

Original Paper

An Investigation of the Effect of Different Feedback-based Assessments on Oral Performance and Attitudes of Iranian EFL Learners

Zahra Karimi Valoojerdy¹

¹ Faculty of Humanities, Tarbiat Modares University, Tehran, Iran

Received: August 12, 2022 Accepted: August 25, 2022 Online Published: September 3, 2022

doi:10.22158/jecs.v6n3p75

URL: <http://dx.doi.org/10.22158/jecs.v6n3p75>

Abstract

Feedback is a core component of the learning process which has three levels: teacher feedback, peer feedback and self-feedback. The role of teacher feedback has received considerable attention from second language acquisition (SLA) researchers for the past three decades. Assessment is another factor in improving learning. The present study is an attempt to investigate and compare the effects of peer feedback-based assessment, teacher feedback-based assessment, and self-feedback-based assessment on oral performance of EFL Iranian learners and also learners' attitudes toward them. The assessments' rubric criteria used in this study was based on 14 points scale of Yamashiro and Johnson (1997). The participants comprised 84 TEFL students. A mixed method experimentation design was followed with a between-groups comparison in which independent groups have been compared followed by another phase looking for students' views through an attitude survey. This study used descriptive statistics, Pearson product moment correlation, one-way ANOVA and post hoc test. The results revealed that teacher feedback-based assessment group performed significantly better than the other two groups in oral performance.

Keywords

Feedback, Teacher Feedback, Self-Feedback, Peer Feedback, Assessment, Oral Performance, Target Language

1. Introduction

The history of teaching a second language has witnessed changing perceptions of giving feedback. In the majority of classes, teachers decide on giving feedback to students. In teacher feedback-based assessment, the teachers, solely give feedback and assess their students. To move with the time and

observe the needs of learners, language teaching has focused on communicative and collaborative language learning, learners, and their needs. Therefore, peer feedback or self-feedback come into existence. During this long way, measurement and evaluation have also been the concern of educators and the educational system. According to Heaton (1988), like teaching, there have been different approaches for language testing as well, such as essay translation, structuralism, integrative, and communicative approaches. Traditionally, assessment could be described as a mere quantitative device used just for summative purposes. These purposes, according to Farhady, Ja'farpur, and Birjandi (1994) are conducted to gain insights into an individual's knowledge and abilities to make various kinds of decisions, ranging from screening and selecting applicants, to a program of study. Hence, it is determined whether additional instructional time should be allocated to a given topic or not. As the goals of education have become more and more complex, and the number of students has enormously increased, evaluation has, accordingly, become much more difficult (Farhady et al., 1994, p. 1). To be in the same line with language teaching, while students and their needs started to gain more and more attention, a shift also started from quantitative methods of language testing to more qualitative ones. According to Brown (2005), a shift occurred from summative forms of assessment to more formative ones. Heaton (1988, p. 1) also approved that the qualitative forms of assessment are —superior to and of —a considerable benefit for language learners. In this newly accepted convention, when people are to be assessed, great effort should be put into minimizing any intervening factors and making sure that the assessment procedure is the same for everyone. In line with shifts to more formative assessments, then the students took more and more responsibility for their learning, monitoring their promotion, and assessing their work. They also, based on the findings of different studies Parti (2002), Roskams (1999), Warren and Cheng (2005), started to be counted as an outstanding source of feedback for their friends in cooperative environments and play a greater role in their learning. However, the teachers still had an undeniable role in the educational system and classrooms. Social psychologists have long been interested in issues of performance feedback and goals (Lewin et al., 1944; cited in Henrich R. Greve, 2003), making the traditional studies emphasize different parts of the process of setting and pursuing goals. One of the parts of this process is to give feedback and assess the students. The term “feedback” has been used to describe the linguistic and metalinguistic information that target language (TL) speakers (principally native speakers (NSs) and teachers) teach learners about the accuracy of their spoken interlanguage (IL) and modify the IL grammar by the learners. The role of teacher feedback has received considerable attention from second language acquisition (SLA) researchers for the past three decades. Numerous studies have explored its incidence and/or effectiveness in developing second language (L2) capacities (Re've'sz & Gurzynski Weiss, 2012).

Second language acquisition researchers hypothesized that the benefit of feedback provision was confirmed by empirical research.

Moreover, several meta-analyses indicate that interactional feedback can indeed facilitate L2 learning, Keck, Iberri-Shea, Tracy-Ventura and WaMbaleka (2006), Li (2010), Lyster and Saito (2010), Mackey

and Goo (2007). As discussed, teacher feedback can improve learning. However, there are other kinds of feedback as well e.g.: peer feedback and self-feedback. A way to engage students actively with feedback processes is to implement peer review or peer-feedback Liu and Carless (2006), Cartney (2010), Nicol (2011). Today's mainstream educational concentration is more on improving speaking abilities Min, H. T. (2005), Narciss, S. (2008), Nunan, D. (1988), Nunan, D. (2004), Parti, M. (2002), Sadler, P., and Good, E., (2006). Therefore, students should be equipped with factors such as communication, cooperation, him/herself, teacher, and peer-assessment as well as feedback by taking part in interactive classrooms and finally improving their speaking ability. Oral performance is one of the most important factors for showing English knowledge and achieving a better oral performance. There are some ways and strategies in this regard including self-and peer- and teacher assessment. Generally, research about self-and peer assessment of oral presentation skills revealed under-explored areas and diverging views, AlFallay (2004), Patri (2002), Campbell et al. (2001), Sadler, P., and Good, E., (2006). To shed light on the areas of self-and peer- and teacher assessment and feedback, this study was conducted. However, it may not be that much easy because giving feedback must be a continuous trend in all activities and any teaching endeavor. As a result, teachers should be able to use different methods and techniques to integrate them into their classrooms. Carrying out empirical studies can help teachers in choosing the best means of assessment and giving feedback. The present investigation can make Iranian EFL teachers aware of the best ways of giving feedback to students, whether to give feedback on their own or take advantage of using peer feedback or self-feedback. It can help students to find out the best source of knowledge. It seems necessary to investigate differences in the oral performance of learners. In this study, the effects of peer feedback-based assessment, teacher feedback-based assessment, and self-feedback-based assessment on the oral performance of EFL learners were investigated and compared in the case of a presentation sheet based on Yamashiro and Johnson (1997). The research is of great importance for language learning and teaching settings in general, and Iranian EFL learners and teachers in particular. The present findings provide insights into a more effective assessment and feedback method incorporating to increase students' willingness not to miss even one session of the class. Therefore, these methods help them to get better speaking abilities. The study is different from the previous ones concerning filling the gap of having no idea about the best way of conducting feedback-based assessment in improving the oral performance of EFL learners in speaking classes.

2. Literature Review

2.1 Whole Persons

Language teaching and learning atmosphere have been recently altered more or less in the educational systems of every place. Moreover, the old approaches and methods paved the way for newer learner-centered and communicative-based methods. After so many trials and errors performed over this long run, learners were concentrated and considered, according to Larsen- Freeman (2000), as

“whole persons” who were little by little taking the responsibility for their own learning and becoming more autonomous.

Then, the notions of peer and self-feedback were introduced paying closer attention to the nature of interaction and processes occurring within groups of learners. As Larsen-Freeman (2000) mentioned, “students learn from each other in groups, however, it is not the group configuration making the cooperative learning distinctive. It is the way that students and teachers work together which is important, the way that teachers manage their teaching” (p. 164). Different approaches to assessment are mainly adopted to provide descriptions of individuals’ performance over time or gain insights into their knowledge to make different decisions. In these approaches, great effort is made to assure that the assessment procedure meets the criteria such as reliability, validity, and generalizability. Farhady (1994) mentioned that “as the goals of education have become more and more complex, and the number of students has enormously increased, evaluation has become much more difficult” Farhadi et al. (1994, p. 1). When people are assessed, great effort is made to minimize any intervening factors such as hints and assistance deviating from the true picture and assure that the assessment procedure is the same for everyone. These traditional approaches to language proficiency are based on the dualistic perspective inherited from psychology. Vygotsky proposed that the human mind is neither exclusively biological, nor social. However, it is caused by the unity resulting from the interplay of biologically specified functions (e.g., memory, attention, and learning) and socially created symbolic means (primarily linguistic) to intentionally control these functions. This dialectic perspective gave rise to a qualitatively different way of thinking about assessment from how it is traditionally understood by classroom teachers and researchers.

2.2 *Alternative Assessments*

Learners’ self-assessment of their language performance or proficiency is often referred to as a kind of alternative assessment Alderson & Banerjee (2001) or alternative in assessment Brown & Hudson (1998). Advantages of alternative assessments including self-and peer-assessment are (a) quick administration; (b) students’ involvement in the assessment process; (c) enhancement of students’ autonomy of language learning utilizing involvement; and (d) increase of students’ motivation toward language learning Blanche & Merino (1989), Brown & Hudson (1998), cited from Manami Suzuki (undated). Disadvantages of alternative assessment are concerned with reliability and validity Blanche (1988), Blanche & Merino (1989), Blue (1988), Jafarpur (1991), cited in Manami Suzuki (undated). The number of studies comparing different kinds of assessments at the same time is still limited. Many authors have proposed self-assessment as a way to escape the reliance on judgments by others, engage students in understanding and internalizing standards and success criteria, and empower students to make judgments about their own performance for themselves Biggs & Moore (1993), Boud (1995, 2002). The evidence for the effectiveness of self-assessment is strong Hattie (2009). However, self-assessment is incompatible with summative assessment. There is evidence suggesting that

self-assessments are weakly to moderately correlated with the judgments of other experts Lew, Alwis, & Schmidt (2010), Ward et al. (2002).

Thus, peer assessment (PA) comes into play. PA can be described generally as a process whereby students evaluate or are evaluated by their peers. In educational practice, this occurs in many different forms.

There are several types of PA including grading a peer's research report, providing qualitative feedback on a classmate's presentation, or evaluating a fellow trainee's professional task performance. Peer assessment has become increasingly popular in education. As a learning tool, assessing their peers can provide students with skills to form judgments about what constitutes high-quality work Topping (1998). PA as an assessment tool can provide teachers with a more accurate picture of individual performance in group work Cheng & Warren (2000). Van Lehn et al. (1995) suggested that peer assessment demands cognitive activities such as reviewing, summarizing, clarifying, giving feedback, diagnosing errors, and identifying missing knowledge or deviations from the ideal. In peer assessment, students have more opportunities to view assignments of peers than in usual teacher assessment settings. Instead of modeling a teacher's cognitive product or process, students learn through cognitive modeling of peers' work. Moreover, peer assessment emphasizes providing and receiving feedback. Previous studies Crooks (1988), Kulik & Kulik (1988), Bangert-Drowns et al. (1991) indicated that receiving feedback is correlated with effective learning. Receiving abundant and immediate peer feedback can prevent some errors and provide hints for making progress. In addition to the positive effects of peer assessment, Lin et al. (2001a) observed that some students had negative feelings about this learning strategy. Some students disliked peer assessment because raters were also competitors. In one case, students could change their previous scores during a certain period. Upon receiving an unexpectedly low score from peers, students often reduced the previous scores given to others as a form of retaliation. Moreover, students often believe that only teachers have the ability and knowledge to evaluate and provide critical feedback Zhao (1998). They may suspect peers' ability, in particular, those who receive lower scores regard peer assessment as inaccurate McDowell (1995). Furthermore, many educators refuse to adopt peer assessment owing to the possibility of over-marking or under-marking peers' performance. It has been evidenced that peer assessment is compromised by existing peer relationships in class.

Topping (2009, p. 24) wrote: "Social processes can influence and contaminate the reliability and validity of peer assessments. Peer assessments can be partly determined by friendship bonds, enmity, or other power processes, the popularity of individuals, perception of criticism as socially uncomfortable or even collusion to submit average scores, leading to lack of differentiation". Falchikov (2001) distinguishes peer assessment from peer feedback; in the former, involvement marks or grades are assigned while in the latter, only comments are provided. Wiliam (2011) presents evidence from teacher assessments that the provision of grades undermines the positive influence of any feedback. Applying this point to peer-assessment, it seems that peers should not be asked to assign grades but to

focus on providing comments. It was revealed that a way for engaging the students actively in feedback processes is to implement peer review which is beginning to receive more attention in higher education Liu and Carless (2006), Cartney (2010), Nicol (2011). This focus on teacher feedback has mainly stemmed from the Interaction Hypothesis Gass (1997), Long (1996), Pica (1994) cited in Gorzynski-Weiss and R'ev'esz (2012), which claims that feedback obtained during interaction may be beneficial and, at times, essential for L2 learning. Interactional feedback may take a variety of forms, including explicit techniques such as overt corrections and metalinguistic explanations, as well as more implicit recasts (i.e., reformulations of an erroneous utterance) and other forms of negotiation (e.g., clarification requests), which do not overtly signal the unacceptability of a non-target like an utterance. Teacher-student feedback moves, in particular, have also often been classified into the categories of explicit correction (i.e., clear signaling of error plus target like reformulation), recasts, and prompts, where prompts refer to feedback techniques that encourage learners to repair their own utterances, instead of supplying the correct forms Lyster & Ranta (1997). It is positively presumed that feedback provides an opportunity for learners noticing of mismatches between problematic interlanguage constructions and those of the target language, potentially leading to restructuring Gass (1997), Schmidt & Frota (1986) cited in Gorzynski-Weiss and R'ev'esz (2010). Derived mostly from the works of Vygotsky (1978), two major learning theories - psycholinguistic theory and sociocultural theory-support collaboration in learning and claim that learning is a social activity. The notions of interaction hypothesis and ZPD are also included in these approaches which highlight the importance of collaboration as well as social interaction in learning. These movements brought about some new roles for teachers. They needed to investigate the best methods and techniques to incorporate into their classrooms to improve autonomy, reflectivity, and critical thinking practices among learners Vygotsky 1978, Brown, H. D. (2004). Regarding the notion of performance-based assessment, different studies have been conducted Parti (2002), Reinersten & Wells (1993), Roskams (1999), Stout (1993), Warren & Cheng (2005) advocating the role of collaborative learning, peer feedback, self and peer-assessment, as well as other factors in improving students' speaking ability.

3. Methodology

The present study aimed at investigating the effects of different kinds of feedback-based assessments on enhancing Iranian EFL learners' oral skills by comparing the peer, self, and teacher feedback-based assessments of oral feedback in different classes. Moreover, the potentially different influences of these assessments on the speaking abilities of different groups were investigated. Moreover, different priorities of criteria used by students were assessed considering the teacher as an interesting topic for exploration. Finally, students' perceptions towards these kinds of feedback-based assessments according to their answers on survey sheets were explored. This study is based on a quasi-experimental design with a between-groups comparison in which independent groups were compared followed by another phase looking for students' views through an attitude survey proposed by Jacobs et al. (1998), Zhang (1995)

and Eddy White (2009).

3.1 Participants

The participants of this study included 6 classes of Iranian EFL learners taught by the teacher and researcher. The number of participants was 84 in speaking courses with approximately the age range of 22 to 25 years and an average of 23. The age considerations may be worthy of notice concerning student views of assessment. As the sample, 100 students were considered, of which 84 were chosen to participate in this study according to the results of their selection test held by the institute and a TOEFL exam held by the researcher. Forty-five of the students were male and 39 were female. The primary learning objectives of the speaking course focused on developing student skills in planning, organizing, and delivering effective oral presentations. All of the participants enrolled in the Sharif language institute in Tehran, Iran.

This institute was chosen because the researcher was teaching English there and she was able to conduct her research more easily with her colleagues and manager's cooperation. The students were categorized into 3 groups. Each group included 2 classes and each class consisted of 14 students of intermediate level. As the classes were all conversational classes, a level of language proficiency (intermediate) was considered the best for this research to remove the likely effect of the variable of proficiency level. All of the participants were full-time university students of different majors. The research was set up during one semester. As mentioned, the students were classified into 2 categorical groups: 1) A control group receiving teacher-feedback-based assessment; 2) An experimental group which was further divided into two sub-groups, i.e., the group receiving peer feedback-based assessment and the one receiving the self-feedback-based assessment.

3.2 Instruments

Four instruments were used in the present study including learner background information, a general proficiency test, a teacher-peer- and Self-assessment sheet of students' performance, and a questionnaire based on students' attitudes toward peer- teacher- self-feedback-based assessment. First, the learner background questions on age, sex, native language, the duration of learning English, and their class level in the institute were written at the beginning of an English proficiency test, which was the Test of English as a Foreign Language (TOEFL) to homogenize participants. The learner's background information was added just because of probable needs for future studies. Questions related to the learners' background information were added by the author at the beginning of the TOEFL exam. Second, to ensure the homogeneity of the groups, a practice TOEFL test (retrieved from www.nbe.ir) was administered apart from a placement test held by the institute itself. TOEFL test contains 3 sections: Section one comprises 50 listening comprehension tests. The audio file was obtained from the manager of the Mirdamad branch of the Iranmehr institution. It consists of 2 parts. In part A, the learner will hear short conversations between two speakers. At the end of each conversation, a third speaker will ask a question about what the first two speakers said. Each conversation and question will be played only one time. In part B, the learner will hear longer conversations. After hearing the

questions, the learner should choose the best answer through four possible items. Section two includes 40 structure and written expression tests with 25 min for answering. It consists of 2 parts: questions 1-15 are partial sentences with 4 multiple choice options. In questions 16-40, every sentence has four underlined words or phrases, that are marked as (A), (B), (C), and (D). The sentence should be changed to become correct. Section three includes 50 reading comprehension questions with a time of 55 min to be answered. In this part, there are three passages, each followed by four or five questions and the need to select the best answer. C) In the assigned speaking classes, no textbook was used for these classes which were held for 7 weeks, two sessions per week. However, the theoretical and practical frameworks were based on a journal article by Yamashiro and Johnson (1997) entitled Public Speaking in EFL: Elements of Course Design. In this article, Yamashiro and Johnson introduced a Public Speaking course developed and used at both secondary and tertiary levels in Japan. A key element of their speaking course is a reference list (Table 1) of speaking elements covered in the course. The assessment entire rubric criteria used in this study were based on these 14 points.

Table 1. Fourteen Points for Public Speaking (Yamashiro & Johnson, 1997, p. 14)

Speaking Area	Comments
Voice Control	
1. Projection	Speaking loud enough (not too loud nor too soft)
2. Pace	Speaking at a good rate (not too fast nor too slow)
3. Intonation	Speaking using proper pitch patterns and pauses
4. Diction	Speaking clearly (no mumbling or interfering accent)
Body Language	
5. Posture	Standing with back straight and looking relaxed
6. Eye Contact	Looking each audience member in the eye
7. Gesture	Using few, well-timed gestures, nothing distracting
Content of Oral	
Presentation	
8. Introduction	Including an attention-getting device, thesis statement
9. Body	Using academic writing structure and transitions
10. Conclusion	Including restatement/summation & closing statement
Effectiveness	
11. Topic Choice	Picking a topic that is interesting to the audience
12. Language Use	Varying types of clear and correct sentence forms
13. Vocabulary	Using vocabulary appropriate to the audience
14. Purpose	Fulfilling the purpose of the speaking task

These 14 elements of public speaking provided criteria, based on which the students had to focus, practice, and assess their own or peers' speaking skills. Based on the recommendations of an expert in this field and the needs of the course, number 11 was removed because the topic was given by the teacher. It should be mentioned that these 13 points rubrics were distributed in 3 groups. However, the control group just had to practice the rubric. Any other activity like giving feedback and assessing should be left for just the teacher herself. D) A questionnaire based on students' attitudes toward peer feedback-based assessment and self-feedback/assessment based on Eddy White (2009) was distributed and completed by 56 members of the experimental groups in the final session. The survey was divided into three sections 1) a rater/ rated by peers, 2) the PA process/ the SA process, and 3) additional comments (open-ended). The open-ended questions in this study allow the students to explain their beliefs in more detail. For gauging the students' perceptions of the teacher feedback-based assessment process, a student survey was distributed and completed in the final session by 28 members of the control group Jacobs et al. (1998), Zhang (1995) cited in Yang Miaoa, Richard Badger b, Yu Zhen (2006). Some points were modified according to the requirements of the present study.

3.3 Procedure

The purpose of this study was to investigate and compare the effects of peer feedback-based assessment, teacher feedback-based assessment, and self-feedback-based assessment on oral performance of EFL learners and gauging learners' attitudes towards these different ways of feedback and assessment. The following procedures were carried out to conduct the research: First, the researcher asked all of the participants who passed placement test held by the institute to take part in a TOEFL exam in order to ensure their homogeneity in general proficiency level. This TOEFL exam was obtained from Iranmehr language institute, Tehran, Iran.

Due to the fact that the researcher needed a qualified practice TOEFL test, she obtained it through her familiarity with the managers and stakeholders of one of Iranmehr branches. This practice TOEFL test included three sections: Section one: 50 listening comprehension tests Section two: 40 structure and written expression tests with 25 minutes to be answered. Section three: 50 reading comprehension questions with time of 55 minutes to be answered. Those students who took part in the exam and had approximately the same results been considered as a population of intermediate general proficiency level. Before the study began, all intermediate students were given an identifying number. 84 numbers were then drawn at random to participants in this study. 28 of these numbers were drawn and assigned to group A, 28 to group B and 28 to group C. groups B and C were considered as subset of a more general group which is group B (experimental groups). With a toss of a coin between group A and B, group A was designated the control group and groups C and B, the experimental groups. This random assignment of students to control and experimental groups means that the groups had an equal chance of being assigned to control and experimental status. The control group received teacher feedback-based assessment and experimental groups received peer- and self-feedback-base assessments separately. Random selection helped ensure equivalence of groups. Each of the three speaking classes

met weekly for 180 minutes, approximately 14 times during the semester. Class time involved such things as: examining and practicing elements of effective public speaking: their voice, intonation, body language, grammar, vocabulary, etc. The 13 Points rubric formed the backbone of course syllabus. It was copied and distributed to the 6 speaking classes on the first day of each class. Most subsequent classes involved examining these points and completing tasks focusing on each one. This rubric clarified the purpose of the course in details. It should be mentioned that these 13 points rubrics was distributed in 3 groups but the control group just had to practice the rubric and any other activity like giving feedback, assessing, etc. should be left for just the teacher herself. Then, a tutorial session was conducted for experimental group students to practice assessing and correcting their peers' spoken performance and their own performance. The training session was an in-class demonstration & modeling which lasted 2 hours per week for a total of about 2 hours. The teacher researcher taught them how to make comments by modeling the procedure. For example, in the peer-feedback group, Students were told about PA, provided with a rationale for why it would be included in the course, and were given the 13 criteria that would be used by peers to assess and score their classmate presentation. In the next phase, students were given just one presentation topic which is the same for all three groups. As the semester progressed, students became very familiar with the differing aspects comprising the key elements of public speaking the course focused on (voice control, body language, content, and effectiveness). After 4 sessions passed, students delivered a 2- or 3-minutes presentation and the teacher researcher tape-recorded students' voices of each presentation in all groups using a tape-recorder. In control group, after each presentation, teacher comments orally and gives oral feedback to the whole class. Then recorded file has been transcribed for content analysis. The teacher researcher returns the transcriptions to the students. In peer-feedback (group B) of experimental groups, after each presentation, students begin giving oral feedback to the presenter and assess him/her on the peer rating sheet anonymously. The recorded file of the whole session remained with the teacher researcher for transcribing main points and analyzing its content and to be given to the presenters if they wished. Each session consisted of 7 presentations, having 2 sessions for only delivering presentation. Classmates give a mark out of 5 to the presenter.

Students were assessed on a five-point scale (5=excellent, 1=poor) by the teacher in group A and by peers in group B. Afterwards, sheets and marks are given to the teacher anonymously and teacher hand it to the presenter. During the treatment, it was announced to the students that their given feedbacks and marks would be anonymous and remain confidential. The teacher also asked students not to write their names on the peer-rating sheet. The procedures in the self-feedback of experimental group are somewhat the same with peer-feedback group. The only thing which is different is that after giving presentation, presenters should assess and give oral feedbacks to their own presentation. The recorded file was given to the presenters if they wished. Students give a mark out of 5 to themselves. Then, sheets and marks are given to the teacher. The same procedure has been carried out for students' second presentations which were delivered on 12th and 13th sessions. At the end of the term, an

attitude questionnaire was conducted. at the end of the semester and after completing the survey, students had to be interviewed by another teacher who was given the structured feedback sheet to mark students on the basis of determined features and then the teacher researcher compared the results of groups. Interviewer is an independent rater, an experienced English teacher in the same institute who was asked to score students using the same guide but without being told about the purpose of study or mentioning the different groups students are in.

4. Data Analysis and Results

In the present study, a quasi-experimental design was used and the data were analyzed quantitatively. The first step was to find out any significant difference in the oral performance of EFL Iranian students as run via peer feedback-based assessment, teacher feedback-based assessment, and self-feedback-based assessment in speaking classes. To answer this question, oral performance was considered the dependent variable. Different kinds of the assessment referred to as teacher-, peer- and self-feedback-based assessments were considered as one independent variable. The teacher feedback-based group became the control group and other groups were considered the experimental groups. The mean score of the last interview was compared among these groups. As stated in Julie Pallant (2010), the data were analyzed quantitatively: “One-way analysis of variance involves one independent variable possessing several different levels. These levels correspond to the different groups or conditions.” Julie Pallant (2010, p. 249). In this study, three kinds of giving feedback and assessment were considered as three levels of independent variable “assessment”. To find out any significant difference among these groups, a one-way ANOVA analysis (between-groups ANOVA/independent group design) was run using the 16th version of the Statistical Package for Social Sciences (SPSS). The dependent variable in this study was scored on oral performance. A significant F test indicates that the author can reject the null hypothesis revealing that the groups’ means are equal. However, no difference was represented among the groups. Thus, the researcher needed to conduct tests. The data collected from all groups were carefully coded and voices were transcribed investigating 13 points of Yamashiro & Johnson 1997. To improve intercoder reliability, coding was done independently by 2 other teachers as well in the search for 13 points. When there was disagreement among the teachers regarding the manner of classifying and coding the points, any disagreement was discussed to reach a consensus and raise the level of agreement.

4.1 The Frequency of Feedbacks

The researcher analyzed the collected data using one-way ANOVA and Post-hoc tests for the quantitative questions. For content analysis of oral feedback, some frequency tables were computed using the 16th version of SPSS (Table 3). Finally, the researcher tried to find out the preferences and perceptions of students towards different feedbacks by analyzing and comparing students’ comments. The total number of the 13 points used by participants (516 times in 1080 minutes of giving feedback) in each group is presented in Table 2.

Table 2. The Statistics for the Frequency of 13 Points in Peer-feedback, Self-feedback, and Teacher-Feedback Groups

Feedback		projection	pace	intonation	diction	posture	Eye contact	gesture	introduction	body	conclusion	Languages	vocabulary	purpose
Peer-feedback	N valid	1	1	1	1	1	1	1	1	1	1	1	1	1
	missing	0	0	0	0	0	0	0	0	0	0	0	0	0
	Mean	17	2	72	4	2	5	0	0	0	15	43	7	0
	Range	0	0	0	0	0	0	0	0	0	0	0	0	0
	Sum	17	2	72	4	2	5	0	0	0	15	43	7	0
Self-feedback	N valid	1	1	1	1	1	1	1	1	1	1	1	1	1
	missing	0	0	0	0	0	0	0	0	0	0	0	0	0
	Mean	0	0	0	0	5	0	0	0	0	0	12	2	0
	Range	0	0	0	0	0	0	0	0	0	0	0	0	0
	Sum	0	0	0	0	5	0	0	0	0	0	12	2	0
Teacher-feedback	N Valid	1	1	1	1	1	1	1	1	1	1	1	1	1
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0
	Mean	20	12	92	21	13	5	0	14	35	24	62	27	5
	Range	0	0	0	0	0	0	0	0	0	0	0	0	0
	sum	20	12	92	21	13	5	0	14	35	24	62	27	5

As displayed in Table 2, the teacher-feedback group exceeded other groups in providing the frequency of oral feedback on oral performances on the whole. It is also worth mentioning that Intonation and Language use components were used above all other components in peer- and teacher-feedback groups. It was also revealed that no Gesture component was used in any group. In other words, the participants and also teacher gave no feedback under the category of Gesture.

Besides, the students in the self-feedback group were not willing to give oral feedback on their own presentations. According to the table, the whole oral feedbacks given by this group were 19 in 360 minutes of class.

Conversely, the peer-feedback group liked this part of the discussion and gave 167 oral feedbacks in 360 min which was significantly more than the self-feedback group and lower than feedback given by the teacher in teacher-feedback classes. To compare the frequency of overall scores in presentations 1 and 2, another frequency table (table 2.1 & 2.2) is shown here:

Table 2.1. The Overall Score in Presentation 1

Assessment			Frequency	Percent
peer-feedback	Valid	weak	2	7.1
		average	22	78.6
		good	3	10.7
		very good	1	3.6
		Total	28	100.0
self-feedback	Valid	average	1	3.6
		good	25	89.3
		very good	2	7.1
		Total	28	100.0
		poor	1	3.6
teacher-feedback	Valid	weak	5	17.9
		average	13	46.4
		good	8	28.6
		very good	1	3.6
		Total	28	100.0

Table 2.2. The Overall Score in Presentation 2

assessment			Frequency	Percent
peer-feedback	Valid	average	5	17.9
		good	19	67.9
		very good	4	14.3
		Total	28	100.0
self-feedback	Valid	good	7	25.0
		very good	21	75.0
		Total	28	100.0
teacher-feedback	Valid	average	4	14.3
		good	16	57.1
		very good	8	28.6
		Total	28	100.0

As displayed in Table 2.1, in the first presentation, a total of 46% of students (13 out of 28) were considered by the teacher as Average in the teacher-feedback group while at the same time, the majority of students, 89%, were graded as Good in the self-feedback group. In the peer-feedback group, 22% of the students were considered Average.

Similarly, the highest percentage in the peer-feedback group was about 68% for good score, 75% for Very Good in the self-feedback group, and 16% for Good in the teacher-feedback group. The students did not consider any Poor or Weak points for their presentations, while the results of the other two groups were more authentic. This may be due to students' willingness to score their peers as authentically as possible. Chang et al. (2012) and Sadler and Good (2006) reported that peer-raters are stricter than self-raters (Table 3). Here, the dependent variable was the oral performance of students (score in the final interview) and the independent variable was assessment with 3 levels.

Table 3. The Descriptive Analysis

final interview	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
peer-feedback	28	86.5	5.3	1	84.5	88.6	75	100
self-feedback	28	76.4	3.8	.7	74.9	77.9	70	82
teacher-feedback	28	88.6	7.9	1.5	85.5	91.7	70	99
Total	84	83.8	7.9	.87	82	85.6	70	100
Model		Fixed Effects	5.9	.65	82.5	85.1		
		Random Effects		3.7	67.6	100		

According to Tables 3 and 2.1 and 2.2, the comparison of the teacher and peer assessment scores about the scores at an interview shows that the peers and teachers still interpret the criteria and indicators of the rubric differently. This can be explained by differences in the width and depth of their experience basis. Also, based on the content analysis of the recorded sessions, within the group of peers, not all students could apply the same criteria in a comparable and/or consistent way.

Moreover, the finding that peers gave higher marks as compared to teachers is in agreement with the results of other studies Langan et al. (2008). Furthermore, the self-assessment scores are higher than the marks given by the teacher, which is consistent with the results reported in the literature Patri (2002).

4.2 Final Interview

As represented in Table 4, the F ratio is 33.783 which is largely higher than 1. Thus, there is some effect of the treatment. It also shows ($P < 0.05$, $df = 81$) a statistically significant difference among groups. Accordingly, the null hypothesis is rejected because sig is smaller than 0.05. A large amount of F ratio was also incorporated to support the fact that the results of this study are statistically significant. The groups did not perform in the same way on the test because we could reject the H_0 at the .05 level. Therefore, the null hypothesis was rejected because of the large gap between the groups. However, we

still do not know exactly where the difference lies. To discover the precise location of the difference, we need to perform a post hoc (Table 5) comparison of the means using the Tukey test.

Table 4. ANOVA: Final Interview

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2402.214	2	1201	33.783	.000
Within Groups	2879.821	81	35.5		
Total	5282.036	83			

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgements and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, non-italic) in addition to the style provided by the drop-down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic.

If there are two or more sub-topics, the next level head should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

The table 5 shows the difference between the groups. The precise location of difference lies in the self-feedback group which is highly different from the other two groups. The values of 10.14286 and 12.25000 in the self-feedback group showed that the mean difference is significant at the level of 0.05. If the reliability and validity studies are to be conclusive, the related investigations must cover larger samples, sufficient course units, and participants from diverse educational backgrounds.

According to Sung et al. (2003) and Sung et al. (2005), repeated practice of self-assessment will help to narrow down the gap between student-based and teacher-based scores. In this view, it was assumed that the students are more likely to sharpen rating abilities when they are provided with sufficient practice thus making a higher validity possible.

To answer the second question of this study, the author tried to discover student perceptions about peer-feedback-based assessment, self-feedback-based assessment, and teacher feedback-based assessment by investigating their written comments and survey items. This discussion is separated into three parts, following the format of the survey. The beginning section deals with student views as a peer assessor (survey items 1-8, modified items for self-assessment group). The second part denotes the student views of peers and self-assessment (items 9-10).

Finally, student comments from section three of the survey were used to share the student perspectives in their own words.

Table 5. The Multiple Comparisons: Final Interview Tukey HSD

(I) assessment	(J) assessment	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
peer-feedback	self-feedback	10.14*	1.59	.000	6.3	13.9
teacher-feedback	peer-feedback	2.10	1.59	.387	-1.6	5.9
	self-feedback	12.25*	1.59	.000	8.4	16

*. The mean difference is significant at the 0.05 level.

Table 6. The Peer-feedback Group Survey

Question			Frequency	Percent	Valid Percent	Cumulative Percent
Question 1	Valid	Agree	25	89.3	89.3	89.3
		Tend to agree	3	10.7	10.7	100.0
		Total	28	100.0	100.0	
Question 2	Valid	Agree	24	85.7	85.7	85.7
		Tend to agree	4	14.3	14.3	100.0
		Total	28	100.0	100.0	
Question 3	Valid	Agree	6	21.4	21.4	21.4
		Tend to agree	2	7.1	7.1	28.6
		Tend to disagree	7	25.0	25.0	53.6
		Disagree	13	46.4	46.4	100.0
		Total	28	100.0	100.0	
Question 4	Valid	Agree	19	67.9	67.9	67.9
		Tend to agree	5	17.9	17.9	85.7
		Tend to disagree	4	14.3	14.3	100.0
		Total	28	100.0	100.0	
Question 5	Valid	Agree	22	78.6	78.6	78.6
		Tend to agree	3	10.7	10.7	89.3
		Disagree	3	10.7	10.7	100.0
		Total	28	100.0	100.0	
Question 6	Valid	Agree	17	60.7	60.7	60.7
		tend to agree	3	10.7	10.7	71.4
		tend to disagree	7	25.0	25.0	96.4

		Disagree	1	3.6	3.6	100.0
		Total	28	100.0	100.0	
Question 7	Valid	Agree	25	89.3	89.3	89.3
		tend to agree	3	10.7	10.7	100.0
		Total	28	100.0	100.0	
Question 8	Valid	Agree	15	53.6	53.6	53.6
		tend to agree	10	35.7	35.7	89.3
		tend to disagree	3	10.7	10.7	100.0
		Total	28	100.0	100.0	
Question 9	Valid	Agree	3	10.7	10.7	10.7
		tend to agree	1	3.6	3.6	14.3
		tend to disagree	6	21.4	21.4	35.7
		Disagree	18	64.3	64.3	100.0
		Total	28	100.0	100.0	
Question 10	Valid	Agree	22	78.6	78.6	78.6
		tend to agree	3	10.7	10.7	89.3
		tend to disagree	2	7.1	7.1	96.4
		Disagree	1	3.6	3.6	100.0
		Total	28	100.0	100.0	

4.3 Peer Feedback-based Assessment Group

The 13 key points for public speaking were considered for the rating sheets used in the course, from Yamashiro and Johnson's (1997) syllabus, related to voice control, body language, content, and effectiveness. Table 6 shows the percentage of each Likert scale in each question counted using the 16th version of the Statistical Package for Social Sciences (SPSS).

Owing to the space limitation, we ignored the details. Generally, the findings for the second research hypothesis show significant differences in the learners' attitudes toward self-feedback, teacher-feedback, and peer-feedback. According to Table 6, the friendship did not influence giving scores in the peer-feedback group which was caused by anonymity in giving scores and telling students those scores did not affect their final grades. About 70% of students indicated that peer ratings were fair and reasonable. They also felt comfortable being judged and assessed by their classmates. This is inconsistent with the result of Falchikov (2000) study indicating that students often dislike being assessed by peers, which have power over them. This study's conclusion is due to the students' comfort in judging peers in the class as a whole, which is possibly related to the situation and setting of this study, condition of the treatment, atmosphere of the classes, and the new experience of students, and some other factors.

Table 6 shows that one of the key reasons for using peer assessment is to provide a way of getting much more feedback from students as compared to a sole teacher assessment as stated by Topping “swifter feedback in greater quantity” Topping (1998, p. 255). Therefore, the students’ attitudes toward peer feedback were positive and they preferred to be involved in their own learning, giving feedback, assessing their peers, and being assessed by them.

4.4 Self Feedback-based Assessment Group

As mentioned before, the original peer rating sheet was modified to the new context for the self-assessment group. Modifications include removing questions 3, 4, 7, 10, and 11 and modifying the wording of questions 5, 6, 8, 9, and 12. It also consists of three parts in the same way as a peer-assessment group.

According to Table 7, in the self-feedback-based assessment group, students’ preferences had their speaking scored and corrected by just their teacher. One reason is that self-assessment imposes a heavy burden on students. Similarly, a meta-analysis of Falchikov (2005) indicated that few students can assess the way teachers apply assessment criteria. This is confirmed by a study by Kruger and Dunning (1999) where novices and low performers overestimate their performance level and lack related metacognitive abilities (monitoring, evaluation). The results of the survey in peer feedback-based assessment and teacher feedback-based assessment groups show that over 80% of the students found peer feedback useful (Table 6) compared to 57% for teacher feedback (Table 8.1).

Table 7. The Self-feedback Group Survey

Questions		Frequency	Percent	Valid Percent	Cumulative Percent
	Agree	24	85.7	85.7	85.7
Question 1	Valid Tend to agree	4	14.3	14.3	100.0
	Total	28	100.0	100.0	
	Agree	6	21.4	21.4	21.4
	Tend to agree	1	3.6	3.6	25.0
Question 2	Valid Tend to disagree	15	53.6	53.6	78.6
	Disagree	6	21.4	21.4	100.0
	Total	28	100.0	100.0	
	Agree	10	35.7	35.7	35.7
	Tend to agree	2	7.1	7.1	42.9
Question 3	Valid Tend to disagree	1	3.6	3.6	46.4
	Disagree	15	53.6	53.6	100.0
	Total	28	100.0	100.0	
	Agree	11	39.3	39.3	39.3
Question 4	Valid Tend to agree	13	46.4	46.4	85.7

		Tend to disagree	4	14.3	14.3	100.0
		Total	28	100.0	100.0	
		Agree	26	92.9	92.9	92.9
Question 5	Valid	Tend to agree	2	7.1	7.1	100.0
		Total	28	100.0	100.0	
		Agree	26	92.9	92.9	92.9
Question 6	Valid	Tend to agree	2	7.1	7.1	100.0
		Total	28	100.0	100.0	
		Tend to agree	1	3.6	3.6	3.6
		Tend to disagree	20	71.4	71.4	75.0
Question 7	Valid	Disagree	7	25.0	25.0	100.0
		Total	28	100.0	100.0	

4.5 Teacher Feedback-based Assessment Group

Table 8. Items 3 & 4

Question		Frequency	Percent	
		very useful	23	82.1
question 3	Valid	useful	5	17.9
		Total	28	100.0
		useful	18	64.3
		a little useful	7	25.0
question 4	Valid	not useful at all	3	10.7
		Total	28	100.0

Table 8.1. Item 5

Question		Frequency	Percent	
		I prefer to receive only teacher-feedback	16	57.1
		I prefer to receive only peer-feedback	1	3.6
question 5	Valid	I prefer to receive both teacher and peer feedback	9	32.1
		I prefer to receive no feedback	2	7.1
		Total	28	100.0

They preferred to be involved in their own learning, give feedback, assess their peers, and be assessed by them. The students' scores in the final interview are to some extent correlated with the teacher-feedback group. Just 3 out of 28 students thought that giving scores should remain just for the teacher. This may indicate that experience of having peer feedback has a positive impact on student perceptions and their preference for peer feedback. Moreover, by training students after a while, the teacher feedback-based assessment can be replaced by peer-feedback-based assessment.

5. Discussion and Conclusion

In the present work, three groups of speaking classes of Iranian EFL were studied, the learners with the same proficiency level, age range, and the same teacher, teacher feedback-based assessment group performed significantly better in oral performance. Peer feedback-based assessment group was correlated with the teacher feedback-based assessment group to some degree. There was no significant difference in the scores of the final interview. By training the students after a while, teacher-feedback-based assessment can be replaced by peer-feedback-based assessment. On the other hand, the self-feedback-based assessment group performed highly differently from the other two groups. Students in this group were mostly over-grade in comparison with the two other groups. As a result, self-assessment is not a valid way of assessment for oral skills and few students can assess in the way teachers apply assessment criteria. Another finding is that teacher feedback-based assessment groups exceeded other groups in providing oral feedback. Peer feedback-based assessment group was mainly concerned with intonation, language use, projection, conclusion, vocabulary, eye contact, diction, posture, and pace, respectively. While the teacher in the teacher feedback-based assessment group highlighted intonation, language use, body, vocabulary, conclusion, diction, projection, introduction, posture, pace, and eye contact, respectively. On the other hand, the self-feedback-based assessment group just paid attention to language use, posture, and vocabulary, respectively. The study shows that if the criteria of assessment are clear and students have a full understanding of each item, any kind of assessment will be more effective. Anonymity should be taken into account as anonymity in giving a score. Moreover, not having anxiety about the effect of the score on the final result helps the students to ignore friendship or any other kind of relationship with their classmates while making them give fair scores. Last but not least is that cultural elements are so important in students' preferences for having peer-, self-, or teacher assessment. In a culture with the teacher-only model of assessment and feedback, it takes too much time to move towards other kinds of assessment and feedback. This is in line with Allaei and Connor (1990) who found that students' culture had a significant impact on determining the amount of effectiveness of feedback and students' perceptions. As a result, more practice and more experience are needed to sharpen rating abilities in students and make them active learners. Generally, necessary training and experience of peer feedback have a positive impact on student perceptions and their preference to be involved in their own learning, giving feedback, assessing their peers, and being assessed by them.

The above findings can indeed be added to a growing body of literature on effective techniques through which students can improve their own speaking skills. The current research tried to open a new window of opportunity for students to try other kinds of assessment in a teacher-oriented assessment culture. This study revealed some interesting results about the self-, peer, and teacher assessment field of oral presentation skills. These discussions led the researcher to question “how a teacher can improve the quality of self-and peer assessment approaches and apply all three kinds of assessment and give feedback in a class?”

In conclusion, the researcher hopes that these suggestions are effective for the teachers, no matter how experienced they are, as a starting point for the development of their own teaching and incorporating all these strategies in their own classes. In respect of some limitations and delimitations, other studies can cover the variables that the researcher could not control. For instance, this study exclusively focused on group assessment, whole-class feedback, the age range of the young people, and intermediate level of English proficiency. Future studies could test the impact of gender, paired assessment instead of group assessment, using other data collection procedures than recording the voices, studying the effect of advanced level of proficiency of students in such studies as well as performing these three kinds of assessment in just one class, not in separate classes. In the present study, the participants also had no experience or knowledge of peer, self-feedback-based assessments, which caused some difficulties for the researcher in the data collection. Moreover, it was time-consuming to familiarize the students with these methods, which should be considered by the researchers in other similar studies. Further studies should be also performed in larger groups to facilitate generalization and incorporate the support or questions in the present study. It is also suggested to perform another replication set of studies to corroborate and validate this study in another context. Since the results of this study were drawn on the data collected through recorded voices and a final interview at the end of the term and attitude surveys, similar studies could be carried out by triangulating the results through further methods of data collection such as observations, video-recording, and interviews instead of attitude survey. A combination of these methods can also give a more trustworthy set of data to the researchers.

This comparative study was performed among three groups with three different kinds of treatment. Further research may be also needed to examine the influence of combining these three methods in one group and its effect on the same students. To the best knowledge of the researcher, no study was performed on the long-term effects of such techniques in an Iranian EFL setting. Accordingly, another line of longitudinal research is recommended concerning the lasting effects of these techniques.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Alderson, JC & Banerjee, J. (2001). Language testing and assessment (Part 1): state-of-the-art review. *Language Teaching*, 34(4), 213-236. <https://doi.org/10.1017/S0261444800014464>
- AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, 32, 407-425. <http://dx.doi.org/10.1016/j.system.2004.04.006>
- Allaei, S. K., & Connor, U. M. (1990). Exploring the dynamics of cross-cultural collaboration in writing classrooms. *The Writing Instructor*, 10, 19-28. <http://hdl.handle.net/1805/2659>
- Bangert-Drowns, R., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213-238. <https://doi.org/10.3102/00346543061002213>
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39(3), 313-340. <https://doi.org/10.1177/003368828801900105>
- Brown, A. (2005). Self-assessment of writing in independent language learning programs: The value of annotated samples. *Assessing Writing*, 10, 174-191. <https://doi.org/10.1016/j.asw.2005.06.001>
- Brown, D., & Hudson, T. (1998). The Alternatives in Language Assessment. *Tesol Quarterly*, 32(4), 653-675. <https://doi.org/10.2307/3587999>
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*, San Francisco. Longman Publications.
- Campbell, K., Mothersbaugh, D., Brammer, C., & Taylor, T. (2001). Peer versus self-assessment of oral business presentation performance. *Business Communication Quarterly*, 64(3), 23-42. <https://doi.org/10.1177/108056990106400303>
- Cartney. (2010). comparative study of peer and teacher feedback and of various peer feedback forms in a secondary school writing curriculum. *British Educational Research Journal*, 36(1), 143-162. <https://doi.org/10.1080/01411920902894070>
- Chang, C.-C. et al. (2013). Is learner self-assessment reliable and valid in a Web-based portfolio environment for high school students? *Computers & Education*, 60, 325-334. <https://doi.org/10.1016/j.compedu.2012.05.012>
- Cheng, W., & Warren, M. (1997). Having second thoughts: Students perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2), 233-239. <https://doi.org/10.1080/03075079712331381064>
- Falchikov, N. (2005). Improving assessment through student involvement. *Practical solutions for language learning in higher and further education*. New York: Routledge.
- 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: 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, 287-322. <https://doi.org/10.3102/00346543070003287>
- Farhady, H., Ja'farpur, A., & Birjandi, P. (1994). *Testing language skills: From theory to practice*. Tehran: SAMT Publications.
- Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Erlbaum.
- Heaton, J. B. (1988). *Writing English language tests*. Cambridge: Longman Publications.
- Henrich R., Greve. (2003). *Organizational Learning from Performance Feedback: A Behavioral Perspective on Innovation and Change*. Cambridge: Cambridge University Press.
- Jacobs, G. M., Curtis, A., Braine, G., & Huang, S.-Y. (1998). Feedback on student writing: Taking the middle path. *Journal of Second Language Writing*, 7(3), 307-317. [https://doi.org/10.1016/S1060-3743\(98\)90019-4](https://doi.org/10.1016/S1060-3743(98)90019-4)
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence led to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134.
- Kulik, J., & Kulik, C. (1988). Timing of Feedback and Verbal Learning. *Review of Educational Research*. 58(1). <https://doi.org/10.3102/00346543058001079>
- Langan, M., Shuker, D., Cullen, R., Penney, D., Preziosi, R., & Wheeler, P. (2008). Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. *Assessment & Evaluation in Higher Education*, 33(2), 179-190. <https://doi.org/10.1080/02602930701292498>
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics*, 1, 49-74. <https://doi.org/10.1177/1362168810383328>
- Larsen-Freeman, D. (2000). *Techniques and Principles in Language Teaching*. Oxford: Oxford University Press.
- Li (2010). The Effectiveness of Corrective Feedback in SLA: A Meta-Analysis. *Language learning*, 60(2), 309-365. <https://doi.org/10.1111/j.1467-9922.2010.00561>
- Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279-290. <https://arrow.tudublin.ie/ltecpdprp/27>
- Lydia Wen, M., & Tsai, C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education*, 51, 27-44. <https://doi.org/10.1007/s10734-004-6375-8>
- Gurzynski-Weiss, L., & Revesz, A. (2012). Tasks, teacher feedback, and learner modified output in naturally occurring classroom interaction. *Language Learning*, 62(3), 851-879. <https://doi.org/10.1111/j.1467-9922.2012.00716.x>
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.). *Handbook of second language acquisition*. San Diego, CA: Academic Press.

- Lyster, R., & Ranta, L. (1997). Corrective Feedback and learner uptake: Negotiation form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37-66. <https://doi.org/10.1017/S0272263197001034>
- Min, H.-T. (2005). Training students to become successful peer reviewers. *System*, 33(2), 293-308. <https://doi.org/10.1016/j.system.2004.11.003>
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125-143) New York: Erlbaum.
- Nicol, D. (2010). From Monologue to Dialogue: Improving Written Feedback in Mass Higher Education. *Assessment & Evaluation in Higher Education*, 35(5), 501-517. <https://doi.org/10.1080/02602931003786559>
- Nunan, D. (1988). *The learner-centered curriculum*. Cambridge: Cambridge University Press.
- Nunan, D. (2004). *Task-based language teaching*. Cambridge: Cambridge University Press.
- Pallant, J., (2010). *Spss survival manual: a step-by-step guide to data analysis using Spss*. Australia: Open University press.
- Parti, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing Journal*, 19(2), 109-131. <https://doi.org/10.1191/0265532202lt224oa>
- Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning*, 44, 493-527. <https://doi.org/10.1111/j.1467-1770.1994.tb01115>
- Reinertsen, S. P., & Wells, C. M. (1993). Dialogue journals and critical thinking. *Teaching Sociology*, 21(2), 182-186. <https://doi.org/10.5829/idosi.hssj.2012.7.1.1104>
- Roskams, T. (1999). Chinese EFL students' attitudes to peer feedback and peer assessment in an extended pair work setting. *RELC Journal*, 30(1), 141-157. <https://doi.org/10.1177/003368829903000105>
- Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31. https://doi.org/10.1207/s15326977ea1101_1
- Stout, C. J. (1993). The dialogue journal: A forum for critical consideration. *Studies in Art Education*, 35(1), 34-44. <https://doi.org/10.2307/3193811>
- Topping, K. (2009). Peer assessment. *Theory into Practice*, 48(1), 20-27. <https://doi.org/10.1080/00405840802577569>
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249-276. <https://doi.org/10.2307/1170598>
- Twardy, R. C. (2005). Argument Maps Improve Critical Thinking. *Teaching Philosophy*, 27(2), 95-116. <https://doi.org/10.5840/teachphil200427213>
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge: Harvard University Press.

- Warren, N., & Cheng, W. (2005). Peer assessment of language proficiency. *Language Testing Journal*, 22(1), 93-121. <https://doi.org/10.1191/0265532205lt298oa>
- Yamashiro, A., & Johnson, J. (1997). *Public Speaking in EFL: elements for course design*. *The Language Teacher*. Retrieved September 10, 2014 from <http://www.jaltpublications.org/ltt/files/97/apr/yamashiro.html>
- Yang Miao, Richard Badger, & Yu Zhen. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15, 179-200. <https://doi.org/10.1016/j.jslw.2006.09.004>