



One, Two, ... , Many: The Outcomes of Paired Peer Assessment, Group Peer Assessment, and Teacher Assessment in EFL Writing

Esmat Babaii

Kharazmi University, Iran

Aylar Adeh

Kharazmi University, Iran

The learning potential of pair/team work in language assessment has been demonstrated in the literature, albeit with little concern about challenges faced by the educators to implement it. To contribute to this line of research, the present study aimed to investigate the impact of paired peer assessment, teacher feedback, and group peer assessment on students' writing performance. To do so, 69 second-year university students of the English language were assigned to three classes. The first class received regular teacher feedback, the second class was instructed to work in pairs and assess each other's work, and students in the third class were divided into groups of four members and a group leader was assigned to each group by the teacher. Team leaders' responsibility was assessing and providing feedback on their fellow members' writing samples. The results showed a growth in the performance of students in paired peer assessment class in comparison to the other two classes. In group peer assessment class, the group heads outperformed their peers but the overall performance of the class was lower than paired peer assessment class. The findings suggest careful consideration of the circumstances under which peer assessment can be employed and call for more research on the contribution of non-pedagogical factors to assessment designs.

Keywords: collaborative learning, teamwork, peer assessment, EFL writing, teacher assessment

Introduction

The learning potential of assessment as an essential component of teaching and its support of learning have been emphasized in recent approaches to assessment (Taras, 2008). Yet, regardless of the integrated nature of assessment, instruction, and curriculum, and the many important roles of assessment in the classroom, it is quite often undervalued (Gareis & Grant, 2015). According to Boud (1995) assessment is the most fundamental prompt for learning and a good assessment can potentially provide students with the expected learning outcomes. The substantial advantages of assessment of student learning, and particularly, formative assessment's significant role, have been highlighted in a series of studies by Black and his colleagues (Black, Harrison, & Lee, 2003; Black & Wiliam, 1998, 2009). Assessment for learning or formative assessment, that is, the integration of assessment into teaching with the aim of providing feedback on learning outcomes and to support and increase learning (Sadler, 1998), provides teachers with information about students' learning that can be used for adjusting teaching accordingly (Black et al., 2003).

Stefani (1998) discusses the effect of participation in the assessment procedure on students' learning strategies and their engagement with their learning, and points to the inadequacy of the unilateral teacher assessment. Empowering students with feedback on their performance, and permitting them to participate in discussions with teachers about the assessment criteria and expected learning outcomes enhance learning and provides the students with a chance to become independent, autonomous, and reflective (Stefani, 1998; Wen & Tsai, 2006).

The role of feedback in learning is emphasized particularly in peer assessment as a formative assessment approach. It is believed that peer feedback supports learning through an intermediate check on the learner's performance, which is adapted to their level and is complemented with comments on students' strengths, weaknesses, and perhaps advices for greater improvements (Falchikov, 1996). Peer assessment can be seen as a learning tool since as stated by Topping (2013), through peer assessment, students develop different skills such as making judgments about what might be considered as good work as well as reflecting upon their own learning. According to Chen (2008), through reflection on their learning with the help of peers and the teacher, students' abilities are constructed and reconstructed. Students may gain the learning benefits of peer assessment while they are provided with peer assessment or they are assessing their peers' work (White, 2018). Boud and Falchikov (2006) believe that, in addition to the immediate needs of students, assessment should contribute to their future learning. They argue that as peer assessment and feedback provision involve making complex judgments on the work of their peers and themselves and require essential critical thinking and problem solving skills, students' active involvement and participation in the assessment process will contribute to their lifelong learning.

Nevertheless, there are still some questions regarding the impact of different ways peer assessment can be done and whether contextual factors, such as cultural background, may moderate the beneficial effects of this mode of assessment. As reminded by Bearman et al. (2017), many complex factors which may not be pedagogical in nature can sabotage the implementation of assumingly fool-proof assessment designs. Existing research (Adachi, Tai, & Dawson, 2018; Liu & Carless, 2006) already warns about possible challenges for successful use of peer assessment, questioning the unconditional prescription of this mode of assessment.

In an attempt to cater for these sensible concerns, the current study set out to investigate and compare the effects of peers assessing each other's work in pairs with peers assessing their fellow peers work in teams, and teacher feedback on Iranian EFL students' writing performance through comparing the performance of students on a series of writing tasks. Considering the nature of writing task which provides the students with a good chance for discovery, reformulation and revision, especially in process-oriented writing courses, it seems to be an appropriate area for practicing peer assessment. Our major impetus in conducting this study comes from the learners' resistance to using peer assessment as well as group work we have observed in many Iranian classes, especially before their actual classroom implementation. We suspect that such a widespread resistance should be something more serious than an initial, temporary unease when encountering something new. Besides, in this study, we added another form of peer assessment and intervention inspired by the 'monitorial' system which has an old history in the Eastern educational systems, especially Iran. This approach to education which was introduced to Western countries in the 19th century by educators such as Andrew Bell and Joseph Lancaster (cf. Encyclopedia Britannica Online) encourages more advanced students to contribute to the learning of their weaker peers. However, we avoided making distinctions between weak and advanced students and just incorporated the idea of being responsible for the improvement of peers to our study. More clarification about the nature of this kind of assessment will be provided in the Method section. In what follows, the related literature on peer assessment will be reviewed and then the current study and its findings will be presented.

Literature Review

Assessment as a substantial component of students' learning (Shepard, 2000) has an important impact on the learning process by improving the quality of learning and helping students to become reflective and responsible (Boud, Cohen, & Sampson, 1999; Orsmond, Merry, & Reiling, 2000). Despite being considered as a somewhat innovative approach to assessment, peer assessment and the advantages of employing peer assessment for writing has been discussed centuries ago (Topping, 2009). Yet, as shown by the substantial volume of research in this regard, the interest in peer assessment has been revitalized during the past decades. Peer assessment has been defined by Topping (1998, 2009, 2013) and Falchikov (2007) as a set of activities in which students participate in assessing and specifying the quality and value, as well as the strengths and weaknesses of the learning outcomes, including writing assignments, oral presentations, and test performance of their fellow peers who are at the equal level with them in terms of skills and proficiencies. It intends to inform and redirect students' efforts toward expected learning outcomes and to involve students in helping and planning their peers' learning and assessing the quality of their work (Topping, 2013).

The comparability and consistency of evaluations assigned by peers and teachers (Falchikov & Goldfinch, 2000), and the reliability and validity of peer assessment have been confirmed through several studies and the overall impact of peer assessment on students' achievement and attitude have been shown to be positive, or better than the effects of traditional assessments conducted by classroom teachers (Topping, 1998, 2009).

According to Topping (2013), peer assessment can be confirmatory, corrective, or suggestive. It is qualitative in nature and involves explanation of good content and the parts needing more improvements, and may involve specifying the errors and weaknesses as well as suggestions about how to improve the quality of work. While such feedback which is provided by equal status peers to each other may be less accurate than teacher feedback, but, it is available more readily and in much greater quantity than teacher feedback and can be individualized to each learner. The immediacy and quantity of feedback provided by peers increase the chances of enhanced performance, greater self-monitoring and self-regulation, and allows learners to undertake deep learning which results from demanding questions and explanations that happen during feedback provision and clarification. Through peer assessment, students receive the opportunities for reflecting upon their own perceptions, drawing on their prior knowledge, and fixing any misinterpretations, or making new inferences. As a result, they can develop their conceptual understanding and content knowledge, and improve their communication skills as well as self-monitoring and assessment skills (Falchikov, 2005; Reinholz, 2016; Roscoe & Chi, 2007).

It is believed that the participation in peer assessment has benefits for both the assessed and the assessor. As stated by Topping (2013) peer assessment is often a reciprocal process and the majority of participating students are both assessed and assessor. It is important to communicate the degree of detail and type of judgment and feedback that students should provide to their peers and if it should involve grading or feedback or both. Besides, the receivers of feedback should be prepared to react to the feedback thoughtfully and use the information to improve their work while deciding on which points to accept and which aspects to disregard (Topping, 2013).

In designing peer assessment in the classroom, some other variables may be considered, such as deciding whether the assessment is quantitative (with assigning marks), qualitative (detailed feedback provision) or both, and whether the students are paired or put into groups randomly or they are matched according to any particular criteria (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Topping, 1998, 2013). Development of clear and detailed criteria for assessment, such as rubrics, grids, and checklists, is recommended to ensure the mutual understanding between teacher and students about the abilities under assessment and the kind of preferred feedback (Gielen et al., 2010; Orsmond et al., 2000). Besides, according to Topping (1998), students' involvement in the process of development and explanation of these criteria brings a sense of ownership and responsibility to the students.

Perception of students about themselves and their peers have significant impact in peer assessment as most often students may not have full confidence in other students' knowledge and therefore, they may doubt their peers' ability to evaluate and assess their work (Ballantyne, Hughes, & Mylonas, 2002; Le Hebel et al., 2018; Strijbos, Narciss, & Dünnebier, 2010). Students may consider their teacher as the only knowledge authority and reserve the role of assessor for the teacher (Brown, Irving, Peterson, & Hirschfeld, 2009). The perceived status of peers and teacher as the knowledge authority depends on individual differences between students and it may impact the process of peer assessment (Le Hebel et al., 2018). However, the findings of Yang, Badger, and Yu (2006) have shown that in spite of this issue, peer feedback proved to be more beneficial in the revision of students' works. As they discuss, students considered the teacher as the knowledge authority and accepted his/her comments without questioning while they had discussions and confirmation checks with peer assessors and therefore had more chance for deep learning. The inadequacy of simply receiving feedback which communicates the correct answer to the students has been also stated by Alevan, Ogan, Popescu, Torrey, and Koedinger (2004).

Considering the significance of peer assessment in education, different studies were undertaken to investigate the use, influence and students' perception of this process. Studies on implementation of peer assessment and teacher feedback in EFL/ESL writing classes revealed mixed results: while some studies reported no significant difference between peer, self, and teacher assessment (e.g., Azarnoosh, 2013), other studies reported the beneficial effects of peer assessment (Khonbi & Sadeghi, 2012; Topping, 1998). The findings of the study by Tsui and Ng (2000) showed that students used more teacher feedback in comparison to peer feedback and perceived teacher feedback to be significantly more effective. However, while some students had doubts about peer feedback and did not take care of them, others stated that peer assessment helped them to find their strengths and weaknesses in writing, give them a sense of real audience, encourage collaborative learning and encourage their ownership of the writings by taking responsibility of their own learning. Investigating the effect of teacher's and senior students' feedback on the writings of a group of Thai EFL students, Sritrakarn (2018) reports beneficial results, while emphasizing the necessity of teacher intervention to ensure the positive impact of senior students' comments.

In an intriguing study, Lundstrom and Baker (2009) attempted to examine whether the assessed and the assessor benefit differently during peer assessment process. In their research, the feedback providers did not receive any peer feedback and the feedback receivers did not review other students' work. The analysis of students' writing samples before and after the course revealed that compared to feedback receivers, feedback givers made more significant improvements in specific areas of own writing over the course of study. The findings were explained based on socio-cultural theory and it was argued that the feedback givers' greater improvement may be because of the fact that they had the opportunity to determine the level at which feedback provision occurs which fell within their own ZPD. However, when this level differed from the feedback receivers' ZPD, then, they did not benefit as much from the feedback that was expected to scaffold their learning.

Peer assessment has shown to be treated differently in different cultures. Birjandi and Hadidi Tamjidi (2012), for instance, investigated the effect of peer, self, or a combination of both of them on Iranian EFL students' writing performance. Their results revealed that in classes where self-assessment and peer assessment in combination with teacher assessment were employed students had the highest degree of improvement. However, in the group where peer assessment and self-assessment were employed without teacher assessment, no significant improvement was shown. The authors explain this issue by the fact that in the Iranian context, teacher feedback is one of the main factors in developing writing skills since students primarily rely on their teachers. Their study suggests that a combination of teacher assessment with peer or self-assessment is the most beneficial.

Berggren (2015) investigated the impact of feedback provision through negotiation of criteria, training, and group peer reviewing on Swedish EFL students' own writing ability. Participation in peer assessment improved their awareness of audience and genre and it also promoted subsequent modifications and revisions to their writing at the macro level. In another context, Zou, Schunn, Wang, and Zhang (2017)

investigated Chinese students' attitudes toward peer assessment and factors that influence their participation. Positive attitude toward peer assessment, interpersonal concerns over the negative effects on interactions, and procedural negative doubts about the rationality of peer assessment were the main three factors that influenced the students' preference to engage in peer assessment. In Chinese culture, teachers are considered as the knowledge authority and the proper assessor, and therefore, the majority of participants stated that they preferred teacher feedback and therefore were reluctant to take responsibility of peer assessment. Students' lack of prior experience with peer assessment was also another concern for employing peer assessment. Similar results were found by Park (2018) in a Korean tertiary setting, where the students felt somehow reluctant to accept their peers' comments as compared to those by the teacher.

The above-mentioned studies reveal the context-sensitive nature of peer assessment and call for further studies investigating the effect of contextual factors and the mode of peer assessment on the beneficial outcomes of this approach. To make a contribution to this area of research, the current study compares EFL learners' writing performance in the Iranian context where teacher assessment, paired peer assessment, and group peer assessment were employed in three writing classes. The study also explores the impact of being a feedback provider as opposed to being a receiver of peer feedback in the third class. To be more precise, the study poses the following research questions:

1. Is there any statistically significant difference in students' performance on writing tasks before and after the course of study?
2. How do paired peer assessment, group peer assessment, and teacher feedback influence students' writing performance? Is there any significant difference in students' performance in three classes?
3. Is there any impact of being feedback provider as opposed to simply a receiver of peer feedback on students' writing performance in the group assessment class?

Method

Participants and Context of the Study

This study was conducted in three EFL writing classes at a state university in Iran. The data for this study were collected during a period of three months and consisted of students' performance on four writing tasks. The participants of the study were 69 Iranian students in their third academic semester for a BA degree in the English language (literature, teaching, and translation). The design of the study was quasi-experimental in which three intact classes were selected and each experienced a different assessment approach. Each session lasted about an hour and half. During the study, one class ($N = 27$) received regular teacher feedback on their writing assignments. The teacher collected their drafts and provided written comments on their work and assigned a mark. In the other class ($N = 22$), students worked in pairs, each individual receiving as well as providing writing feedbacks. In the third class ($N = 20$), students were assigned into five groups consisting of four members, one of them selected as the leader of the group who had the responsibility of feedback provision to the writing samples of their fellow peers and also self-assessment of their own work. To avoid bias and favoritism, the heads were selected randomly. The medium of instruction was English and all classes were taught by the same instructor. In general, process writing and formative assessment approaches were adopted in writing instruction in all classes. Besides, students in all three classes were permitted to ask for further clarification and elaboration once they felt the comments were vague or irrelevant. The writing proficiency of all students was tested four times, at the beginning of the course, during the course, and after the course. The details of these writing assessments are reported in the Results section.

Data Collection Procedure

All three classes worked according to the normal syllabus of the class and different writing tasks were practiced in each session. The first class (N = 27) had its regular formative assessment routine and assessments were conducted by the teacher. Such that, each session, classes' normal writing lessons and assignments were followed and at the end teacher collected the assignments, corrected them, and provided a written feedback on students' writings at the beginning of the next session.

The other class in which peer assessment was carried out (N = 22), the teacher divided the whole class into pairs randomly. During the first couple of sessions, the teacher discussed with students the concept of peer assessment and the idea of reviewing and giving feedback on their peers' works. To ensure that all students have understood the criteria for assessing and how to practice them, the teacher arranged some time to discuss the process and students could ask their questions and fix any misunderstandings or problems. The concept was gradually introduced to the class and then the process started with discussions about assessment criteria. While taking into consideration all students' comments, the teacher and students agreed to focus their feedbacks on the criteria of linguistic accuracy and content relevance. The process of assessment and feedback provision was modeled by the teacher and all students were asked to perform a similar process of assessment on a sample writing assignment. Then the students' assessment was analyzed by the teacher and necessary notes and recommendations to be considered in students' upcoming assessments were provided to them. However, this introductory stage lasted for the first couple of sessions to get students used to the concept. All through the next sessions up to the end of the course, each session started with regular lessons, then students had time to work on the new task which often involved writing a 150-200 word essay, diary, argument, or any other writing tasks according to the syllabus. After completing their own assignments, students exchanged their assignments with peers and completed their peer assessments according to the specified assessment criteria of linguistic accuracy and content relevance in almost 30 minutes. They had to read and review each other's work and provide a written feedback about strengths and weaknesses, grammatical errors, and possibly some revision suggestions by highlighting the areas and with question marks or briefly indicating the problem. During this time, the teacher checked students' progress and helped them. Subsequently, after assessing each other's assignments and providing written feedback, the students exchanged their feedbacks and discussed the comments. When necessary, they provided explanations about their comments to each other and decided if any revision was necessary or not. At the end of the class time, all works were collected by the teacher to be assessed and marked as part of classes' ongoing formative assessment.

The last class (N = 20) was divided into five teams of four students, and for each group a group head was assigned by the teacher. The introductory phases of introducing peer assessment and the assessment criteria that were practiced in second class were in the same way performed in this class. The only difference was that, in this class, only group heads were responsible for assessing and providing feedback on their fellow team members' writing samples and the other members were simply receivers of peer assessment. During each session, after completing their assignments individually, group heads had time to assess their peers' work and communicate their assessment and feedback with fellow team members both written and orally. Meanwhile, other group members were encouraged to examine their own drafts, check reference books, and ask questions from the teacher. Group heads' self- and peer-assessments were collected by the teacher at the end of the session and these students received feedback on their own work from the teacher at the beginning of the next session.

While students were provided with assessment criteria to prevent difficulty and misunderstanding, to avoid erroneous judgments or misleading feedbacks, the teacher also analyzed feedbacks provided by students and commented on their appropriateness and validity. The teacher's analysis of students' work was helpful in deciding on parts or aspects that students needed more clarification and help.

In addition to do writing tasks each session, the students were asked to write four samples at the beginning, during and at the end of the course so that the researchers could examine their writing ability and their progress. The four writing samples were assessed and marked on a five level scale, one being

the lowest level of achievement and five the highest. Lumely's (2002) four sets of criteria, that is, (a) task fulfillment and appropriacy, (b) conventions of presentation, (c) cohesion and organization, and (d) grammatical control, were used to rate students' writing. The two raters in this study were two PhD students of applied linguistics with solid background in language assessment. The average spearman correlation for rating the tasks was .87, suggesting a reasonable degree of inter-rater reliability for the scoring procedure.

Results

In this section the findings of the study are presented. First, findings concerning the impacts of pair and group peer assessment and teacher assessment in developing students' writing ability are presented and then differences in students' performance according to the assessment approach will be explored. At the beginning of the study, all students participated in a writing pretest which aimed at assessing students' current writing proficiency. To observe if there existed any significant differences between three groups in their writing skills at the beginning of the study one-way analysis of variance (ANOVA) was run. As shown in Table1, there was no statistically significant difference in the three groups ($p = .962$) in terms of their writing proficiency which indicates that students of all classes had roughly comparable ability in writing.

TABLE 1
ANOVA Results for Writing Test 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.016	2	.008	.039	.962

Subsequently, students' writing performance was examined three more times. Half way through the course, and at the end of the course, students were administered writing tests that included a writing task somewhat longer than usual class assignments. To answer the first research question of the study, and to see how students performed on four writing tasks during and after the course of study we performed a repeated measures ANOVA. The descriptive statistics of these tests are presented in Table 2.

TABLE 2
Students' Performance on Writing Tasks at Four Stages of Time

	Treatment	Mean	Std. Deviation	N
Time1	Teacher Feedback	1.62	.49	27
	Paired Peer Assessment	1.63	.41	22
	Group Peer Assessment	1.60	.41	20
	Total	1.62	.44	69
Time2	Teacher Feedback	2.38	.75	27
	Peer Assessment	2.81	.93	22
	Group Peer Assessment	2.15	.36	20
	Total	2.45	.77	69
Time3	Teacher Feedback	3.25	.93	27
	Peer Assessment	3.95	1.0	22
	Group Peer Assessment	3.17	.97	20
	Total	3.45	1.0	69
Time4	Teacher Feedback	3.75	1.0	27
	Peer Assessment	4.59	.61	22
	Group Peer Assessment	3.92	.83	20
	Total	4.07	.94	69

A look at the table reveals that regardless of the assessment approach employed, all classes showed improvements. However, to find out if these changes were statistically significant, repeated-measures ANOVA was run and its results are presented in Table 3.

TABLE 3
Repeated-measures ANOVA for the Effects of Time and Treatment

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	242.54	3	80.84	341.74	.000
	Greenhouse-Geisser	242.54	2.26	107.29	341.74	.000
time *	Sphericity Assumed	6.25	6	1.04	4.40	.000
	Greenhouse-Geisser	6.25	4.52	1.38	4.40	.001
Error (time)	Sphericity Assumed	46.84	198	.23		
	Greenhouse-Geisser	46.84	149.20	.31		

The repeated measures ANOVA with a Greenhouse-Geisser correction results revealed a significant effect of time on mean scores, as the mean scores for students writing skills were statistically different in different time points of the study ($F(2.26, 149.20) = 341.74, p < .000$). As the overall ANOVA results were significant, to see where the differences occurred, the pairwise comparison table which is the result of the Bonferroni post hoc test is presented.

TABLE 4
Results of Bonferroni Post-hoc Test

(I) Treatment	(J) Treatment	Mean Difference (I-J)	Std. Error	Sig.
Teacher Feedback	Paired Peer Assessment	-.49*	.19	.03
	Group Peer Assessment	.04	.19	1.00
Paired Peer Assessment	Teacher Feedback	.49*	.19	.03
	Group Peer Assessment	.53*	.20	.03
Group Peer Assessment	Teacher Feedback	-.04	.19	1.00
	Paired Peer Assessment	-.53*	.20	.03

Post hoc tests using the Bonferroni correction revealed significant differences in students' writing performance between the three groups. Students of the class in which peer assessment in pairs was employed outperformed both of the teacher assessment class ($p = .03$), and group peer assessment class ($p = .03$). However, no significant difference was found in writing performance of students in teacher assessment and group peer assessment class ($p = 1.00$). The difference between teacher assessment and group peer assessment class is also revealed in the mean difference of the two groups which is very trivial (.04). Therefore, we can conclude that while the other two classes experience improvements in their writings, employing peer assessment has produced a statistically significant development in students' writing ability.

The study was also concerned with students' performance in group peer assessment class and whether there was any impact of being feedback provider as opposed to simply a receiver of peer feedback on students' writing performance. The number of students in this class was 20 and they were assigned into five groups with four members. So, we had 5 group heads who performed peer assessment and 15 students as group members and receivers of peer assessment. Mann-Whitney U Tests were run to see if there was any significant difference in students' performance on four writing tests. The results of tests presented in Table 5 show that at the beginning of the study ($U = 30.50, Z = -.65, p = .51$) and the second tests ($U = 25.00, Z = -1.76, p = .07$), the difference between group heads and their fellow team members' performance was not significantly different. However, after a few weeks of practicing peer assessment, the differences in performance increased as in the third test the difference was significant and group heads outperformed their group members ($U = .00, Z = -3.46, p = .000$). In the last test, too, the difference between the writing performance of group heads and group members was statistically significant ($U = 2.500, Z = -3.21, p = .001$).

TABLE 5
Mann-Whitney U Test Results

	Time1	Time2	Time3	Time4
Mann-Whitney U	30.50	25.00	.00	2.50
Wilcoxon W	150.50	145.00	120.00	122.50
Z	-.65	-1.76	-3.46	-3.21
Asymp. Sig. (2-tailed)	.51	.07	.001	.001
Exact Sig. [2*(1-tailed Sig.)]	.55	.30	.000	.001

From this finding, it can be concluded that developments in writing performance of students who were group heads was significantly higher than their peers who simply received peer assessment and did not perform assessment.

Discussion

An increasing number of studies are corroborating the many potential benefits of peer assessment and peer feedback on students' learning and their later academic achievement. This study investigated the impact of peer assessment, group peer assessment, and teacher assessment on Iranian university students' EFL writing performance. The results of this study evidenced the positive impact of students' participation in one to one peer assessment in developing their writing abilities. Consistent with the literature (Berggren, 2015; Birjandi & Hadidi Tamjid, 2012; Khonbi & Sadeghi, 2012), it has provided statistical evidence of the effectiveness of peer assessment in EFL writing classes as the results have shown that, the students in peer assessment class demonstrated a higher degree of development in writing skill than the other two groups. Students' gains in terms of language skills, communication abilities, and autonomy and thoughtfulness can easily be justified by the benefits of employing peer assessment as a tool in enhancing student learning and educational success that has already been discussed.

Another finding of this study was that students who were group heads and provided peer assessment and feedback, outperformed their fellow team members who were receiving peer assessment without having the chance to assess others' writing samples. The findings confirm the findings of Lundstrom and Baker (2009), according to which, feedback givers had greater developments in their writing proficiency in comparison to feedback receivers.

The lesser development in group members' writing performance may be explained in terms of students trust or, rather mistrust, in their peers' language proficiency and assessment knowledge and therefore the accuracy of their feedback. In addition, students' persistent experience with teacher-centered educational system of Iran can be another explanation for students' smaller improvement after being the receiver of peer assessment for a period of time. Students' belief in teacher as the knowledge authority and peers being unqualified to assess other students' work has been reported in several studies (Kaufman & Schunn, 2011; Mangelsdorf, 1992; Mendonça & Johnson, 1994; Zhao, 2014). Likewise, according to Nelson and Murphy (1993), ESL students viewed teacher as the knowledge authority and they did not trust their peers' competence as English was not their first language. So they considered peers' comments with doubt and mistrust and did not incorporate all of these comments in their revisions. Students' reluctance to peer comments and preference for teacher feedback has also been reported by Nelson and Carson (2006). Teachers were considered the only legitimate source of feedback. In this regard Tsui and Ng (2000) argue that teacher comments may be more effective or that, they may simply be assumed to be so, even if they are not. This issue can be addressed by teacher intervention and their support of peer feedbacks. If the teacher provides a supportive comment on the appropriateness of the peer feedback and its legitimacy, it will improve students' perception of it and their incorporation of such feedback in their work (Zhao, 2014).

The remarkable point about the current study is that although both peer assessment classes were supported by the teacher, the outcomes of paired assessment class were more satisfactory. It may be

argued that in paired peer assessment class, each student was given the chance to be both feedback receiver and provider, and this created an atmosphere of equality and partnership. While in the other class, the students had to accept the leadership of another fellow student. That is, they needed to deal with the question of power, assumed 'unprivileged' at times, exercised by their classmate. This may not be a serious issue in many cultural contexts but it tends to provoke some negative reactions on the part of Iranian students. Iranian culture, as mentioned by Zonis (1971) and Beeman (1986), is characterized by some degree of mistrust in interpersonal relationships. According to Beeman (1986, p. 32), "mistrust, seen as an attitudinal feature, implies that other people are so unpredictable that one never knows what malevolence they are going to inflict on one, or how they are going to fail to support one's interests". The point is that in paired peer assessment, there is a possibility of catering for another person's unpredictable behavior (here assessment), since the power of one student as the feedback provider is counterbalanced by his/her role as the feedback receiver. However, this mechanism is not at work in group peer assessment because the feedback provider role of the group leader is non-reciprocal. The result is an atmosphere of anxiety, mistrust and lack of cooperation. All this, should remind testing professionals of the importance of contextual factors in assessment. As Bachman (1990, p. 279) aptly puts: "tests are not developed and used in a value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or of society at large". Therefore, the success or failure of a given test and/or assessment approach cannot be reliably predicted without some understanding about socio-cultural background of the people who will serve as the participants. Peer assessment is not an exception.

To put it another way, to enjoy the cognitive and meta-cognitive gains of peer assessment, the students need to learn to trust other student's abilities and judgment. They might also need to be assured that they would not fall victims to their classmates' possible biases and misjudgments. As noted by Topping (2009), while teachers may already be aware of the importance of peer assessment and peer feedback in the classroom, students should be gradually educated in this regard and the process should be introduced and incorporated into classes with patience and thought. In fact, peer assessment can be added as a supplement to the existing procedures, and be used together with teacher feedback. Students should be advised about giving feedback and the fact that there is no stress on them to be always right (Topping, 2013). Likewise, Saito (2008) argues that a successful implementation of peer assessment requires training of the students and checking their ability of the assessment or otherwise, the advantages of peer assessment cannot be assured. Considering the central role of students as both assessor and assessed in peer assessment, more attention and care is needed concerning their perceptions and attitude toward the process.

Conclusion

In this study, developments in students' writing were compared in three classes where different approaches to feedback and formative assessment, including paired peer assessment, group peer assessment, and teacher feedback were employed. As it was discussed earlier, the outcomes of paired peer assessment were superior or at least comparable to classes where assessment was non-reciprocal or the main role was given to teacher feedback. Through paired peer assessment, students had active roles in their learning and as they worked in pairs, their fellow classmates and their peers had the role of rather real audience and students had to discuss and negotiate their writings with them, which in turn gave them a sense of ownership of their work which led to their relative autonomy and taking control of their learning.

As for peer assessment in teams, the findings call for careful considerations of students' attitude and cultural background. As discussed by Christison and Murray (2009), considering the vagueness of the concept for most professionals and a lack of precision about main constituents of teamwork, it might be used in contexts where it is rather inappropriate to use. It can be argued that due to the students' mistrust of their peers' judgments and the non-reciprocal feedback provider role of team leaders, a desired level of

cooperation was not achieved in group peer assessment class. This finding gains momentum if we look at it from the perspective of fairness in assessment. As reminded by Wallace (2018), fairness is a subjective phenomenon (at least in part) and we cannot confidently consider our assessment procedure to be fair if the majority of examinees think otherwise. Therefore, to make peer assessment beneficial to the whole class and also increase fairness of the procedure, devising strategies to guarantee the reciprocity of feedback provision is recommended, especially in socio-cultural settings with the (past) experience of a hierarchical social system.

The present research had some limitations that should be stated. As the study was conducted in a university, there were faculty expectations and predefined compulsory course outline for the classes had to be accomplished and therefore class time allocated for peer assessment processes were limited to 90 minutes in each session. The small size of the sample and the use of intact classes were also other limitations of the study. The time that could be allocated to training students for peer assessment at the beginning of the research was fairly restricted and it is assumed that in studies with larger samples, more time should be given to training students and teachers for precise implementation of peer assessment.

The Authors

Esmat Babaii (corresponding author) is associate professor of applied linguistics at Kharazmi University, Iran, where she teaches research methods, language assessment and discourse analysis to graduate students. She has served on the editorial boards and/or the review panels of several national and international journals. She has published articles and book chapters dealing with issues in Systemic Functional Linguistics, Appraisal theory, test-taking processes, discursive analysis of textbooks, and critical approaches to the study of culture and language.

Department of Foreign Languages
Faculty of Humanities
Kharazmi University
Tehran 15719-14911, Iran
Phone & Fax: +98 021 88304896
Email: babai@khu.ac.ir

Aylar Adeb is a PhD candidate in applied linguistics at Kharazmi University, Iran. She works on language assessment, English as an International Language (EIL) and teacher identity.

Department of Foreign Languages
Faculty of Humanities
Kharazmi University
Tehran 15719-14911, Iran
Phone & Fax: +98 021 88304896
Email: aylar23@gmail.com

References

- Aleven, V., Ogan, A., Popescu, O., Torrey, C., & Koedinger, K. (2004). Evaluating the effectiveness of a tutorial dialogue system for self-explanation. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Intelligent tutoring systems* (Vol. 3220, pp. 443-454). Berlin: Springer.
- Azarnoosh, M. (2013). Peer assessment in an EFL context: attitudes and friendship bias. *Language Testing in Asia*, 3(1), 11. doi:10.1186/2229-0443-3-11
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

- Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, 27(5), 427-441. doi:10.1080/0260293022000009302
- Beeman, W. (1986). *Language, status and power in Iran*. Indiana: Indiana University Press.
- Berggren, J. (2015). Learning from giving feedback: a study of secondary-level students. *ELT Journal*, 69(1), 58-70. doi:10.1093/elt/ccu036
- Birjandi, P., & Hadidi Tamjid, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education*, 37(5), 513-533. doi:10.1080/02602938.2010.549204
- Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Maidenhead: Open University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education: Principles, policy & practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5. doi:10.1007/s11092-008-9068-5
- Boud, D. (1995). Assessment and learning: Contradictory or complementary. In P. Knight (Ed.), *Assessment for learning in higher education* (pp. 35-48). London: London: Kogan Page.
- Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education*, 24(4), 413-426. doi:10.1080/0260293990240405
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413.
- Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction*, 19(2), 97-111. doi:https://doi.org/10.1016/j.learninstruc.2008.02.003
- Chen, Y.-M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12(2), 235-262. doi:10.1177/1362168807086293
- Christison, M., & Murray, D. E. (2009). Building effective teams. In M. Christison & D. E. Murray (Eds.), *Leadership in English language education* (pp. 200-218). New York: NY: Routledge.
- Falchikov, N. (1996). *Improving learning through critical peer feedback and reflection*. Paper presented at the Herdsa Conference: Theory and Practice in Higher Education, Perth, Australia.
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. New York, NY: Routledge Falmer.
- Falchikov, N. (2007). The place of peers in learning and assessment. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 128-143). New York, NY: Routledge.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322. doi:10.3102/00346543070003287
- Gareis, C. R., & Grant, L. W. (2015). *Teacher-made assessments* (2nd ed.). New York, NY: Routledge.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), 304-315. doi:https://doi.org/10.1016/j.learninstruc.2009.08.007
- Hounsell, D. (2003). Student feedback, learning and development. In M. Slowey & D. Watson (Eds.), *Higher education and the lifecourse* (pp. 67-78). Berkshire Open University Press.
- Jhangiani, R. S. (2016). The impact of participating in a peer assessment activity on subsequent academic performance. *Teaching of Psychology*, 43(3), 180-186. doi:10.1177/0098628316649312
- Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science*, 39(3), 387-406. doi:10.1007/s11251-010-9133-6

- Khonbi, Z. A., & Sadeghi, K. (2012). The effect of assessment type (self vs. peer vs. teacher) on Iranian university EFL students' course achievement. *Language Testing in Asia*, 2(4), 47. doi:10.1186/2229-0443-2-4-47
- Le Hebel, F., Constantinou, C. P., Hospesova, A., Grob, R., Holmeier, M., Montpied, P., Žlábková, I. (2018). Students' perspectives on peer assessment. In J. Dolin & R. Evans (Eds.), *Transforming assessment: Through an interplay between practice, research and policy* (pp. 141-173). Cham: Springer International Publishing.
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? *Language Testing*, 19, 246-276. DOI: 10.1191/0265532202lt230oa
- Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, 18(1), 30-43. doi:https://doi.org/10.1016/j.jslw.2008.06.002
- Mangelsdorf, K. (1992). Peer reviews in the ESL composition classroom: What do the students think? *ELT Journal*, 46(3), 274-284. doi:10.1093/elt/46.3.274
- Mendonça, C. O., & Johnson, K. E. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, 28(4), 745-769. doi:10.2307/3587558
- Nelson, G., & Carson, J. (2006). Cultural issues in peer response: Revisiting "culture". In F. Hyland & K. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 42-59). Cambridge: Cambridge University Press.
- Nelson, G., & Murphy, J. M. (1993). Peer response groups: Do L2 writers use peer comments in revising their drafts? *TESOL Quarterly*, 27(1), 135-141. doi:10.2307/3586965
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23-38. doi:10.1080/02602930050025006
- Park, J. (2018). Effectiveness of Teacher and Peer Feedback: Through the Lens of Korean Tertiary Writing Classroom. *The Journal of Asia TEFL*, 15(2), 429-444. doi: 10.18823/asiatefl.2018.15.2.11.429
- Reinholz, D. (2016). The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 41(2), 301-315. doi:10.1080/02602938.2015.1008982
- Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4), 534-574. doi:10.3102/0034654307309920
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in education: Principles, policy & practice*, 5(1), 77-84. doi:10.1080/0969595980050104
- Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, 25(4), 553-581. doi:10.1177/0265532208094276
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14. doi:10.3102/0013189x029007004
- Sritrakarn, N. (2018). A comparison of teacher's and senior students' feedback: Student attitudes and their writing improvement. *The Journal of Asia TEFL*, 15(2), 329-348. doi: 10.18823/asiatefl.2018.15.2.5.329
- Stefani, L. A. J. (1998). Assessment in partnership with learners. *Assessment & Evaluation in Higher Education*, 23(4), 339-350. doi:10.1080/0260293980230402
- 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. doi:https://doi.org/10.1016/j.learninstruc.2009.08.008
- Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education*, 9(2), 172-192. doi:10.1177/1469787408091655
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276. doi:10.3102/00346543068003249

- Topping, K. (2009). Peer assessment. *Theory Into Practice*, 48(1), 20-27. doi:10.1080/00405840802577569
- Topping, K. (2013). Peers as a source of formative and summative assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 395-412). Thousand Oaks: Thousand Oaks, California: SAGE Publications, Inc.
- The Editors of Encyclopaedia Britannica (2015). Monitorial system. Encyclopædia Britannica, inc. <https://www.britannica.com/topic/monitorial-system>. (Published January 16, 2015; accessed March 03, 2019).
- Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, 9(2), 147-170. doi:[https://doi.org/10.1016/S1060-3743\(00\)00022-9](https://doi.org/10.1016/S1060-3743(00)00022-9)
- Wallace, M. (2018). Fairness and justice in L2 classroom assessment: Perceptions from test takers. *The Journal of Asia TEFL*, 15(4), 1051-1064. doi: 10.18823/asiatefl.2018.15.4.11.1051
- Watanabe, Y., & Swain, M. (2007). Effects of proficiency differences and patterns of pair interaction on second language learning: collaborative dialogue between adult ESL learners. *Language Teaching Research*, 11(2), 121-142. doi:10.1177/136216880607074599
- Wen, M. L., & Tsai, C.-C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education*, 51(1), 27-44. doi:10.1007/s10734-004-6375-8
- White, E. (2018). Does the feedback feed forward? Student response to and views of teacher feedback in an EFL academic writing class. In S. Hidri (Ed.), *Revisiting the assessment of second language abilities: From theory to practice* (pp. 109-132). Cham: Springer International Publishing.
- Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15(3), 179-200. doi:<https://doi.org/10.1016/j.jslw.2006.09.004>
- Zhao, H. (2014). Investigating teacher-supported peer assessment for EFL writing. *ELT Journal*, 68(2), 155-168. doi:10.1093/elt/cct068
- Zonis, M. (1971). *The political elite of Iran*. Princeton, New Jersey: Princeton University Press.
- Zou, Y., Schunn, C. D., Wang, Y., & Zhang, F. (2017). Student attitudes that predict participation in peer assessment. *Assessment & Evaluation in Higher Education*, 1-12. doi:10.1080/02602938.2017.1409872