"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

José Carlos G. Ocampo[1,2], Ernesto Panadero[1, 3, & 4], Iván Sánchez-Iglesias[5], & Fernando Diez Ruiz[1]

[1] Faculty of Education and Sports, University of Deusto, Bilbao, Spain

[2]Institute of Education, Tübingen Center for Digital Education, University of Tübingen, Germany, [3]Centre for Assessment Research Policy and Practice in Education (CARPE), School of Policy and Practice, Institute of Education, St. Patrick's Campus, Dublin City University, Ireland, [4]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain, [5]Psicobiología y Metodología de las Ciencias del Comportamiento, Universidad Complutense de Madrid, Madrid, España

Recommended Citation

Ocampo, J.C.G., Panadero, E., Sánchez-Iglesias, I., & Diez Ruiz, F. (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*, 1-12. https://doi.org/10.1016/j.stueduc.2025.101455

Author Note

José Carlos G. Ocampo 0000-0002-2180-7618

Ernesto Panadero 0000-0003-0859-3616

Iván Sánchez-Iglesias 0000-0002-6934-7905

Fernando Díez Ruiz 0000-0002-7522-5189

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 Nº 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.

Acknowledgements: Our sincerest gratitude to Dr David Kofoed Wind for the platform access, Ms Patricia Mae Taba, Dr Abigail Capay, and Ms Trish Pangilinan for the assistance in data collection. Thanks to David Zamorano for the graphics in this paper.
We have no conflicts of interest to disclose.

This is a post-peer-review, pre-copyedit version of an article published in *Studies in Educational Evaluation*. The final authenticated version is available online at: https://doi.org/10.1016/j.stueduc.2025.101455. 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, Emai: carlos.ocampo@uni-tuebingen.de

Abstract

This study explores the impact of peer feedback content (concise general vs elaborated specific), assessee gender (male vs female), and fictitious assessor gender (male vs female vs anonymous) on essay revisions and perceptions of peer feedback. A total of 284 undergraduate students (n_{Men} = 138, n_{Women} = 146) from two private universities in the Philippines participated in this study. Participants submitted an argumentative essay draft (pre-test), received concise general or elaborate specific peer feedback from a fictitious male or female or anonymous peer assessor, and submitted a revised argumentative essay (post-test). We found that the pre-test and post-test essay scores were unaffected by feedback content, assessee, or assessor gender. However, a significant triple interaction emerged between assessee gender, feedback content, and assessment time. Similarly, triple interactions were found for trust in peers as assessors and perceived adequacy of feedback. No interactions were observed for perceptions of (dis)comfort and motivation.

Keywords: gender, peer feedback content, essay revisions, intrapersonal and interpersonal factors, randomised-controlled design

"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

Peer assessment has been widely applied as an instructional strategy in writing instruction due to its learning benefits and flexibility. Instructors may require students to provide qualitative (e.g., peer feedback) and/or quantitative (e.g., peer scores) information to their peers (Topping, 1998). Various web-based peer-assessment platforms can also be used to ease the implementation of peer assessment ((Ocampo & Panadero, 2023), and evidence has shown desirable outcomes given the right conditions and support (Huisman et al., 2019). Nonetheless, the implementation of peer assessment is also constrained by certain limitations. For instance, because many web-based peer assessment platforms rely on text-based messaging, the transmission of non-verbal cues that face-to-face interactions provide may be limited (Ocampo & Panadero, 2023; Phielix et al., 2010). Additionally, student characteristics such as gender have also been documented to influence how students provide, process, and react to peer feedback in web-based peer assessment platforms (see Kerman et al., 2023). Thus, a careful examination of how these variables interact is warranted to improve how peer assessment is implemented. Hence, this study aims to examine the effects of peer feedback content, assessee gender, and fictitious assessor gender on students' essay revisions and perceptions of peer feedback in a web-based peer assessment platform.

Gender and Peer Feedback

One of the most extensive reviews on peer assessment has revealed growing evidence of the impact of gender on peer assessment (Alqassab et al., 2023). Recent studies have suggested that women tended to offer more peer feedback that is positively phrased and that contains more suggestions than the feedback provided by men when engaged in a writing-focused web-based peer assessment task (Ocampo, Panadero, Zamorano, et al., 2023). Additionally, women provided higher quality, better justified, and more detailed peer

feedback than men did in computer-facilitated writing tasks (Noroozi et al., 2020, 2022). On the other hand, men tend to make general comments but are also capable of identifying correct and problematic areas of an output, explaining why an area is correct, and providing specific suggestions to improve the work (Leung et al., 2010).

Another important aspect to explore when investigating peer feedback is whether students implement the feedback they receive from their peers. This has been the subject of many peer assessment studies with writing-related tasks, which usually measure peer feedback *implementation* or *uptake* through the quality of textual revisions made in students' essays (e.g., Patchan et al., 2016; Wichmann et al., 2018). In the context of this study, we use the term *peer feedback implementation* and *essay revisions* interchangeably to refer to changes in the essay after receiving peer feedback. While it can be assumed that women produce better quality revisions after peer feedback because they have better writing abilities (Al-Saadi, 2020) and respond to feedback in online learning environments better than men do (Prinsen et al., 2009), findings still tend to report similar writing performance from both genders. For instance, both genders were comparable in overall revised essay scores after receiving peer feedback, but women showed an advantage in incorporating and forming counterarguments (Banihashem et al., 2023; Noroozi et al., 2022). While men and women did not differ in their overall argumentative essay scores, male assessees showed a slight advantage in integrating pros and cons in essays (Noroozi et al., 2020). In another study, gender was the only variable that affected essay revisions, where women implemented more peer feedback than men did when it was stated frequently (Wu & Schunn, 2020). Despite these differences in peer feedback quality and revisions, there is little evidence regarding the difference or similarity in peer feedback implementation between male and female assessees (Noroozi et al., 2022).

Culture, Gender, and Peer Feedback

Another important factor to consider when looking at gender factors in receiving peer feedback is the potential impact of the culture-related gender stereotypes that students hold. The case study of Cao et al. (2019) on Chinese undergraduate writing students describes how women tended to be disengaged from the peer feedback given by men because of differences in ideas and the poor quality of the peer feedback they provided (i.e., excessive praise and inhibition to provide critical feedback due to assessee's gender). Similarly, men tended to devalue their assessors and refused to use the peer feedback they received. Noroozi et al. (2020) mentioned that men and women were comparable in their argumentative peer feedback due to cultural reasons; they attributed this to cultural aspects of interaction, where participants (i.e., Dutch students) were accustomed to discussing their thoughts and sorting out conflicts regardless of gender because they are from a low-power distance society. On the contrary, because the respondents in the study by Cao et al. (2019) were from a high-power distance society (i.e., China), where instructors are seen as authorities and masculinity is valued highly (Hofstede et al., 2010), assessees' approach to dealing with the peer feedback might also have been affected by such social ideologies.

These culturally related propositions were also evident in Ocampo et al. (2023), who found that male and female assessors were more likely to provide positively phrased peer feedback than negative and neutrally phrased feedback. They mentioned that this might be attributable to cultural values, as the participants in their study (i.e., Filipino university students) are from a highly collective society where providing negatively or neutrally phrased feedback may be viewed as a sign of social disruption. This trend aligns with earlier studies that found Chinese students were hesitant to deliver negatively and critically phrased feedback and preferred to give positively phrased feedback to preserve social harmony (Carson & Nelson, 1996). Given that men and women vary both in terms of providing and

receiving peer feedback, it is also important to examine if gender issues exist in the implementation of peer feedback during essay revisions. Investigating if assessees are more likely to implement peer feedback provided by a specific gender is crucial, because such gender stereotypes affect peer assessment (Torres-Guijarro & Bengoechea, 2017).

Peer Feedback Content, Gender, and Essay Revisions

Several studies have sought to identify the key content characteristics that should be incorporated into peer feedback messages, and such studies often develop detailed coding schemes to classify the content of peer feedback in writing tasks (e.g., Noroozi et al., 2020; Patchan et al., 2016). For instance, there are available peer feedback coding schemes for analytical writing tasks (Patchan et al., 2016), academic writing (Gielen & De Wever, 2015), and argumentative writing (Noroozi et al., 2020), among others. Most of these peer feedback content frameworks are anchored in the general feedback literature (see Lipnevich & Panadero, 2021), which describes and classifies the contents of peer feedback. For example, a widely investigated peer feedback content classification proposed by Strijbos et al. (2010) was based on the earlier work of Narciss (2008, 2013). This widely investigated model has provided much support in online learning settings—such as the one we used in this study.

Narciss (2008, 2013) proposes that feedback should contain the following components: function, content, and form. Function refers to individual factors that affect a learner's cognitive, metacognitive, and motivational meaning of feedback. Content can be classified into the verification component (i.e., outcome-related data—right or wrong, pass or fail, etc.) and the elaboration component (i.e., further information). These two classifications are further divided into simple feedback types (e.g., "knowledge of result", rightness or wrongness in a task) and elaborated feedback types (e.g., how to improve performance). Lastly, form deals with how the feedback is presented to the learner in relation to timing (e.g., immediate vs delayed), adaptivity (e.g., adaptive vs non-adaptive), and modality (e.g., written

vs spoken). Taken together, the combination of these components generates diverse feedback types that may influence feedback perceptions and performance. It also outlines the complex factors and processes involved in both external and internal feedback loops, thus illustrating how their potential interactions might affect the overall effects of feedback (e.g., through implementation or essay revisions).

As mentioned earlier, this feedback model was adapted in peer feedback settings by Strijbos et al. (2010), who developed a classification for providing feedback in an academic writing task by fictitious peers of high or low competence. This classification captures the verification (i.e., concise general feedback, CGF) and elaboration (i.e., elaborate specific feedback, ESF) components mentioned in Narciss (2008, 2013). This classification has also been used in scenario-based research, similar to the one we used in this study. Strijbos et al. (2010) found that ESF provided by a highly competent fictional peer was considered adequate by assessees, whereas CGF positively affected performance. Berndt et al. (2018) also found that ESF provided by a highly competent fictional peer was also seen to be adequate, and contrary to earlier findings, assessees who received ESF in this study had a higher positive affect towards peer feedback compared to assessees who received CGF. An appropriate amount of justification in an ESF led to positive effects on academic writing revisions, while an ESF with too many justifications may impose a cognitive load on the students, which may affect the processing of feedback as well as student performance (Berndt et al., 2022).

In relation to gender and peer feedback, recent evidence mentions that male assessees are more likely to receive elaborate feedback with more suggestive components than females (Ocampo, Panadero, Zamorano, et al., 2023). They hypothesised that this might be because of gender stereotypes about male students, notably that they are less conformist than female students and that they will not use the feedback if it is not detailed enough (Gray & Leith,

2004; Marshall & Smith, 1987). However, evidence mentions that males may pay less attention to verbose feedback (e.g., ESF; Narciss et al., 2014). It is thus worth examining if gender and peer feedback content affect students' essay revisions. Specifically, it is important to investigate if varying peer feedback content provided by a specific gender is likely to be implemented by assessees because students may hold gender stereotypes that may perpetuate in implementing peer feedback (Cao et al., 2019). Additionally, because different peer feedback content also influences students' perceptions of peer feedback (Strijbos et al., 2010), it is also important to investigate how male and female assessees are affected by different peer feedback types intrapersonally and interpersonally.

Interpersonal and Intrapersonal Factors of Peer Assessment

The review conducted by Panadero et al., (2023) thoroughly examined the bi-directional influence of interpersonal and intrapersonal factors on peer assessment, identifying 11 variables and re-evaluating previous categorisations (see Panadero, 2016). Interpersonal factors are the social variables (e.g., trust in the peer) that affect students during the peer assessment process, while interpersonal factors are the variables (e.g., motivation) that relate to or affect the learner from within during peer assessment. The bidirectional view of how peer assessment may influence intrapersonal and interpersonal factors (or vice-versa) is crucial given the intricate nature of implementing peer assessment (Panadero et al., 2023). For instance, gender is a growing area of interest in this regard. Ocampo et al. (2023) found that males felt discomfort and less trust in their abilities as assessors when assessing females. As mentioned by Cao et al. (2019), assessees did not implement the peer feedback provided by opposite-sex peers because they thought it lacked quality. On the other hand, men appeared to show less concern about the negative interpersonal issues in peer assessment, whereas women were worried about the negative effects of peer assessment on social

connections (Zou et al., 2018). Given these negative social effects, anonymity modes may be beneficial for peer assessment, especially for women (Rotsaert et al., 2017).

However, findings that explore the effects of anonymity on gender tend to be mixed. Lane et al. (2018) found that female assessors appear to be more affected by an anonymous intervention than male assessors. For example, anonymous female assessors provided balanced positive and negative peer feedback, while non-anonymous female assessors were more positive and constructive in their peer feedback. From the perspective of the assessees, Bloom and Hautaluoma (1987) found that female assessees had more positive reactions when feedback came from a supervisor than from an anonymous peer. It thus remains unclear if anonymity—a common feature in many web-based peer assessment platforms (Ocampo & Panadero, 2023)—can be a useful scaffold to minimise gender issues in peer assessment. This is important given that previous reviews have noted the positive effects of anonymity on peer assessment and that gender is an understudied variable in this area (Panadero & Alqassab, 2019).

The Present Study: Aim and Research Questions

This study is part of a larger project that aims to unpack the role of gender in the peer assessment process. In this study, we aim to examine the effects of peer feedback content, assessee gender, and fictitious assessor gender on essay revisions and perceptions of peer feedback in a web-based peer assessment platform. Specifically, we examined the following three independent variations (3 × 2 × 2) using a randomised controlled trial: (a) male versus female versus anonymous assessors, (b) male versus female assessees, and (c) CGF versus ESF. The research questions (RQ) of this study are as follows:

RQ1: Does assessor gender, assessee gender, and peer feedback content affect essay revisions?

RQ2: Does assessor gender, assessee gender, and peer feedback content affect interpersonal factors (i.e., trust and perceived adequacy of feedback)?

RQ3: Does assessor gender, assessee gender, and peer feedback content affect intrapersonal factors (i.e., (dis)comfort and motivation)?

Method

Participants

A total of 284 (n_{Men}= 138; n_{Women}= 146) students participated in this study. Their age ranged from 18 to 23 (M= 19.02; SD=.94), and most participants were first-year students (95.4%). Participants majored in diverse fields, including computer studies (e.g., computer science; 38.7%, n_{Men}=89 n_{Women}=21); health and welfare (e.g., nursing; 16.9%, n_{Men}=10 n_{Women}=38); social sciences (e.g., psychology; 23.9%, n_{Men}=21 n_{Women}=47); and education (e.g., elementary teaching; 20.4%, n_{Men}=18 n_{Women}=40). Participants belong to two Philippine private universities in Metro Manila (45.8%) and Western Visayas (54.2%). The distribution of participants by category of each independent variable can be seen in Figure 2.

Course Context

Participants were enrolled in the mandatory course *Understanding the Self*. The course covers personal and social identity theories, and participants are required to submit regular writing outputs (e.g., argumentative essays, position papers, and reflection logs). The course was taught in a hybrid mode (i.e., face-to-face and online) by two lecturers (one for each university) with English as the mode of instruction. The study was conducted in Weeks 14–17 of a 20-week semester, when the classes discussed the topic “Digital and Political Self”.

Materials

Web-Based Peer Assessment Platform

This study used the Eduflow peer assessment platform, which allows students to perform various assessment activities, such as peer and self-assessment. Additionally, this platform offers assessor and assessee anonymity. Figure 1 shows each of the Eduflow pages that the participants used.

Argumentative Essay Rubric

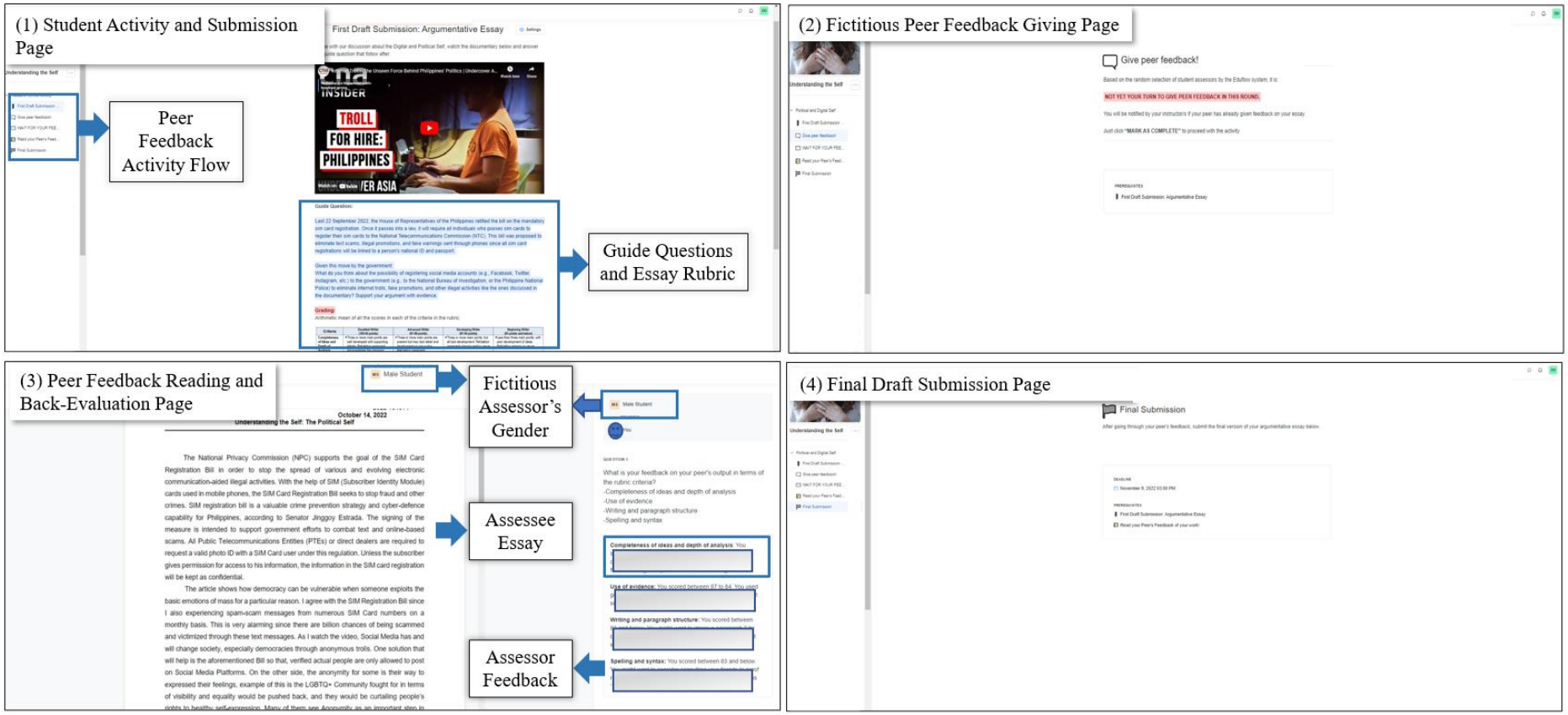
A modified rubric for argumentative essays was used to score the essays in this study (Schwalm, 2007). After consultation with the instructors, the original criterion names were changed to terms the participants were accustomed to for recognition. For example, “Use of Sources” was changed to “Use of Evidence”. The criteria in the rubric are (a) Completeness of Ideas and Depth of Arguments, (b) Use of Evidence, (c) Writing and Paragraph Structure, and (d) Spelling and Syntax.

Peer Feedback Content Template

A standard peer feedback template, designed to be adaptable to the quality of essays submitted by the students, was used to give feedback on the essays (See Supplementary Material 1). Based on the work by Strijbos et al. (2010, 2021) on peer feedback content, this template covers the classification of feedback components developed by Narciss (2008), where each essay would receive feedback in each rubric criterion.

Figure 1

Participants' Eduflow Activities.



Measures

Essay Scores

Essays were rated using an argumentative essay rubric, where each criterion was rated from 0 (lowest) to 100 points (highest); these scores were averaged to obtain the overall essay score. The first draft of the argumentative essays was rated by the first author (Rater 1), while the final draft was rated by one of the course instructors (Rater 2). The instructor is a licenced professional teacher with a postgraduate degree in education and has been teaching the course for two academic years. The two raters exhaustively discussed the rubric and rated a random sample of 45 essays (15.7%) for interrater reliability, which showed an acceptable intraclass correlation between the two raters in all rubric criteria (.78 to .89).

Intrapersonal and Interpersonal Factors Questionnaires

We operationalised students' perceptions of peer feedback using three questionnaires that measured interpersonal and intrapersonal variables (see Supplementary Material 3). For interpersonal variables, we used Ching and Hsu's (2016) trust in their peer as an assessor scale (4 items; α=.89). The composite score from three subscales (i.e., fairness, usefulness, and acceptance) of the peer feedback perceptions questionnaire developed by Strijbos et al. (2010, 2021) was used to measure the perceived adequacy of the feedback dimension (9 items; α=.73). For intrapersonal variables, the assessee's level of (dis)comfort was assessed based on a single-item question developed by Panadero et al. (2013). Finally, motivation was measured using the willingness to improve dimension (3 items; α=.72) of Strijbos et al. (2010, 2021).

In all items, participants responded on a 5-point Likert scale (1=*strongly disagree* to 5=*strongly agree*). It is also important to note that some questionnaire items were adapted to the task performed in this study. For example, "I would be willing to work on further *business letter* assignments" was changed to "I would be willing to work on further

argumentative essay assignments." The English versions of the questionnaires were used, because English is the medium of instruction in both universities.

Procedure

The procedures undertaken in this study are summarised in Figure 2. In the pre-experimental phase, the instructors briefed the participants that they would participate in a peer feedback activity for their argumentative essay requirement. During this phase, instructors discussed the rationale of the peer feedback activity, the task instructions and assessment criteria, and the steps for submitting their essays in Eduflow. Afterwards, participants watched a documentary produced by *Channel News Asia* about online trolls (Ong, 2022). Participants then submitted a 1 to 1½-page double-spaced argumentative essay (pre-test) about the prospect of registering social media accounts with the government to eliminate online trolls and other illegal activities.

In the experimental phase, participants were block randomised based on gender into 12 subgroups to establish the conditions. It is also important to note that participants were divided into subgroups based on the biological sex they reported when they applied to the university. Participants were also asked to self-report their gender in an open-ended textbox. We did not find discrepancies in the biological sex reported in their university records and their self-reported gender for all our included participants. Two participants were excluded from the final sample as they did not complete all phases of the study.

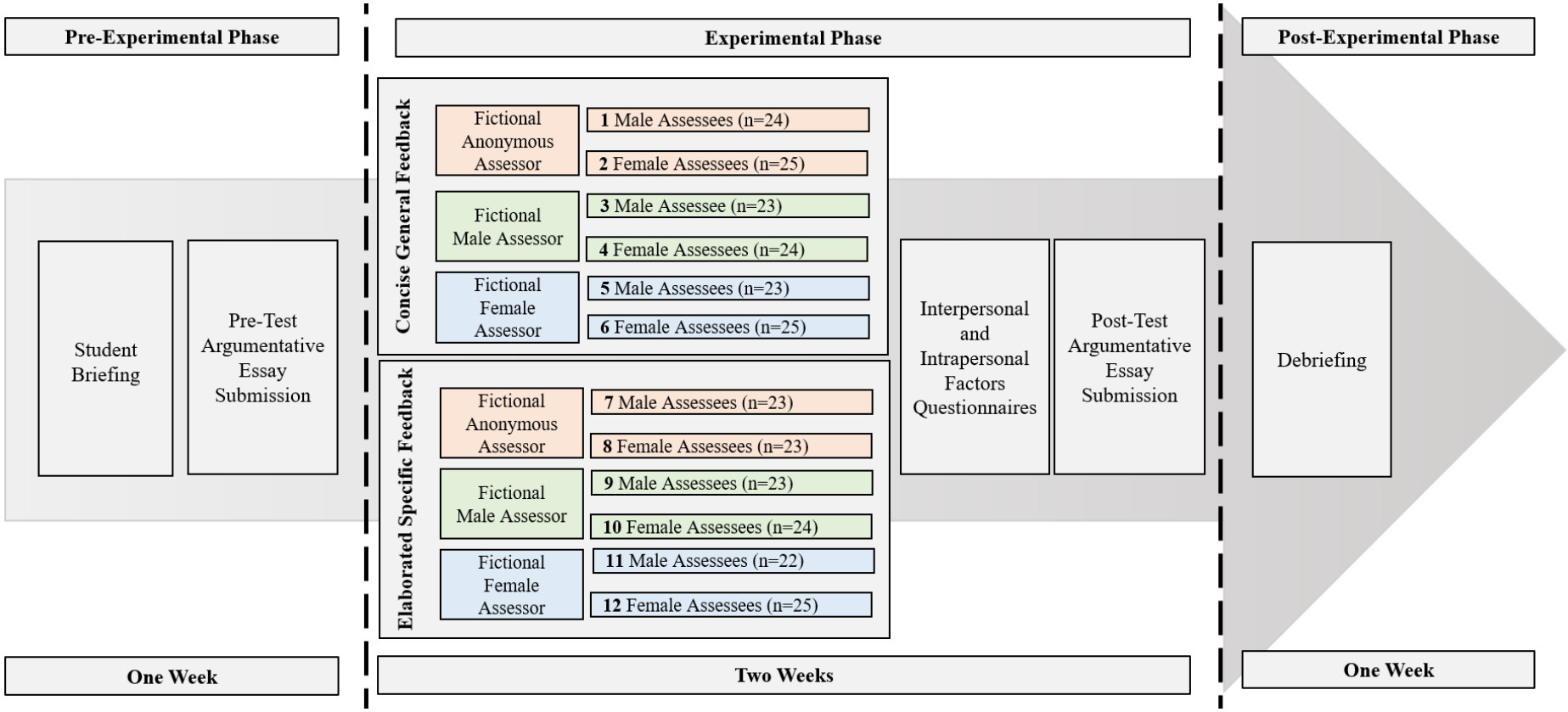
Participants were told that their identities would be anonymous. They were made to believe that random students taking the same course would be selected by Eduflow to give peer feedback on the essays and that the assessors' identities would be replaced by their gender (i.e., male, female, and anonymous) for instructor identification. Participants were unaware that the first author gave peer feedback to their essays. We did this to maintain coherence in the peer feedback components that the participants received in their respective

subgroups using the peer feedback content template (see Supplementary Material 1). In other words, none of the participants provided peer feedback in this study. The participants did not know that the male, female, and anonymous peer assessor who gave peer feedback to their essay was the first author. The participants received peer feedback on their essays within one week of their essay submission. More importantly, they were told not to share any details about their peer assessor or the peer feedback they received with anyone to maintain confidentiality. Subsequently, they answered questionnaires in Eduflow. The instructors told the participants to submit a revised version of their essay (post-test) based on the peer feedback within one week.

In the post-experimental phase, the first author debriefed the participants about the study's design via online meetings to avoid interfering with their face-to-face sessions. Participants were informed that the peer feedback on their work was provided by the first author and not by a random male, female, or anonymous peer assessor. It is important to note that the participants were allowed to withdraw from the study in the debriefing session. Supplementary Material 2 provides a more thorough explanation of the peer assessment design elements we used in this study.

Figure 2

Study Procedure Summary



Research Design and Data Analysis

We used a randomised controlled trial design in this study, where participants were blocked randomised to the subgroups based on the gender they reported when they enrolled at the university. For RQ1, we calculated descriptive statistics for essay scores based on the studied factors (i.e., assessee gender, assessor gender, peer feedback content, and assessment time). Four-way ANCOVA (mixed, as assessment time was a within-subjects factor) was computed to study the effect of the factors on essay scores, including university as a covariate to discount its potential effect on the dependent variables. Similarly, RQ2 and RQ3 were analysed using factorial ANCOVA, where the ratings in the questionnaire were treated as the dependent variables and the university as a covariate. We checked ANCOVA assumptions using Levene's test for homogeneity of variances and Box's M test for equality of variance-covariance matrices. Simple effects were computed and interpreted for significant interaction effects. The significance level for all inferential tests was set at alpha = .050. We used SPSS version 25 for all analyses.

We accounted for the influence of participants' academic ability (i.e., GPA) on their argumentative essay quality. Regardless of the gender of the assessee and type of peer feedback received, those who received peer feedback from males (M = 3.29, SD = 0.34), females (M = 3.27, SD = 0.39), and anonymous (M = 3.25, SD = 0.27) assessors were comparable in academic ability, $F(2, 281) = .30$, $p = .738$. Without regard for assessor gender and type of peer feedback received, assessors who rated fictitious male (M = 3.26, SD = 0.35) and female assessees (M = 3.28, SD = 0.33) were also comparable in academic ability, $t(282) = -.59$, $p = .344$. Finally, regardless of assessor and assessee gender, those who received CGF (M = 3.27, SD = 0.34) and ESF (M = 3.27, SD = 0.34) were comparable in academic ability, $t(282) = -.070$, $p = .623$.

The participants were also asked an open-ended question, *"Who provided peer feedback to your work?"* to determine if they recognised the gender of their peer assessor, which is found in multiple parts of the peer feedback reading page in Eduflow (see Figure 1). A total of 229 (80.6%) participants correctly identified the gender of their assessor.

Results

RQ1. Does assessor gender, assessee gender, and peer feedback content affect essay revisions?

Four-way mixed ANCOVA was conducted to assess the effects of assessee gender × assessor gender × peer feedback content × assessment time (pre-test vs post-test) and their interactions on the essay scores. Descriptive statistics can be found in Table 1. The Box's M test result was nonsignificant, $F = 1.41$, $p = .060$. Levene's tests were nonsignificant at both assessment time points: pre-test, F = 1.47, $p = .142$; post-test, $F = 1.19$, $p = .294$.

Table 1

Essay Scores: Descriptive Statistics, by Assessee Gender, Assessor Gender, Peer Feedback Content, and Assessment Time

Assessee Gender	Assessor Gender	Peer Feedback Content		Pre-Test	Post-Test
			N	*M (SD)*	*M (SD)*
Men	Men	CGF	23	82.3 (2.2)	85.0 (3.8)
		ESF	23	81.4 (2.9)	86.8 (3.6)
		Total	46	81.9 (2.5)	85.9 (3.8)
	Women	CGF	23	81.0 (4.0)	84.7 (3.8)
		ESF	22	81.0 (2.6)	85.9 (3.3)
		Total	45	81.0 (3.4)	85.3 (3.6)
	Anonymous	CGF	24	81.9 (2.8)	85.5 (3.7)
		ESF	23	81.5 (3.8)	86.8 (3.0)
		Total	47	81.7 (3.3)	86.1 (3.4)
	Total	CGF	70	81.8 (3.1)	85.1 (3.8)
		ESF	68	81.3 (3.1)	86.5 (3.3)
		Total	138	81.5 (3.1)	85.8 (3.6)
Women	Men	CGF	24	82.1 (2.9)	86.2 (3.8)
		ESF	24	82.7 (3.1)	88.2 (3.5)
		Total	48	82.4 (3.0)	87.2 (3.8)
	Women	CGF	25	82.5 (3.0)	87.0 (2.9)
		ESF	25	82.2 (3.2)	86.4 (4.4)
		Total	50	82.4 (3.1)	86.7 (3.7)
	Anonymous	CGF	25	82.2 (3.6)	87.4 (3.2)

		ESF	23	81.6 (2.7)	86.4 (2.5)
		Total	48	81.9 (3.2)	86.9 (2.9)
	Total	CGF	74	82.3 (3.1)	86.9 (3.3)
		ESF	72	82.2 (3.0)	87.0 (3.6)
		Total	146	82.2 (3.1)	86.9 (3.5)
Total	Men	CGF	47	82.2 (2.5)	85.6 (3.8)
		ESF	47	82.1 (3.0)	87.5 (3.6)
		Total	94	82.1 (2.8)	86.6 (3.8)
	Women	CGF	48	81.8 (3.5)	85.9 (3.6)
		ESF	47	81.6 (3.0)	86.1 (3.9)
		Total	95	81.7 (3.3)	86.0 (3.7)
	Anonymous	CGF	49	82.1 (3.2)	86.5 (3.6)
		ESF	46	81.6 (3.3)	86.6 (2.7)
		Total	95	81.8 (3.2)	86.5 (3.2)
	Total	CGF	144	82.0 (3.1)	86.0 (3.6)
		ESF	140	81.8 (3.1)	86.7 (3.5)
		Total	284	81.9 (3.1)	86.4 (3.6)

Table 2

ANCOVA Summary Table for Essay Scores

Effect	*F*	*df*	*p*	η^2
Assessee Gender	2.266	1, 271	0.133	0.008
Assessor Gender	0.860	2, 271	0.424	0.006
Peer Feedback Content	0.454	1, 271	0.501	0.002
Assessment Time	25.716	1, 271	< .001	0.087
Assessment Time × Assessee Gender	0.250	1, 271	0.618	0.001
Assessment Time × Assessor Gender	0.327	2, 271	0.721	0.002
Assessment Time × Peer Feedback Content	6.632	1, 271	0.011	0.024
Assessment Time × Assessee Gender × Assessor Gender	0.282	2, 271	0.754	0.002
Assessment Time × Assessee Gender × Peer Feedback Content	4.744	1, 271	0.030	0.017
Assessment Time × Assessor Gender × Peer Feedback Content	1.679	2, 271	0.189	0.012
Assessment Time × Assessee Gender × Assessor Gender × Peer Feedback Content	0.154	2, 271	0.857	0.001

Note. $N = 284$.

Table 2 summarises the effects of the factors and their potential interactions, after discounting the significant effect of the university on essay scores, $F(1, 271) = 4.79$, $p = .029$, $\eta^2 = .017$. We found significant main effects for assessor gender and assessment time and for the assessment time × peer feedback content interaction. However, we were particularly interested in the assessment time × assessee gender × peer feedback content interaction, $F(1,272) = 4.7$, $p = .030$, $\eta^2 = .017$, which implies a two-way interaction that varies across

levels of a third variable. Figure 3 shows the pre-test and post-test essay scores considering the assessee's gender and peer feedback content. All simple effects possible between pre-test and post-test means were statistically significant, *p*s < .001. Moreover, Table 3 shows a significant difference in the post-test scores between men and women, where female assessees who received CGF outperformed male assessees in their post-test essays. Similarly, Table 4 shows the differences in the post-test essay scores of assessees who received CGF and ESF, where male assessees who received ESF outperformed male assessees who received CGF in their post-test essay scores.

Table 3

Essay Scores: Statistical Simple Effects Between Assessee Genders, by Peer Feedback Content and Assessment Time.

Assessment Time	Peer Feedback Content	M_{Men}	M_{Women}	*p*
Pre-Test	CGF	81.8	82.3	.640
	ESF	81.3	82.2	.236
Post-Test	CGF	85.1	86.9	.031
	ESF	86.5	87.0	.981

Notes. *N* = 284. All simple effects use Bonferroni correction.

Table 4

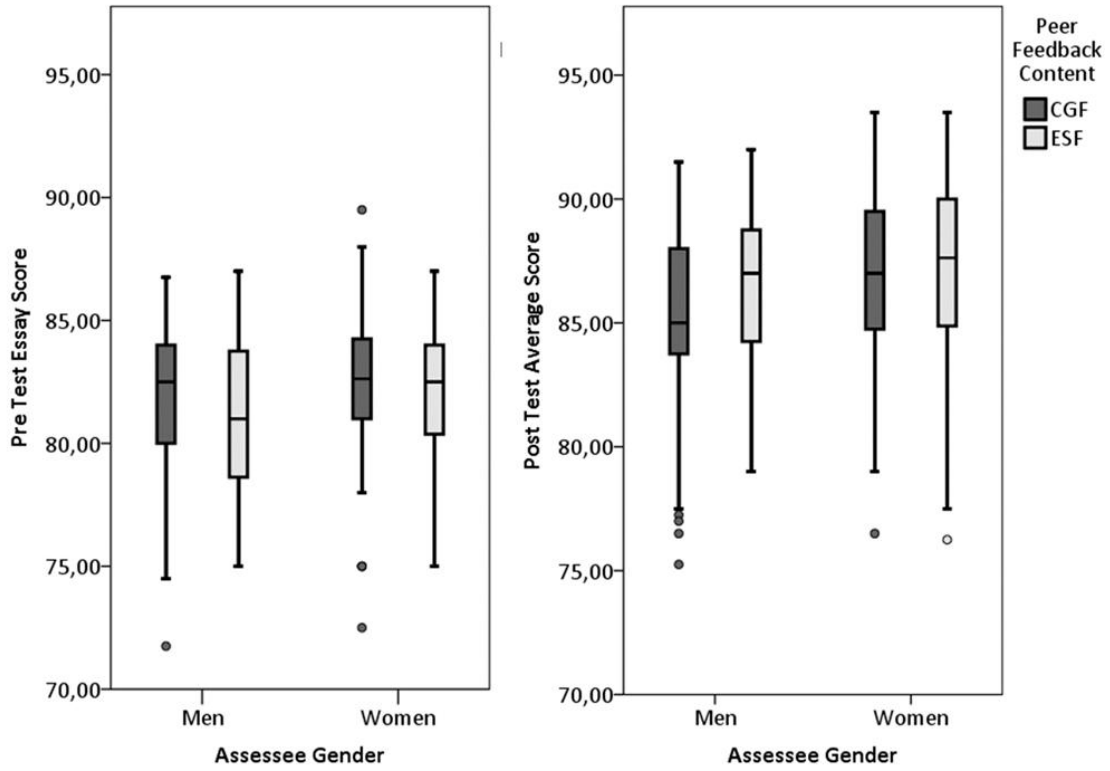
Essay Scores: Statistical Simple Effects Between Peer Feedback Content by Assessee Gender and Assessment Time.

Assessment time	Assessee Gender	M_{CGF}	M_{ESF}	*p*
Pre-Test	Men	81.8	81.3	.365
	Women	82.3	82.2	.876
Post-Test	Men	85.1	86.5	.020
	Women	86.9	87.0	.894

Notes. *N* = 284. All simple effects use Bonferroni correction.

Figure 3

Pre-test and Post-test Essay Scores by Assessee Gender and Peer Feedback Content



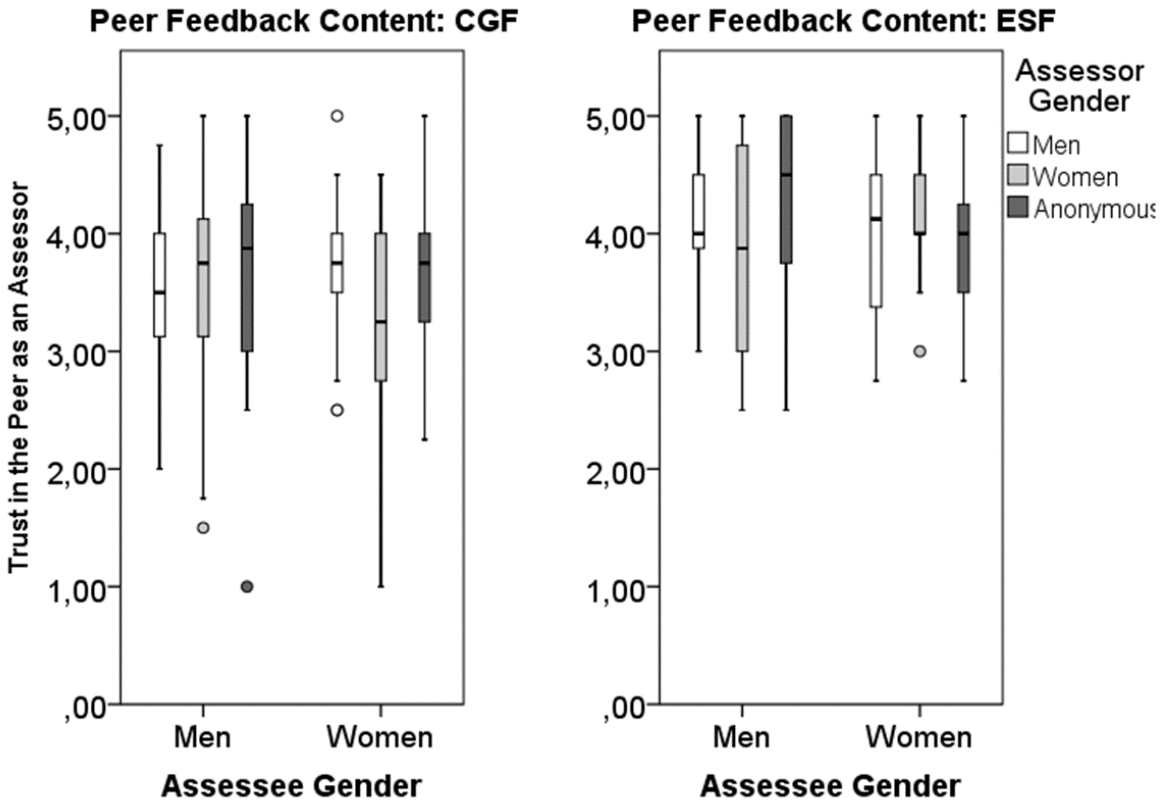
RQ2: Does assessor gender, assessee gender, and peer feedback content affect interpersonal factors?

Descriptive statistics for the interpersonal variable trust in the peer as an assessor and perceived adequacy of feedback are summarised in Table 5, while the effects of the factors and their potential interactions are found in Table 6. In the ANCOVA model, the covariate university was non-significant for trust in the peer as an assessor, $F(2,271) = 3.86$, $p = .051$, and for adequacy of feedback, $F(2,271) = 0.32$, $p = .206$. Also, the results showed a significant triple interaction between assessee gender × assessor gender × peer feedback content for trust in the peer as an assessor, $F(2,271) = 4.31$, $p = .018$, $\eta^2 = .029$. The simple effects analysis showed that male assessees trusted male ($p = .014$) or anonymous ($p = .005$) assessors who gave them ESF more compared to female assessors. Female assessees also trusted female assessors who gave them ESF ($p = < .001$). Moreover, female assessees trusted

the CGF more when given by a male assessor than by a female assessor (p = .027). Finally, there was a significant main effect for peer feedback content, where assessees trusted ESF more than CGF. Figure 4 shows the scores for trust in the peer as an assessor considering the assessee's and assessor's gender in two panels, one for each peer feedback content.

Figure 4

Trust Scores by Assessee and Assessor Gender Across Peer Feedback Content Types



Similarly, the ANCOVA results showed a significant triple interaction between assessee gender × assessor gender × peer feedback content for perceived adequacy of feedback, $F(2,271) = 4.75$, p = .009, η^2 =.034. Simple effects analysis showed that male assessees perceived that the ESF provided by male assessors was more adequate (p = .032), while female assessees perceived that the ESF provided by female assessors was also more adequate (p = <.001). Additionally, female assessees also perceived the CGF provided by male (p = .044) or anonymous (p = .042) assessors as more adequate than that provided by female assessors. Finally, there is a significant main effect for peer feedback content, where

assessees perceived ESF as more adequate than CGF. Figure 5 shows the perceived adequacy of feedback scores considering the assessee's and assessor's gender in two panels, one for each peer feedback content.

Figure 5

Perceived Adequacy of Feedback Scores by Assessee and Assessor Gender Across Peer Feedback Content Types

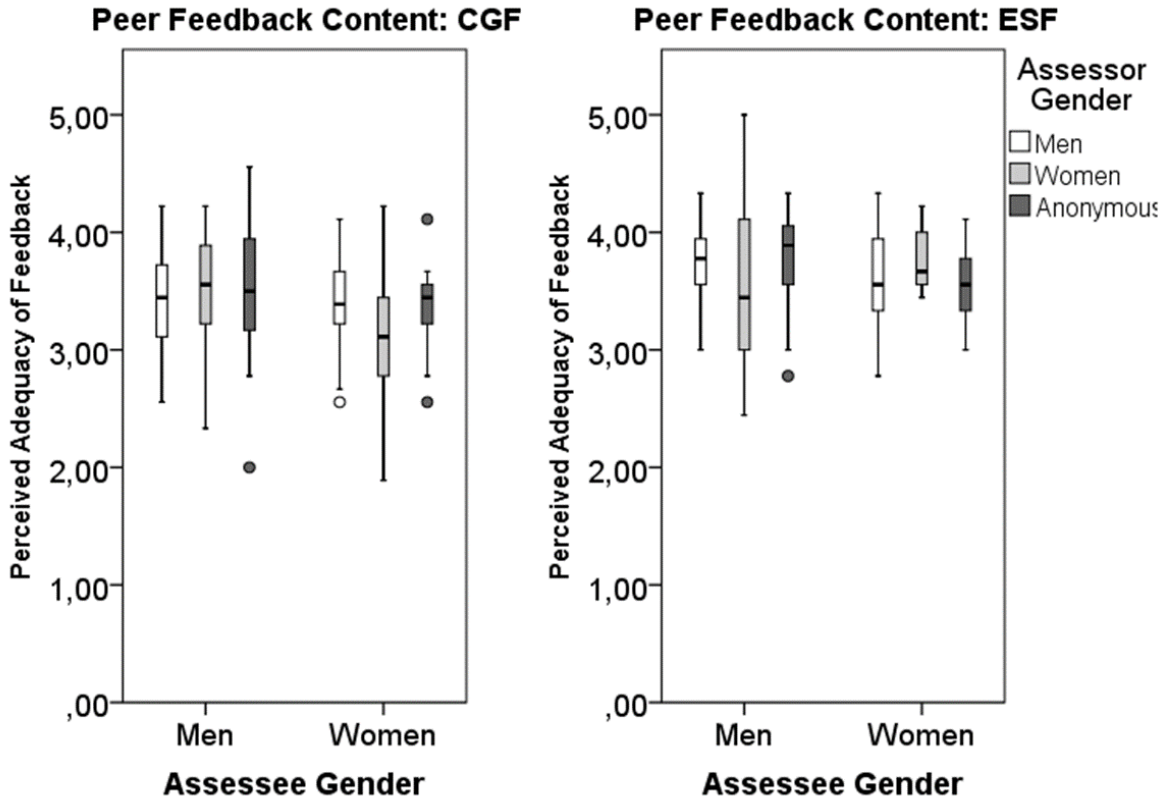


Table 5

Descriptive Statistics for Interpersonal and Intrapersonal Factors

				Interpersonal Factors		Intrapersonal Factors	
Assessee Gender	Assessor Gender	Peer Feedback Content		Trust in the Peer as an Assessor	Perceived Adequacy of Feedback	(Dis)Comfort	Motivation
			N	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Men	Men	CGF	23	3.58 (0.65)	3.43 (0.43)	3.61 (0.78)	3.87 (0.68)
		ESF	23	4.14 (0.50)	3.72 (0.36)	3.96 (0.64)	4.00 (0.59)
		Total	46	3.86 (0.64)	3.57 (0.42)	3.78 (0.73)	3.93 (0.63)
	Women	CGF	23	3.60 (0.94)	3.50 (0.51)	3.78 (0.90)	4.07 (0.59)
		ESF	22	3.84 (0.79)	3.56 (0.60)	4.00 (0.87)	4.17 (0.60)
		Total	45	3.72 (0.87)	3.53 (0.55)	3.89 (0.88)	4.12 (0.59)

	Anonymous	CGF	24	3.64 (0.91)	3.52 (0.59)	3.92 (0.93)	3.93 (0.67)
		ESF	23	4.25 (0.80)	3.75 (0.45)	4.13 (0.76)	4.16 (0.83)
		Total	47	3.94 (0.90)	3.63 (0.53)	4.02 (0.85)	4.04 (0.75)
	Total	CGF	70	3.60 (0.83)	3.48 (0.51)	3.77 (0.87)	3.96 (0.64)
		ESF	68	4.08 (0.72)	3.68 (0.48)	4.03 (0.75)	4.11 (0.68)
		Total	138	3.84 (0.81)	3.58 (0.50)	3.90 (0.82)	4.03 (0.66)
Women	Men	CGF	24	3.72 (0.60)	3.41 (0.36)	3.87 (0.68)	4.11 (0.45)
		ESF	24	4.00 (0.76)	3.60 (0.43)	3.83 (0.87)	4.19 (0.57)
		Total	48	3.86 (0.69)	3.50 (0.40)	3.85 (0.77)	4.15 (0.51)
	Women	CGF	25	3.15 (0.94)	3.08 (0.59)	3.44 (0.92)	4.13 (0.66)
		ESF	25	4.21 (0.52)	3.78 (0.29)	4.04 (0.84)	4.21 (0.64)
		Total	50	3.68 (0.92)	3.43 (0.58)	3.74 (0.92)	4.17 (0.65)
	Anonymous	CGF	25	3.64 (0.60)	3.40 (0.32)	3.80 (0.82)	4.15 (0.41)
		ESF	23	3.91 (0.56)	3.55 (0.30)	3.91 (0.85)	4.17 (0.44)
		Total	48	3.77 (0.59)	3.47 (0.31)	3.85 (0.82)	4.16 (0.42)
	Total	CGF	74	3.50 (0.77)	3.30 (0.46)	3.70 (0.82)	4.13 (0.51)
		ESF	72	4.05 (0.63)	3.64 (0.36)	3.93 (0.84)	4.19 (0.55)
		Total	146	3.77 (0.75)	3.47 (0.45)	3.82 (0.84)	4.16 (0.53)
Total	Men	CGF	47	3.65 (0.62)	3.42 (0.39)	3.74 (0.74)	3.99 (0.58)
		ESF	47	4.07 (0.64)	3.66 (0.40)	3.89 (0.76)	4.10 (0.58)
		Total	94	3.86 (0.66)	3.54 (0.41)	3.82 (0.75)	4.05 (0.58)
	Women	CGF	48	3.36 (0.96)	3.28 (0.59)	3.60 (0.92)	4.10 (0.62)
		ESF	47	4.04 (0.68)	3.68 (0.47)	4.02 (0.85)	4.19 (0.62)
		Total	95	3.70 (0.89)	3.48 (0.57)	3.81 (0.90)	4.15 (0.62)
	Anonymous	CGF	49	3.64 (0.76)	3.46 (0.47)	3.86 (0.87)	4.04 (0.56)
		ESF	46	4.08 (0.71)	3.65 (0.39)	4.02 (0.80)	4.17 (0.66)
		Total	95	3.85 (0.76)	3.55 (0.44)	3.94 (0.84)	4.10 (0.61)
	Total	CGF	144	3.55 (0.80)	3.39 (0.49)	3.74 (0.84)	4.05 (0.58)
		ESF	140	4.06 (0.67)	3.66 (0.42)	3.98 (0.80)	4.15 (0.62)
		Total	284	3.80 (0.78)	3.52 (0.48)	3.86 (0.83)	4.10 (0.60)

Table 6

ANCOVA Summary Table for Interpersonal Factors

	Trust in the Peer as an Assessor				Perceived Adequacy of Feedback			
Effect	*F*	*df*	*p*	η^2	*F*	*df*	*p*	η^2
Assessee Gender	2.432	1, 271	.120	.009	5.861	1, 271	.016	.021
Assessor Gender	1.585	2, 271	.207	.012	.707	2, 271	.494	.005
Peer Feedback Content	33.933	1, 271	<.001	.111	25.291	1, 271	<.001	.085
Assessee Gender × Assessor Gender	0.521	2, 271	.612	.004	.302	2, 271	.740	.002
Assessee Gender × Peer Feedback Content	0.078	1, 271	.701	.001	1.899	1, 271	.169	.007
Assessor Gender × Peer Feedback Content	0.859	2, 271	.425	.006	1.265	2, 271	.284	.009
Assessee Gender × Assessor Gender × Peer Feedback Content	4.306	2, 271	.018	.029	4.751	2, 271	.009	.034

Note. $N = 284$.

RQ3. Does assessor gender, assessee gender, and peer feedback content affect intrapersonal factors?

Descriptive statistics for the intrapersonal variable (dis)comfort and motivation are summarised in Table 5, while Table 7 summarises the effects of the factors and their potential interactions. The results of the ANCOVA for the intrapersonal variable (dis)comfort showed that the covariate (i.e., university) was significant, $F(2, 271) = 5.49$, $p = .020$, $\eta^2 = .020$. After accounting for the covariate, we did not find any significant interactions. However, there was a significant main effect for peer feedback content, where those who received ESF felt more comfortable than those who received CGF, $F(1,271) = 5.968$, $p = .015$, $\eta^2 = .022$ (small effect). Similarly, there was no significant effect for the covariate when using motivation as the dependent variable, $F(1,271) = 1.632$, $p = .202$. We did not find any significant interaction or main effects in this ANCOVA model.

Table 7

ANCOVA Summary Table for Intrapersonal Factors

	(Dis)comfort				Motivation			
Effect	*F*	*df*	*p*	η^2	*F*	*df*	*p*	η^2
Assessee Gender	3.139	1, 271	.078	.011	1.155	1, 271	.284	.004
Assessor Gender	0.589	2, 271	.555	.004	.612	2, 271	.543	.004
Peer Feedback Content	5.968	1, 271	.015	.022	2.178	1, 271	.141	.008
Assessee Gender × Assessor Gender	0.742	2, 271	.477	.005	.447	2, 271	.640	.003
Assessee Gender × Peer Feedback Content	0.028	1, 271	.868	.000	.362	1, 271	.548	.001
Assessor Gender × Peer Feedback Content	0.927	2, 271	.397	.007	.008	2, 271	.992	.000
Assessee Gender × Assessor Gender × Peer Feedback Content	0.819	2, 271	.442	.006	.184	2, 271	.202	.006

Note. $N = 284$.

Discussion

Increasing work in the area of gender and peer feedback has shown that male and female assessors may provide peer feedback with varying characteristics (e.g., Ocampo et al., 2023). More importantly, evidence also mentions that male and female assessees may inhibit

the implementation of peer feedback they receive from opposite-sex peers (Cao et al., 2019). However, evidence explaining if male and female assessees differ in implementing the peer feedback from male, female, or anonymous assessors remains unexplored. This study therefore investigated the effects of peer feedback content, assessee gender, and fictitious assessor gender on students' essay revision and perceptions of peer feedback.

Effects on Essay Revisions

Our study found that men and women achieved comparable post-test essay scores, regardless of the assessor's gender and the content of the peer feedback received. This finding aligns with prior studies suggesting the absence of difference in the overall essay scores between men and women after receiving peer feedback (Noroozi et al., 2020, 2022). In other words, male and female assessees are likely to use any feedback they received, regardless of who provided it. More importantly, this finding contradicts earlier studies that found assessees inhibit implementation of peer feedback from opposite-sex peers (Cao et al., 2019).

Additionally, we found a significant triple interaction for assessment time, assessee gender, and peer feedback content. Specifically, female assessees who received CGF had higher post-test scores than male assessees who also received CGF. On the other hand, male assessees who received ESF also had higher post-test scores than male counterparts who received CGF. Although they received shorter feedback without elaboration components, women performed better than men in their post-test essay scores, possibly due to their proficiency in language-related activities (e.g., essay writing) compared to men (Voyer & Voyer, 2014). On the contrary, men may have performed better when they received more verbose and detailed feedback because such feedback shows the aspects of their essay that needed improvement. This finding partially contradicts Narciss et al. (2014), who reported that men tend to pay less attention to lengthy feedback, while this reinforces our earlier

finding that peer assessors are more inclined to provide more detailed feedback to men, as men may pay less attention to unelaborated feedback (Ocampo, Panadero, Zamorano, et al., 2023).

Effects on Interpersonal Factors

We found that male assesses trusted and perceived the adequacy of ESF when delivered by male and anonymous assessors, while female assesses showed a preference for male assessors delivering CGF. Conversely, female assesses trusted and found adequacy in ESF when provided by female assessors and in CGF when delivered by male assessors. Both these interpersonal variables indicate a common theme where men and women tend to associate longer and more detailed feedback with same-sex peers while attributing shorter feedback to opposite-sex peers. This finding may be due to a preference for same-sex friendship (see Rotenberg, 1986), where assessees reported more confidence in the ability and adequacy of peer feedback provided by same-sex peers.

A notable alignment was observed among female assesses who received CGF, as they perceived and trusted feedback from male assessors in CGF to be adequate, which coincided with improvements in their post-test essay scores. Interestingly, despite showing better post-test essay improvements when receiving ESF, regardless of the gender of the assessor who provided it, male assesses trusted and perceived that the ESF provided by male assessors was more adequate. This is similar to earlier findings, where women were more accurate in assessing the performance of both genders, while men favoured same-sex peers (Langan et al., 2005). Once again, we believe that the same-sex friendship preference affected men's perceptions of the ESF they receive. Finally, our finding supports earlier studies that students perceived ESF as more trustworthy and adequate (Berndt et al., 2018; Strijbos et al., 2010). Participants may have preferred ESF due to its verbosity, elaborate explanation of problems in the essay, and potential strategies for improvement.

Effects on Intrapersonal Factors

In terms of intrapersonal factors, regardless of their gender, the gender of the assessor, and the type of feedback they received, students provided similar ratings for comfort and motivation. However, those who received ESF had a higher level of comfort than those who received CGF, which aligns with Berndt et al. (2018) but contradicts Strijbos et al. (2010). Two things can be inferred from this finding. First, the online nature of the peer feedback activity may have helped students feel more comfortable receiving peer feedback because it may have reduced the discomfort generally experienced in face-to-face peer assessment (Lim et al., 2021). Second, students who received ESF may have felt more comfortable given that this feedback type explains the student's performance in the essay in length. It is thus unsurprising that assessees also trusted and perceived that this feedback was more adequate because they also reported more comfort when they received it. Because the feedback was specific and detailed, assessees may have felt a sense of comfort that their peers devoted time and effort to read and give feedback on their essay.

Our finding also coincides with earlier studies that found that men and women reported comparable levels of motivation after receiving feedback (Narciss et al., 2014). Additionally, this finding supports the large body of evidence showing the positive influence of peer assessments on motivation (Panadero et al., 2023). We believe that the use of the web-based peer assessment platform may have added value to this finding, because the platform was organised in a systematic manner that shows a step-by-step guide on the activities that the students had to perform (see Figure 1), the rubric used to score their essay, and the interface of the peer feedback reading page. Moreover, we also found that students who received CGF or ESF were comparable in their levels of motivation—a finding also observed by Strijbos et al. (2010). We believe that the components for knowledge of results and knowledge of mistakes provided in both the CGF and ESF may have equalised students'

level of motivation. These components may have been sufficient for those in the CGF condition to stay motivated in revising their essays.

Limitations and Future Lines of Research

First, the participants of this study were limited to university students in the Philippines. Hofstede et al. (2010) mentioned that masculinity is strongly valued in the Philippines, and it is common for men to restrict their emotional expression to exhibit strong masculinity. Future studies could explore gender and peer feedback in diverse cultural contexts to investigate socio-cultural stereotypes related to gender in peer assessments, as also suggested by previous research in this area (e.g., Cao et al., 2019; Noroozi et al., 2020). Second, the measurement of the intrapersonal and interpersonal variables was limited to self-report measures. Prior studies have mentioned that participants may modify their answers to questions based on what is socially desirable and acceptable (Holtgraves, 2004). However, given the web-based design of this study, where participants are online and may perform the tasks asynchronously, the self-report approach was the best way to measure the intrapersonal and interpersonal factors in our study. Future studies could look at physiological measures to measure students' reactions to varying qualities of peer feedback (see Panadero, 2023). Third, the peer assessors that provided the peer feedback in this study were fictitious, as the first author provided the feedback. This strategy has been used in several peer assessment studies (e.g., Berndt et al., 2022; Peters et al., 2018; Strijbos et al., 2010), and we adapted this strategy to control for potential confounds in our study (i.e., varying peer feedback content). Simulated interventions (e.g., fictional assessors) have been shown to reduce gender-related bias in other gender studies (e.g. Moss-Racusin & Rabasco, 2018), which may have also affected our results. Additionally, although participants were told that the peer feedback came from another peer taking the same course, we recommend that future studies investigate gender in a more ecological method by having students perform the roles of both assessor and

assessee. Fourth, we only considered two genders in this study (i.e., male and female). While we adhere to the recommendations of education and gender researchers, such as Kube et al. (2022), by incorporating multiple methods to account for gender (i.e., university records and self-reported gender), we encourage future researchers to include a broader range of genders across the spectrum. Fifth, while we found significant main and interaction effects in our research questions, it is also important to note that most of these effect sizes (i.e., η^2) may be considered 'small' based on the conventional thresholds proposed by Cohen (1988). Although some findings are similar to previous studies, the practical implications may be limited. In other words, while significant relationships were observed, this may not necessarily translate to extensive differences between men and women in classrooms. Nonetheless, teachers should still be cognisant of gender's influence on peer assessment, given previous findings on peer feedback implementation and perceptions of peer feedback (Kerman et al., 2023).

Pedagogical Recommendations

First, we recommend implementing gender sensitivity and awareness workshops to minimise any biases that students may have towards their same-sex or opposite-sex peers. Although we did not find any sign of bias in implementing peer feedback, the results show that males and females may have different interpersonal reactions to varying peer feedback content, especially when it is given by a certain gender. Implementing gender sensitivity and awareness sessions may make students aware of the unconscious gender biases that they hold towards their peers, which we have recommended in our previous studies (Ocampo, Panadero, & Díez, 2023; Ocampo, Panadero, Zamorano, et al., 2023). Second, although we used anonymity in a limited manner in this study, we recommend that teachers maximise this feature, which is available in many web-based peer assessment platforms (Ocampo & Panadero, 2023), given that it may potentially help mitigate any intrapersonal or interpersonal

issues that might occur in peer assessments. However, teachers should also be aware of the limitations of anonymity. For example, it may inhibit a more dialogic and collaborative process that peer assessment should foster (see Wood, 2022).

Conclusions

The first half of our title is “Due to the composition of the feedback, I think it’s a girl”. This specific phrase came from the manipulation check of one participant when we asked them to report who they thought gave them the peer feedback. We believe this response perfectly captures some of the students’ preconceived stereotypes attached to gender during the peer assessment process, thus supporting earlier observations made by Torres-Guijarro and Bengoechea (2017). Specifically, we unpacked these preconceived gender stereotypes by examining the effects of peer feedback content, assessee gender, and fictitious assessor gender on students’ essay revision and perceptions of peer feedback in a web-based peer assessment platform with a gender-balanced sample. We found that students improved their essays, regardless of whether they received shorter or longer feedback from a male, female, or anonymous assessor. This may mean teachers should not worry about the effects of gender on students’ reception of peer feedback, given circumstances similar to the ones in our study. While this suggests the absence of gender issues regarding the results of students’ post-test essay scores, the results for trust in the peer as an assessor and perceived adequacy of feedback suggest that men and women may prefer certain peer feedback content, especially when given by a specific peer. These findings magnify the claims of previous reviews (e.g., Kerman et al., 2023; Panadero et al., 2023) on the importance of considering individual difference factors, like gender, intrapersonal and interpersonal factors, given that it may affect the peer assessment process.

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

During the preparation of this work, the authors used ChatGPT-4 to shorten long sentences and improve readability. After receiving suggestions from the tool, the authors further reviewed and edited the content as needed, taking full responsibility for the final version of the publication.

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

Al-Saadi, Z. (2020). Gender differences in writing: The mediating effect of language proficiency and writing fluency in text quality. *Cogent Education*, *7*(1), 1770923. https://doi.org/10.1080/2331186X.2020.1770923

Banihashem, S. K., Noroozi, O., Biemans, H. J. A., & Tassone, V. C. (2023). The intersection of epistemic beliefs and gender in argumentation performance. *Innovations in Education and Teaching International*, *61*(4), 716–734. https://doi.org/10.1080/14703297.2023.2198995

Berndt, M., Strijbos, J.-W., & Fischer, F. (2018). Effects of written peer-feedback content and sender's competence on perceptions, performance, and mindful cognitive processing. *European Journal of Psychology of Education*, *33*(1), 31–49. https://doi.org/10.1007/s10212-017-0343-z

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

Bloom, A. J., & Hautaluoma, J. E. (1987). Effects of message valence, communicator credibility, and source anonymity on reactions to peer feedback. *The Journal of Social Psychology*, *127*(4), 329–338. https://doi.org/10.1080/00224545.1987.9713712

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

Ching, Y.-H., & Hsu, Y.-C. (2016). Learners' interpersonal beliefs and generated feedback in an online role-playing peer-feedback activity: An exploratory study. *The International Review of Research in Open and Distributed Learning, 17*(2). https://doi.org/10.19173/irrodl.v17i2.2221

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

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

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

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 3rd ed.). McGraw-Hill. http://catdir.loc.gov/catdir/enhancements/fy1009/2010010437-b.html

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin, 30*(2), 161–172. https://doi.org/10.1177/0146167203259930

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

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*, *29*, 763–813. https://doi.org/10.1007/s10639-023-12273-8

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

Lane, J. N., Ankenman, B., & Iravani, S. (2018). Insight into gender differences in higher education: Evidence from peer reviews in an introductory STEM course. *Service Science*, *10*(4), 442–456. https://doi.org/10.1287/serv.2018.0224

Langan, A. M., Wheater, C. P., Shaw, E. M., Haines, B. J., Cullen, W. R., Boyle, J. C., Penney, D., Oldekop, J. A., Ashcroft, C., Lockey, L., & Preziosi, R. F. (2005). Peer assessment of oral presentations: Effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment & Evaluation in Higher Education*, *30*(1), 21–34. https://doi.org/10.1080/0260293042003243878

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

Lim, W., Son, J.-W., & Kang, S.-H. (2021). How reducing discomfort impacts peer assessments of preservice teachers. *Sustainability*, *13*(11), 6435. https://doi.org/10.3390/su13116435

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education*, *6*, 720195. 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

Moss-Racusin, C. A., & Rabasco, H. (2018). Reducing gender identity bias through imagined intergroup contact. *Journal of Applied Social Psychology*, *48*(8), 457–474. https://doi.org/10.1111/jasp.12525

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In D. Jonassen, M. J. Spector, M. Driscoll, M. D. Merrill, J. van Merriënboer,& 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. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review*, *23*(1), 7–26.

Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, 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

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*, *31*(10), 6302–6316. https://doi.org/10.1080/10494820.2022.2034887

Noroozi, O., Hatami, J., Bayat, A., van Ginkel, S., Biemans, H. J. A., & Mulder, M. (2020). 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

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

Ocampo, J. C., Panadero, E., Zamorano, D., Sanchez-Iglesias, I., & Díez, F. (2023). The effects of gender and training on peer feedback characteristics. *Assessment & Evaluation in Higher Education*.

Ong, J. (2022, September 4). *Trolls for hire in Philippines: The concealed political weapon used in a social media war* [Documentary]. Channel News Asia (CNA). https://www.channelnewsasia.com/cna-insider/paid-troll-army-hire-philippines-social-media-elections-influencers-2917556

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. (2023). Toward a paradigm shift in feedback research: Five further steps influenced by self-regulated learning theory. *Educational Psychologist*, *58*(3), 193–204. https://doi.org/10.1080/00461520.2023.2223642

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., 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., Romero, M., & Strijbos, J.-W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, *39*(4), 195–203. https://doi.org/10.1016/j.stueduc.2013.10.005

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

Peters, O., Körndle, H., & Narciss, S. (2018). Effects of a formative assessment script on how vocational students generate formative feedback to a peer's or their own performance. *European Journal of Psychology of Education*, *33*(1), 117–143. https://doi.org/10.1007/s10212-017-0344-y

Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. *Computers in Human Behavior*, *26*(2), 151–161. https://doi.org/10.1016/j.chb.2009.10.011

Prinsen, F. R., Volman, M. L. L., Terwel, J., & van den Eeden, P. (2009). Effects on participation of an experimental CSCL-programme to support elaboration: Do all students benefit? *Computers & Education*, *52*(1), 113–125. https://doi.org/10.1016/j.compedu.2008.07.001

Rotenberg, K. J. (1986). Same-sex patterns and sex differences in the trust-value basis of children's friendship. *Sex Roles*, *15*(11), 613–626. https://doi.org/10.1007/BF00288218

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

Schwalm, K. (2007). *Rubric for the assessment of the argumentative essay*. https://web.gccaz.edu/english/assessment/spring07/final/argumentrubric.htm

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

Strijbos, J.-W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the Feedback Perceptions Questionnaire. *Studies in Educational Evaluation*, *68*, 100980. https://doi.org/10.1016/j.stueduc.2021.100980

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

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

Wichmann, A., Funk, A., & Rummel, N. (2018). Leveraging the potential of peer feedback in an academic writing activity through sense-making support. *European Journal of Psychology of Education*, *33*(1), 165–184. https://doi.org/10.1007/s10212-017-0348-7

Wood, J. (2022). Making peer feedback work: The contribution of technology-mediated dialogic peer feedback to feedback uptake and literacy. *Assessment & Evaluation in Higher Education*, *47*(3), 327–346. https://doi.org/10.1080/02602938.2021.1914544

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

Zou, Y., Schunn, C. D., Wang, Y., & Zhang, F. (2018). Student attitudes that predict participation in peer assessment. *Assessment & Evaluation in Higher Education*, *43*(5), 800–811. https://doi.org/10.1080/02602938.2017.1409872

Supplementary Material 1: Peer Feedback Content

Type of Feedback	Rubric Criteria	Example Feedback Statements	Feedback Component
Concise General Feedback	Completeness of Ideas and Depth of Arguments	You scored between (SCORE) to (SCORE). You discussed your arguments and views excellently.	KR + KM
		You scored between (SCORE) to (SCORE). The views and arguments presented are acceptable.	KR + KM
		You scored between (SCORE) to (SCORE). The arguments and ideas were vague.	KR + KM
		You scored between (SCORE) to (SCORE). You did not answer the guide question.	KR + KM
	Use of Evidence	You scored between (SCORE) to (SCORE). You cited excellent evidence.	KR + KM
		You scored between (SCORE) to (SCORE). Some relevant and reliable evidence was used.	KR + KM
		You scored between (SCORE) to (SCORE). Some references were questionable.	KR + KM
		You scored between (SCORE) to (SCORE). No evidence was used.	KR + KM
	Writing and Paragraph Structure	You scored between (SCORE) to (SCORE). The structure of your paper is great.	KR + KM
		You scored between (SCORE) to (SCORE). Your paper's structure is mostly concise.	KR + KM
		You scored between (SCORE) to (SCORE). The ideas in the paper are disorganised.	KR + KM
		You scored between (SCORE) to (SCORE). The essay is unclear and disorganised.	KR + KM
	Spelling and Syntax	You scored between (SCORE) to (SCORE). Spelling and grammar are excellent.	KR + KM
		You scored between (SCORE) to (SCORE). There are a few grammar and spelling mistakes.	KR + KM
		You scored between (SCORE) to (SCORE). There are several grammar and spelling errors in the essay.	KR + KM
		You scored between (SCORE) to (SCORE). There are numerous spelling and grammar mistakes.	KR + KM
Elaborated Specific Feedback	Completeness of Ideas and Depth of Arguments	You scored between (SCORE) to (SCORE). You answered the guide question really well, but I think you could still improve the ideas in your argument about …. in paragraph … to strengthen your paper.	KR + KH
		You scored between (SCORE) to (SCORE). Your arguments and counter arguments are well developed, but you might want to summarise these points before you move to the next paragraph (see for example paragraph…).	KR + KH + KM
		You scored between (SCORE) to (SCORE). Your arguments about social media registration are acceptable, but you might want to acknowledge the opposing views to your argument (see for example…).	KR + KH + KM
		You scored between (SCORE) to (SCORE). You did not answer the guide question, so I suggest you watch the documentary again so you can pick up ideas about how social media registration can be used to prevent trolls.	KR+ KM + KH

Use of Evidence	You scored between (SCORE) to (SCORE). You used reliable and relevant refences in your paper, but you might want to integrate them with your arguments to form a more solid paper.	KR + KH
	You scored between (SCORE) to (SCORE). You mentioned some evidence about registering one's social media accounts; perhaps you can write a short discussion about its potential effects if not controlled.	KR + KH
	You scored between (SCORE) to (SCORE). The evidence you used should be checked to see if it is reliable (especially …); also use the given format in citing your refences.	KR + KH
	You scored between (SCORE) to (SCORE). Because you did not use any evidence or references, I suggest you find similar cases online and discuss their effects (registration of social media accounts) to strengthen your arguments.	KR+ KM + KH
Writing and Paragraph Structure	You scored between (SCORE) to (SCORE). Your structure is already concise and organised, but I suggest that you use transitional words for better paragraph flow.	KR + KH
	You scored between (SCORE) to (SCORE). The connection between your first and second paragraph is a bit confusing, so try restructuring it by using …	KR + KH + KM
	You scored between (SCORE) to (SCORE). Paragraphs … and … are a bit long, try to get straight to the point and be specific.	KR + KH + KM
	You scored between (SCORE) to (SCORE). You might want to use an outline when you revise your paper because the ideas were unclear and disorganised (See …)	KR + KM + KH
Spelling and Syntax	You scored between (SCORE) to (SCORE). Your grammar and spelling were consistent, but you could improve them further by using fewer passive sentences (see paragraph …, sentence …) .	KR + KH + KM
	You scored between (SCORE) to (SCORE). Spelling and syntax were good, so I would suggest proofreading or double checking if there are any mistakes (ex. … sentence in paragraph … sounds incorrect).	KR + KH + KM
	You scored between (SCORE) to (SCORE). Your spelling and syntax were okay, so I would suggest being consistent with your spelling (some words were in British English and some were in American English).	KR + KH + KM
	You scored between (SCORE) to (SCORE). You might want to consider consulting your friends to proofread your paper, because there were major grammatical errors.	KR + KM + KH

Legend (Narciss, 2008; Strijbos et al., 2010):

- **KR (Knowledge of Result/Response)** — Provides learners with information on the correctness of their actual response output (e.g., correct or incorrect)
- **KM (Knowledge of Errors/Mistakes)** — Provides learners with information on errors or mistakes in the output.
- **KH (Knowledge on how to Proceed**) – Provides learners with procedural knowledge relevant for task processing

Supplementary Material 2: Peer Assessment Design Characteristics Questionnaire

Instrument to report the characteristics of peer assessment designs

Designed by Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education*. **Use that citation if you include the instrument in your publication.**

TITLE OF THE ARTICLE

"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

Authors

ANONYMISED

Our study investigates:

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

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

	Category	Description [i]	Our study
Context			
1	**Subject domain**	Subject domain the study was done in (e.g., mathematics, instructional sciences, accounting, etc.)	Compulsory general education course: Understanding the Self
2	**Place/Time**	Where was the PA conducted? (In class or out of class?)	☐ In class/class time ☒ Out of class/free time
3	**Setting**	Formal or informal education setting?	☒ Formal ☐ Informal
4	**Requirement**	Was PA compulsory or voluntary for assessor/assessee?	☒ Compulsory ☐ Voluntary
5	**Alignment**	Was the PA activity aligned to curriculum, learning goals or teaching?	☒ Yes ☐ No
Instructional design			
6	**Purpose**	What was the assessment purpose of the PA activity? (Formative, summative or both?)	☒ Formative ☐ Summative ☐ Both
7	**Object**	What was assessed? (e.g., written assignment, oral presentation, contribution to group work)	Argumentative essays.
8	**Product/Output**	What was the output of the PA? (e.g., score, written feedback, oral feedback, or a combination)	Written peer feedback.
9	**Relation to staff assessment**	Was PA done without staff assessment (substitutional) or in addition to staff assessment (supplementary)?	☐ Substitutional ☒ Supplementary
10	**Official weight**	Did participation in the PA activity or the grade given by peer(s) contribute to learners' final grades?	☐ No ☒ Yes-for participation in PA ☐ Yes-for PA grade ☐ Yes-both (PA & participation) ☐ Other: Click here to add text
11	**Reward**	Was there a reward for participation in PA?	☐ No ☒ Yes-course credit ☐ Yes-incentives (e.g., free time, money, etc.) ☐ Other: Click here to add text
12	**Directionality**	Was the learner assessing another without being assessed (unidirectional) or acting as both assessor and assessee (bidirectional)?	☒ Unidirectional ☐ Bidirectional
13	**Degree of interactivity**	How did the assessee demonstrate engagement and response to PA?	☒ Reactive: assessee responds to assessor ☐ Reciprocal: same learners assess each other on same task ☐ Negotiated: PA was done more than once on the same task and both parties negotiated it
14	**Frequency**	How often was the PA of the same task done? (Once, twice, etc.)	Once.
15	**Group constellation**	Did members of the same group assess each other (intragroup) or peers from another group (intergroup) or both?	☒ Intragroup ☐ Intergroup ☐ Both
16	**Constellation assessor**	The number of assessors assigned to each assessee	The first author acted as a fictious man, woman, and anonymous peer assessor to all assessees. One fictitious assessor to 22 to 25 asessees (depending on the condition assignment).
17	**Constellation assessee**	The number of assessees per assessor	Approximately 22 to 25 assessees per fictious assessor (depending on the condition assignment).

18	**Unit of assessment (assessor)**	At what level did the assessor(s) perform PA? Individual, group, or both?	☒ Individual ☐ Group ☐ Both
19	**Unit of assessment (assessee)**	At what level did the assessee(s) experience the PA? Individual, group, or both?	☒ Individual ☐ Group ☐ Both
20	**Privacy**	Did assessor and assessee know the identity of each other? (Was PA public, single-blind, or double-blind?)	☐ Public ☐ Single-blind (for assessor) ☐ Single-blind (for assessee) ☐ Double blind (anonymous) ☒ Other: Fictitious assessor gender was known to the assessee
21	**Contact**	Was PA done face-to-face or online? How was the contact between assessor and assessee?	☐ **Face-to-face synchronous:** same time same place (ST SP) ☐ **Online synchronous:** same time different place (ST DP), e.g., skype, chat, etc. ☒ **Online asynchronous:** different time different place (DT DP), e.g., e-mail, SWoRD, etc. ☐ **Other:** Click here to add text
22	**Matching**	How were assessor and assessee matched for the PA activity? (e.g., random, skill, self-select, gender, friendship, performance, etc.)	Gender matched. (Fictional anonymous assessor: men and women assessees; fictional man assessor: men and women assessees; fictional woman assessor: men and women assessees).
23	**Format**	How was the PA guided?	☐ Freestyle (no instruction provided to learers) ☐ Guided instruction ☐ Guided prompts ☐ Guided criteria/rubrics ☒ Guided criteria/rubrics and prompts ☐ Other: Click here to add text
24	**Training**	Did the learners receive PA training at any time? If yes, describe the moment in which they received it and the type of training.	Students did not receive training. As students only received peer feedback from fictional assessors, an overview of peer assessment, which explained its benefits and the rubrics that will be used in assessing their work were discussed to them.
25	**Revision**	Did learners revise their work after receiving or providing PA?	☐ No ☒ Yes
26	**Scope of involvement**	How were learners involved in the PA activity?	☐ Development of assessment criteria ☐ Gave PA only ☒ Received PA only ☐ Gave and received PA ☐ Additional self-assessment ☐ Other: Click here to add text
Outcomes			
27	**PA Outcomes**	These variables are directly measured as outcomes of the PA activity (i.e., why was the PA activity conducted?) Select all the options that apply to your study from the right column.	☒ **Beliefs & perceptions:** including perceptions of learning capacity to perform PA or any perceptions related to the PA processes (e.g., fairness, usefulness), metacognition and self-regulation, attitudes and beliefs (e.g., self-efficacy), teachers' perceptions/conceptions. ☒ **Emotions and motivation:** emotions experienced by learners (e.g., achievement emotions, social emotions, etc.) & all motivational beliefs (e.g., learning motivation). ☒ **Performance:** academic/domain specific performance, achievement, improved draft/work (i.e., revision). ☐ **Skills:** quality of contribution to the group, professional behaviour, problem solving skills, work habits, interpersonal skills, metacognitive & self-regulatory skills. ☐ **Reliability of PA:** consistency of PA scores compared to other peer assessors, or PA over several rounds. ☐ **Validity of PA:** accuracy of PA compared to teachers/tutor/expert's assessment. ☐ **PF content:** characteristics of the feedback messages including type, focus, quality, frequency of comments/posts (i.e., participation), etc. ☒ **PF processing:** Includes implementation, reactions to PF, seeking PF, coping with PF, etc. ☐ **Other:** Click here to add text
Moderators/mediators			
28	**PA Moderators/mediators**	Variables that are not usually manipulated but are taken into account when investigating PA. Select the variables that have been explored in your study from the right column.	☒ **Gender:** of assessor/assessee. ☐ **Ability & Skills:** includes prior knowledge, prior performance, achievement level, GPA, finished high school, previous level of education, year of enrolment, etc. ☐ **Skills:** reviewing ability, computer skills, etc. ☐ **Age:** of assessor/assessee. ☐ **Culture, ethnicity, nationality or race** ☐ **Other:** Click here to add text
This tool is based on: Alqassab, M., Strijbos, J., Panadero, E., Fernández Ruiz, J., Warren, M., & To, J. (2023). A systematic review of peer assessment design elements. *Educational Psychology Review*.			

Note. PA = peer assessment; PF = peer feedback

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

Supplementary Material 3: Questionnaire Items

Interpersonal Factors

A. Trust in the peer as an assessor —4 items (Ching & Hsu, 2016; 1 = Strongly Disagree to 5 = Strongly Agree)

1. I am satisfied with the overall quality of the feedback I've received from peers.
2. My peers provided sufficient amount of feedback on my essay.
3. The peer feedback I received was helpful to improve my essay.
4. Peers have adequate knowledge to comment on my essay.

B. Perceived Adequacy of Feedback (Composite score: Fairness, usefulness, acceptance)—9 items (PFPQ; Strijbos et al., 2010, 2021)

1. I would be satisfied with this feedback. (Fairness)
2. I would consider this feedback fair. (Fairness)
3. I would consider this feedback justified. (Fairness)
4. I would consider this feedback useful. (Usefulness)
5. I would consider this feedback helpful. (Usefulness)
6. This feedback would provide me a lot of support. (Usefulness)
7. I would accept this feedback. (Acceptance)
8. I would dispute this feedback. (Acceptance)
9. I would reject this feedback. (Acceptance)

Intrapersonal Factors

C. Discomfort—1 item (Panadero et al., 2013)

1. I felt comfortable after receiving feedback on my argumentative essay from my peer.

D. Motivation (Willingness to Improve)—3 items (PFPQ; Strijbos et al., 2010, 2021)

1. I would be willing to improve my performance.
2. I would be willing to invest a lot of effort in my revision.
3. I would be willing to work on further argumentative essay assignments.