2007). Gender-related differences in computer-mediated communication and computer-supported collaborative learning. *Journal of Computer Assisted Learning, 23*(5), 393–409. https://doi.org/10.1111/j.1365-2729.2007.00224.x
Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education, 33*(1), 75–99. https://doi.org/10.1007/s10212-017-0339-8
Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. *Studies in Educational Evaluation, 53*, 29–40. https://doi.org/10.1016/j.stueduc.2017.02.003
Senden, M., & Coertjens, L. (2024). The impact of peer assessment design on interpersonal processes: a systematic review. *Studies in Educational Evaluation, 83*, Article 101405. https://doi.org/10.1016/j.stueduc.2024.101405
Seroussi, D.-E., Peled, Y., Sharon, R., Rothschild, N., Halperin Barlev, O., Weissblueth, E., & Harpaz, G. (2023). A different Voice' in peer feedback: gender specificity in Students' willingness to provide peer feedback. Article 7. *Education Sciences, 13*(7). https://doi.org/10.3390/educsci13070654
Sluijsmans, D., Brand-Gruwel, S., van Merriënboer, J. J. G., & Martens, R. L. (2004). Training teachers in peer-assessment skills: effects on performance and perceptions. *Innovations in Education and Teaching International, 41*(1), 59–78. https://doi.org/10.1080/1470329032000172720
Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: what are they talking about? *Computers Education, 46*(1), 29–48. https://doi.org/10.1016/j.compedu.2005.04.002
Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: are they critical for feedback perceptions and efficiency? *Learning and Instruction, 20*(4), 291–303. https://doi.org/10.1016/j.learninstruc.2009.08.008
To, J., & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assessment & Evaluation in Higher. Education, 44*(6), 920–932. https://doi.org/10.1080/02602938.2018.1548559
Torres-Guijarro, S., & Bengoechea, M. (2017). Gender differential in self-assessment: a fact neglected in higher education peer and self-assessment techniques. *Higher Education Research Development, 36*(5), 1072–1084. https://doi.org/10.1080/07294360.2016.1264372
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychological Bulletin, 140*, 1174–1204. https://doi.org/10.1037/a0036620
Wu, Y., & Schunn, C. D. (2023). Passive, active, and constructive engagement with peer feedback: a revised model of learning from peer feedback. *Contemporary Educational Psychology*. , Article 102160. https://doi.org/10.1016/j.cedpsych.2023.102160
Zou, Y., Schunn, C. D., Wang, Y., & Zhang, F. (2018). Student attitudes that predict participation in peer assessment. *Assessment Evaluation in Higher Education, 43*(5), 800–811. https://doi.org/10.1080/02602938.2017.1409872

José Carlos G. Ocampo is a Postdoctoral Researcher at the Eberhard Karls University of Tübingen (Germany), where he works in the Teaching and Learning with Digital Media group at the Institute of Educational Science and the Hector Research Institute of Education Sciences and Psychology. He obtained his PhD in Education from the Universidad de Deusto in Bilbao (Spain). His research interests include using technology to enhance teaching and learning practices, and examining how individual differences influence feedback processes.

Ernesto Panadero is a Professor and Director of the Centre for Assessment Research, Policy and Practice in Education (CARPE) at Dublin City University (Ireland), and an Ikerbasque Professor as well as the international leader of the Education, Regulated Learning & Assessment group (ERLA) at the University of Deusto (Spain). His research applies educational psychology to the study of self-regulated learning, educational assessment, and the interplay between the two. He examines how formative assessment practices (e.g., self-assessment, peer assessment, the use of rubrics, and feedback) affect self-regulated learning, self-efficacy, and academic performance. In addition, he investigates the psychological, interpersonal, and social factors that shape and are shaped by these assessment practices.

David Zamorano is a PhD student at the University of Deusto (Spain) and a member of the ERLA research team. His research focuses on formative assessment, with particular emphasis on peer assessment, as well as feedback, rubrics, and self-assessment. He has extensive experience with experimental and quasi-experimental methods and employs process measures such as eye tracking, think-aloud protocols, and electrodermal activity to investigate the cognitive processes and emotional responses of students when engaging in assessment activities.

Iván Sánchez-Iglesias is a researcher in the Department of Psicobiología y Metodología en Ciencias del Comportamiento at the Faculty of Psychology, Universidad Complutense de Madrid (Spain). He works in the area of of Behavioural Science Methodology, and he is an active member of the research groups Bienestar psicológico, salud, envejecimiento y trabajo and Teamwork, Leadership and Healthy Organizations / Equipos de trabajo, liderazgo y organizaciones saludables. He holds a PhD from the Universidad Autónoma de Madrid. His research interests include psychological well-being, occupational health, team dynamics, leadership, and the promotion of healthy organizational environments.