THE EFFECT OF SELF- AND PEER-ASSESSMENT VERSUS TEACHER- ASSESSMENT TO ENHANCE WRITING SKILL OF IRANIAN EFL LEARNERS (A MIXED METHOD APPROACH)

Afsaneh Bahraini[1], Shohreh Raftari[2]*, Hossein Bagheri[3]

Abstract

This study was a mixed-method design to explore the efficiency of alternative assessments in boosting language performance. To achieve this objective, 60 Iranian Language learners participated in the study. The participants were advanced students of three intact classes in an Iranian online language center in Sirjan, Kerman (A province in South East Iran). The study focused on the effect of Self- and Peer-Assessment versus Teacher-Assessment on enhancing the Writing Skill of Iranian EFL Learners in the online context. To conduct the quantitative (true-experimental) phase, the three classes were randomly assigned to two experimental groups (N= 19, N=19) and a control group (N=22). The classes were randomly chosen to be studied as self-, peer- and teacher-assessment groups. To answer the research questions of the quantitative phase of the study, a mixed design ANCOVA with one within-subject variable (writing performance), one between-subject variable (group), and two covariates (OOPT and Pre-test scores) were launched. According to the results, the self-assessment group in both ratings obtained higher post-test scores compared with peer- and teacher-assessment groups. All groups showed significant gains from pre-test scores to post-test scores. However, the self-assessment group showed higher gains compared to the other two groups. The results of the qualitative (exploratory, descriptive, and interpretive) phase of the study revealed that nearly all the students believed that self-assessment processes in the light of the teacher's guidance led to their improvement in writing. The results indicated that students did better in realizing their problems and improving their writing when they received their peers' feedback.

Keywords: Assessment, Online Assessment, Self-assessment, Peer-assessment, Teacher-assessment

[1] M.A. in Applied Linguistcs, Kerman Bahonar University, Kerman, Iran.

[2] Assistant Professor, Department of General Education, Afzalipour Faculty of Medicine, Kerman University of Medical Sciences, Kerman, Iran. (Corresponding Author)

[3] Assistant Professor, Department of General Education, School of Medicine, Zahedan University of Medical Sciences, Zahedan, Iran.

Eur. Chem. Bull. **2023**,*12(3), 2371-2382*

2371

1. Introduction

During the late 1990s, language assessment witnessed a kind of paradigm shift by which the positivist norm-referenced paper-and-pencil tests were replaced by more contextualized learner-focused assessments whose primary purpose was to enhance learning rather than objectively measuring the learnt things (Gipps, 1994). Therefore, much attention was directed toward alternative assessments and its sub-categories as appropriate and viable alternatives to traditional tests (Hamp-Lyons, 2009). Two major types of alternative assessment, self-assessment, and peer assessment have received much attention from L2 scholars and practitioners (Sambell, McDowell, & Montgomery, 2012).

As far as L2 wiring is concerned, teachers of foreign languages should be able to train their learners with good writing habits so that they can produce high-quality written drafts (Baker, 2016). One such good habit is the fact that longer pieces of writing should be divided into smaller pieces for each of which the learners receive continuous feedback (Bean, 2011). Although such regular feedback provision can be very helpful for students' planned writings and thoughtful revisions, providing continuous feedback is demanding and timeconsuming for teachers as they should devote much time to reading the learners' drafts carefully and expressing their opinions on various mechanical, rhetorical, and linguistic dimensions of the writing tasks (Baker, 2016; Herrington & Cadman, 1991).

In the meantime, learners are less likely to answer all the feedbacks they receive consistently and may feel uncomfortable or frustrated to be given continuous teacher feedback (Jonsson, 2013). In addition, given the limited class hours dedicated to L2 writing courses, teachers may be unable to provide feedback on various drafts of learners in relatively bigger classes (Kumar, et al., 2022). As a result, of these seeming warranted hurdles for L2 writing practitioners, numerous writing researchers have turned to self-assessment and peer-assessment as viable and alluring alternatives for the traditional teacher-fronted writing assessment classes (Jensen & Fischer, 2005; Topping, 1998).

In online settings, there have been many reports on the problems of online writing assessment including computer integration training for writing teachers, the absence of computer-based writing instruction, and teachers' reluctance to transfer face-to-face activities to virtual spaces; moreover, since the breakout of the pandemic, writing assessment has required digital writing skills and fundamentals (Anasse & Rhandy, 2021) (Balkanski, et al., 2021).

This study was an attempt to explore the efficiency of alternative assessments in boosting language performance. In this regard, it focused on the effect of Self- and Peer-Assessment versus Teacher-Assessment on enhancing the Writing Skill of Iranian EFL Learners in an online context.

2. Materials and Methods

2.1. Participants

To fulfill the purpose of this research, 60 Iranian Language learners participated in the present study. The participants were advanced students of three intact classes in an Iranian online language center in Sirjan, Kerman. To homogenize the participants in terms of general English proficiency, Oxford Online Placement Test (OOPT) was administered to the students of the three groups (self, peer- and teacher assessment). The students were all studying English as advanced students in the institute, but they received different marks from the OOPT. It should be noted that according to the test results they were all considered advanced. Moreover, the range of their marks approved their proficiency level. They were both male and female students whose ages varied from 16 to 25 with a mean age of 20.5. The three classes were randomly assigned to two experimental groups (N= 19, N=19) and a control group (N=22). The classes were randomly chosen to be studied as self-, peer- and teacher-assessment groups. At the time of the study, the students were learning English writing as an obligatory part of their course.

2.2. Instruments

2.2.1. Phase 1: Quantitative (true experimental design)

The following instruments were used in phase one, quantitative (true experimental) design:

- Language Proficiency Test
- Timed-writing Essays (pretest and posttest)
- The Writing Scoring Scale

2.2.1.1. Language Proficiency Test

To homogenize the participants in terms of general English proficiency, Oxford Online Placement Test (OOPT) was administered to the students of the three groups (self-, peer- and teacher assessment) (Tuleubayeva, et al., 2021).

*Eur. Chem. Bull. **2023**,12(3), 2371-2382*

2372

OPT is considered a reliable and valid standard English proficiency test that can be administered to a different number of learners with various proficiency levels (Allen, 2004). The version of OPT that was used in this study was online and contained multiple-choice items measuring vocabulary, grammar, and reading. There was also an optional writing task. According to the candidate's scores, they were benchmarked on a continuous numerical scale.

2.2.1.2. Timed-writing Essays

To measure the writing performance of the participants before (i.e., as the pretest) and after the treatment (i.e., as the posttest), two 45-minute writing essays were administered to the participants of the three groups. To this end, two general topics which did not require any particular technical background knowledge were assigned for each administration. It should be noted that students took part in the test through adobe connect, so it was possible for the teacher to test them during a special time limit.

Topic 1: Some people argue that it might be better to raise children in the countryside than in big cities. What is your opinion?

Topic 2: Some people argue that success is the result of hard work and luck has no effect on someone's success. What is your opinion?

2.2.1.3. The Writing Scoring Scale

To score the participants' essays, Jacobs et al.'s (1981) writing scale was used which is mainly based on an analytical scoring procedure. The effectiveness of this writing scoring rubric has been verified by Brown and Baily (1984). According to this scale, a written text or an essay should be evaluated against a set of five criteria or subcategories such as content, organization, vocabulary, language use, and mechanics. This rubric consists of a 100-point scheme based on which 30 points are devoted to the content, 25 points to language use (mainly syntax), 20 points to organization, 20 points to vocabulary use, and 5 points to mechanics. To ensure the inter-rater reliability of the given scores to the essays, the written essays for both topics in the pre-test and post-test were rated by two independent researchers who were familiar with the scoring rubric. The given scores of the two raters, as well as those of the researcher's, were evaluated by Cohen's Kappa's interrater reliability test (Livingston, 2018).

2.2.2. Phase 2: Qualitative

The following instruments were used in phase two (the qualitative design) of the study:

- The students' self-assessment of Pre-test and Post-test

- Their final self-assessment of the Whole Process

- Semi-structured interview questions

2.2.2.1. Students' self-assessment of Pre-test and Post-test

After each test, and without knowing the researcher's report on their performance, learners filled out a self-assessment form of their work in that particular task. It should be noted that students filled out this form in English since they were competent enough to express themselves in English. The form had six generic questions about whether learners reached the expected objectives or not and their possible ways to overcome the difficulties they faced. The questions were designed by the researcher after reviewing the literature and analyzing the research context.

2.2.2.2. Students' Final Self-assessment of the Whole Process

At the end of the course, students took an institutional final exam at the end of each level. Later, students filled out a final self-assessment form of their communicative competence, plans to improve, and whether or not they considered themselves to have achieved the class objectives. The form consisted of eleven parts designed by the researcher. The researcher designed the parts according to the research goals. It worths mentioning that the students filled out this form according to the instructions. They were asked to mark the options and also to write their comments preferably in English.

2.2.2.3. Semi-structured Interview Questions

A semi-structured interview, consisting of twelve questions designed by the researcher, was used to assess students' views of the process. The interview was conducted in Persian and the quotes from this interview were translated into English. Finally, a conference was held, whose objectives were to find out about the participants' perceptions of their learning process and their ideas towards the usefulness of implementing self- and peer-assessment during the course, and their insights about the impacts of these implementations (self- and peer-assessment).

2.3. Data Collection Procedure

This study was a mixed-method one including a true experimental design and a qualitative exploratory, descriptive, and interpretive approach. The two phases of the research were

conducted by the researcher simultaneously.

2.3.1. Phase 1: Quantitative (true experimental design)

2.3.1.1. Randomization

Before the experiment, the language proficiency test (i.e., OOPT) was administered to the participants of the three intact classes to get sure about their homogeneity.

2.3.1.2. Pre-test

Then a timed-writing essay (Topic 1) was given to the participants of the three groups (self-assessment, peer-assessment and teacher-assessment groups) as the pre-test to measure their writing performance before starting the interventions. It should be mentioned that students took part in an online test using the adobe connect platform; as a result, it was possible for the teacher to test them during a special time limit.

2.3.1.3. Instruction and Treatment

The instruction i.e., the online writing course (20 two-hour sessions in 20 weeks) was held via the adobe connect platform. During the 20-session course, the basics of paragraph writing were instructed. To fulfill the objectives of the present study, the same materials and course book were employed by the same instructor, i.e., the researcher, in the three classes. Moreover, the teacher assessment (the control group), the self-assessment, and the peer assessment (the experimental groups) received the proper practices. In addition, the students of self- and peer-assessment groups were trained on how to use the writing scoring rubric to assess the written tasks and essays of themselves and their peers. To ensure that the students understood how to score their peers and also their pieces of writing, every week the instructor planned an online workshop and checked all the necessary points regarding scoring the pieces of writing with the members of self- and peer-assessment groups. The students of the self-assessment and peer-assessment groups were required to assess the written tasks i.e., their own written paragraphs (self-assessment group) and the written paragraphs of their peers (peer-assessment group) regularly (every week); whereas, the students in the teacher-assessment group were assessed only by their teacher.

2.3.1.4. Post-test

When the online treatment sessions completed, another timed writing essay (Topic 2) was administered to the three groups to measure their writing ability as the post-test of the study again using the adobe connect platform, and during a special time limit.

2.3.2. Phase 2: Qualitative

The researcher had three online conferences whose objectives were to find out about the participants' perceptions of the process (self-, peer- or teacher-assessment) during the online course and insights about the impacts of the implementation before, while, and after the online course.

After each timed-writing essay, and without knowing the teacher's report on their performance, learners filled out a self-assessment form of their work in that particular writing task. It should be noted that students filled out the form in English since they were competent enough to understand the questions. They could also write their comments in English. The form had generic questions about whether learners accomplished the expected objectives or not, possible ways to overcome their difficulties, and their insights about the implementation of a special kind of assessment (self-, peer- or teacher-assessment). Each student requested a short conference with the teacher, during class time, to compare what they had self-assessed, peer-assessed, and what the teacher had assessed on their performance. These sessions were held online using Skype or what's app video calls. After this short talk, some students, by their initiative, decided to do some remedial work if they considered it necessary.

At the end of the online course, students took an institutional final exam at the end of the term. Later, students filled out a final self-assessment form of their communicative competence, plans to improve, whether or not they considered they had passed the class objectives, and their insights about the implementation of a special kind of assessment (self-, peer- or teacher-assessment). It should be reminded that students filled out this form in English since they mainly had to mark options. They could also write comments in English. The interview was in Farsi and the quotes from this interview were translated into English for this study. Finally, the researcher had another online conference whose objectives were to find out about their perceptions of the learning process during the online course and insights about the impact of the implementation of the three different types of assessment (self-, peer- or teacher-assessment).

It should be noted that during the whole process of data collection the teacher, the researcher,

*Eur. Chem. Bull. **2023**,12(3), 2371-2382*

2374

herself provided the students with online conferences and designed online meetings via the adobe connect platform. As a result, students of the three assessment groups didn't face any difficulties in getting connected and communicating their ideas together. It should be noted that the teacher was always accessible to the students. All the online sessions were also recorded, so the researcher had the chance to check them as many times as she needed to record the students' ideas during online conferences and interviews.

2.4. Data Analysis Procedure

2.4.1. Phase 1: Quantitative (true experimental design)

To compare the OOPT mean scores of the three groups, an independent-sample t-test was performed.

Then, to examine the effects of peer- and self-assessment activities on the writing performance of the participants, two paired-sample t-tests were performed to evaluate the change in the mean scores of both groups from the pretest to the post-test.

Afterwards, a One-Way Analysis of Covariance (ANCOVA) was performed on writing performance scores to evaluate the effects of the three types of alternative assessments that were employed in the present study on the L2 writing performance. The independent variable was the type of intervention (i.e. teacher- assessment versus self-assessment and peer assessment), and the dependent variable was the participant's scores on the post-test of the timed-writing essay. The pre-test scores of writing performance were considered as the covariate in the ANCOV analysis.

2.4.2. Correlation Phase 2: Qualitative

The analysis of data combined inductive and deductive approaches in terms of the presence of each category in the participants' perceptions.

3. Results

3.1. Phase 1: Quantitative (true experimental design)

There were 22 participants in the Teacher-Assessment Group. The other two groups had 19 participants each.

The results of descriptive statistics regarding the three groups' scores on the Oxford Online Placement Test (OOPT), Pre-test, and Post-test assessments scored by two raters showed that among 60 participants, the minimum score obtained on OOPT was 80 and the maximum score was 98 (M = 89.38, SD = 5.49).

Among 60 participants of the present investigation, the minimum score given by Rater 1 on the Pre-test was 60 and the maximum score was 81 (M = 72.06, SD = 6.2). The minimum score given by Rater 2 on the Pre-test was 60 and the maximum score was 80 (M = 71.93, SD = 5.86). On the Post-test, the minimum score given by Rater 1 was 80 and the maximum score was 98 (M = 92.05, SD = 3.61). Also, the minimum score given by Rater 2 on the Post-test was 81 and the maximum score was 98 (M = 92.08, SD = 3.48).

The assessment of the normality of the data was analyzed via Kolmogorov-Smirnov and Shapiro-Wilk estimates. The results did not show significant deviations from normality. Therefore, it was safe to conduct parametric analyses in terms of the normality assumption.

The reliability analyses for the pre-test scores, as well as post-test scores by the two raters for each, were launched using the intraclass correlation coefficient (ICC). A two-way mixed model with the consistency type was used to analyze the data. According to the results, a high level of reliability was found in measurements in both pre and post-test assessments. The average measure ICC for the pre-test was .996 with a 95% confidence interval from .993 to .997, F (59,59) = 224.892, p <.001. The average measure ICC for the post-test was .991 with a 95% confidence interval from .984 to .994, F (59,59) = 105.642, p <.001. Taken together, the results showed a high level of reliability between raters in both pre-and Post-test assessments.

To ensure the homogeneity of the three groups of the study in terms of language proficiency and the scores on the pre-test assessments by raters, three one-way ANOVAs were conducted. According to the results, concerning the OOPT, there were statistically significant differences among the groups, F (2, 57) = 5.470, p = 007. The analysis of the Tukey's b post hoc test showed that the Self-Assessment group (M = 92.36) obtained a higher mean score compared to the Teacher Assessment (M = 87.09) and Peer Assessment (M = 89.05) groups. There was no statistically significant difference between the Teacher and Peer Assessment groups. Considering the groups' scores on the pretest scores given by the two raters, the results in both cases showed statistically significant differences among all the three groups, FRater 1 (2, 57) = 32.08, p = 000 and FRater 2 (2, 57) = 30.36, p = 000. Specifically, in both rating measurements, the

Eur. Chem. Bull. 2023,12(3), 2371-2382

2375

Self-assessment group (M = 77.21) outperformed the other two groups. Moreover, the Peer-Assessment group (M = 72.26) outperformed the Teacher-Assessment group (M = 67.09).

To answer the research questions of the first phase of the study (true experimental phase), a mixed design ANCOVA with one within-subject variable (writing performance), one between-subject variable (group), and two covariates (OOPT and Pre-test scores) was launched. Before the main analyses, assumptions of normality, homogeneity of variance, and sphericity were investigated. The results obtained from the Kolmogorov-Smirnov test, Levene's test, and Mauchly's test of sphericity did not show any violations of the assumptions. The descriptive statistics of the variables are presented in Table 1.

Table 1. Descriptive Statistics of the variables in the Mixed design ANCOVA analysis.

	Group	**Mean**	**Std. Deviation**	**N**
	Teacher-Assessment	67.0000	4.15188	22
Pre-test Rater 1	Peer-Assessment Group	72.4737	5.43004	19
	Self-Assessment Group	77.5263	2.52473	19
	Total	72.0667	6.02500	60
	Teacher-Assessment	90.5000	3.46066	22
Post-test Rater 1	Peer-Assessment Group	91.1053	3.54173	19
	Self-Assessment Group	94.7895	2.12339	19
	Total	92.0500	3.61459	60

The analyses indicated that there was a significant main effect on writing performance, $F(1, 55) = 103.33$, $p < .001$, partial $n^2 = 653$, (M Pre-test = 72.06, M post-test = 92.05).

The analyses of the within-subjects effects indicated that the writing performance * OOPT contrast was not statistically significant, F (1, 55) = 1.370, p = .247, n2 = .024. This implied that while the proficiency levels of the participants were statistically different in the three groups, the differences in students' post-test writing scores could not be attributed to differences in language proficiency. The contrast writing performance * group [F (1, 55) = 8.08, p = .001, p2 = .227] and writing performance * pre-test [F (1, 55) = 120.169, p = .000, p2 = .686] were statistically significant.

These results indicated that there were statistically significant within-group differences in groups' performances. Specifically, all the groups obtained higher scores on the post-test assessment compared with their scores on the pre-test writing performance. The teacher-assessment group [(t (21) = -19.580, p < .000, Mpre = 67, Mpost = 90.50)], the peer-assessment group [(t (18) = -11.284, p < .000, Mpre = 72.47, Mpost = 91.10)], and the self-assessment group [(t (18) = -29.182, p < .000, Mpre = 77.52, Mpost = 94.78)].

Considering the between-subjects effects, the results showed significant differences in group, pretest, and posttest assessments (Table 2).

Table 2. Tests of Between-Subjects Effects.

Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Interce	540.059	1	540.059	99.748	.000	.645
OOPT	10.956	1	10.956	2.023	.161	.035
Pre2	388.304	1	388.304	71.719	.000	.566
Group	103.375	2	51.687	9.547	.000	.258
Error	297.782	55	5.414			

Eur. Chem. Bull. **2023**, *12(3)*, 2371-2382

2376

This indicated that while the participants in each group differed in terms of language proficiency, the differences in the post-test writing performance could not be attributed to language proficiency differences. The results also indicated statistically significant differences among the groups in terms of post-test writing scores, $F_{(2, 57)} = 10.84$, $p = .000$.

Since the two raters scored each assessment, the multiple comparisons related to each rater were presented in Table 3.

Table 3. Multiple Comparisons related to the two raters on post-test assessment.

Dependent Variable	(I) Group	(J) Group	Mean Std. Differenc Error e (I-J)	Sig.	95% Confidence Interval Upper Bound	Lower Bound
	Teacher-Assessment	Peer-Assessment Group	-.60526 .98030	.539	-2.5683	1.3577
	Group	Self-Assessment Group	-4.28947* .98030	.000	-6.2525	-2.3265
		Teacher-				
Post-test LSD	Peer-Assessment Group	Assessment Group	.60526 .98030	.539	-1.3577	2.5683
Rater 1		Self-Assessment Group	1.0155 -3.68421* 3	.001	-5.7178	-1.6507
	Self-Assessment Group	Teacher-Assessment Group	4.28947* .98030	.000	2.3265	6.2525
		Peer-Assessment Group	1.0155 3.68421* 3	.001	1.6507	5.7178
	Teacher-Assessment	Peer-Assessment Group	-.55981 .93107	.550	-2.4242	1.3046
	Group	Self-Assessment Group	-4.29665* .93107	.000	-6.1611	-2.4322
		Teacher-				

Eur. Chem. Bull. 2023,12(3), 2371-2382

2377

Post-test LSD	Peer-Assessment Group	Assessment Group	.55981 .93107		.550	-1.3046	2.4242
Rater 2		Self-Assessment Group	-3.73684* .96453		.000	-5.6683	-1.8054
	Self-Assessment Group	Teacher-Assessment Group	4.29665* .93107		.000	2.4322	6.1611
		Peer-Assessment Group	3.73684* .96453		.000	1.8054	5.6683

*The mean difference is significant at the 0.05 level.

According to the results, the self-assessment group in both ratings obtained higher post-test scores compared with peer- and teacher-assessment groups. There were no statistically significant differences between teacher and peer assessment groups regarding the post-test writing scores.

3.1.1. Summary of Inferential Statistics

In light of the results of the inferential statistics of phase 1 of the study, the research questions are answered.

Which type of assessment (teacher-, peer-, or self-assessment) would promote the best achievement in the writing course?

All groups showed significant gains from pre-test scores to post-test scores. However, the self-assessment group showed higher gains compared to the other two groups.

Would the differences in assessment methods (self-, peer-, or teacher-assessment) be reflected in students' final examination scores?

Both within-group and between-group effects were significant and different experiments showed significant differences.

3.2. Phase 2: Qualitative

The second phase of this study followed a qualitative-exploratory, descriptive, and interpretive approach (Bonilla & Rodriguez, 1997; Leedy & Ormrod, 2001) characterized by an intervention. The study aimed at observing, understanding, and interpreting the role of formative assessment (FA) on students' views of their learning and their views on this kind of assessment. One of the researcher's interests throughout this study was to inquire about how FA may influence students' perceptions of their learning. Another main concern the researcher had throughout this study was to explore students' opinions about the implementation of FA and if it fits its purposes.

3.2.1. The Role of Formative Assessment (self- and peer-assessment) on Students' Views of Their Learning

One of the researcher's interests throughout this study was to inquire about how FA may influence students' perceptions of their learning. All data collection instruments served to build these perceptions and draw conclusions about the effectiveness of this kind of assessment for enhancing their metacognition and consequently their autonomous learning. Mainly, students reported that FA helped them become aware of their weaknesses and strengths in their communicative competence and of the situations in which this awareness arose, and thanks to FA they experienced a sense of achievement because they realized how much they had learnt. It should be added that for the analysis of students' self-assessment of their communicative competence Bachman and Palmer (2002) were used as a reference.

3.2.2. Strengths and Weaknesses in Communicative Competence

All instruments but the final self-assessment task did not directly request students to report the weaknesses and strengths of their communicative competence. Then, students' accounts of their learning were more a by-product of their reflective skills encouraged by very generic prompts. Data depicted a recurrent tendency to refer to strengths in pragmatic knowledge with a far much higher frequency than those in

*Eur. Chem. Bull. **2023**,12(3), 2371-2382*

2378

organizational knowledge. Conversely, there was a more recurrent tendency to refer to their weaknesses in terms of organizational knowledge than pragmatic knowledge.

The fact that students provided specific information on their strengths and weaknesses in relationship to what was being learnt supported the idea that students were able to assess their learning by making explicit knowledge and those abilities through procedures that helped them think and reflect on what was or was not being learnet.

It was not surprising that they referred more often to pragmatic knowledge as their strength and organizational knowledge as their weakness because during teaching there was far more stress on communication rather than on a descriptive study of the language as the main learning goal. Organizational knowledge was presented as a part of what was required to achieve communication.

3.2.3. Awareness of Situations in Which Students Realized Their Strengths and Weaknesses

Only data gathered from students' interviews confirmed the situations that resulted in this awareness. Their awareness came to light through different situations which not necessarily involved FA procedures such as assessment tasks, self-assessments, conferences, or remedial tasks, but also classroom situations different from FA procedures, some others outside the classroom and others they did not specify.

Regardless of the situations in which students' insights about their learning process emerged, the fact that students reported them through self-assessment instruments and the interviews suggests that the habit of having students exposed to FA and making explicit what they are learning may have triggered an awareness of what they were able to do or not to do with the language and what they knew about the language in various situations.

3.2.4. Sense of Success through Formative Assessment (self- and peer-assessment)

Most students also experienced a sense of success through FA because they were aware of having achieved something, and this awareness arose during FA procedures. In the interviews, fifty-five students reported a degree of satisfaction because they realized they had learnt a lot thanks to FA.

Students reported some learning outcomes and

a kind of satisfaction during different stages of the cycle of FA. This success was demonstrated in the different instances used to assess their communicative competence: In the first writing task, fifty-three students achieved the objectives of the task and seven agreed on the fact that they did not achieve the task successfully but requested a chance to remediate, and in the remedial work they evidenced they had corrected their problems and achieved the learning outcomes of the task. In the second writing task, all the students completed the task and all of them corrected the minor language problems they had experienced after the conferences. Finally, in the final exam, students demonstrated they were able to use the language properly. These students passed the course with the necessary language tools to use the language properly; although, they evidenced and self-reported having some language difficulties yet.

3.2.5. Students' Views of Formative Assessment

Another main concern the researcher had throughout this study was to explore students' opinions about the implementation of FA and if it fitted its purposes. Analysis of data from interviews with students generated three main findings. Firstly, students reported FA as a procedure that helped them identify their weaknesses and strengths. Secondly, students considered FA as a transparent process in terms of how their learning outcomes and difficulties were reported and the accuracy of the reports of what they learned.

3.2.6. Purposes of Formative Assessment

Students repeatedly declared the purposes of FA as a procedure that mainly helped them identify their weaknesses and enhance their learning. They barely indicated FA was intended to spot their strengths. However, in different stages in the course, students were able to report their strengths and weaknesses through self-assessment procedures along the course, namely, self-assessments, conferences, and the final interview which evidenced some of the purposes of FA, that is, to detect students' strengths and weaknesses related to their communicative competence and enable them to make informed decisions about their learning.

The fact that most students did not see FA as a procedure that helped them identify their problem areas even though they were asked to do so in different stages of the cycle was unexpected. This could be because in their previous learning experiences, students were used to being assessed and tested and the results of these procedures

Eur. Chem. Bull. **2023**,*12(3), 2371-2382*

2379

highlighted what they mainly did wrong or the main difficulties they had. It was as if learning was "measured" as to what students could not learn first instead of what they learned.

3.2.7. Formative Assessment as a Transparent Process

In the interviews, all students identified FA as a transparent process. This was mainly because they built their visions of their communicative competence and then they compared them with their teacher's vision to agree on their learning outcomes and difficulties. Although the reliability of students' self¬assessment results about their communicative competence was not the objective of this research. The fact that students in all stages self-reported their learning outcomes and difficulties honestly, made the process transparent. Moreover, the fact that later through conferences they compared what they self-assessed and what the teacher had observed to agree on what they did or did not learn made the procedure far more transparent and the assessment results more accurate and reliable.

In the light of study results, the research questions were answered as follows:

What is the role of the implementation of self-assessment on students' views?

Nearly all students believed that self-assessment processes in the light of the teacher's guidance led to their improvement in writing.

What is the role of the implementation of peer assessment on students' views?

According to the results of the study, students did better in realizing their problems and improving their writing when they received their peers' feedback.

4. Discussion

4.1. Quantitative Phase

According to the results, the self-assessment group in both ratings obtained the highest post-test scores compared with peer- and teacher-assessment groups. There were no statistically significant differences between teacher- and peer-assessment groups regarding the post-test writing scores. It was also highlighted that both within-group and between-group effects were significant and different experiments showed significant differences.

The first question sought whether there was any difference between the writing performance of the control group (teacher assessment) and the

experimental groups (i.e., self and peer assessment). According to the results, the self-assessment group in both ratings obtained the highest post-test scores compared with peer- and teacher-assessment groups. There were no statistically significant differences between teacher- and peer-assessment groups regarding the post-test writing scores. This finding was in line with the findings of a significant bulk of previous studies (Iraji, Enayat, & Momeni, 2016; Liu & Brantmeier, 2019; Mazloomi & Khabiri, 2018; Ratminingsih et al., 2018). Furthermore, this finding was consistent with a substantial body of research suggesting that learners' involvement in the assessment and feedback process enhances the quality of the written drafts (e.g., Jensen & Fischer, 2005).

The second question sought which type of assessment (self-, peer-, or teacher-assessment) would promote the best achievement in the writing course. According to the results, all groups showed significant gains from pre-test scores to post-test scores. However, the self-assessment group showed the highest gains compared to the other two groups. The results of data analysis indicated that the self-assessment group outperformed the other two groups in writing performance measured by time-writing essays. This finding concured with the findings of a significant number of previous studies which found self-assessment to be more effective than peer- assessment in enhancing EFL writing performance (Panadero, Brown, & Strijbos, 2016; Bhatti, 2020).

The third question sought to find if the differences in assessment methods (self-, peer-, or teacher-assessment) would be reflected in students' final examination scores. The results of the study revealed that both within-group and between-group effects were significant and different experiments showed significant differences. The analyses indicated that there were statistically significant within-group differences in groups' performance. Specifically, all the groups obtained higher scores on the post-test assessment compared with their scores on the pre-test writing performance. This finding was also in line with the findings of a significant bulk of previous studies (Iraji, Enayat, & Momeni, 2016; Liu & Brantmeier, 2019; Mazloomi & Khabiri, 2018; Ratminingsih et al., 2018).

4.2. Qualitative Phase

According to the results of the study, nearly all students believed that self-assessment processes in the light of the teacher's guidance led to their

Eur. Chem. Bull. **2023**,*12(3), 2371-2382*

2380

improvement in writing. Moreover, the results of the study revealed that students did better in realizing their problems and improving their writing when they received their peers' feedback.

Even though teachers all provide FA in many ways, the lack of systematization and clear focus, as regards its purpose, is self-defeating. At some points in teachers' careers, they have failed to recognize the formative nature of assessment, its underlying principles, and the most suitable alternatives for implementing it in the classroom as the findings suggested by the results of Arias and Maturana (2005), Bernal and Lopez (2009), Black and Wiliam (1998). This implementation is not a model but an example of the principles that guide FA.

5. Conclusion

In conclusion, formative assessment (FA) did help learners form a picture of their learning regarding the strengths and problems of their communicative competence and perceive the situations in which they developed that awareness including those that exposed them to FA; additionally, it endowed them to experience success in their learning. Also, learners labeled the purposes of this kind of assessment as mainly to identify their weaknesses and improve, and barely to identify their strengths. Furthermore, learners observed FA as a transparent procedure.

References

Allen, L. Q. (2004). Implementing a culture portfolio project within a constructivist paradigm. Foreign language annals, 37(2), 232-239.

Anasse, K., & Rhandy, R. (2021). Teachers' Attitudes towards Online Writing Assessment during Covid-19 Pandemic. International Journal of Linguistics, Literature and Translation, 4(8), 65-70.

Arias, C., & Maturana, L. (2005). Evaluation in foreign languages: Practice and discourse. fkala, Revista de Lenguaje y Cultura, 10(16), 63¬91.

Baker, K. M. (2016). Peer review as a strategy for improving students' writing process. Active Learning in Higher Education, 17(3), 179-192.

Bachman, L. F., & Palmer, A. (1996). Language testing in practice. Oxford: Oxford University Press.

Balkanski, S. V., Lebanova, H. V., Grigorov, E.

E., & Getov, I. N. (2021). In-depth Assessment after 18 Months of Distance E-learning of Pharmacists in Bulgaria. Archives of Pharmacy Practice, 12(2), 12-16.

Bean, T. W. (2011). The context of English language arts learning: The high school years. In Handbook of research on teaching the English language arts (pp. 72-80). Routledge.

Bernal, R., & Lopez, A. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. PROFILE Issues in Teachers' Professional Development, 11(2), 55-70.

Bhatti, A. (2020). Impact of Learner-learner Rapport on L2 Learning: A Study of Public Sector Universities in Sindh, Pakistan. The Asian EFL Journal, 27(4.6), 204-226.

Black, P., & Wiliam, D. (1998). Assessment for learning in the classroom. In J. Gardner (Ed.), Assessment and learning (pp. 9-25). London, UK: Sage Publications Ltd.

Bonilla, E., & Rodriguez, P. (1997). Beyond methods: Research in social sciences. Bogota, CO: Norma.

Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. Language learning, 34(4), 21-38.

Gipps, C. (1994). Developments in educational assessment: what makes a good test?. Assessment in Education: Principles, Policy & Practice, 1(3), 283¬292.

Hamp-Lyons, L. (2009). Principles for large-scale classroom-based teacher assessment of English learners' language: An initial framework from school-based assessment in Hong Kong. Tesol Quarterly, 43(3), 524-530.

Herrington, A. J., & Cadman, D. (1991). Peer review and revising in an anthropology course: Lessons for learning. College Composition and Communication, 42(2), 184-199.

Iraji, H. R., Enayat, M. J., & Momeni, M. (2016). The effects of self-and peer- assessment on Iranian EFL learners' argumentative writing performance. Theory and Practice in Language Studies, 6(4), 716-722.

Jensen, W., & Fischer, B. (2005). Teaching technical writing through student peer-evaluation. Journal of technical writing and communication, 25(1), 95-100.

Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. Educational research review, 9, 129-144.

Eur. Chem. Bull. **2023**,*12(3), 2371-2382*

2381

CG, S. K., Alshammari, Q. T., Reddy, R. S., Kandakurti, P. K., & Amaravadi, S. K. (2022). Stress among Teachers in Virtual Classrooms Working in the Higher Education Sector during the COVID-19 Outbreak. *International Journal of Pharmaceutical Research and Allied Sciences, 11*(3), 55-59.

Leedy, P., & Ormrod, J. (2001). Practical research: Planning and design (7th ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Liu, H., & Brantmeier, C. (2019). "I know English": Self-assessment of foreign language reading and writing abilities among young Chinese learners of English. System, 80, 60-72.

Livingston, S. A. (2018). Test reliability—Basic concepts (Research Memorandum No. RM-18-01). Princeton, NJ: Educational Testing Service.

Mazloomi, S., & Khabiri, M. (2018). The impact of self-assessment on language learners' writing skill. Innovations in Education and Teaching International, 55(1), 91-100.

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. Educational Psychology Review, 28(4), 803-830.

Ratminingsih, N. M., Marhaeni, A. A. I. N., & Vigayanti, L. P. D. (2018). Self-Assessment: The Effect on Students' Independence and Writing Competence. International Journal of Instruction, 11(3), 277-290.

Sambell, K., McDowell, L., & Montgomery, C. (2012). Assessment for learning in higher education. Routledge.

Tuleubayeva, S., Tleuzhanova, G., Shunkeyeva, S., Turkenova, S., & Mazhenova, R. (2021). Functional ranking of English in multilingual education in Kazakhstan (on the example of high school students). *Journal of Advanced Pharmacy Education and Research, 11*(4), 143-148. https://doi.org/10.51847/dwQur0yalt

Topping, K. (1998). Peer assessment between students in colleges and universities. Review of educational Research, 68(3), 249-276.

Eur. Chem. Bull. **2023,***12(3), 2371-2382*

2382