

The Effects of Gender and Training on Peer Feedback Characteristics

José Carlos G. Ocampo¹, Ernesto Panadero², David Zamorano¹, Iván Sánchez-Iglesias³, &
Fernando Diez Ruiz¹

¹Faculty of Education and Sports, University of Deusto, Bilbao, Spain, ²Center for Assessment Research Policy and Practice (CARPE), School of Policy and Practice, Dublin City University, Ireland, ³Psicobiología y Metodología de las Ciencias del Comportamiento, Universidad Complutense de Madrid, Madrid, España

Recommended Citation

Ocampo, J.C.G., Panadero, E., Zamorano, D., Sánchez-Iglesias, I., & Diez, F. (2023). The effects of gender and training on peer feedback characteristics. *Assessment and Evaluation in Higher Education*, 1-17. <https://doi.org/10.1080/02602938.2023.2286432>

Funding: The first author of this project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska Curie grant agreement N° 847624. In addition, a number of institutions backed and co-financed his project. 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. The third author is funded by the Formación de Personal Investigador, reference PRE2020-092280.

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

This is a post-peer-review, pre-copyedit version of an article published in *Assessment and Evaluation in Higher Education*. The final authenticated version is available online at:

<https://doi.org/10.1080/02602938.2023.2286432>. This manuscript may not exactly replicate the published version due to editorial changes and/or formatting and corrections during the final stage of publication. Interested readers are advised to consult the published version. The authors declare to not having any conflict of interest regarding this manuscript.

Correspondence concerning this article should be addressed to José Carlos G. Ocampo, ERLA Research Group, Universidad de Deusto, Bilbao, España. Email: jc.ocampo@deusto.es

Abstract

Previous research has demonstrated the benefits of peer feedback for improving student work. Gender, as an individual characteristic, is now receiving increased attention due to its influence on the peer feedback process. This study examined the effects of gender and peer assessment training on the amount and content of peer feedback provided by assessors for poor, average and excellent writing samples, using a randomised controlled design. A total of 240 undergraduate psychology students ($n_{\text{Men}} = 120$) participated in the study. Half of the participants ($n_{\text{Men}} = 60$, $n_{\text{Women}} = 61$) received peer assessment training, while the other half received task instructions only ($n_{\text{Men}} = 60$, $n_{\text{Women}} = 59$). Participants were assigned to eight subgroups, providing peer feedback to writing samples attributed to fictitious male or female assessee. Analysis of 3,017 feedback segments revealed that women provided a greater amount of peer feedback compared to men. Women also offered more positive verifications and suggestive elaborations for average and poor writing samples, respectively. Additionally, male assessees received more suggestive elaborations, while trained assessors provided more positive verifications. These findings suggest the need for a multifaceted training programme to bridge the gap between gender-based differences in peer feedback characteristics.

Keywords: gender, peer feedback content, peer feedback amount, training, randomised-controlled design

The Effects of Gender and Training on Peer Feedback Content

Peer assessment is one instructional method instructors use to evaluate students' learning outputs (Alqassab et al., 2023). Peer assessment is an umbrella term for activities in which students judge and make decisions about their peer's performance, and instructors may opt to require students to give qualitative (e.g. peer feedback) and/or quantitative (e.g. peer scores) judgments of their classmate's work (Topping, 1998). When peer assessment is implemented for formative purposes, students may engage in interpersonal exchanges with their peers when they provide and receive feedback (Panadero, 2016; Panadero et al., 2023; van Gennip et al., 2009). Studies have suggested that for a more successful implementation of peer assessment, teachers can train students to be better assessors (Panadero et al., 2018).

It has also been reported that individual characteristics, such as gender, may affect peer assessment. For example, while some studies have reported that men and women are comparable in the amount of peer feedback they give (Cheng et al., 2014; Hamer et al., 2015), the content of their peer feedback may vary (Johnson & Roen, 1992; Leung et al., 2010; Noroozi et al., 2018, 2022). The characteristics of peer feedback provided by assessors are important, as assessees may tend to be selective in implementing peer feedback based on the gender of the assessor (Cao et al., 2019). Thus, more research is needed in this area to establish how gender may affect the peer feedback males and females give to same- or opposite-sex peers. Therefore, using a randomised controlled design, this study explored how peer assessment training may influence male and female assessor's peer feedback to male or female assessees' writing samples.

Peer Feedback and Its Characteristics

Peer feedback has been applied in a wide range of disciplines and levels due to its potential to enhance learning (Liu & Carless, 2006). Importantly, peer feedback has been demonstrated to have a positive influence on cognitive and non-cognitive outcomes (Double

et al., 2020; Huisman et al., 2019; Usher, 2023). During peer feedback, assessors provide qualitative information about the assessee's performance, which assessees may use to improve their performance (Panadero et al., 2018). Thus, to reap the benefits of peer feedback, it is important to look at the structure of peer feedback.

At the surface level, one of the first things that assessees notice from the peer feedback provided to them is its amount. Studies usually measure the amount of peer feedback via the number of words that a peer feedback message contains or the number of segments that can be extracted from the peer feedback (Patchan et al., 2016; Zong et al., 2021). It is important to investigate the amount of peer feedback that students provide, because studies have shown that it may relate to assessor's engagement (Jin et al., 2022), assessor's subsequent peer feedback performance (Zong et al., 2021), and assessee's peer feedback uptake (Wu & Schunn, 2020). Given this, peer assessment support strategies have been offered in web-based peer assessment platforms, where many peer assessment activities now take place (Ocampo & Panadero, 2023). Some require a minimum number of words in peer feedback messages, while some incentivise students who provide lengthier peer feedback (Patchan et al., 2018).

Although providing lengthy peer feedback may be a sign of behavioural engagement (e.g., Jin et al., 2022), we echo Zong et al.'s (2021) comment that requiring a minimum word count might force assessors to provide superficial peer feedback just to reach the word count or to attain the grade incentive. Also, lengthier peer feedback may not necessarily translate into improvement at all times, because students may experience overload when receiving verbose feedback (Berndt et al., 2022; Patchan et al., 2016). While investigating the amount of peer feedback may have some limitations, the evidence reviewed shows its potential to improve peer assessment processes and outcomes.

The contents of peer feedback messages are also critical. In writing-focused tasks, several studies have attempted to deconstruct peer feedback content by developing coding schemes with detailed categories and descriptions. Among others, studies have attempted to unpack the content of peer feedback for argumentative essays (e.g., Nelson & Schunn, 2009), academic writing tasks (e.g., Gielen & De Wever, 2015a), expository essays (e.g. Cho et al., 2006) and analytical writing tasks (e.g. Wu & Schunn, 2020). These studies demonstrate the importance of disentangling students' peer feedback to ascertain features that can potentially affect performance. Some of these coding schemes suggest that effective feedback should contain verification and/or elaboration components (Gielen & De Wever, 2015a; Kulhavy & Stock, 1989; Narciss, 2008). Verification refers to feedback that contains a dichotomous evaluation of whether a work is correct or incorrect, while elaboration refers to feedback that further helps the student by providing additional information to improve his or her work (Hattie & Gan, 2010). The work of Gielen and De Wever (2015a) further extended this in peer feedback settings, where they included the types of peer feedback based on a given peer feedback style. For verification, they mentioned that it can be positively, negatively or neutrally phrased, while elaboration can contain informative or suggestive statements. Although feedback with verification or elaboration components may differ in content and structure, both have been documented as helping to improve students' writing performance in peer feedback settings (Strijbos et al., 2010). Given this, it is important to understand how students provide peer feedback, as it may have an effect on performance.

Apart performance-related outcomes, these peer feedback coding schemes also show that certain variables (e.g. peer assessment design, prior experience, personal characteristics) may potentially influence the content of the peer feedback that students provide (Cho et al., 2006; Gielen & De Wever, 2015b, 2015a; Nelson & Schunn, 2009). However, gender, a variable that has been documented to potentially affect writing performance and peer

feedback needs to be further explored given that gender difference may exist between men and women in writing abilities (Voyer & Voyer, 2014) and peer feedback uptake (Wu & Schunn, 2020).

Gender Effects on Peer Feedback Characteristics

Increasing evidence on gender's influence on peer assessment has been documented in one of the largest reviews in peer assessment (Alqassab et al., 2023). For instance, studies found no difference when comparing men and women in terms of the amount of peer feedback they provide (Cheng et al., 2014; Hamer et al., 2015). A recent study with a predominantly women sample found that when the assessor's gender differed from that of the assessee, the assessor provided less and less implementable peer feedback (Wu & Schunn, 2021). Although it should be noted that the peer feedback activity was conducted anonymously in this study. Studies have also found that female assesseees are more likely to implement peer feedback, especially when it was mentioned multiple times by multiple peers (Wu & Schunn, 2020). Some studies have also suggested that men may tend to be inattentive to verbose feedback (Narciss et al., 2014). Given this, the amount of peer feedback provided is important since it may have an impact on peer feedback implementation.

Looking at the effects of gender on peer feedback content, Johnson and Roen (1992) considered the patterns of compliment use among graduate students. Their findings revealed that women were more likely to use intensifiers (e.g. "I *really*..." "It's *very*..."), personal references (e.g. use of first-person pronouns) and compliment pairing strategies (e.g. opening compliments + criticisms and suggestions + closing compliments) than men. On the other hand, the study by Leung et al. (2010) examined gender difference in peer feedback in relation to the peer feedback model developed by Nelson and Schunn (2009). They found that men and women assessors were able to localise and explain specific parts of an output. Specifically, men were more inclined to do this for correct areas in the output, whereas

women were more likely to do this for problematic areas of the output. Additionally, both genders showed the ability to offer solutions, but women were more likely to give direct error corrections and suggestions, while men provided more specific suggestions. In terms of feedback scope and use of affective language, Leung et al. (2010) found that both genders tended to give general comments, but women were more likely to give specific comments, while men were more inclined to give praise feedback. Noroozi et al. (2018) looked at how both genders differed in terms of the quality of their argumentative feedback. They found that women provided higher quality peer feedback in terms of providing an intuitive opinion and justifications than males. This difference in peer feedback supporting female assessors was again found in a recent study, which demonstrated that women provide better peer feedback than men, because they tended to provide better justifications for problematic areas in the essay and delivered constructive and feedforward-type feedback (Noroozi et al., 2022).

Altogether, these findings suggest that men and women may differ in the characteristics of the peer feedback they provide to assessees, and female assessors may tend to provide more detailed and suggestive peer feedback than male assessors. Thus, it is important to look at ways to bridge the gap between differences in peer feedback content between men and women, because it may be a cause of gender issues in classrooms. To illustrate, Cao et al. (2019) found that male assessees were unwilling to implement peer feedback from females, because they doubted its quality, while female assessees felt that they learned less from receiving peer feedback from males as it contained too much praise. Therefore, crucial to monitor students' peer feedback, as well as devise interventions to minimise potential gender influences on peer feedback to reduce possible deterrents of peer assessment.

Peer Assessment Training

One challenge noted about peer feedback is the underdeveloped ability of students to give accurate and quality judgements of peer's work (e.g., Liu & Carless, 2006). To mitigate such challenges, scaffolds to develop students' peer feedback skills have been suggested by previous studies (Panadero et al., 2018). Several frameworks have also been developed to train students' peer feedback skills (e.g., Lam, 2010). Often, these training frameworks include discussion and demonstration by the instructor on how to give peer feedback, as well as the use of different assessment scaffolds. Studies have demonstrated that these carefully planned scaffolds improve students' peer feedback skills (Gielen & De Wever, 2015b, 2015a; Sluismans et al., 2004).

Studies examining the effects of peer assessment training on gender have received little attention, even if prior studies have suggested examining the potential of training for reducing gender issues during peer assessment (e.g., Ocampo et al., 2023). A study that looked at the effects of peer assessment training in minimising gender issues in peer scoring contexts found that even after receiving an extensive peer assessment training, women tended to be biased against female assessees and favoured male assessees more by giving higher peer scores to men, while men tended to be biased against male assessees and favoured female assessees more by giving higher peer scores to women (Aryadoust, 2016). Thus, there is thus a need to examine the potential of peer assessment training in improving men and women's peer feedback due to documented differences.

Aim and Research Questions

This publication is part of a larger study exploring the effects of gender on peer assessment. In this publication, our aim is to investigate the effects of gender and training on the amount and content of peer feedback. Investigating the effects of assessor gender (men vs women), assessee gender (men vs women) and peer assessment training (with vs without) on

the amount of peer feedback and the contents of peer feedback will help to reveal the influence of gender in the peer feedback process. Hence, our research questions are as follows:

RQ1: Does assessor gender, assessee gender and training affect the amount of peer feedback?

RQ2: Does assessor gender, assessee gender and training affect verification peer feedback type?

RQ3: Does assessor gender, assessee gender and training affect elaboration peer feedback type?

Method

Participants

The participants of this study were 240 second year undergraduate psychology students enrolled in research writing courses (i.e. experimental research and qualitative research) from two campuses of a private university in the Philippines. There were 120 men and 120 women (based on the gender they reported when they enrolled to the university) with an age range from 18 to 27 years ($M = 20.8$; $SD = 1.16$). To establish the training conditions, 121 students (60 men, 61 women) received peer assessment training, while 119 students (60 men, 59 women) did not receive peer assessment training. We created two identical classes (one for each training condition) with a total of eight subgroups in the online peer assessment platform. Participants were randomly divided into the subgroups in which they gave peer feedback to fictitious peers of the same or the opposite gender (see Figure 3 for subgroup distribution). In all conditions and subgroups, participants were unaware that they were giving peer feedback to fictitious male and female assessees. Regardless of the training condition, 61.7% of the participants reported that they had experienced giving peer feedback in other courses (i.e. 51.4% in the peer assessment training condition and 48.6% in the no peer assessment training condition). Data were collected in the first and second semester of two academic years (i.e. 2021–2022 and 2022–2023). Sensitivity analysis was conducted using the G*Power software (v. 3.1.9.7; Faul et al., 2009) for the Wilcoxon–Mann–Whitney test for two groups, with an alpha of .05, a power of .80 and sample size of 120 for Group 1 and a sample size of 120 for Group 2. The results showed that the minimum effect size that should exist in our sample to get a significant result is $d = 0.372$.

Materials

Writing Samples

The participants provided peer feedback to three writing samples, which were of poor, average or excellent quality, based on the rating of three independent experts (i.e. a psychology instructor, educational assessment expert and academic writing instructor) using the same rubric that the participants used in assessing the writing samples, with good agreement ($ICC = .984$). To increase the ecological validity of the study, writing samples from former students in past research courses were provided by the instructors (see Supplementary Material 1).

Writing Rubric

Participants used a rubric for research projects recommended by the psychology department where the data collection took place. In this study, assessors in the training condition were told to use the rubric as a guide in composing their peer feedback (see Supplementary Material 2).

Online Peer Assessment Platform

Students provided peer feedback via the Eduflow platform, a web-based educational platform that has peer and self-assessment as one of its major functions. Each student was assigned one account to perform the activity (see Figure 1).

Figure 1*Participant Page in Eduflow.*

1) Participant Activity Page

Peer Assessment of Research Papers

Welcome!

Literature Review Drafts

- 1 First Paper
- 2 Second Paper
- 3 Third Paper

Writing Samples

1 First Paper

In this activity, you will peer score and give peer feedback to a writing draft written by one of your peers from another block.

Your peer's name was substituted with his gender to maintain his privacy.

Download it and read it carefully.

IF YOU CANNOT DOWNLOAD THE FILE, HERE'S A GOOGLE DOCS LINK VERSION:

MIC1-2022 .DOCX

MALE STUDENT PAPER

Fictitious Assessee Gender

2) Participant Peer Feedback Page

Peer Feedback Prompts

QUESTION 1

What did this student do well in his or her literature review draft?

the paper is well written

QUESTION 2

What can this student improve on in his or her literature review draft?

the student should focus more checking his spellings, in addition his reference does not really makes sense and comes from unreliable source like wikipedia

Sample Peer Feedback From an Assessor

Measures

Amount of Peer Feedback

The amount of peer feedback provided was calculated using two approaches. The first approach was by computing the total word count assessors provided for the three writing samples, which ranged from 13 to 1,007 words ($M = 183.93$; $SD = 153.57$). The second was by computing the total segment count, which represented the total number of segments that the assessor included in their peer feedback for the three writing samples after it was divided into syntactical units. Total segment count in the present sample ranged from 5 to 57 segments ($M = 12.57$; $SD = 8.54$).

Peer Feedback Content

The peer feedback style and type categories from Gielen and De Wever (2015a) were used to code the peer feedback in this study. Peer feedback style has three categories: verification peer feedback style, elaboration peer feedback style and general peer feedback style. Verification peer feedback style is further divided to three peer feedback types: positive, negative and neutral. Similarly, the elaboration peer feedback style is further divided to two peer feedback types: informative and suggestive. Finally, the general peer feedback style encompasses statements that do not contain verification or elaboration characteristics. Given the nature of the general peer feedback style, we opted not to include it in the analysis. Supplementary Material 3 shows the coding scheme used, as well as the respective definitions and examples from our data.

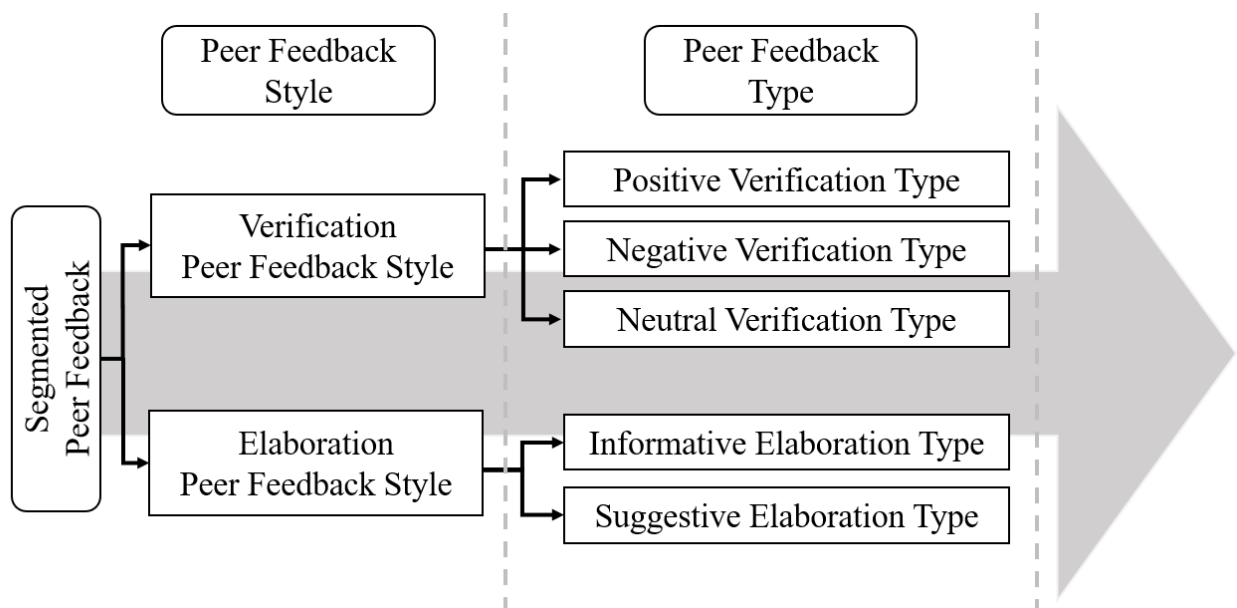
Based on the recommendations of Strijbos et al. (2006), our peer feedback data were segmented into syntactical units, which yielded 3,017 segments for content analysis. In terms of the reliability of the segmentation process, random peer feedback messages were segmented by the first and third author to become acquainted with the process and the peer feedback messages, which resulted in a proportion agreement of 93% (124 of 133 segments). The peer feedback content coding scheme was then collaboratively discussed by the authors, which proceeded with an independent coding of the segmented peer feedback. Each segment was coded for the presence (1) or absence (0) of peer feedback categories based on the definitions in our coding scheme. Deliberation about the coding strategies for each category was done to clarify the understanding of the categories and the coding process. Subsequently, a total of 455 random segments were double coded for interrater reliability. All Cohen's Kappa values were above the conventional benchmarks of .80, which suggests reliability between the two coders (Landis & Koch, 1977; see Supplementary Material 3). It is important to note that the peer feedback coding strategy involved an ongoing process in

which each segment could be categorised as belonging to either the verification or the elaboration peer feedback style, leading to their respective peer feedback type categories.

Figure 2 illustrates the peer feedback coding strategy performed.

Figure 2

Peer Feedback Coding Strategy



Participants' Academic Ability

To address the potential differences in ability among participants when giving feedback, participants' cumulative grade point average (GPA) was obtained from the university in adherence with the university's data privacy guidelines. The grading scale used by the university ranges from 1.0 (lowest) to 4.0 (highest). In this study, the participants' GPA averaged at 3.19 (SD = .37).

Procedure

Figure 3 summarises the study procedure flow. Research writing courses were chosen because students had to complete a group research paper as a major requirement in their respective courses. As instructors used peer assessment as one of their assessment strategies to assess the group research paper, the students were informed that they would receive peer feedback training via web-conferencing delivered by the first author. Participants in the peer

assessment training condition underwent the modelling and exploring phases of Lam's (2010) peer assessment workshop, which included a discussion about the purpose of peer feedback and a demonstration on how to give feedback. Participants then practised giving peer feedback on a writing sample and discussed their judgements collectively to clarify any misunderstandings about the process. Finally, instructions were given on how to submit their peer feedback in Eduflow. The entire peer assessment training lasted for approximately 2 hours and 30 minutes. Participants in the no peer assessment training condition only received task instructions and a demonstration on how to submit their peer feedback in Eduflow, which took around 30 minutes. Both the peer assessment training and no peer assessment training lesson guides we followed in this study are uploaded as Supplementary Material 4. The study was conducted from week 7 through week 10 of a 13-week semester. Regardless of the training condition, participants answered the following feedback prompts to elicit their 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?

Participants were informed that they would individually assess literature review drafts by peers from other research classes. To set up the assessee gender manipulation, participants were told that the assessee's name had been replaced with his or her gender to maintain their privacy and for the instructor's identification. Participants were told that their identities would remain anonymous and that they were not allowed to share the topics of the writing samples with anyone to maintain the confidentiality of the process. Instructors told the participants to complete the activity within two weeks. After everyone had submitted their peer feedback, the researcher debriefed the participants about the design of the study, and participants assigned to the no peer assessment training condition were also given the same

peer assessment training. The peer assessment design characteristics of the present study have been uploaded as Supplementary Material 5.

Research Design and Data Analysis

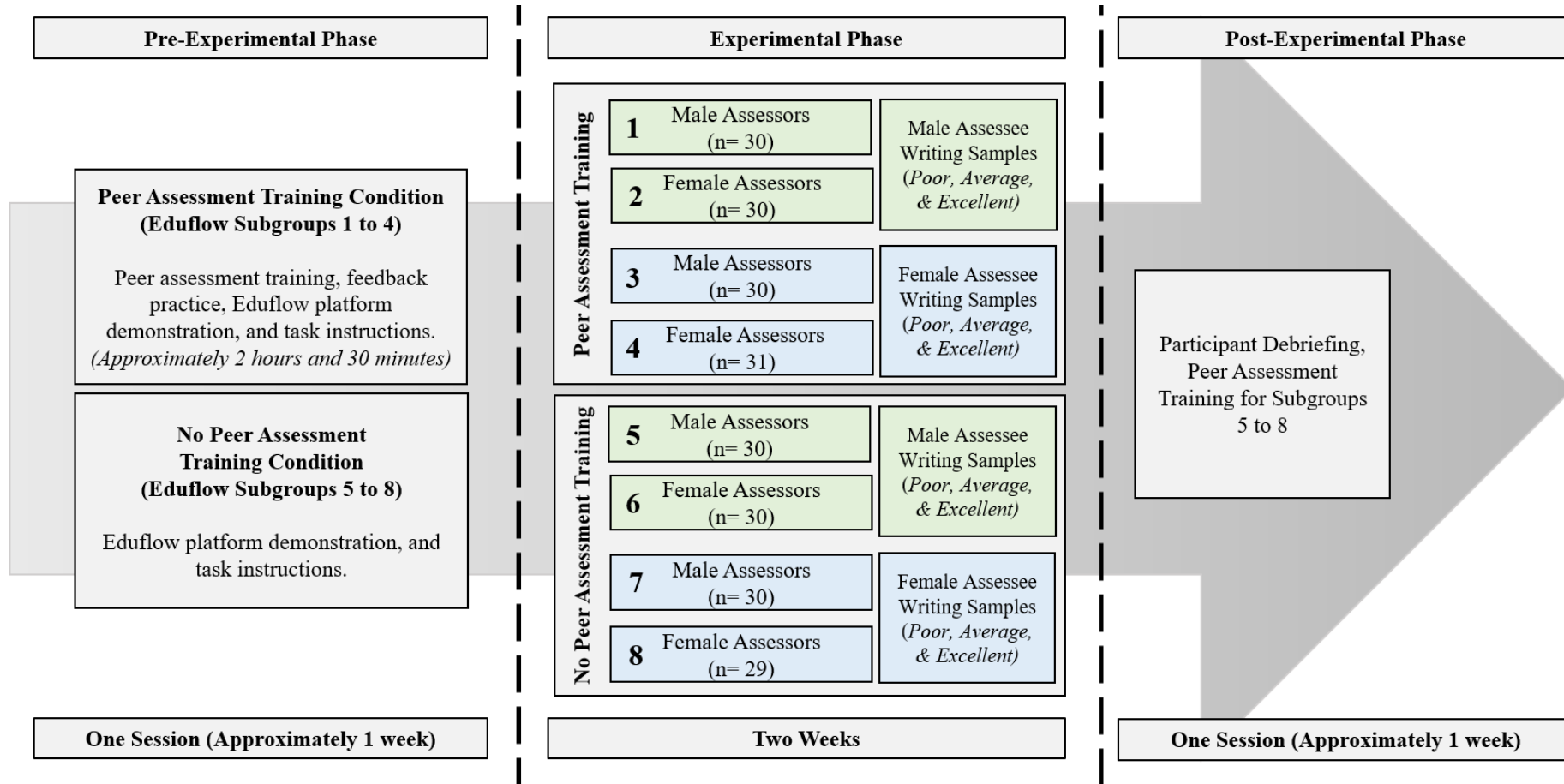
In this study, we used a $2 \times 2 \times 2$ randomised controlled design with three independent variables: assessor gender (men vs women), fictitious assessee gender (man vs woman) and training condition (with vs without). Our dependent variables were: amount of peer feedback provided – (1) total word count and (2) total segment count; the proportion of verification peer feedback types – (3) positive, (4) negative and (5) neutral; and elaboration peer feedback types – (6) informative and (7) suggestive.

Because we were interested in comparing the effects of assessor gender, assessee gender and peer assessment training on peer feedback content, we opted to use the proportion of peer feedback content for each assessor and per writing sample as the main dependent variable. Proportions were computed by dividing the raw count of a specific peer feedback content dimension to the segment count per writing sample, and we used the raw counts to present descriptive information about the content of participants' peer feedback for each dimension.

As academic ability (i.e., measured via GPA) could potentially influence peer feedback content, we conducted an independent *t*-test and we found that participants' academic ability was comparable for each independent variable, with *p*-values ranging from .151 to .873 (see Supplementary Material 6 for more information). Because the word count, segment count and the proportions in each of the peer feedback content dimensions were not normally distributed, we used the Mann–Whitney U non-parametric test in SPSS software version 28.

Figure 3

Study Procedure Flowchart



Results

RQ 1. Does assessor gender, assessee gender and training affect the amount of peer feedback?

Table 1 provides a summary of the descriptive statistics and Mann–Whitney U values pertaining to the amount of peer feedback received. The Mann–Whitney U test showed that there was a significant difference between male and female assessors' total word count for the three writing samples, $U = 4312.5$, $z = -5.369$, $p < .001$, $r = -.35$. More specifically, female assessors had a greater total number of words in their peer feedback than male assessors did. On the other hand, a non-significant result was found for male and female assessees, which indicates that the fictitious male and female assessees received the same amount of words in their peer feedback. Finally, the total word count of the trained assessors was not significantly different from that of the untrained assessors.

Similar to the results for word count, male and female assessors differed in the total number of segmented peer feedback per assessor $U = 3944.5$, $z = -6.10$, $p = < .001$, $r = -.39$ (see Table 1). Once again, female assessors' peer feedback on the three writing samples had more total segments than did the male assessors. Also, the results were non-significant for male and female assessees, which means that both assessee genders received a comparable total segment frequency. Finally, the number of segments of peer feedback between trained and untrained assessors was comparable.

Table 1

Amount of Peer Feedback

	Groups	Amount of Peer Feedback							
		M	SD	Min	Max	<i>U</i>	<i>z</i>	<i>p</i>	<i>r</i>
Total	Male Assessor	140.36	132.31	13	1007	4312.5	-5.36	<.001***	-0.34
Word Count	Female Assessor	227.51	161.34	30	854				
	Male Assessee	184.7	155.7	22	1007	7186	-0.02	0.979	-0.06
	Female Assessee	183.21	152.1	13	854				
	Training	190.8	148.9	13	854	6505.5	-1.29	0.191	-0.08
	No Training	176.91	158.5	17	1007				
	Male Assessor	10.13	7.15	5	57	3944.5	-6.08	<.001***	-0.40

Total	Female Assessor	15.01	9.13	6	51				
Segment	Male Assessee	12.40	8.30	6	57	6981.5	-0.40	0.683	-0.02
Count	Women Assessee	12.74	8.80	5	51				
	Training	13.23	8.34	6	51	6176.5	-1.91	0.056	-0.12
	No Training	11.90	8.70	5	57				

*** $p < .001$

RQ2. Does assessor gender, assessee gender and training affect verification peer feedback type?

Results for verification peer feedback type are shown in Table 2 for the poor writing sample, Table 3 for the average writing sample and Table 4 for the excellent writing sample. The proportion of positive verifications between male and female assessors in the poor and excellent writing sample did not show significant differences. However, male and female assessors differed in the proportion of their positive verifications provided in the average writing sample, $U = 6024.5$, $z = -2.51$, $p = .012$, $r = -.150$, where female assessors provided more positive verifications than male assessors did. No significant differences were observed between male and female assesseees in the three writing samples. In terms of the effects of training, a non-significant result was observed between trained and untrained assessors in the poor and excellent writing samples. However, trained assessors tended to provide more positive verifications in the average writing sample than did untrained assessors, $U = 6117.5$, $z = -2.31$, $p = .021$, $r = -.150$.

The proportion of negative verifications were comparable for male and female assessors and were similar for the three writing samples. Similarly, assessee gender was not a factor when receiving negative verifications for the three writing samples. Training was also a non-significant factor when giving negative verifications in the three writing samples. Results for the proportion of neutral verifications also showed that male and female assessors were similar in their neutral verifications in three writing samples. Similarly, non-significant findings were also observed for the three writing samples in terms of assessee gender.

Finally, training was not a significant factor when giving neutral verifications for the three writing samples.

RQ3. Does assessor gender, assessee gender and training affect elaboration peer feedback type?

Results for elaboration peer feedback type are shown in Table 2 for the poor writing sample, Table 3 for the average writing sample and Table 4 for the excellent writing sample. Male and female assessors were comparable in the proportion of informative elaborations in the peer feedback provided for the three writing samples. Likewise, assessee gender was not a factor when receiving informative feedback for the three writing samples. In terms of training condition, there were no significant differences between trained and untrained assessors for informative elaborations in the feedback to the three writing samples.

Regarding the proportion of suggestive elaborations, significant differences were seen for the poor writing sample. Specifically, female assessors provided significantly more suggestive feedback than did male assessors for the poor writing sample, $U = 5767.0$, $z = -2.70$, $p = .007$, $r = -.174$. However, this was not the case for the average and excellent writing samples. Similarly, male and female assesseees received the same amount of suggestive feedback for the poor and excellent writing samples, but male assesseees received a significantly greater proportion of suggestive feedback for the average writing sample than did female assesseees, $U = 6039.0$, $z = -2.20$, $p = .027$, $r = -.144$. Finally, the suggestive feedback of trained and untrained assessors was comparable in the three writing samples.

Overall, this shows that, across the three writing samples, male and female assessors provided the same amount of informative feedback to male and female assesseees. However, female assessors provided more suggestive elaboration feedback to the poor sample than did male assesseees, while male assesseees were more likely to receive suggestive feedback on the average sample than female assesseees. Additionally, trained and untrained assessors provided

the same proportion of informative and suggestive elaboration feedback to the poor, and average and excellent writing samples.

Table 2*Verification and Elaboration Peer Feedback for the Poor Writing Sample*

Groups		Poor Writing Sample							
		M	SD	Min	Max	<i>U</i>	<i>z</i>	<i>p</i>	<i>r</i>
Positive Verification	Male Assessor	0.24	0.56	0	3	6416.5	-1.90	.057	-.122
	Female Assessor	0.38	0.62	0	3				
	Male Assessee	0.32	0.62	0	3	7147.5	-.128	.898	-.008
	Female Assessee	0.30	0.57	0	3				
	Training	0.36	0.64	0	3	7018.0	-.441	.659	-.028
	No Training	0.27	0.55	0	3				
Negative Verification	Male Assessor	0.07	0.31	0	2	6843.0	-1.45	.146	-.093
	Female Assessor	0.18	0.96	0	10				
	Male Assessee	0.20	0.98	0	10	6836.5	-1.48	.139	-.095
	Female Assessee	0.05	0.21	0	1				
	Training	0.13	0.93	0	10	6821.0	-1.54	.123	-.010
	No Training	0.12	0.37	0	3				
Neutral Verification	Male Assessor	0.03	0.18	0	1	7020.5	-1.35	.177	-.087
	Female Assessor	0.01	0.09	0	1				
	Male Assessee	0.02	0.12	0	1	7138.5	-.462	.644	-.029
	Female Assessee	0.03	0.15	0	1				
	Training	0.03	0.18	0	1	7022.5	-1.33	.183	-.085
	No Training	0.01	0.09	0	1				
Informative Elaboration	Male Assessor	1.58	1.81	0	12	6375.0	-1.55	.120	-.100
	Female Assessor	2.23	2.71	0	17				
	Male Assessee	1.80	2.09	0	13	6587.0	-1.15	.248	-.074
	Female Assessee	2.02	2.54	0	17				
	Training	2.09	2.48	0	17	6748.5	-.849	.396	-.548
	No Training	1.72	2.15	0	13				
Suggestive Elaboration	Male Assessor	1.58	1.92	0	15	5767.0	-2.70	.007**	-.174
	Female Assessor	2.60	2.08	0	15				
	Male Assessee	2.04	2.05	0	15	7187.5	-.024	.981	-.001
	Female Assessee	2.13	2.09	0	15				
	Training	2.17	1.80	0	10	7086.0	-.215	.830	-.013
	No Training	2.00	2.32	0	15				

** $p < .01$

Table 3*Verification and Elaboration Peer Feedback for the Average Writing Sample*

Groups		Average Writing Sample							
		M	SD	Min	Max	<i>U</i>	<i>z</i>	<i>p</i>	<i>r</i>
Positive Verification	Male Assessor	0.39	0.77	0	5	6024.5	-2.51	.012*	-.162
	Female Assessor	0.80	1.06	0	4				
	Male Assessee	0.54	0.96	0	5	6752.5	-.955	.339	-.061
	Female Assessee	0.65	0.94	0	4				
	Training	0.72	0.97	0	4	6117.5	-2.31	.021*	-.150

	No Training	0.47	0.91	0	5				
Negative Verification	Male Assessor	0.05	0.25	0	2	7141.5	-.331	.741	-.021
	Female Assessor	0.05	0.31	0	3				
	Male Assessee	0.05	0.31	0	3	7141.5	-.331	.741	-.021
	Female Assessee	0.05	0.25	0	2				
	Training	0.08	0.37	0	3	6908.5	-1.64	.100	-.105
	No Training	0.02	0.12	0	1				
Neutral Verification	Male Assessor	0.01	0.09	0	1	7199.5	-.006	.995	0.00
	Female Assessor	0.01	0.09	0	1				
	Male Assessee	0.02	0.12	0	1	7080.0	-1.41	.156	-.091
	Female Assessee	0.00	0.00	0	0				
	Training	0.00	0.00	0	0	7078.5	-1.43	.153	-.092
	No Training	0.02	.129	0	1				
Informative Elaboration	Male Assessor	1.47	1.70	0	13	6832.5	-.699	.484	-.045
	Female Assessor	2.02	2.20	0	12				
	Male Assessee	1.61	1.71	0	8	6721.0	-.911	.362	-.058
	Female Assessee	1.59	1.71	0	13				
	Training	1.89	2.16	0	13	6692.5	-.965	.335	-.062
	No Training	1.59	1.73	0	11				
Suggestive Elaboration	Male Assessor	1.16	1.50	0	13	6792.0	-.775	.438	-.050
	Female Assessor	1.84	1.86	0	13				
	Male Assessee	1.88	2.19	0	13	6039.0	-2.20	.027*	-.142
	Female Assessee	1.41	1.66	0	13				
	Training	1.36	1.20	0	7	6248.5	-1.80	.071	-.116
	No Training	1.64	2.06	0	13				

* $p < .05$ **Table 4***Verification and Elaboration Peer Feedback for the Excellent Writing Sample*

		Excellent Writing Sample							
Groups		M	SD	Min	Max	<i>U</i>	<i>z</i>	<i>p</i>	<i>r</i>
Positive Verification	Male Assessor	0.57	0.99	0	6	7024.0	-.366	.714	-.023
	Female Assessor	0.64	0.90	0	5				
	Male Assessee	0.54	0.85	0	5	6896.0	-.632	.527	-.040
	Female Assessee	0.67	1.03	0	6				
	Training	0.74	1.11	0	6	6549.5	-1.35	.176	-.087
	No Training	0.47	0.72	0	4				
Negative Verification	Male Assessor	0.03	0.20	0	2	7140.0	-.451	.652	-.029
	Female Assessor	0.03	0.15	0	1				
	Male Assessee	0.04	0.23	0	2	7022.0	-1.33	.181	-.085
	Female Assessee	0.01	0.09	0	1				
	Training	0.03	0.18	0	1	7023.0	-1.32	.185	-.012
	No Training	0.02	0.18	0	2				
Neutral Verification	Male Assessor	0.02	0.12	0	1	7080.0	-1.41	.156	-.091
	Female Assessor	0.00	0.00	0	0				
	Male Assessee	0.01	0.09	0	1	7200.0	.000	1.00	0.00
	Female Assessee	0.01	0.09	0	1				
	Training	0.01	0.09	0	1	7198.5	-.012	.991	0.00
	No Training	0.01	0.09	0	1				
Informative Elaboration	Male Assessor	1.47	1.66	0	10	6345.0	-1.61	.107	-.103
	Female Assessor	2.30	1.80	0	10				
	Male Assessee	1.79	1.75	0	10	6692.5	-.958	.338	-.061

	Female Assessee	1.97	1.82	0	10				
	Training	2.01	1.91	0	10	6774.0	-.803	.422	-.051
	No Training	1.76	1.63	0	10				
Suggestive Elaboration	Male Assessor	.83	.941	0	6	6498.0	-1.33	.181	-0.85
	Female Assessor	1.43	1.36	0	7				
	Male Assessee	1.23	1.22	0	6	6279.0	-1.75	.079	-.112
	Female Assessee	1.03	1.20	0	7				
	Training	1.07	1.03	0	6	6537.0	-1.26	.207	-.081
	No Training	1.18	1.36	0	7				

Discussion

We investigated the effects of assessor gender, assessee gender and peer assessment training on the amount and content of peer feedback using a randomised controlled design. In this study, trained and untrained male and female assessors gave peer feedback to poor, average and excellent writing samples by fictitious male and female assessees. Next, we discuss the results of our three research questions.

Effects on the Amount of Peer Feedback

In our RQ1, we investigated if gender and peer assessment training had an effect on the amount of peer feedback provided. We found that female assessors wrote more words and had more segments in their peer feedback. This finding is in contrast with prior evidence that men and women did not differ in the amount of peer feedback they provided (Cheng et al., 2014; Hamer et al., 2015). The contrasting result could be because of the difference in the gender distribution in our studies. To illustrate, the study by Cheng et al. (2014) involved a women dominant sample (58 women and 7 men), while the study by Hamer et al. (2015) involved a men dominant sample (161 men and 47 women); our study had an equal number of men (120) and women (120). Although gender was not the central variable explored in their studies, our findings provide insight into women's tendency to provide more peer feedback than men. Additionally, while our study and the studies by Cheng et al. and Hamer et al., were conducted in web-based environments, our finding resonates with existing evidence suggesting that women may provide more messages in web-based environments

than men (Caspi et al., 2008). This suggests that it is important to consider gender when implementing web-based peer assessments, as women may provide a greater amount of peer feedback than men. Another important finding is that our assesseees received the same amount of peer feedback, regardless of whether they were a man or a woman. This is an encouraging finding, given that earlier studies found that assessee's gender may be a source of bias in peer scoring (Aryadoust, 2016). While peer scoring is important in higher education due to its positive influence on performance (Double et al., 2020), our study sheds light on the fact that male and female assesseees are likely to receive the same amount of peer feedback from same-sex/opposite-sex assessors.

When it comes to the effects of training, trained and untrained assessors were similar in the amount of peer feedback provided, which is contrary to other findings that examined training's effects on peer feedback length (Darvishi et al., 2022). This finding may be attributed to two reasons. First, they may be credited to the short duration of the peer assessment training, which was only delivered for one session of about two and a half hours. Second, it might be due to the fact that all assessors received peer assessment scaffolds (i.e. a rubric, prompts) in the peer assessment platform, regardless of whether they received the training or not. These tools have previously been shown to have positive effects on peer assessment outcomes (Gielen & De Wever, 2015a; Panadero et al., 2016); therefore, the presence of rubrics and prompts might have reduced the effect of the training. Both of these reasons resonate with the findings of our earlier study, which found no difference in the peer scores provided by male and female assessors due to the duration of training and the availability of rubrics in the peer assessment platform (Ocampo et al., 2023).

Effects on Verification Peer Feedback Type

In our RQ2, we investigated if gender and training had an effect on positive, negative and neutral verification peer feedback types. We found that female assessors were more

likely to give positive verifications than male assessors for the average writing sample. This finding adds value to earlier studies that reported that women are more likely to give positive feedback than men (Johnson & Roen, 1992), and it contradicts earlier findings that found men provided more positive feedback than women (Leung et al., 2010) or findings that found no difference between men and women (Hamer et al., 2015). The proportions of negative and neutral verifications provided by male and female assessors did not differ significantly. Additionally, male and female assesseees received similar proportions of neutral and positive verifications, further corroborating previous findings (Hamer et al., 2015; Leung et al., 2010). We also found that trained assessors provided more positive verifications for the average writing sample.

Be it positively, negatively or neutrally phrased, peer feedback content can affect students intrapersonally and interpersonally (or vice-versa; Panadero et al., 2023). The task of crafting peer feedback not only requires cognitive abilities, but also a careful use of affective language (Nelson & Schunn, 2009). A hypothesis to be tested is that female assessors in our study may have expressed more positively phrased peer feedback for the average writing sample due to a possible female tendency to be more emotionally expressive than men (Chaplin, 2015), while male assessors may have restricted their emotional expression in their peer feedback due to socio-cultural values attached to masculinity, which was found to be strongly valued in the culture of where the study took place (i.e. the Philippines; Hofstede et al., 2010). Moreover, training may only have affected the average writing sample with regards to positive verifications, because the poor and excellent writing samples had noticeable differences in quality. The poor writing sample had obvious problems in quality, while the excellent writing sample had obvious merits in quality. Trained assessors may have capitalised on the training they received to detect the positive aspects of the output. Although there were prompts available in the peer assessment platform and the rubric was a department

recommended tool that the students use in their other courses, untrained assessors may have struggled to use these tools to detect the positive aspects of the average sample, because they were not thoroughly trained on how to use them.

Why was there no variation in the proportion of negative and neutral verification peer feedback types? We hypothesise that this might be due to cultural influences on peer assessment. Our sample belong to a highly collective society where smooth interpersonal relations are valued more highly than confrontations (Lajom et al., 2009). Therefore, our participants may have avoided providing negative or neutral verifications as it might create interpersonal conflicts with peers, although trained assessors were taught to provide constructive peer feedback (i.e. explaining errors and providing suggestions) and all participants were assured that their identities were anonymous. This pattern is similar to the findings of Carson and Nelson (1996) with Chinese students (i.e. another collectivist society), who exhibited a reluctance to give feedback, particularly negative and critical remarks, preferring instead to provide positive feedback to maintain social harmony.

Effects on Elaboration Peer Feedback Type

In our RQ3, we investigated the effects of gender and training on informative and suggestive elaboration peer feedback types. We found that male and female assessors did not differ in the proportion of the informative elaborations for the three writing samples, which is similar to what Noroozi et al. (2022) found in the identification aspect of their coding scheme. In the same light, male and female assesseees were likely to receive the same amount of informative elaborations, while trained and untrained assessors were also able to provide the same proportion of informative elaborations. Once again, the prompts and rubric available in the peer assessment platform may have played an instrumental role in assisting assessors in crafting informative peer feedback regarding what went well or wrong in the writing sample. Also, these non-significant results might be due to the comparable peer feedback experience

of the participants, as more than half of the participants had experienced giving peer feedback in previous courses. Untrained assessors may have used their prior experience in providing feedback to inform the writer of the paper of the correct or incorrect areas in the three writing samples.

For the suggestive elaboration type, women provided more suggestive elaborations for the poor writing sample, which is in line with the findings of earlier studies (Leung et al., 2010; Noroozi et al., 2018, 2022). Male assesseses, meanwhile, were more likely to receive suggestive elaborations in the average writing sample. There are two plausible explanations for these findings. First, there is evidence that women tend to perform better than men in overall scholastic ability, especially in language courses (Voyer & Voyer, 2014). Women may have used their linguistic abilities to conscientiously read and identify the potential limitations of the work to be able to provide more suggestions – a phenomenon also observed by Noroozi et al. (2022). Second, assessors may have given more suggestive elaborations to male assesseses due to preconceived gender stereotypes about male students. Marshall and Smith's (1987) earlier study found that teachers provide more detailed feedback to male students compared to female students, who only received verification feedback, because men have the tendency to be less conforming than women (Gray & Leith, 2004). In the case of our study, assessors might have provided more suggestive feedback to men due to pre-existing perceptions that men may not adhere to the feedback if it is not detailed enough. Given this, socio-cultural gender stereotypes may have affected how assessors provided peer feedback to assesseses, a case also seen in peer scoring contexts (Aryadoust, 2016).

Limitations and Future Lines of Research

The first limitation in our study is the short duration of the peer assessment training we provided. Although the training delivered was well-structured (see Supplementary Material 4), future studies should consider implementing longer interventions. Second, we

only considered two genders (i.e. men and women) in this study. Although we used participant's self-reported gender when they enrolled to the university as the main basis of the conditions, future studies should also account for other genders (see Kube et al., 2022). Third, as we have argued that socio-cultural stereotypes may have affected participants' peer feedback, we recommend that other studies should consider replicating this study in other cultural contexts to explore further how culture may affect peer assessment. Fourth, due to the non-parametric nature of the analyses, we were unable to thoroughly explore the potential interaction between factors. Subsequent studies, if they obtain normal distributions for their dependent variables, could employ parametric techniques to test for interactions.

Pedagogical Recommendations

Because our results suggest that women are more likely to provide peer feedback with better characteristics (e.g. longer, positively phrased and suggestive), we reiterate our previous recommendation about implementing a multifaceted training programme to address potential gender-related issues that might arise during peer assessment (Ocampo et al., 2023). First, students should undergo a well-designed peer assessment training (see Panadero et al., 2016 for guidelines) to enhance their assessment capabilities to bridge the gap between men and women's peer feedback capabilities. Second, because most peer assessment activities are now conducted online, teachers should utilise the features embedded in these platforms to train students to optimise peer feedback provision (Ocampo & Panadero, 2023). Third, because our results also suggests that socio-cultural gender stereotypes may play a role in the peer feedback received by the assessees, gender-awareness workshops should also be considered to eliminate gender stereotypes and biases that students might unconsciously hold, as also recommended by prior gender and peer assessment studies (e.g., Aryadoust, 2016; Ocampo et al., 2023).

Conclusions

This research built upon our previous findings, which investigated gender peer scoring contexts (Ocampo et al., 2023). In this study, we achieved gender balance by ensuring an equal distribution of male and female assessors within each subgroup, while also maintaining an overall gender balanced sample. Using a randomised controlled design, we investigated the effects of gender and training on the amount and content of peer feedback provided for writing samples of varying quality. In general, we found that women provided more peer feedback than men. Women also provided a greater amount of positive verifications for the poor writing sample, as well as more suggestive elaborations for the average writing sample. Male assessors received more suggestive peer feedback for the average sample. We also found that trained assessors had more positive verifications in their peer feedback on the average sample. Notably, we found that male and female students were likely to provide and receive the almost the same peer feedback content in a number of aspects based on our coding scheme. This suggests that gender issues are becoming less of a problem in some peer assessment contexts, because our results indicate that the gender of the assessee did not affect the characteristics of the peer feedback, but the gender of the assessor did affects feedback characteristics on some occasions.

References

- 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>
- Berndt, M., Strijbos, J.-W., & Fischer, F. (2022). Impact of sender and peer-feedback characteristics on performance, cognitive load, and mindful cognitive processing. *Studies in Educational Evaluation, 75*, 101197. <https://doi.org/10.1016/j.stueduc.2022.101197>
- 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>
- Carson, J. G., & Nelson, G. L. (1996). Chinese students' perceptions of ESL peer response group interaction. *Journal of Second Language Writing, 5*(1), 1–19. [https://doi.org/10.1016/S1060-3743\(96\)90012-0](https://doi.org/10.1016/S1060-3743(96)90012-0)
- Caspi, A., Chajut, E., & Saporta, K. (2008). Participation in class and in online discussions: Gender differences. *Computers & Education, 50*(3), 718–724. <https://doi.org/10.1016/j.compedu.2006.08.003>
- Chaplin, T. M. (2015). Gender and Emotion Expression: A Developmental Contextual Perspective. *Emotion Review : Journal of the International Society for Research on Emotion, 7*(1), 14–21. <https://doi.org/10.1177/1754073914544408>

- Cheng, K.-H., Hou, H.-T., & Wu, S.-Y. (2014). Exploring Students' Emotional Responses and Participation in an Online Peer Assessment Activity: A Case Study. *Interactive Learning Environments*, 22(3), 271–287.
<https://doi.org/10.1080/10494820.2011.649766>
- Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on Writing: Typology and Perceived Helpfulness of Comments from Novice Peer Reviewers and Subject Matter Experts. *Written Communication*, 23(3), 260–294.
<https://doi.org/10.1177/0741088306289261>
- Darvishi, A., Khosravi, H., Rahimi, A., Sadiq, S., & Gasevic, D. (2022). Assessing the Quality of Student Generated Content at Scale: A Comparative Analysis of Peer Review Models. *IEEE Transactions on Learning Technologies*, XX(XX), 1–15.
<https://doi.org/10.1109/TLT.2022.322>
- 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>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Gielen, M., & De Wever, B. (2015a). 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, M., & De Wever, B. (2015b). Structuring the peer assessment process: A multilevel approach for the impact on product improvement and peer feedback quality: Structuring the peer assessment process. *Journal of Computer Assisted Learning*, 31(5), 435–449. <https://doi.org/10.1111/jcal.12096>

- 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>
- Hattie, J., & Gan, M. (2010). Instruction based on feedback. In P. Alexander & R. E. Mayer (Eds.), *Handbook of Research on Learning and Instruction* (pp. 249–271). Routledge.
- 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>
- Huisman, B., Saab, N., van den Broek, P., & van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: A Meta-Analysis. *Assessment & Evaluation in Higher Education, 44*(6), 863–880.
<https://doi.org/10.1080/02602938.2018.1545896>
- Jin, X., Jiang, Q., Xiong, W., Feng, Y., & Zhao, W. (2022). Effects of student engagement in peer feedback on writing performance in higher education. *Interactive Learning Environments, 0*(0), 1–16. <https://doi.org/10.1080/10494820.2022.2081209>
- Johnson, D. M., & Roen, D. H. (1992). Complimenting and Involvement in Peer Reviews: Gender Variation. *Language in Society, 21*(1), 27–57.
- Kube, D., Weidlich, J., Jivet, I., Kreijns, K., & Drachsler, H. (2022). “Gendered differences versus doing gender”: A systematic review on the role of gender in CSCL. *Unterrichtswissenschaft, 50*(4), 661–688. <https://doi.org/10.1007/s42010-022-00153-y>

- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*(4), 279–308.
<https://doi.org/10.1007/BF01320096>
- Lajom, J. A., Canoy, N., Amarnani, R., Parcon, A. M., & Valera, P. M. (2009). Barkadahan: A study of peer group norms and values among Filipino adolescents. *Philippine Journal of Psychology*, *42*(2), 195–211.
- Lam, R. (2010). A Peer Review Training Workshop: Coaching Students to Give and Evaluate Peer Feedback. *TESL Canada Journal*, 114–114.
<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* (Vol. 6248, pp. 149–159). Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-642-14657-2_15
- Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, *11*(3), 279–290.
<https://doi.org/10.1080/13562510600680582>
- 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>
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In van Merriënboer, J. M. Spector, M. D. Merrill, & M. P. Driscoll (Eds.), *Handbook of research on*

- educational communications and technology* (3rd ed., pp. 125–144). Lawrence Erlbaum. <https://doi.org/10.4324/9780203880869.ch11>
- Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Gogvadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education, 71*, 56–76. <https://doi.org/10.1016/j.compedu.2013.09.011>
- 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>
- Noroozi, O., Hatami, J., Bayat, A., van Ginkel, S., Biemans, H. J. A., & Mulder, M. (2018). Students' online argumentative peer feedback, essay writing, and content learning: Does gender matter? *Interactive Learning Environments, 28*(6), 698–712. <https://doi.org/10.1080/10494820.2018.1543200>
- 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. [10.1007/978-3-031-29411-2_8](https://doi.org/10.1007/978-3-031-29411-2_8)
- Ocampo, J. C., 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 & Evaluation in Higher Education.

<https://doi.org/10.1080/02602938.2022.2130167>

- Panadero, E. (2016). Is It Safe? Social, Interpersonal, and Human Effects of Peer Assessment: A Review and Future Directions. In *Handbook of Human and Social Conditions in Assessment* (pp. 247–266). Routledge.
- Panadero, E., Alqassab, M., Ruiz, J. F., & Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education.*
- <https://www.tandfonline.com/doi/abs/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., Jonsson, A., & Strijbos, J.-W. (2016). Scaffolding Self-Regulated Learning Through Self-Assessment and Peer Assessment: Guidelines for Classroom Implementation. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the Challenge of Implementation* (Vol. 4, pp. 311–326). Springer International Publishing. https://doi.org/10.1007/978-3-319-39211-0_18
- Patchan, M. M., Schunn, C. D., & Clark, R. J. (2018). Accountability in peer assessment: Examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education, 43*(12), 2263–2278.
- <https://doi.org/10.1080/03075079.2017.1320374>
- 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>

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>

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>

Usher, N. (2023). Self-regulating writers' uses and non-uses of peer feedback. *Assessment & Evaluation in Higher Education*, 0(0), 1–13.

<https://doi.org/10.1080/02602938.2023.2179970>

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>

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. (2020). When peers agree, do students listen? The central role of feedback quality and feedback frequency in determining uptake of feedback.

Contemporary Educational Psychology, *62*, 101897.

<https://doi.org/10.1016/j.cedpsych.2020.101897>

Wu, Y., & Schunn, C. D. (2021). From plans to actions: A process model for why feedback features influence feedback implementation. *Instructional Science*, *49*(3), 365–394.

<https://doi.org/10.1007/s11251-021-09546-5>

Zong, Z., Schunn, C. D., & Wang, Y. (2021). Learning to improve the quality peer feedback through experience with peer feedback. *Assessment & Evaluation in Higher Education*, *46*(6), 973–992. <https://doi.org/10.1080/02602938.2020.1833179>