

Does anonymity matter? Examining quality of online peer assessment and students' attitudes

Michiko Kobayashi

Southern Utah University

The study investigated the effects of anonymity on online peer assessment and compared three different conditions. Fifty-eight preservice teachers at a mid-size US university engaged in a series of online peer assessments during fall 2017. Peer assessment was embedded in a blended course as a required asynchronous activity using the Canvas learning management system. Students were randomly assigned to three different peer assessment conditions: anonymous, partially anonymous, and identifiable. They were asked to provide feedback comments and rate the quality of peers' work. The researcher examined to what extent three different conditions had influenced the quality of feedback comments, measured quantitatively through the number of words and negative statements. At the end of the semester, a survey that included a 5-point Likert scale and several open-ended questions was also distributed to analyse students' perceptions about peer assessment and anonymity. The results indicate that although students prefer anonymity, it may not be a necessary condition for increasing student engagement.

Implications for practice or policy:

- Instructors should provide several opportunities for online peer assessment to reinforce the skill of writing effective feedback throughout a course.
- Students may be given an option for anonymous peer feedback to ease their anxiety.
- Providing specific grading criteria for feedback quality is strongly recommended.

Keywords: peer assessment, anonymity, online feedback, quantitative, teacher education

Introduction

Assessment is an integral part of teaching and learning. In traditional classrooms, only instructors assess students' work, provide feedback, and assign grades. Peer assessment has changed the old view. By adopting peer assessment, students become active participants in their own learning processes (Falchikov, 2013). Reviewing others' work also motivates students to improve the quality of their own work (Nicol, Thomson, & Breslin, 2014). Peer assessment is particularly beneficial for pre-service teachers. To improve teaching skills, they are encouraged to critique each other's lessons and share ideas (Özdemir, 2016). Wilkins, Shin, and Ainsworth (2009) also define peer assessment as "reciprocal teaching in which paired teacher candidates provide assistance to one another as they incorporate new teaching skills" (p. 80). Thus, for pre-service teachers, peer assessment is not merely evaluating others' work, but rather viewed as collaborative learning. Research has shown that peer assessment training for pre-service teachers develop their professionalism and improve teaching performance (Koç, 2011). Now, a variety of learning management systems (LMS) allow instructors to integrate peer assessment easily (Sullivan & Watson, 2015). The advancement of technology has boosted the potential of peer assessment and its integration into college classrooms.

While many researchers recognise the pedagogical value of online peer assessment, not all students perceive its benefit (Cheng, Liang, & Tsai, 2015). Past studies have shown that anonymity in online peer assessment can reduce students' negative attitudes and increase critical feedback (Howard, Barrett, & Frick, 2010; Vanderhoven, Raes, Montrieux, Rotsaert, & Schellens, 2015). Using the current technology, anonymous peer reviews can be easily assigned. However, earlier research has reported several concerns regarding anonymity (Postmes, Spears, Sakhel, & de Groot, 2001). Anonymous comments may promote distrust among their peers, especially when negative reviews are received. This raises the question of whether or not anonymous peer assessments are necessarily the best instructional approach. To date, many researchers have studied the effects of anonymity in peer assessment using both quantitative and qualitative methods (Panadero, 2016). However, there is still not sufficient evidence to generalise the findings, as assessment formats and learning environments are becoming increasingly varied. Hence, the present study

aims to extend existing literature in online peer assessment. The following section will provide a review of past research findings in relation to online peer assessment and the effects of anonymity.

Literature review

Online peer assessment is gradually replacing traditional face-to-face peer assessment as LMSs have become widely available in universities. While audio or video recorded peer feedback is emerging, most peer assessment is implemented in an asynchronous writing format (Van Popta, Kral, Camp, & Martens, 2017). Unlike peer assessment in physical classrooms, asynchronous peer assessment allows students more time for reflecting and organising their thoughts (Pena-Shaff, Altman, & Stephenson, 2005). This can produce higher quality feedback than those provided in real-time. Research shows that in an undergraduate classroom, students tend to prefer online peer feedback over face-to-face feedback because peer feedback received in-person are less critical or helpful than online feedback (McCarthy, 2017). Online peer assessment is also beneficial to graduate students taking online courses as they generally have less opportunity for participating in real-time interaction with other students (Yang, 2016). Furthermore, Van Popta et al. (2017) maintain that online peer assessment benefits not only the recipient of feedback but also the provider of feedback, if students are required to use high level of cognitive skills, involving “an evaluative judgement, a suggestion for improvement, and an explanation” in their written feedback (p. 32).

Along with LMSs, social media are also used as a platform for online peer assessment (Cheng et al., 2015; Shih, 2011). Unlike LMSs, social media allow students to access people outside the classroom and promote global collaboration (McCarthy, 2012). Demir (2018) examined student teachers’ perceptions about peer assessment using Facebook and found that Facebook promoted objective feedback and students’ engagement. Furthermore, Cheng, Hou, and Wu (2014) used YouTube for peer assessment in a college class and examined students’ emotional responses and participation. They found that students receiving positive feedback demonstrate a higher level of participation in a peer feedback activity. Wikis are another online program used for asynchronous peer assessment (Gielen & De Wever, 2015; Peled, Bar-Shalom, & Sharon, 2014). Peled et al. (2014) found that females are less comfortable with giving and receiving feedback in Wiki environments than males. In contrast, a study by Cheng et al. (2014) showed that gender does not influence students’ level of participation in an online peer feedback activity.

To date, many studies have documented the positive impact of online peer assessment on student learning (Panadelo, 2016). However, when it comes to students’ attitudes, they are mostly negative. Wilson, Diao, and Huang (2015) found that many students do not trust in the peer’s ability of assessing their work and feel that peer assessment is unfair. Another study showed that students often give lower marks to peers after they received lower scores than they expected (Lin, Liu, & Yuan, 2001). Workload also contributes to students’ negative feelings; students tend to perceive peer assessment as just additional work, rather than meaningful activities (Wilson et al., 2015). In addition, lack of training/experience inhibits peer assessment (Kilickaya, 2017). For example, some students with limited experience or knowledge are not comfortable with evaluating other students’ work. Another concern is over-scoring due to a friendship bias (Panadero, Romero, & Strijbos, 2013). Students are reluctant to give lower scores to their peers because they don’t want to lose their friendship (Kilickaya, 2017). Panadero et al. (2013) found that the use of rubrics increased the validity of peer assessment and students’ performance. However, the rubrics only reduced biases of low and moderate level friendships. Students with a higher level of friendship with the assessors still tended to over-score significantly more than those with a lower level friendship (Panadero et al., 2013). To ease such social pressure and students’ negative attitudes, researchers’ interest has been directed towards investigating the effects of anonymity/identifiability on online peer assessment (Li, 2017).

Anonymity enhances perceived psychological safety (Zhang, Fang, Wei, & Chen, 2010). Psychological safety refers to “a feeling able to show and employ one’s self without fear of negative consequences to self-image, status, or career” (Kahn 1990, p. 703). This explains why anonymous peer feedback tends to bring more favourable outcomes than non-anonymous conditions. Consistent with this theory, past research shows that the anonymity of assessors increases students’ positive perceptions about peer assessment (Lin, 2018; Vanderhoven et al., 2015) and reduces social pressure, which helps students focus on their task (Howard et al., 2010). Furthermore, several studies showed that students in an anonymous condition tend to produce more critical feedback (Lu & Bol, 2007) and demonstrate a higher level of academic performance (Li, 2017) than those in non-anonymous settings. Lin (2018) recently reported that students’ perceived learning in the anonymous group was significantly higher than those in the non-anonymous group.

Despite the benefits reported in past research, anonymity can also cause problems. In online communications, anonymity has a potential risk of promoting anti-social behaviors (Postmes et al., 2001). Zhao (1998) also argues that anonymity increases social loafing, which allows students to put less effort into their assigned tasks. Moreover, research showed that students' perceived fairness in anonymous peer assessment is significantly lower compared to that in a non-anonymous condition (Lin, 2018). In addition, Yu and Wu (2011) examined fifth graders' perception about assessors in different identity modes: real name, anonymity, nickname, and user self-choice. They found that students viewed assessors in the real-name and self-choice groups more positively than those in the anonymous and nickname groups. Thus, the attitude towards anonymity may vary depending on the age group.

While the debate on anonymity continues, the following studies suggest that under certain conditions, the effect of anonymity may be weakened or may not be necessarily for promoting effective online peer feedback. Liang and Tsai (2010) have examined a series of online peer assessment activities in an anonymous condition and found that the validity of peer scores improved (became closer to the instructor's scores) as students had gone through more rounds of peer assessment. Furthermore, Rotsaeart, Panadero, and Schellens (2018) found an anonymous condition in peer assessment at the beginning stage can serve as a scaffold and improve the quality of feedback over time. In their study, students engaged in several peer assessment activities for 4 weeks using a mobile response technology in face-to-face classrooms. The first 2 weeks were anonymous and then feedback switched to a non-anonymous condition. Rotsaeart et al. (2018) found the quality of student feedback does not decrease even after switching to non-anonymous condition. Moreover, students became less concerned about anonymity after completing several anonymous peer assessment sessions. These two studies together indicate that repeated practices improve the quality of online peer feedback and the assessor's skills for assessing peers, however, the scaffolding effect of anonymity on the quality of feedback needs to be examined further because in the study by Rotsaeart et al. (2018), data were collected from a single group.

This study further investigates the effect of anonymity on the quality of feedback and the student's attitude towards peer assessment. More specifically, based on the study by Rotsaeart et al. (2018), the researcher re-examined the scaffolding effect of anonymity in an asynchronous setting where students engage in peer assessment individually outside class. In contrast to the Rotsaeart et al. (2018) study, in this study students were assigned to three different groups. One group used only anonymous (A) peer evaluations. The partially anonymous (PA) group used a combination: only the first two peer assessments were anonymous, then they switched to identifiable. The other group used only identifiable (non-anonymous) (ID) peer evaluation to compare the quality of feedback among these three groups. It is hoped that this study helps to fill in the gap in the current literature and contributes to research on online peer assessment strategies.

Research questions

The main purpose of this study was to investigate how varied peer assessment conditions affect students' quality of peer feedback. The attitudes towards peer assessment and anonymity, as well as their relationships with demographic factors and the quality of feedback were also examined. Specific research questions included:

1. Does the quality of online peer feedback differ significantly among three conditions: anonymous, partially anonymous, and identifiable?
2. Does the quality of online peer feedback change significantly across four different data collection points during the semester?
3. What are the students' attitudes towards peer assessment and anonymous feedback?
4. Are there significant relationships among the students' attitudes towards online peer assessment and anonymity, the quality of feedback, and demographic factors?

Methods

Participants of the study were education major students enrolled in two sections of a multicultural education course at a rural, mid-size US university. In addition to face-to-face class lectures, the instructor also used Canvas, the university's LMS. The total of 58 students participated in the study. Of the 58, only three students were male. About two-thirds were aged 18-20 years, and either sophomores or juniors. More than

70% of students had prior experience with peer assessment (Table 1). The researcher obtained approval from the University Institutional Review Board (IRB) for conducting this study.

Table 1
Demographic characteristics of participants

Demographic Characteristics	Number (Percentage)	
Gender	Male	3 (5.2)
	Female	55 (94.8)
Age	18-20	38 (65.5)
	21-23	16 (27.6)
	24-26	1 (1.7)
	27-29	1 (1.7)
	30-32	0 (0)
	33-35	0 (0)
	36 or older	2 (3.4%)
Academic standing	Freshman	15 (25.9)
	Sophomore	17 (29.3)
	Junior	23 (39.7)
	Senior	3 (5.2)
Peer assessment experience in college courses	0	13 (22.4)
	1	6 (10.3)
	2-3	30 (51.7)
	4 or more	9 (15.5)

Video/article reflections and online peer assessment

Video/article reflections were part of the required course assignments. Using a Canvas built-in tool, students were randomly assigned to three different conditions: anonymous (A), partially anonymous (PA), and identifiable (ID). There were 17 in the A group, 17 in the PA group, and 19 in the ID group. To avoid possible effects on the quality of their peer reviews, students were not informed which group they were assigned to. In the video/article reflection assignments, students were asked to view online videos and read articles related to the course content, then write reflections based on the instructor's prompts. After they posted their reflections, the instructor randomly assigned each student two peer reviews through Canvas's auto feature. Canvas has a capability to automatically assign peer reviews and options to choose anonymous or non-anonymous peer review settings. For the PA group, the first two peer reviews were anonymous, and the last two were identifiable. At the beginning of the semester, students received instructions on how to enter scores in the grading rubrics and provide constructive feedback. Peer assessment was required and included in the grading component for these assignments. A new topic was posted every 2 or 3 weeks. Four video/article reflection topics were given throughout the semester, and each student completed eight peer assessments in total.

Measuring quality of online peer assessment

The success of peer assessment depends on the quality of peer feedback (Lin, 2018). There are several ways to assess the quality of peer feedback. In this study, the quality of peer feedback was measured based on the total number of words in feedback comments obtained from the video/article reflection assignments and the extent of critical feedback. The total number of words in feedback comments is considered as a quality indicator because it reflects the level of engagement based on the amount of time and effort students spent to construct meaningful feedback (Howard et al., 2010). The instructor did not assign a minimum number of words. The extent of critical feedback is assessed using the percentages of students' negative comments: identifying weaknesses and/or providing suggestions for improvement. An example of students' negative comments is: "You could have included more details about how you could touch on this subject even more." As discussed earlier, previous research shows that students tend to provide fewer critical comments in a non-anonymous condition (Lu & Bol, 2007). As such, this study used the percentages of critical feedback as another quality indicator. In addition, the percentages of positive and neutral comments from each

feedback session were also examined to see if the ratio of those two types of comments differs across the three groups. Positive comments ranged from simple praise, such as “Good job!” to more descriptive comments like “Your responses to the questions were very well thought out and enjoyable to read.” Neutral comments are relevant to the topic, but convey neither positive nor negative reactions of the assessor. Some might state their own opinions and share personal experiences related to the topic. An example of neutral comments is: “It is so important that as teachers we don’t go into the classroom with pre-assumptions.” The researcher and a research assistant coded students’ feedback comments independently. After the first set of data with 434 utterances coded, the inter-rater reliability was assessed using Cohen’s Kappa (Stemler, 2001). There was a substantial level of agreement between the two sets of coded data, $\kappa = .833$, $p < .001$. Thus, the raters continued coding the remaining data.

Measuring attitudes towards peer assessment

To assess students’ attitudes towards peer assessment, a survey was distributed at the end of the semester. This survey was originally created by McGarr and Clifford (2013) and used to assess both undergraduate and graduate students’ attitudes towards peer assessment. Their survey consisted of 17 Likert items and four open-ended questions. The researcher adopted 16 Likert items and 3 open-ended questions that fitted the peer assessment format in the current study. To examine students’ attitudes towards anonymity, three new items were also added to the Likert survey. The scale used in this study was in a 5-point format, ranging from *strongly disagree* to *strongly agree*. This survey also included several questions about students’ demographics. All 58 students completed the survey online. The Cronback’s alphas for the attitude towards peer assessment (16 items) and the attitudes towards anonymity (3 items) were .85 and .76 respectively.

Data analysis

This study adopted an experimental design and a mixed methods approach for data analysis. Quantitative data were analysed using the Statistical Package for the Social Sciences (SPSS) IBM version 22. To examine the effect of three condition groups: A, PA, and ID, across four peer assessment sessions, a two-way mixed design ANOVA was employed. The extent of positive, negative, and neutral feedback was compared to answer research questions 1 and 2. This statistical procedure is suitable for the current study because it allows comparison between two or more independent groups over time and examines the interaction between two independent variables: group conditions and time (Kirk, 2013). To answer research question 3, the descriptive analysis of the attitude survey was conducted. Students’ narrative responses to the three open-ended questions were also analysed to augment the quantitative data. For research question 4, correlation analysis was conducted to examine the relationship between students’ demographic factors and quality of feedback, and attitudes.

Results

Comparison of the quality of online peer assessment

Table 2 shows descriptive analysis of the number of words in online feedback. Overall, PA demonstrated the highest number of words in feedback among all three groups ($M = 126.4$; $SD = 38.5$). The number of words in feedback varied depending on the feedback session, and all groups showed the highest number of words in Session 3. The number of words and percentages of each type of feedback: positive, negative, and neutral were listed in Table 3. The results showed that in all three groups, the amount of negative comments was less than positive and neutral comments. Students wrote more negative comments in Sessions 1 and 2 than they did in Sessions 3 and 4. Overall, Group ID produced the highest percentages of negative feedback.

Table 2
Number of words in online feedback

	Topic 1 <i>M (SD)</i>	Topic 2 <i>M (SD)</i>	Topic 3 <i>M (SD)</i>	Topic 4 <i>M (SD)</i>	Total <i>M (SD)</i>
A ($n = 17$)	106.6 (58.8)	124.7 (65.6)	128.4 (43.7)	126.4 (57.9)	121.5 (49.5)
PA ($n = 17$)	106.5 (58.0)	125.3 (59.1)	147.9 (52.5)	126.1 (39.0)	126.4 (38.5)
ID ($n = 19$)	107.7 (62.6)	111.1 (50.2)	136.6 (50.7)	116.4 (50.0)	117.9 (44.6)
Total ($N = 53$)	107.0 (58.8)	120.0 (57.6)	137.6 (48.9)	123.0 (48.8)	121.8 (43.7)

Note. A: anonymous group; PA: partially anonymous group; ID: identifiable group.

Table 3
 Total number of words and percentage of positive, negative, and neutral online feedback

		Topic 1	Topic 2	Topic 3	Topic 4	Total
A (<i>n</i> = 17)	Positive	922 (50.9%)	1134 (53.2%)	947 (42.9%)	1047 (48.7%)	4050 (48.8%)
	Negative	81 (4.5%)	67 (3.1%)	26 (1.2%)	12 (.6%)	186 (2.2%)
	Neutral	809 (44.6%)	932 (43.7%)	1233 (55.9%)	1089 (50.7%)	4063 (49.0%)
PA (<i>n</i> = 17)	Positive	776 (42.8%)	1247 (58.5%)	1049 (41.7%)	1171 (54.6%)	4243 (49.3%)
	Negative	156 (8.6%)	278 (13.1%)	92 (3.7%)	116 (5.4%)	642 (7.5%)
	Neutral	879 (48.5%)	605 (28.4%)	1373 (54.6%)	856 (39.9%)	3713 (43.2%)
ID (<i>n</i> = 19)	Positive	960 (46.9%)	1196 (57.2%)	1092 (42.1%)	1190 (53.8%)	4438 (49.6%)
	Negative	266 (13.0%)	317 (15.2%)	121 (4.7%)	131 (5.9%)	835 (9.3%)
	Neutral	821 (40.1%)	577 (27.6%)	1382 (53.3%)	891 (40.3%)	3671 (41.0%)

Note. A: anonymous group; PA: partially anonymous group; ID: identifiable group. The percentages were calculated based on the number of words in each type of comment.

Two-way mixed design ANOVA

A two-way mixed-design ANOVA was also conducted to examine the effect of group conditions (A, PA, and ID) on the number of words in online feedback, across four data collection points. Prior to the analysis, required assumptions for variables were inspected for each group. The original data for the number of words in online feedback failed to meet the normality assumption, and several outliers were also found. Because the sample size was small, the researcher decided to transform the data, instead of removing the outliers. The transformed data met the assumption of normal distribution, as assessed by Shapiro-Wilk test ($p > .05$) and Normal Q-Q Plot. The z scores calculated using skewness and kurtosis were also within or close to the acceptable ranges, ± 2 (Gravetter & Wallnau, 2005). Mauchly's test of sphericity indicated that the assumption of sphericity was violated, $\chi^2(5) = 12.100, p = .033$; therefore, the Greenhouse-Geisser estimate was used for analysing within-subject effects. There was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices ($p = .332$). Homogeneity of variances for the number of words in feedback was confirmed based on Levene's test for equality of variances ($p > .05$).

The results showed that there was no significant interaction between the group conditions and the data collection points, $F(5.147, 128.671) = .384, p = .864$, partial $\eta^2 = .015$. There was a significant main effect for the data collection points, $F(2.573, 128.671) = 9.410, p < .001$, partial $\eta^2 = .158$. Post hoc comparisons revealed that there were significant differences between Sessions 1 and 3 ($p = .001$); Sessions 1 and 4 ($p = .025$); Sessions 2 and 3 ($p = .022$). This indicated that students wrote more feedback in later sessions than they did in earlier sessions. However, there was not a significant main effect for the group conditions, $F(2, 50) = .178, p = .838$, partial $\eta^2 = .007$ suggesting no difference of the three condition groups, in terms of the number of words in online feedback.

For the extent of critical feedback (negative comments), ANOVA was not conducted because few negative comments were found in students' feedback across all groups, and the same students tended to write negative feedback throughout the four online feedback sessions. Nevertheless, based on the descriptive data shown in Table 2, anonymity did not appear to increase the amount of critical feedback.

Attitudes towards peer assessment and anonymity

Students' attitudes towards peer assessment and anonymity were assessed using a survey that consisted of a 5-point Likert scale and three open-ended questions. For the Likert scale, Items 1, 2, 4, 7, 8, 9, 11, 12, 13, 14, and 18 were reverse-coded, and higher scores denoted more positive attitudes. Table 4 shows the frequencies, means, and standard deviations for each item. Overall, students' attitudes towards peer assessment were slightly positive ($M = 3.39$, $SD = .062$). About 75% of students felt that peer assessment has an educational value. An 80% was confident in their skills and knowledge to assess peers. However, about half of the students expressed that they were reluctant to be critical of their peers and give low marks, and also felt that peer assessment should not affect their overall grades too much. For the attitude towards anonymity, about 80% of students expressed that peer assessment should be anonymous.

Table 4
Frequencies of responses in the attitude survey

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	<i>M</i> (<i>SD</i>)
1. Nervous about peer assessment.	20.7%	46.6%	24.1%	8.6%	0%	3.79 (.87)
2. Limited educational value.	12.1%	63.8%	13.8%	6.9%	3.4%	3.74 (.89)
3. Enjoyed being peer assessed.	3.4%	17.2%	24.1%	53.1%	1.7%	3.33 (.91)
4. Reluctant to be critical to peers.	0%	29.3%	19.0%	43.1%	8.6%	2.69 (.99)
5. A fairer assessment method.	3.4%	13.8%	46.6%	34.5%	1.7%	3.17 (.82)
6. Enjoyed assessing peers.	5.2%	13.8%	27.6%	50.0%	3.4%	3.33 (.94)
7. I did not have the skills and knowledge to assess peers.	19.0%	60.3%	17.2%	1.7%	1.7%	3.93 (.77)
8. Reluctant to give my peers low marks.	1.7%	22.4%	19.0%	48.3%	8.6%	2.60 (.99)
9. Did not like being assessed by peers.	0%	62.1%	25.9%	8.6%	3.4%	3.47 (.80)
10. Peer assessment made the assessment more accurate.	3.4%	15.5%	41.4%	36.2%	3.4%	3.21 (.87)
11. Prefer the instructor grade only.	3.4%	39.7%	36.2%	15.5%	5.2%	3.21 (.93)
12. My peers did not assess my work accurately.	8.6%	63.8%	22.4%	5.2%	0%	3.76 (.68)
13. The task of peer assessment was difficult.	15.5%	65.5%	15.5%	3.4%	0%	3.93 (.67)
14. Peer assessment is unfair.	13.8%	65.5%	15.5%	5.2%	0%	3.88 (.73)
15. Peer assessment is valuable exercise.	1.7%	10.7%	25.9%	51.7%	10.3%	3.59 (.88)
16. My peers should have a greater say in mark.	5.2%	44.8%	41.4%	6.9%	1.7%	2.55 (.78)
17. Should be anonymous.	0%	8.6%	12.1%	46.6%	32.8%	4.03 (.89)
18. Anonymous makes me uncomfortable.	19%	56.9%	19.0%	3.4%	1.7%	3.88 (.82)
19. Prefer anonymous.	1.7%	8.6%	17.2%	37.9%	34.5%	3.95 (1.01)

Note. $N = 58$. Items 1-16 assess the attitude towards peer assessment. Items 17-19 assess the attitude towards anonymity.

Students' narrative comments

In the first question: "Is there any value for the assessor in peer assessment? Please explain.", most students expressed that the assessor can learn different perspectives and deepen their understanding about the topic. Several students also mentioned that assessing peers is particularly valuable for education majors. One student said: "In this specific class, many of us are future educators who will be assessing our own students' work. So grading one another's work is good practice." Another student also commented: "As educators, we will need to provide feedback and criticism. Doing peer assessments is a great way for us to strengthen that skill in a kind manner." Although peer assessment is valuable experiences for them, giving critical feedback to peers seems to be a challenging task. One student pointed out that "[i]t is often hard to give accurate feedback in fear of offending someone or grading them based on other aspects, not just their work".

For the second question: "Is there any value for the student being assessed by his/her peers? Please explain.", students' responses were also mostly positive. One student stated: "[In peer assessments], you get input from someone who is in the same stage of life as you. If I understand something at a similar level as one of my peers then I know I am progressing at an okay rate." This suggests that peer feedback can promote self-confidence in their academic abilities. Another student commented: "Sometimes it is more nerve racking when the teacher grades, because you feel they may grade more harshly and your work vs. the whole class. Being assessed by a peer may take away some of this stress." Reducing students' anxiety about grades is also another benefit of peer assessment, which may increase student involvement in the activities. On the other hand, several students' comments suggested that peer feedback is not always helpful. One student said: "90% of the time the student doing the peer review is just trying to be nice, not actually saying anything critical."

The last open-ended question asked: "Overall, do you think that peer assessment should be included in all college level courses? Please explain why." While most students supported peer assessment, half of them said that it should not be required for all college courses. For example, they felt that peer assessment is not appropriate in math and science classes, and that it is suitable for courses that involve a lot of writing tasks. Several students also pointed out that peer assessment can be included in any college level courses, if it is correctly done. They emphasised that peer assessment should be anonymous and it should not affect the overall grades too much because not all peers are knowledgeable about the content.

Demographic factors, attitudes, and the number of words in feedback

To examine the relationships among students' demographic factors, the total number of words in online feedback, and the attitude towards peer assessment and anonymity, Kendall's tau_b and Pearson correlation tests were conducted. Data for the attitude variables were transformed due to violation of the normality assumption. The transformed data met the assumption, as assessed by Shapiro-Wilk test ($p > .05$).

The results of Kendall's tau_b tests showed that the academic standing (e.g., freshman, sophomore, junior, and senior) and the age group were significantly correlated with the total number of words in online peer feedback ($\tau_b = .285, p = .008$; $\tau_b = .282, p = .011$). Higher level of academic standing and older age students wrote more in peer feedback compared to lower academic standing and younger students. The other demographic factors, such as prior experiences with peer assessment and GPAs, were not significantly related to the total number of words in peer feedback. In addition, the results of Pearson correlation tests indicated that the attitudes towards peer assessment (items 1-16) and anonymity (items 17-19) were not related to the total number of words in peer feedback, $r = .134, p = .338$; $r = -.088, p = .530$. Lastly, it is important to note that Kruskal-Wallis H tests confirmed that the academic standing and the age group were equally distributed across three condition groups ($p > .05$). Therefore, these two demographic factors did not affect the ANOVA result reported earlier.

Discussion

Does the quality of online peer assessment differ significantly among three group conditions?

In this study, anonymity did not influence students' engagement in online peer assessment. Consistent with a previous study (Rotsaeart et al., 2018), the level of engagement in the PA group did not drop after

switching to an anonymous condition. However, there seems to be other factors that have affected student engagement and possibly moderated the effect of anonymity. In this study, online peer feedback was part of the video/article reflection assignment and included in the grading rubric. If students' feedback comments were poorly constructed, they would have received lower marks for that grading category. Only the instructor graded the quality of online peer feedback after students entered peer feedback in each session. Özdemir (2016) maintains that rewarding students for the quality feedback motivates them to engage in online peer assessment. Thus, the researcher suggests that instructors should include the quality of peer feedback in the grading rubric, which is assessed only by the instructors, and make explicit their expectation, prior to the activity.

As expected, few students provided critical feedback. The result was consistent with past research showing that students tend to provide more positive feedback than negative (Howard et al., 2010). This study was conducted in a blended classroom where students were required to attend face-to-face class meetings 2 days a week and engage in a variety of group activities and discussions. These physical meetings can bolster the social connection among students (Postmes et al., 2001). Moreover, because they were all elementary education majors at the same college, many of them might have already known each other personally prior to taking this course. Therefore, students may have felt social pressure even in an anonymous condition, which inhibited negative comments. Previous research also has shown that a strong friendship tends to inflate over-scoring in peer assessment (Panadero et al., 2013).

Furthermore, the ID group wrote more negative comments than the other two groups. This is a note-worthy finding as previous research has reported the opposite results (Howard et al., 2010). In this study, the gap might be caused due to personality differences because the same students tended to provide negative comments across four online feedback sessions. Another possible explanation is that some students in the ID group may have provided critical comments as instructed because their identities were disclosed, they did not want to embarrass themselves by posting poor quality feedback. Thus, a moderate level of peer pressure may contribute to the quality of online peer assessment.

Does the quality of online peer assessment change significantly across four data collection points?

The results showed that regardless of the group conditions, repeated practices tend to increase the number of words in feedback. In this study, there were four occasions for online peer assessment of two peers' work, and the increase was observed up to three feedback sessions. Research shows that students have little experience or lack of confidence in providing feedback to peers (Kilickaya, 2017). In this study, the first two online feedback sessions may have served as a scaffolding period. After several practices, students have become more comfortable with online peer feedback and built confidence, which contributed to the increase in the number of words in later feedback sessions. Li (2017) also has reported that students who received training prior to online peer assessment, demonstrated a higher level of performance and positive attitudes towards peer assessment than those in the non-training group. Although in this study, students did not receive intensive training for how to provide feedback, the instructor explained her expectation about the feedback quality at the beginning of the semester. Thus, based on this finding, instructors are strongly encouraged to provide students with training for peer assessment, and then assign them several peer assessment activities throughout the semester.

Although no significant difference was found, the number of words in feedback decreased from Topic 3 to Topic 4 sessions for all three groups. Topic 4 was the last topic for this assignment and assigned towards the end of the semester. One possible explanation for the decrease is that students were busy doing final projects or assignments in other classes, which resulted in the reduced number of words in feedback. Previous research shows that students tend to view online peer assessment as time-consuming work (Wilson et al., 2015). When students have time constraints, requiring online peer assessment may discourage students to make full efforts on writing quality feedback.

In addition, for all three groups, descriptive data showed that the number of negative comments decreased in later online feedback sessions. There was a considerable drop from Topic 2 to Topic 3. This may not necessarily indicate that the quality of feedback declined. Even though negative comments decreased, neutral comments increased from Topic 2 and Topic 3. As mentioned earlier, neutral comments are neither positive nor negative. These comments can include expressing own ideas or sharing personal experiences

related to the topic, which indicates their engagement in reflective thinking. Promoting self-reflection has been identified as one of the benefits in peer assessment (Saito & Fujita, 2009). Therefore, the increase in neutral comments may suggest that students were engaged in the activity of self-reflection. In their feedback comments, several students also mentioned that they could not find anything that needed improvement, which indicates that at least they tried to provide constructive or critical feedback. To become able to provide effective online feedback, constant reinforcement is required. LMSs, and many other online programs allow monitoring of students' feedback entries, which helps instructors detect any issue promptly. Research shows that instructor's intervention improves students' perceived value of peer assessment (Zhao, 2014). When students recognise the value of peer assessment, their level of engagement will increase. Thus, it is recommended that instructors re-teach students how to provide effective feedback several times, not only at the beginning of the semester.

What are the students' attitudes towards peer assessment and anonymous feedback?

The results of the attitude survey were similar to past research findings (McGarr & Clifford, 2013). In general, students perceived peer assessments to be helpful, but they were not strongly supportive ($M = 3.39$, $SD = .062$). The open-ended questions also revealed that although the majority recognised its value and benefits, a few students expressed strong resistance to peer assessment.

As mentioned earlier, all students were already familiar with the LMS used for peer assessment, which probably eased their anxiety for technical issues, as they did not have to learn a new computer program. In addition, participants in this study were highly confident in their skills for assessing peers. This was because most of them already had experience with peer assessment in other education classes. Several items that lowered the overall average were related to the concerns identified in past studies: social pressure and reliability/validity of peer scores (Özdemir, 2016). Students' responses to open-ended questions also revealed that negative attitudes were stemmed from these two factors.

Both the Likert survey and narrative comments indicate that students preferred anonymity. This is mostly due to the fear of hurting peers' feelings. In this study, anonymity did not affect the level of student engagement. However, earlier research suggests that anonymity increases psychological safety (Zhang et al., 2010), which will promote students' positive learning experiences (Barrett, 2010). Therefore, as suggested by past studies (Roberts & Rajah-Kanagasabai, 2013; Rotsaert et al., 2018), allowing students an anonymous option at least at the beginning stage seems to be an effective approach for those with limited experiences in online peer assessment.

The Likert survey showed that although most students trust their peer's ability in assessing their work, they do not want online peer assessment to affect their overall grades too much. In their narrative comments, a few students expressed that peers' comments are not very useful because they are mostly positive and not always true. One student stated rigidly that peers are not qualified to assess others. Nevertheless, the large number of positive statements indicate that their perceived benefits outweigh their concerns. One of the major benefits was related to their future profession. Because all participants of this study were preservice teachers, they found online peer assessment activities to be helpful. Furthermore, half of the students expressed that online peer assessment should not be integrated in all college courses. Several students who supported online peer assessment also mentioned that it depends on how it is implemented. Online peer assessment can take a variety of formats. To increase students' positive attitudes, instructors should design an appropriate online peer assessment according to the subject area and the type of assignment.

Are there significant relationships among the students' attitudes towards peer assessment and anonymity, the quality of online feedback, and demographic factors?

The results showed that academic standing and age influence the level of student engagement in online peer assessment. Higher academic standing and older students wrote more feedback comments than lower academic standing and younger students. This is not surprising because junior and senior students are more likely to be confident in their knowledge in their discipline area than freshman and sophomore students. Also, older and more mature students may already have some professional experiences, therefore, they are more likely to recognise the benefits of peer assessment for their future careers and to engage in the activity than younger students. The results suggest that in lower division classes, more training or practice for online peer assessment is required before the assignment is given. Furthermore, attitudes towards peer assessment

and anonymity did not affect student's engagement. In every classroom, there may be some students who are not willing to participate in peer assessment whether it is assigned online or in a physical classroom. As mentioned earlier, online peer assessment gives the instructor and students more flexibility, such as creating anonymous or indefinable conditions, revising text, and inserting rubrics, than paper-based peer assessment (Howard et al., 2010; McCarthy, 2017). Therefore, as long as it is designed and implemented properly, online peer assessment should benefit students as evidenced in numerous past studies (Falchikov, 2013).

Implications for practice and future research

This study indicates that to increase student engagement in online peer assessment, repeated practice is more critical than anonymity. Therefore, instructors should provide several opportunities for online peer assessment to reinforce the skill of writing effective feedback throughout a course. Further, although anonymity did not influence the level of engagement, because students express strong preference for anonymity, providing an option for anonymous feedback may ease their anxiety and enhance positive learning experiences. In this study, neither repeated practice nor anonymity increased the extent of critical feedback. The results of the survey indicate that students don't want to criticise their friends. Thus, especially if the class size is relatively small and students seem to have strong friendships, providing specific grading criteria for feedback quality is strongly recommended.

Lastly, participants in this study were mostly white females and all were elementary education major students. Previous research shows that gender and ethnicity influence peer assessment (Falchikov, 2013). Therefore, the study should be replicated with more diverse subject samples and settings. Future research could examine the extent to which assessing the quality of online feedback in the grading rubric influences the amount of critical feedback. In addition, this study showed that academic standing and age were correlated with the number of words in online feedback, therefore, further research is needed to investigate if grouping configurations, such as grouping by academic standing and mixed versus single sex groups, affect the extent of critical feedback.

References

- Barrett, B. J. (2010). Is "safety" dangerous? A critical examination of the classroom as safe space. *Canadian Journal for the Scholarship of Teaching and Learning*, 1(1). <https://doi.org/10.5206/cjsotl-rcacea.2010.1.9>
- Cheng, K. H., Hou, H. T., & Wu, S. Y. (2014). Exploring students' emotional responses and participation in an online peer assessment activity: A case study. *Interactive Learning Environments*, 22(3), 271-287. <https://doi.org/10.1080/10494820.2011.649766>
- Cheng, K. H., Liang, J. C., & Tsai, C. C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education*, 25, 78-84. <https://doi.org/10.1016/j.iheduc.2015.02.001>
- Demir, M. (2018). Using online peer assessment in an instructional technology and material design course through social media. *Higher Education: The International Journal of Higher Education Research*, 75(3), 399-414. <https://doi.org/10.1007/s10734-017-0146-9>
- Falchikov, N. (2013). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. New York, NY: RoutledgeFalmer.
- Gielen, M., & De Wever, B. (2015). Scripting the role of assessor and assessee in peer assessment in a wiki environment: Impact on peer feedback quality and product improvement. *Computers & Education*, 88, 370-386. <https://doi.org/10.1016/j.compedu.2015.07.012>
- Gravetter, F. J., & Wallnau, L. B. (2005). *Essentials of statistics for the behavioral sciences*. Belmont, CA: Wadsworth.
- Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to promote peer feedback: Pre-service teachers' comments in asynchronous computer-mediated communication. *Journal of Educational Computing Research*, 43(1), 89-112. <https://doi.org/10.2190/EC.43.1.f>
- Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, 33(4), 692-724. <https://doi.org/10.5465/256287>
- Kilickaya, F. (2017). Peer assessment of group members in tertiary contexts. In M. Sowa, & J. Krajka (Eds.), *Innovations in languages for specific purposes - Present challenges and future promises* (pp. 329-343). Frankfurt am Main: Peter Lang.
- Kirk, R. E. (2013). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Thousand

- Oaks, CA: Sage.
- Koç, C. (2011). The views of prospective class teachers about peer assessment in teaching practice. *Educational Sciences: Theory and Practice*, 11(4), 1979-1989. Retrieved from <http://oldsite.estp.com.tr/en/makale.asp?ID=584&act=detay>
- Liang, J. C., & Tsai, C. C. (2010). Learning through science writing via online peer assessment in a college biology course. *The Internet and Higher Education*, 13(4), 242-247. <https://doi.org/10.1016/j.iheduc.2010.04.004>
- Li, L. (2017). The role of anonymity in peer assessment. *Assessment & Evaluation In Higher Education*, 42(4), 645-656. <https://doi.org/10.1080/02602938.2016.1174766>
- Lin, G. (2018). Anonymous versus identified peer assessment via a Facebook-based learning application: Effects on quality of peer feedback, perceived learning, perceived fairness, and attitude toward the system. *Computers & Education*, 116, 81-92. <https://doi.org/10.1016/j.compedu.2017.08.010>
- Lin, S. S., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, 17(4), 420-432. <https://doi.org/10.1046/j.0266-4909.2001.00198.x>
- Lu, R., & Bol, L. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning*, 6(2), 100-115. Retrieved from <http://www.ncolr.org/issues/jiol/v6/n2/a-comparison-of-anonymous-versus-identifiable-e-peer-review-on-college-student-writing-performance-and-the-extent-of-critical-feedback.html>
- McCarthy, J. (2012). International design collaboration and mentoring for tertiary students through Facebook. *Australasian Journal of Educational Technology*, 28(5), 755-775. <https://doi.org/10.14742/ajet.1383>
- McCarthy, J. (2017). Enhancing feedback in higher education: students' attitudes towards online and in-class formative assessment feedback models. *Active Learning in Higher Education*, 18(2), 127-141. <https://doi.org/10.1177/1469787417707615>
- McGarr, O., & Clifford, A. M. (2013). 'Just enough to make you take it seriously': Exploring students' attitudes towards peer assessment. *Higher Education*, 65(6), 677-693. <https://doi.org/10.1007/s10734-012-9570-z>
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation In Higher Education*, 39(1), 102-122. <https://doi.org/10.1080/02602938.2013.795518>
- Özdemir, S. (2016). The opinions of prospective teachers on peer assessment. *Educational Research And Reviews*, 11(20), 1859-1870. <https://doi.org/10.5897/ERR2016.2997>
- Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G.T. L. Brown, & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment*, (pp.1-39). New York: Routledge.
- Panadero, E., Romero, M., & Strijbos, J. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, 39(4), 195-203. <https://doi.org/10.1016/j.stueduc.2013.10.005>
- Peled, Y., Bar-Shalom, O., & Sharon, R. (2014). Characterisation of pre-service teachers' attitude to feedback in a wiki-environment framework. *Interactive Learning Environments*, 22(5), 578-593. <https://doi.org/10.1080/10494820.2012.731002>
- Pena-Shaff, J., Altman, W., & Stephenson, H. (2005). Asynchronous online discussions as a tool for learning: Students' attitudes, expectations, and perceptions. *Journal of Interactive Learning Research*, 16(4), 409-430. Retrieved from <https://www.learntechlib.org/primary/p/5964/>
- Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin*, 27(10), 1243-1254. <https://doi.org/10.1177/01461672012710001>
- Roberts, L., & Rajah-Kanagasabai, C. (2013). "I'd be so much more comfortable posting anonymously": Identified versus anonymous participation in student discussion boards. *Australasian Journal of Educational Technology*, 29(5), 612-625. <https://doi.org/10.14742/ajet.452>
- Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology Of Education*, 33(1), 75-99. <https://doi.org/10.1007/s10212-017-0339-8>

- Saito, H., & Fujita, T. (2009). Peer-assessing peers' contribution to EFL group presentations. *RELC Journal*, 40(2), 149–171. <https://doi.org/10.1177/0033688209105868>
- Shih, R.-C. (2011). Can Web 2.0 technology assist college students in learning English writing? Integrating Facebook and peer assessment with blended learning. In J. Waycott, & J. Sheard (Eds.), *Assessing students' Web 2.0 activities in higher education. Australasian Journal of Educational Technology*, 27(Special issue 5), 829-845. <https://doi.org/10.14742/ajet.934>
- Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation*, 7(17). Retrieved from <http://PAREonline.net/getvn.asp?v=7&n=17>
- Sullivan, D., & Watson, S. (2015). Peer assessment within hybrid and online courses: Students' view of its potential and performance. *Journal Of Educational Issues*, 1(1), 1-18. <https://doi.org/10.5296/jei.v1i1.7255>
- Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, 81, 123-132. <https://doi.org/10.1016/j.compedu.2014.10.001>
- Van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, 20, 24–34. <https://doi.org/10.1016/j.edurev.2016.10.003>
- Wilkins, E. A., Shin, E.-K., & Ainsworth, J. (2009). The effects of peer feedback practices with elementary education teacher candidates. *Teacher Education Quarterly*, 36(2), 79–93. Retrieved from <http://www.jstor.org/stable/23479253>
- Wilson, M. J., Diao, M. M., & Huang, L. (2015). "I'm not here to learn how to mark someone else's stuff": An investigation of an online peer-to-peer review workshop tool. *Assessment & Evaluation in Higher Education*, 40(1), 15-32. <https://doi.org/10.1080/02602938.2014.881980>
- Yang, Y. F. (2016). Transforming and constructing academic knowledge through online peer feedback in summary writing. *Computer Assisted Language Learning: An International Journal*, 29(4), 683–702. <https://doi.org/10.1080/09588221.2015.1016440>
- Yu, F. Y., & Wu, C. P. (2011). Different identity revelation modes in an online peer-assessment learning environment: Effects on perceptions toward assessors, classroom climate and learning activities. *Computers & Education*, 57(3), 2167-2177. <https://doi.org/10.1016/j.compedu.2011.05.012>
- Zhang, Y., Fang, Y., Wei, K., & Chen, H. (2010). Exploring the role of psychological safety in promoting the intention to continue sharing knowledge in virtual communities. *International Journal of Information Management*, 30(5), 425-436. <https://doi.org/10.1016/j.ijinfomgt.2010.02.003>
- Zhao, H. (2014). Investigating teacher-supported peer assessment for EFL writing. *ELT Journal: English Language Teaching Journal*, 68(2), 155–168. <https://doi.org/10.1093/elt/cct068>
- Zhao, Y. (1998). The effects of anonymity on computer-mediated peer review. *International Journal of Educational Telecommunication*, 4(4), 311-345.

Corresponding author: Michiko Kobayashi, kobayashi@suu.edu

Copyright: Articles published in the Australasian Journal of Educational Technology (AJET) are available under Creative Commons Attribution Non-Commercial No Derivatives Licence ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)). Authors retain copyright in their work and grant AJET right of first publication under CC BY-NC-ND 4.0.

Please cite as: Kobayashi, M. (2020). Does anonymity matter? Examining quality of online peer assessment and students' attitudes. *Australasian Journal of Educational Technology*, 36(1), 98-110. <https://doi.org/10.14742/ajet.4694>