Studies in Second Language Learning and Teaching
Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz
SSLLT 3 (1). 87-107
http://www.sslIt.amu.edu.pl

Self-, peer-, and teacher-assessment: An investigation into Iranian EFL students' attitudes

Zainab Abolfazli Khonbi
Urmia University, Iran
z.abolfazli1986@gmail.com

Karim Sadeghi
Urmia University, Iran
ksadeghi03@gmail.com

Abstract

Student-directed teaching and assessment techniques are gradually dominating educational systems almost all over the world. This study investigated a cohort of male and female Iranian EFL students' attitudes toward self-, peer-, and teacher-assessment experiences. Sixty three students at Urmia University and Tabriz Islamic Azad University, in the form of three intact classes, experienced self-, peer-, and teacher-assessment activities for one academic semester (having taken a knowledge pretest, four assessment series, and a course achievement posttest). Of all the participants, 38 completed a 5-point Likert-scale attitude questionnaire. The application of ANOVA, across and within group dependent samples t tests, as well as some qualitative analyses, indicated that the three experimental groups had positive attitudes toward their assessment experiences. While the peer-assessment group was the most positive in this regard, slight differences were found in the groups' attitudes and beliefs. Further findings and implications are discussed in the paper.

Keywords: course achievement, peer-assessment, self-assessment, teacher-assessment, student attitudes

Academic assessment serves several purposes. For example, Boud (1990) argues that assessment of students "improve[s] the quality of learning . . . [and] the accreditation of knowledge or performance" (p. 102). These two purposes can be differentiated in terms of formative assessment, which serves students' learning needs, and summative assessment, which serves the needs of the society to evaluate the end-result of schooling.

Many studies indicate that student learning is positively influenced by assessment (Black & William, 1998; Kennedy, Chan, Fok, & Yu, 2008). However, over the past 50 years, the institutes producing and administering standardized tests have been enjoying a social, political, and organizational power leading to the so-called test driven curricula where educational materials are directed toward the content of the test rather than toward learning what the learners should be learning. This aspect of assessment, known as negative washback, has diminished the learning value of language instruction across the world (Farhady, 2006).

Since it provides feedback, assessment informs students about their strengths and weaknesses and indicates the next steps to take in the learning process. One important condition for assessment to support student learning is the active involvement in the assessment process on the part of students themselves (Black & William, 1998). Alternative assessment methods, while associated with students' learning approaches (Scouller, 1998), include self- and peer-assessment, and are designed to develop active, autonomous, responsible, and reflective learners (Sambell & McDowell, 1998).

Self-assessment (SA) is viewed as an individual's own evaluation of their language ability, generally according to how good they are at particular language skills (e.g., reading and speaking), how well they are able to use the language in different domains or situations (e.g., at the office and at school), or how well they can use different styles of the language (e.g., a formal and an informal style) (Mousavi, 2012). As a variant of SA (Mousavi, 2012), peer-assessment (PA) is defined as "an arrangement in which individuals consider the amount, level, worth, quality of success of the products or outcomes of learning of peers of similar status" (Topping, 1998, p. 250).

Attitude is referred to as the degree of an individual's like or dislike of an institution, situation, event, and the like. In other words, it is the tendency to react consistently favorably or unfavorably to a stimulus (persons, objects, or concepts). In the case of language testing, the stimulus is a language or speakers of a language. Attitudes tend to be quite stable and less subject to factual input than are beliefs and opinions (Mousavi, 2012). Richards and Schmidt (2002) speak of attitude as the language attitudes which speakers of different languages or language varieties have toward each other's languages

or towards their own language. Expressions of positive and negative feelings towards a language may reflect impressions of linguistic complexity or simplicity, ease or difficulty of learning, degree of importance, elegance, social status, and so on (p. 286). In this study, by *attitude* we mean the attitude not toward a language and the people speaking that language, but the attitude toward the experience of a certain type of learning and assessment.

Oscarson (1989) briefly itemized the rationale of SA procedures in language learning (pp. 4-6):

- *Promotion of learning:* SA gives learners training in evaluation, which, in itself, is beneficial to learning.
- *Raised level of awareness:* Training in self-assessment stimulates learners to consider course content and assessment principles in a more discerning way than is usually the case.
- *Improved goal-orientation:* The practice of self-assessment further tends to enhance learners' knowledge of the wide variety of possible goals in most language learning contexts.
- *Expansion of range of assessment:* In certain respects, the learner's own appreciation of his competence in the language is, for natural reasons, superior to that of an outside tester, namely in areas of affective learning (in turn contributing somehow to students' attitudes toward learning and assessment).
- *Shared assessment burden:* It has been pointed out that a further positive aspect of learner participation in assessment is the possibility that it may alleviate the assessment burden on the teacher (Dickinson, 1987).
- *Beneficial postcourse effects:* Teaching students how to carry on learning the language autonomously after the course is universally considered an important objective in foreign language instruction.

As regards PA, Topping (1998) argued that given the many different types of PA, establishing a single overarching mechanism or model of the process seems likely to be difficult. In a literature review, he stated that the literature proposes many hypotheses about the mechanisms through which PA may create its effects. He categorized these mechanisms according to the following domains:

- *Cognition and metacognition:* In a review of the wider literature on peer-assisted learning, Topping and Ehly (1998) noted that cognitively PA might create effects by increasing a number of variables for assessors, assessees, or both. These variables could include levels of time on task, engagement, and practice, coupled with a greater sense of accountability and responsibility. They argued that formative PA is likely to involve intelligent questioning, together with increased self-disclosure and, thereby, assessment of understanding. Furthermore,

they believe that PA could enable earlier error and misconception identification and analysis and could lead to the identification of knowledge gaps and to the engineering of their closure through explaining, simplifying, clarifying, summarizing, reorganizing, and cognitive restructuring. Finally, they asserted that cognitive and metacognitive benefits might accrue before, during, or after PA.

- *Affect:* Both assessors and assesses might experience initial anxiety about the process. However, PA involves students directly in the learning process and may promote a sense of ownership, personal responsibility, and motivation. Giving positive feedback first might reduce assessee anxiety and improve acceptance of negative feedback. PA might also increase variety and interest, activity and interactivity, identification and bonding, self-confidence, and empathy for others.
- *Social and transferable skills:* PA can develop teamwork skills and promote active rather than passive learning. It can also develop verbal communication skills, negotiation skills, and diplomacy (Riley, 1995).
- *Systemic benefits:* PA can give students greater insight into institutional assessment processes (Fry, 1990). Students might thus develop more confidence in these processes and greater tolerance of the inevitable difficulties of discrimination at the margin. Alternatively, if institutional assessment procedures are inadequate, greater awareness of this among students could generate a positive feeling toward improvement.

Studies concerned with PA show that students generally display a liking for PA activities because these activities provide an opportunity for comparison of student work, but, simultaneously, students are much less appreciative of criticism from peers (Brindley & Scoffield, 1998; Smith, Cooper, & Lancaster, 2002; Williams, 1992). However, studies also show a lack of self-confidence in students when they rate their peers (Sullivan, Hitchcock, & Dunnington, 1999), and the need for a pre-existing guideline or rule for the assessment activity (Orsmond & Merry, 1996).

Clifford (1999) devised a learner-controlled learning environment and utilized both PA and SA techniques to help students develop autonomy. The researcher surveyed students' attitudes toward PA and reported that students found PA activities educative, but they also felt frustrated when they had no clear frameworks and guidance. Furthermore, these students viewed grading as the teacher's responsibility because they wished to stay away from "the process of evaluating . . . peers' performance where marks are concerned" (p. 122).

Davies (2000) investigated a group of undergraduate students after they participated in a computerized PA project and found that nonanonymous PA was negatively perceived by these students, and that this could be associated

with the difficulty of criticizing or rating their peers. However, other research has suggested that anonymous assessment could provide more truthful and appropriate attitudes toward the assessment processes (Ballantyne, Hughes, & Mylonas, 2002). Nonetheless, students' resistance to and negative perceptions of PA could discourage teachers from using such technological innovations (Cohen, 1988; McNeil, 1988).

The study conducted by Bullock (2011) on teachers' beliefs about learner SA looked at issues surrounding learner SA and teachers' beliefs. His study was designed to explore teachers' attitudes, beliefs, and behavior with regard to learner SA during the implementation of a revision of assessment procedures for teens aged 14-16 years. He used both quantitative and qualitative methods to discover what teachers understood by SA and in what ways, if any, they had implemented this. The relationship between attitudes, beliefs, and practices was also explored; accordingly, he identified some specific factors responsible for facilitating or obstructing implementation. The findings showed that teachers believed that (a) when supported, learners benefit from assessing their own work, (b) SA raises learners' awareness of their strengths and weaknesses, and (c) SA stimulates motivation and involvement in the learning process. During the interviews with teachers, he identified some other favorable attitudes toward SA: (a) it is a good idea in theory, (b) it is better than tests, (c) it adds structure and context, (d) it works, (e) it helps students see their progress, and (f) it gives students greater ownership.

For the student who receives peer review, studies report deepened subject matter knowledge (Barak & Rafaeli, 2004) and a more positive attitude toward it (Katstra, Tollefson, & Gilbert, 1987). In his review of the PA research, Topping (1998) noted that students' acceptance of PA is quite independent of their knowledge about the demonstrated reliability and validity of that assessment. Davies (2000) and Liu and Carless (2006) noted that some students who have negative perceptions of PA doubt the expertise of their fellow students (as compared to their instructors). Researchers have also hypothesized that students' discomfort and negativity can be traced to the problematic power relations that students associate with assessing their peers (Liu & Carless, 2006). These studies indicate that students may have trouble with the nontraditional idea of their peers assessing their work in place of an instructor. Furthermore, variation in students' attitudes about assessment may rely a great deal on how individual instructors introduce and plan SA and PA. Studies have also advocated certain steps to alleviate students' negative perceptions of PA, including more PA experience (Wen & Tsai, 2006), clarity about PA criteria (Smith et al., 2002), and support and training in regard to the PA process (Cheng & Warren, 1997; Falchikov, 2005, 2007).

Beyond the need to know what attitudes students have about SA and PA in order to design better SA and PA processes, we also need more evidence about how SA and PA, and students' perceptions of them impact the quality of their work, particularly in the realm of higher education where student satisfaction becomes increasingly important for course evaluation and class selection by students. Studies indicate that students with positive perceptions of teaching and course goals often adopt a deep approach to studying, whereas students with negative perceptions about course workload and the appropriateness of assessment often adopt a surface approach to studying (Crawford, Gordon, Nicholas, & Prosser, 1998; Lizzio, Wilson, & Simons, 2002). However, the studies in question are based solely on students' self-reports and not empirical measures of students' work. Some investigations into students' attitudes towards and beliefs about self-, peer-, and teacher-assessment have been conducted either previous to or after the experiments (Richards & Schmidt, 2002). In 38 control group studies measuring tutor achievement, tutors outperformed controls in 33 (Topping, 1996). In addition, improved tutor attitudes and self-concept as a result of peer-tutoring have been reported (Topping, 1996). As mentioned above, research studies carried out in the same areas as the present research have found that students may have different attitudes toward the practices of self-, peer-, and teacher-assessment, and that these differences in attitude may influence their performance (Roskams, 1999; Van Zundert, Sluijsmans, & Van Merriënboer, 2010). It should be added that students' responses and attitudes toward SA also appear to differ depending on the cultural and educational contexts in which teaching and learning take place (Oscarson, 1997).

Still, in the Iranian context, regardless of the issue of learner-centered and individualistic approaches to teaching prevalent elsewhere, most of the evaluations and assessments in formal educational settings are exercised by teachers. To contribute to the thriving body of research in this field, this study sought to investigate undergraduate EFL students' attitudes toward self-, peer-, and teacher-assessment in the form of student-generated tests (Brown, 2004), after they had experienced them for an academic semester in their Teaching Methods course. The pedagogical findings of the present research may be of interest to a variety of individuals as well as groups. They could be regarded as providing insight for curriculum developers, course designers as well as teachers and professors themselves. The following research question was the foundation of this research project: Is there any statistically significant difference among Iranian university EFL students' attitudes toward their self-, peer-, and teacher-assessment experiences?

Method

Participants

The participants were 63 male and female EFL students at Urmia University and Tabriz Islamic Azad University, West and East Azarbaijan Provinces, Iran, working towards a BA in English Language and Literature. They were within the age range of 20 to 22 and had experienced self-, peer-, and teacher-assessment for an academic semester. However, not all the participants completed an attitude questionnaire, and this study is based on the performance of 38 participants (i.e., 11, 15, and 12 candidates in the self-, peer-, and teacher-assessment groups, respectively) who filled out the relevant questionnaires.

Instruments

As was the case with similar investigations (e.g., Bullock, 2011), the main material used for the present research was an attitude questionnaire which was distributed at the end of the semester to the students who had just completed their self-, peer-, and teacher-assessment treatments. The three questionnaires used for the purposes of gathering information on the students' attitudes and beliefs were compiled on the basis of those used in some previous studies of a similar kind. They were checked for validity by scholars and were found to be valid to a satisfactory degree. Their reliability was established by calculating Cronbach's alpha: for the SA group (with 11 cases and 22 items) $r = .73$, for the PA group (with 15 cases and 22 items) $r = .78$, and for the teacher-assessment (TA) group (with 12 cases and 20 items) $r = .86$.

The questionnaires consisted of three parts and had 26 items in the self- and peer-assessment groups, but only 24 items in the TA group. All of the items, except the last four, were answered by means of a 5-point Likert scale ranging from 1 (*strongly agree*) to 5 (*strongly disagree*). The first 16 items in the case of the SA and PA groups, and the first 14 items in the case of the TA group, measured students' attitude toward the practice and the process of assessment, while the next six items were meant to tap into the students' beliefs about possible outcomes of these assessment methods. The last four were open-ended questions asking students for their justifications of their group's performance.

Procedure

Originally, the study followed a semi-experimental intact group design in which the classes were randomly assigned to self-, peer-, and teacher-

assessment groups. Later on, a comparative design was used to compare the students' attitudes toward their assessments and to examine if any significant differences existed among them. At the beginning of the term, all the groups took a pretest which measured the students' existing knowledge of a specific course book. Then, in the SA and PA groups, the students were trained on how to assess themselves as well as their peers, respectively. While in the SA group they were instructed to make, answer and mark their own papers, in the PA group the students were instructed to take tests made and marked by anonymous peers. In the TA group, however, the teacher was asked to design and mark the papers. Implementing the procedures as described to them and having received an assessment every two units, the experimental groups took four assessment series during the term. At the end of the term, all the three groups received an end-of-the-course achievement posttest.

When the treatment was finished, an investigation was carried out as to the extent to which the groups differed in terms of their attitudes toward the assessment types they had experienced. The students were required to respond to the Likert scale questionnaire items and to give up to three main reasons for their performance. Since the students needed to reflect upon their experiences during the relevant treatment, they were not required to complete the questionnaire within a certain time limit; however, it did not take more than 30 min.

Results

Quantitative Analysis

Table 1 shows the descriptive statistics for the groups on the whole questionnaire (consisting of 22 items in the case of self- and peer-assessment groups and 20 items for the teacher-assessment group). In this table, *N* stands for the number of items in the questionnaires, and *M* shows the mean score (obtained from the values of 1 corresponding to *strongly agree*, 2 to *agree*, 3 to *cannot decide*, 4 to *disagree*, and 5 to *strongly disagree*). For instance, it can be inferred that in the PA group, the grand mean for all 22 questions was 2.27, that is, in this group students had an attitude somewhere between *agree* and *cannot decide* toward their PA. This was the most positive attitude, which was followed by TA and then SA groups.

Table 1 Descriptive statistics for each group on the whole questionnaire

	N	*M*	*SD*	*SE*	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
PA	22	2.2791	0.40078	0.08545	2.1014	2.4568	1.33	3.13
SA	22	2.7686	0.34865	0.07433	2.6141	2.9232	2.30	3.55
TA	20	2.6480	0.39238	0.08774	2.4644	2.8316	1.90	3.33
Total	64	2.5627	0.43093	0.05387	2.4550	2.6703	1.33	3.55

In order to analyze the data, a one-way analysis of variance (ANOVA) was applied to compare the mean scores of the groups. Table 1 showed that the groups had generally a more favourable attitude toward the assessments (the means were all below 3, which means *cannot decide*); however, the ANOVA table that follows (Table 2) shows that there were significant differences among the groups' attitudes, with $F(2, 61) = 9.81$, $p < .001$.

Table 2 ANOVA of the groups' mean scores on the whole questionnaire

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	2.848	2	1.424	9.814	.000
Within Groups	8.851	61	.145		
Total	11.699	63			

Table 3 is a report of the post-hoc tests used to identify significant differences among the groups. It was found that the PA group ($M = 2.27$, $SD = 0.40$) differed significantly ($p = .00$) from both the SA ($M = 2.76$, $SD = 0.34$) and TA ($M = 2.64$, $SD = 0.39$) groups in being the most positively disposed towards the assessment experience.

Table 3 Post-hoc Tukey HSD test (multiple comparisons) for the groups' mean scores on the whole questionnaire

(I) group	(J) group	Mean Difference (I-J)	*SE*	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
PA	SA	-.48955*	0.11485	.000	-.7654	-.2136
	TA	-.36891*	0.11769	.007	-.6516	-.0862
SA	PA	.48955*	0.11485	.000	.2136	.7654
	TA	.12064	0.11769	.564	-.1621	.4033
TA	PA	.36891*	0.11769	.007	.0862	.6516
	SA	-.12064	0.11769	.564	-.4033	.1621

* The mean difference is significant at the .05 level

Based on the above analyses, it was found that the groups differed in terms of the level of positive or negative disposition toward the assessments. Therefore, the null hypothesis was rejected. However, in order to get a general

view of the amount of agreement and disagreement across the three experimental groups, a paired samples *t* test was conducted to explore the difference between the students' positive and negative attitudes toward their experiences of self-, peer-, and teacher-assessment. To this end, the average of *strongly agree* (StA) and *agree* (A) options (as an index of agreement) was compared with the average of *disagree* (DA) and *strongly disagree* (StDA) options (as an index of disagreement). Tables 4 and 5 that follow show the results of this analysis. It was found that there was a statistically significant difference across the three groups between the mean scores of both SA and A options, that is, positive attitudes (M = 11.45, SD = 5.07) on the one hand and DA and SDA options, that is, negative attitudes (M = 4.32, SD = 3.63) on the other, in favor of SA and A options, with $t(37) = 5.44$, $p < .00$ (two-tailed). This indicated that the groups had more agreement toward the assessments (eta squared = .44).

Table 4 Inferential statistics: Paired samples *t* test for the StA and A vs. DA and StDA options across groups

		M	*N*	*SD*	*SE* Mean
Pair 1	StA and A	11.45	38	5.076	0.823
	DA and StDA	4.32	38	3.632	0.589

Table 5 Paired samples *t* test for the StA and A vs. DA and StDA options across groups

		Paired Differences					*t*	*df*	Sig. (2-tailed)
		M	*SD*	*SE* Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	StA and A – DA and StDA	7.132	8.081	1.311	4.475	9.788	5.440	37	.000

Within groups, a paired samples *t* test was also applied to see what the story is as to the amount of agreement and disagreement within each group. The result for the SA group is presented in Tables 6 and 7. Based on these two tables, the SA group showed no significant difference (p = .12) between the positive attitudes (M = 10.91, SD = 4.74) and negative attitudes (M = 6.45, SD = 4.54), with $t(10) = 1.65$, $p < .00$, with the eta squared of .21.

Table 6 Inferential statistics: Paired samples *t* test for the SA and A vs. DA and SDA options in SA group

		M	*N*	*SD*	*SE* Mean
Pair 1	StA and A	10.91	11	4.742	1.430
	DA and StDA	6.45	11	4.547	1.371

Table 7 Paired samples *t* test for the StA and A vs. DA and StDA options in the SA group

		Paired Differences					*t*	*df*	Sig. (2-tailed)
		M	*SD*	*SE* Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	StA and A – DA and StDA	4.455	8.904	2.685	-1.527	10.436	1.659	10	.128

Within the PA group, a paired samples *t* test revealed significant difference ($p = .00$) in the mean of agreement options ($M = 14.33$, $SD = 3.83$) and disagreement options ($M = 2.53$, $SD = 1.95$), with $t(14) = 8.46$, $p < .00$., with the eta squared of .83. It can be concluded that the PA groups students expressed significantly more positive (than negative) attitudes toward PA. The results are presented in Tables 8 and 9.

Table 8 Inferential statistics: Paired samples *t* test for the StA and A vs. DA and StDA options in the PA group

		M	*N*	*SD*	*SE* Mean
Pair 1	StA and A	14.33	15	3.830	0.989
	DA and StDA	2.53	15	1.959	0.506

Table 9 Paired samples *t* test for the SA and A vs. DA and SDA options in the PA group

		Paired Differences					*t*	*df*	Sig. (2-tailed)
		M	*SD*	*SE* Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	StA and A – DA and StDA	11.800	5.401	1.395	8.809	14.791	8.462	14	.000

As for the amount of agreement and disagreement in the TA group, paired samples *t* test results yielded no significant difference ($p = .12$) between the means of agreement ($M = 8.33$, $SD = 5.03$) and disagreement options ($M = 4.58$, $SD = 3.42$), with $t(11) = 1.66$, $p < .00$, with the eta squared of .20. Tables 10 and 11 show the results.

Table 10 Inferential statistics: Paired samples *t* test for the StA and A vs. DA and StDA options in the TA group

		M	*N*	*SD*	*SE* Mean
Pair 1	StA and A	8.33	12	5.033	1.453
	DA and StDA	4.58	12	3.423	0.988

Table 11 Paired samples *t* test for the StA and A vs. DA and StDA options in the TA group

		Paired Differences					*t*	*df*	Sig. (2-tailed)
		M	*SD*	*SE* Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	SA and A – DA and SDA	3.750	7.818	2.257	-1.217	8.717	1.662	11	.125

Based on the above analyses, an across group paired samples *t* test found that all the three experimental groups showed more agreement toward the statements in the attitude questionnaires. However, a within group paired samples *t* test found that the difference as to the amount of agreement versus disagreement was significant only in the case of the PA group and not in the SA and TA groups. Although the existence of limitations in the sample size, in the proficiency level of students among other things, reduces the generalizability of results, based on the students' responses to the items in the questionnaires, some tentative conclusions may be drawn concerning students' attitudes toward the assessments.

The students agreed that self-, peer-, and teacher-assessment raise their awareness of their strengths and weaknesses, stimulate motivation and involvement in the learning process, produce independent (autonomous) learners, make them competent at recognizing areas needing further study, make them read more seriously about the subject matter, prepare them for the final examinations, improve the quality of their learning and achievement, provide them with an accurate and fair assessment of themselves, are beneficial to the learning process, facilitate learning, and help learning through enjoying feedback. Students also agreed that these assessment methods are useful means for assessing achievement, and that none of them is a waste of teaching time. In the case of self- and peer-assessment, students also agreed that if they are trained and get experience in these assessment procedures, they can do the assessments effectively, and if given clear criteria for item construction and marking, they can undertake the assessments in a consistent fashion.

Students in the SA group also agreed that SA would lead to surface-level learning, memorization, not taking the task seriously, no study, designing easy items, and being lax in scoring. On the other hand, students in the PA group also agreed that PA may lead to real learning, in-depth study, stricter scoring of peers' achievement, enjoying peer-feedback, feelings of competition, and fear of losing face in front of peers. Students in the TA group agreed, to some extent, that TA can lead to surface-level learning, memorization, not taking the task seriously, no study, enjoying feedback, and feelings of competition.

Qualitative Analysis

In this section, the students' answers to the four open-ended questionnaire items are briefly listed. The students were required to answer the questions in English; however, due to lexical and grammatical errors, their responses were first corrected, redundancies were omitted, and then the responses were summarized. Tables 12, 13, and 14 list examples of the experimental groups' attitudes toward and beliefs about their experience of self-, peer-, and teacher-assessment, respectively. In the first column of the tables, the questions are provided and in the second column, the students' responses are briefly stated.

Table 12 SA group's attitudes toward self-assessment

Questionnaire items	Students' responses
SA group's mean scores were higher than the PA group's in all four series of assessments, because of:	▪ Remembering the questions and answers ▪ Designing easy items ▪ Correcting the papers ▪ Being lax in scoring
SA group failed to outperform PA group on the final achievement test, because of:	• Not valid with questions of little importance • Not getting used to other types of questions, not studying well • Not studying properly, no concern for the scores • PA group's having more study during the term to answer the questions not aware of content • PA group's having more practice
Recommend or reject; in other words, state the advantages and disadvantages of SA.	A ▪ The opportunity to examine oneself; getting a new way of testing ▪ Better understanding ▪ A vision of one's skill in testing ▪ No stress DA • Remembering the questions or also some answers; no guarantee for learning the material, just memorization • Correcting one's own paper, not afraid of the scores; no serious study • Self-selecting parts to study; not studying in detail and completely
If anything else was missing from this investigation of the effect of SA on Iranian university EFL students' course achievement which you think you would like to add, mention it.	▪ Desire to use PA, getting acquainted with many different kinds of questions

A = advantages, DA = disadvantages

Table 13 PA group's attitudes toward peer-assessment

Questionnaire items	Students' responses
PA group's mean scores were lower than the SA group's on all four series of assessments, because of:	▪ Not anticipated questions, their type; not knowing how to study ▪ Somehow strange questions, sometimes not standard, and difficult; problems understanding them ▪ Prejudice in grading ▪ Peers being very strict in scoring ▪ Taking part in such activities for the first time ▪ SA group's memorizing the answers ▪ SA group's cheating, giving high scores to oneself
PA group outperformed SA group on the final achievement test, because of:	• Experiencing different tests, a variety of questions • No knowledge of the questions, PA forcing study • Papers being assigned by peers; deep study • The opportunity to be involved in a give and take practice; remembering the materials easily • Fear of a low score in final exam; more study • Having to study all the materials for designing questions; PA students' being well prepared for the tests • Students' studying the questions when tests finished; aware of weaknesses; trying to improve • Having to study alone, PA being an effective activity • Considering PA a final opportunity (for preparing for the final exam)
Recommend or reject; in other words, state the advantages and disadvantages of PA.	A ▪ Causing better study ▪ Scoring peers each session; encouraging ▪ Good and useful activity ▪ A kind of motivation for study ▪ Review the course book to have a better performance ▪ Familiarity with different kinds of questions ▪ Having to study, strengthening knowledge; advantageous ▪ An effective and motivating activity ▪ Helping more detailed study ▪ Studying deeply; recognizing areas of weakness through groupwork ▪ So beneficial for students, stimulating motivation, preparing for final term examinations; doing better than the usual schedule DA • Needing time
If anything else was missing from this investigation of the effect of PA on Iranian university EFL students' course achievement which you think you would like to add, mention it.	▪ Good, happening in an intimate atmosphere ▪ A most effective criteria for evaluation purposes, if names shown; paying attention in giving scores

A = advantages, DA = disadvantages

Table 14 TA group's attitudes toward teacher-assessment

Questionnaire items	Students' responses
TA group's mean scores were lower than SA and PA group's on all four series of assessments, because of:	▪ Not having real study and enough practice ▪ Not listening carefully to the teacher ▪ Too difficult questions ▪ Cooperation in SA and PA groups ▪ PA and SA groups' taking the tasks seriously and studying harder ▪ Very boring class, not getting the lesson well
TA group failed to outperform SA and PA groups on the final achievement test, because of:	• Having stress • Low motivation; not studying well
Recommend or reject; in other words, state the advantages and disadvantages of TA.	▪ Having feedback, improving students' knowledge ▪ Continuous review
If anything else was missing from this investigation of the effect of TA on Iranian university EFL students' course achievement which you think you would like to add, mention it.	• Perfect, excellent • It was really complete

A categorization of students' attitudes toward the assessment type experiences in the tables above supported students' positive attitudes established by quantitative analysis. Generally speaking, students in the SA group justified their high performance on the assessment series by *making easy items, memorization, being lax in scoring*, and so on and stated that their low performance on the posttest was largely due to *no in-depth and serious study.* However, they recommended this type of assessment, for they believed it leads to *better understanding, no stress,* and *an opportunity to gauge and improve their testing skills*. They also showed a desire for PA in order to *get acquainted with different kinds of questions.*

Peer-assessment group members expressed the most beliefs and ideas about their assessment practice in comparison with the other two groups. They related their low performance on the series of assessments to *not anticipated and difficult questions,* and *peers' strictness in scoring.* As their statements indicated, their good performance on the posttest would be a sign of *deep and detailed study, receiving feedback, independent study,* and *review.* They mentioned such benefits of PA as *causing students to be familiar with different item types, bringing forth better study, an encouraging and beneficial activity, making students recognize their areas of strength and weakness,* and *a motivating and effective practice.*

Although the students in the self- and peer-assessment attributed their performances more or less to the assessment types, students in the teacher-assessment group, for the most part, identified individualistic reasons for their low performance both on the assessment series and on the posttest. They

claimed that *not listening carefully to the teacher, not having enough study, having stress, too difficult items, and low motivation* were the contributing factors. However, they expressed their positive attitudes by stating that TA *leads to continuous review and feedback which improves students' knowledge.* They also provided some reasons for the better performance of students in the self- and peer-assessment such as *cooperation, taking the tasks seriously,* and *harder study.*

Discussion

In this study, undergraduate EFL students experienced self-, peer-, and teacher-assessment exercises in the course of an academic semester in a Teaching Methods course and were investigated in terms of their attitudes toward their assessments at the end of the experiment. The study found that the SA group had the highest mean scores on the series of assessments followed by the PA group, and then the TA group, which gained the lowest mean scores. However, on the course achievement posttest, the PA group outperformed the others and the SA group obtained second best scores followed by the TA group. The analysis of attitude questionnaires indicated that the students in the three experimental groups had positive attitudes toward the associated assessment types. The PA group was found to have the most positive attitudes toward PA, although not differing significantly from the SA group. This finding may be related somehow to the better performance of the PA group on the posttest in comparison with the SA and TA groups and to the outperformance of the TA group by the SA and PA groups on the same test.

It can be concluded that assessment types (SA vs. PA vs. TA) and students' attitudes toward them and their perfomances might have affected each other. The assessments were found to positively affect students' performances on the posttest; however, they may have affected students' attitudes as well. In addition, students' performances on the posttest could be regarded as an influential factor for students' attitudes toward the assessments or the other way round.

Consistent with the findings of the present research regarding students' attitudes, some studies have observed a positive connection between prior training, accurate PA, and a favorable attitude toward the notion of PA (Williams, 1992). Van Zundert et al. (2010) found that the practice of PA improves students' performances and positively affects their attitudes toward its practice. Lupo (as cited in Oscarson, 1989) reports positive effects, as a result of positive attitudes towards the practice of SA, on the basis of the results of a survey among 15 to 16-year-olds. As it was mentioned by the students in the SA and PA groups, because they involve both reflection and evaluation of one's own performance, SA and PA were found to give students an opportuni-

ty to feel a sense of control over their own actions and to develop positive attitudes toward learning, thus increasing motivation (Paris & Paris, 2001).

While some research in the context of higher education indicates that students' satisfaction with school and positive perceptions of their learning environment influences their GPA (Grade Point Average) (Lizzio et al., 2002), studies specifically focused on PA have demonstrated that students' discomfort with peer review does not correlate with their grades (Simkin & Ramarapu, 1997).

Wen and Tsai (2006) in their study on university students' perceptions of and attitudes toward (online) PA found that participating students had a positive attitude toward general PA activities and they thought that PA guidelines were helpful for their understanding of the PA activity. However, in contrast to our findings, the researchers found these students had neutral attitudes toward online PA, indicating that they, on average, were neither positive nor negative toward the use of online PA activities. Similarly, they had neutral perceptions of the negative aspects of PA, demonstrating that these university students on average neither agreed nor disagreed with these negative aspects of PA. Wen and Tsai concluded that, on the one hand, the university students involved in the study thought that a major part, if not all, of the responsibility of grading should belong to the teacher/instructor. On the other hand, they also respected peers' judgments about their own performance. Their finding also concurs with many other studies (Brindley & Scoffield, 1998; Clifford, 1999), suggesting that more effort needs to be placed on giving students responsibility for grading to develop a sense of learner control and ownership of their own learning, especially in higher education.

Also in contrast to this study's findings, Kaufman and Shunn (2011) in their study on students' "perceptions about peer assessment for writing" (p. 387) found that students sometimes regard PA as unfair and often believe that peers are unqualified to review and assess their work. Furthermore, the researchers claimed that students' perceptions about the fairness of PA drop significantly following students' experience in doing PA. They concluded those students' fairness perceptions, and drops in those perceptions, are most significantly associated with their perceptions about the extent to which peers' feedback is useful and positive; however, students' perceptions appear to be unrelated to the extent of their revision work. Other studies on PA suggest that students' continued exposure to PA (Wen and Tsai 2006) will help them to view that assessment more positively.

The degree of care in preparing (often reluctant) students for SA and PA can yield very different effects. The generalizability threat to this research may be the small size and the gender balance that was skewed to females. No control over the proficiency levels of the students was exerted.

Through the use of SA and PA in more classes, students may begin to regard such assessments as a normal part of their education and may also understand more clearly how their own and peers' advice can contribute to their education. Continued use of SA and PA may also encourage students to see themselves as legitimate audience for their own and peers' work and thus a valuable source of feedback about that work. In addition, future research should examine the role that other individual differences play in student acceptance of peer feedback, such as motivational factors or resilience to critical feedback. Some studies across settings and subjects could provide more information about how students' perceptions influence the work they do and allow for some generalizations about how students' attitudes affect their performance. Future research could helpfully focus on how to assist students in providing useful feedback in the process of PA activities and how to enhance the validity of PA. Such studies might also emphasize the instructor's role in SA and PA and help teachers to carefully monitor and manage these processes. Gender and proficiency may be two interesting factors to be investigated. Miller and Ng (1994) stated that proficient and highly motivated L2 learners are able to more realistically assess their peers' language ability. Finally, similar studies could be done with groups of participants of very different social, cultural, and disciplinary backgrounds and L2 proficiency levels for comparison purposes. However, to date, very little systematic investigation has been conducted on how the teaching or learning context interacts with students' responses and attitudes towards SA, PA and TA (Butler & Lee, 2010).

Conclusion

The research project reported here found that students show positive attitudes toward and beliefs about self-, peer-, and teacher-assessment practices, with the PA group expressing significantly more positive than negative attitudes in this regard. The results of this study can be used in all educational centers. They may have direct and indirect implications and applications in teaching, learning, materials development, syllabus design, and curriculum and test development. University professors can benefit from SA and PA techniques (with an eagle eye on the latter) in order to educate more active and autonomous students who are at the same time better communicators.

References

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*, 427-441.

Barak, M., & Rafaeli, S. (2004). On-line question-posing and peer-assessment as a means for web-based knowledge sharing in learning. *International Journal of Human-Computer Studies, 61*(1), 84-103.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*, 141.

Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education, 15*, 110-113.

Brindley, C., & Scoffield, S. (1998). Peer assessment in undergraduate programmes. *Teaching in Higher Education, 3,* 79-89.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Pearson Education.

Bullock, D. (2011). Learner self-assessment: An investigation into teachers' beliefs. *ELT Journal*, *65*(2), 114-125.

Butler, Y., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5-31.

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*, 233-239.

Clifford, V. A. (1999). The development of autonomous learners in a university setting. *Higher Education Research & Development, 18*, 115-128.

Cohen, D. K. (1988). Educational technology and school organization. In R. S. Nickerson & P. P. Zodhiates (Eds.), *Technology in education: Looking toward* 2020 (pp. 231-264). Hillsdale, NJ: Lawrence Erlbaum.

Crawford, K., Gordon, S., Nicholas, J., & Prosser, M. (1998). Qualitatively different experiences of learning mathematics at university. *Learning and Instruction, 8*(5), 455-468.

Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Teaching International, 37*, 346-355.

Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge: Cambridge University Press.

Falchikov, N. (2005). Improving assessment through student involvement. 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.

Farhady, H. (2006). *Twenty five years of living with applied linguistics: Collection of articles*. Tehran: Rahnama.

Fry, S. A. (1990). Implementation and evaluation of peer marking in higher education. *Assessment and Evaluation in Higher Education, 15*, 177-189.

Katstra, J., Tollefson, N., & Gilbert, E. (1987). The effects of peer evaluation on attitude toward writing and writing fluency of ninth grade students. *Journal of Educational Research, 80*(3), 168-172.

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.

Kennedy, K. J., Chan, J. K. S., Fok, P. K., & Yu, W. M. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and cultural issues. *Educational Research for Policy and Practice, 7,* 197-207.

Lee, I. (1998). Supporting greater autonomy in language learning. *ELT Journal, 54*(4), 282-288.

Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279-290.

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education, 27*(1), 27-52.

McNeil, L. M. (1988). *Contradictions of control: School structure and school knowledge*. New York: Routledge.

Miller, L., & Ng, R. (1994). Peer assessment of oral language proficiency. *Perspectives: Working Papers of the Department of English, City Polytechnic of Hong Kong, 6,* 41-56.

Mousavi, S. A. (2012). *An encyclopedic dictionary of language testing*. Tehran: Rahnama.

Orsmond, P. & Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education, 21*, 239-250.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing, 6*(1), 1-13.

Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham, & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 175-187). Dordrecht: Kluwer Academic.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychology, 36*(2), 89-101.

Richards, J. C., & Schmidt, R. (2002). *Dictionary of language teaching and applied linguistics* (3rd ed.). London: Pearson Education.

Riley, S. M. (1995). Peer responses in an ESL writing class: Student interaction and subsequent draft revision. *Dissertation Abstracts International*, *56,* 3031.

Roskams, T. (1999). Chinese EFL students' attitudes to peer feedback and peer assessment in an extended pair work setting. *RELC Journal, 30*(1), 79.

Sambell, K., & McDowell, L. (1998). The values of self and peer assessment to the developing lifelong learner. In Rust, C. (Ed.), *Improving student learning – improving students as learners* (pp. 56-66). Oxford: Oxford Center for Staff and Learning Development.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education, 35*, 453-472.

Simkin, M. G., & Ramarapu, N. K. (1997). Student perceptions of the peer review process in student writing projects. *Journal of Technical Writing and Communication, 27*(3), 249-263.

Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International, 39*, 71-81.

Sullivan, M. E, Hitchcock, M. A., & Dunnington, G. L. (1999). Peer and self assessment during problem-based tutorials. *The American Journal of Surgery, 177*, 266-269.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68,* 249-276.

Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education, 32*(3), 321-345.

Topping, K. J., & Ehly, S. (1998). *Peer assisted learning*. Mahwah, NJ: Erlbaum.

Van Zundert, M., Sluijsmans, D. M. A., & Van Merriënboer, J. J. G. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, *20*(4), 270-279.

Wen, M. L., & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education*, *51*(1), 27-44

Williams, E. (1992). Student attitudes towards approaches to learning and assessment. *Assessment and Evaluation in Higher Education, 17*, 45-58.