Journal of Teaching Language Skills
(JTLS)
39(3.2), Fall 2020, pp. 181-225
DOI: 10.22099/jtls.2021.38753.2899

Online ISSN: 2717-1604
Print ISSN: 2008-8191

Research Paper

The Effect of Blended Online Automated Feedback and Teacher Feedback on EFL Learners' Essay Writing Ability and Perception

Masumeh Salavatizadeh [*] **Abdorreza Tahriri** [**]

Department of English Language and Literature, Faculty of Humanities, University of Guilan, Rasht, Iran

Abstract

Applying both teacher- and computer-generated feedback to foster EFL learners' writing remains a less explored issue. The purpose of this study was two-fold. Firstly, it examined whether blending both online automated feedback (OAF) and teacher feedback had a significant effect on EFL learners' writing ability or not. Secondly, it explored their perceptions on the use of the blended OAF and teacher feedback. To this end, this study employed a mixed-method design using essay writing, questionnaire, and interview to investigate the impact of blended OAF and teacher feedback on EFL learners' writing ability in an essay writing course. Thirty intermediate EFL learners participated in the study; they were given the first writing pre-test, and then were assigned into experimental and control groups randomly. The experimental group was given access to the integrated OAF and teacher feedback on their two-paragraph problem-solution essay writing during the treatment phase while the control group received teacher-only feedback. The results of the independent samples t-test and RM ANOVA showed that the experimental group outperformed the control group in terms of the overall score for the two final essay writings, and also demonstrated a significant improvement in writing scores across four essays during the treatment. The qualitative data also revealed the participants' positive perception towards the effectiveness of

Received: 21/11/2020 Accepted: 08/01/2021

[*] M.A. Student, Email: masumeh.salavatizadeh@gmail.com
[**] Assistant Professor, Email: atahriri@gmail.com, Corresponding Author

the blended OAF and teacher feedback on their writing. Overall, the findings of this study can contribute to a better understanding of the impact of the blended feedback in enhancing EFL learners' writing ability.

Keywords: Online Automated Feedback (OAF), Teacher Feedback, Essay Writing Ability, Computer-generated Feedback, EFL

According to Long and Robinson (1998, as cited in Doughty & Williams, 1998), the target language input should be accompanied by corrective feedback (CF) in second language learning. On the other hand, Sauro (2009, as cited in Doughty & Williams, 1998) stated that providing CF raises EFL learners' awareness about the differences between their output and the target language input. Additionally, CF can lead to noticing which is necessary for learning to occur (Schmidt, 2001; cited in Robinson, 2001). Providing pen and paper-based CF on students' L2 writing, especially when a large number of students are involved, is a demanding task and teachers have to devote a considerable amount of time and energy to evaluate and score students' writings. In recent decades, one possible solution to reduce the heavy workload of evaluating students' written work is making use of computer-generated feedback (Fotoohnejad, 2018).

In our modern era, the attention of numerous researchers has mainly been drawn to the advances of computer technology involving automated essay scoring and computer-generated feedback. The first two waves for writing assessment are introduced as direct assessment and multiple-choice testing (1950-1986) and the third wave (from 1986 up to the present time) is introduced as a modern period for writing assessment, which refers to the possibility of assessing writing online or through computer software (Hamp-Lyons, 2002; Yancey, 1999). It is also suggested that a fourth and future generation will also need to be technological; therefore, the issue of computer-based writing assessment is supposed to be focused on not only at large-scale

assessment by test-takers and raters, but more importantly at classroom-based assessments by educators and teachers. However, it is stated that this generation must be both humanistic and technological, due to the advances both in computer technology and writing assessment which is a complex process that involves various authors and readers (Hamp-Lyons, 2002).

On the other hand, there is consistently a controversy among the researchers and teachers over whether to employ teacher feedback or computer-generated feedback or both of them. As it is frequently repeated by numerous researchers (such as Cheng, 2017, Dikli & Bleyle, 2014, Huang, 2014, Kim, 2014, Li, Link, Ma, Yang, & Hegelheimer, 2014, Liao, 2016, Stevenson & Phakiti, 2014), each of these kinds of feedback in L2 writing is claimed to bear its own advantages and disadvantages which can either foster EFL learners' writing or become problematic for their L2 writing when it is employed alone. Hence, it is frequently expressed by numerous researchers that AWE feedback should be integrated with teacher feedback. Moreover, corrective feedback generated by computers might not be the mere reason for improving the accuracy of EFL learners' writing (Li, Link, & Hegelheimer, 2015); there might be other factors contributing to EFL learners' writing accomplishment.

Since recent studies have mainly investigated the effect of either human feedback or computer-generated feedback on students' writing, there has been almost no research in applying both human- and computer-generated feedback to facilitate students' essay writing. Additionally, no study has empirically examined the effectiveness of utilizing both OAF and teacher feedback simultaneously on enhancing EFL learners' writing improvement in the Iranian context. Additionally, many Iranian instructors have lagged in terms of aptly utilizing different as well as appropriate types of feedback to improve their students' writing ability. In numerous Iranian language learning contexts,

writing skill is not taken into account as much as other skills and instructors do not usually take the trouble to use technology in their classes due to numerous reasons such as lack of technology-related training or lack of access to technological facilities like computer and the Internet. That is, despite technological advances, and given that writing ability is necessary for Iranian EFL learners, computer potentials for learning writing skill have not completely been exploited and the impact of computer-generated feedback on Iranian EFL learners' writing has rarely been taken into account in the Iranian context. Indeed, in Iranian language learning contexts, teacher-generated feedback is merely provided for EFL learners to enhance their writing ability. However, computer-based programs have attracted considerable attention around the world as they can substantially reduce the heavy workload placed upon teachers when dealing with L2 learners' written work.

The present study is intended to answer the researchers' call for blending both computer-generated and teacher feedback for writing ability in the classroom instruction and investigates whether blending both OAF and teacher feedback influences EFL learners' essay writing. The findings of the present study can be significant for teachers, students, and other stakeholders such as test-developers and curriculum designers since instructors and the administrators, being eager to apply technological devices when providing feedback on learners' writing, need more research to make informed choices about investing their time and budgets (Cunningham, 2018). The findings from this study can advance their understanding of how EFL learners respond to the blended feedback and the extent to which they subsequently improve their essay writing ability.

To address the aforementioned gaps in the existing literature, this study offers one potential response to the researchers' call for blending both OAF and teacher feedback concerning instructing and assessing EFL learners'

writing ability as well as investigates their perceptions of the blended feedback. Therefore, to contribute to the existing literature, the current study sought to explore an innovative way in developing Iranian EFL learners' writing ability through blending both teacher-generated and computer-based feedback provided by an OAF tool and to pave the way for further studies in this area.

Literature Review

Corrective Feedback

Richards and Schmidt (2010) defined feedback in EFL/ESL as comments or any other information that EFL learners receive either from the teacher or other sources concerning their L2 learning success on certain tasks or tests. According to Keshavarz (2017), treatment of learners' errors/mistakes is referred to as corrective feedback (CF). Feedback on learners' writing can be produced by various sources like teachers, peers, or computers. Teacher-generated feedback is regarded as a common and traditional approach; however, computer-generated feedback and peer feedback are becoming more popular in recent years in L2 writing classrooms because of growing interest in AWE and learners' collaboration (Cunningham, 2018). Following Ellis's typology, five classifications of CF include direct vs. indirect feedback, metalinguistic feedback, focused vs. unfocused feedback, electronic feedback, and reformulation by a native speaker (Ellis, 2008). It is worth mentioning here that the present study, in blending computer-teacher feedback, exploited the potential of four types of feedback, including direct versus indirect feedback, metalinguistic feedback, and electronic feedback.

Teacher's Feedback on EFL Learners' Writing

Drawing on the ideas of Hyland and Hyland (2006), teacher-written corrective feedback plays a central role in almost every second and foreign language writing classroom, and mainly leads to writing improvement. As a result of Chandler's (2003) study, to increase the accuracy of ESL learners' writing, teachers should give error feedback and students are required to correct the errors. Hence, providing direct correction and raising learners' awareness about their writing errors can unquestionably reduce those errors in the subsequent revisions to a great extent.

Despite EFL learners' positively strong preferences for teacher written feedback in terms of developing their writing, its effect has not assuredly been claimed yet, either on written revisions or on EFL learners' writing ability. The reason for such uncertainty is that EFL learners might not notice the teacher's feedback or use it in the wrong way (Hyland & Hyland, 2006). This is exactly what Ferris, as well as Conrad and Goldstein, would call 'misunderstanding' the teacher's feedback, or understanding it but not being able to apply it in the written revisions appropriately (Conrad & Goldstein, 1999; Ferris, 1995, 1997). As a result, EFL learners may ignore and remove the feedback provided by the teacher (Hyland, 1998). Furthermore, other justifications raise doubts about the effectiveness of the teacher's feedback. Different texts, for instance, demand different ways of feedback provision; that is, which feedback type will be the most appropriate for various texts to be employed is still under question. The role of context, EFL learners' preferences, and factors between student-teacher or student-student (peers), and finally the difficulty of conducting longitudinal studies are regarded as other elements that lead the researchers to be uncertain about the impact of the provided teacher's feedback (Hyland & Hyland, 2006).

Computer-generated Feedback on EFL Learners' Writing

Both computer-generated feedback and OAF are widely known as Automated Writing Evaluation (AWE). On the one hand, the usefulness of AWE programs feedbacks on EFL learners' writing has been under question by some researchers (Ericsson & Haswell, 2006; Hearst, 2000). They criticized such programs for being unreliable (Krishnamurthy, 2005), lacking appropriate pedagogical principles (Chapelle, 2001), and their ability in dealing with long as well as different text types (Hyland & Hyland, 2006; Stevenson & Phakiti, 2014). They believe that such programs are unable to understand meaning like a human since there is no real audience, no meaningful communication between the writer and the AWE tool, and no real-world interaction in the process of evaluating EFL learners' writing (Hyland & Hyland, 2006). Moreover, they worry that AWE tools lead learners to focus only on surface structures and forms rather than meaning in their essay writing (Chen & Cheng, 2008; Ericsson & Haswell, 2006). Such tools fail to measure the meaningfulness of content, argumentation quality, and rhetorical effectiveness (Deane, 2013). Furthermore, employing AWE programs would not be a good option if the writing objective is to convey the writer's voice involving his/her creativity and originality, which is specifically the case when dealing with advanced EFL learners (Chen & Cheng, 2008; Huang, 2014).

On the other hand, some of the studies view AWE tools as useful and handy as nobody seems to firmly reject their practical effects (Ene & Upton, 2014; Whithaus, 2006). For instance, AWE tools would be helpful for EFL learners in fostering their essay writing and gaining satisfactory scores in large-scale writing tests. AWE developers note that these programs demonstrate the ability to assess and provide feedback to EFL learners' writing as a teacher does (Attali & Burstein, 2006; Vantage Learning, 2007; cited in Chen & Cheng, 2008). Hyland and Hyland (2006) hold the position

that when classrooms are largely populated and the learners' expectations are high, AWE tools might also be seen as a cost-effective way of replacing or enhancing teacher's feedback. To this effect, it will be possible for teachers to get rid of spending hours on commenting on students' papers, and focus on other aspects of the teaching since AWE tools can provide more extensive feedback in a much shorter time (Dikli & Bleyle, 2014). The provision of immediate holistic and analytical feedback is regarded as the strength of computer-generated feedback tools which cannot be denied (Hoon, 2006; Yeh, Liou, & Yu, 2007). AWE feedback can motivate learners (Grimes & Warschauer, 2010; Klobucar et al., 2013), especially those with a low computer-anxiety feature, and save time for EFL teachers (Lai, 2010). In terms of applying OAF in L2 writing, EFL learners showed positive attitudes and autonomous learning by accepting more responsibility and taking control of their L2 writing (Cheng, 2017). The convenience of AWE feedback, the capability of finding errors, speed, and allowing EFL learners' to self-edit their writing can be taken into account as advantages of AWE programs (Kim, 2014). AWE tools, indeed, can significantly improve students' writing accuracy in terms of grammatical structures, vocabulary usage, and spelling (Li et al., 2015; Wang, Shang, & Briody, 2013).

Learners' Perception of Computer-generated Feedback

Regarding the EFL learners' perception, the findings of a study showed that AWE feedback can be perceived favorably when it is followed by the teacher and peer feedback to revise their writing (Chen & Cheng, 2008). This study also revealed that AWE feedback could not replace teacher and peer feedback since it led to frustration on the part of EFL learners and limited their learning of writing. However, it is stated that AWE can be exploited as a supplement to teacher and peer feedback. Furthermore, the effectiveness of

AWE can be determined by teachers' attitudes and their skills in terms of using AWE tools, as well as EFL learners' characteristics and goals for learning writing. To implement AWE programs in the classrooms, the teachers' job is to take the inherent limitation of AWE tools into account and then think over well-thought-out pedagogical designs related to the objectives of the learning of writing (Chen & Cheng, 2008). Additionally, due to insufficient information on EFL learners' views of AWE programs or the impacts of AWE feedback on one hand, and increasing developments in AWE software, on the other hand, students' perceptions and use of computer-generated feedback in naturalistic settings are certainly required to be examined in recent studies (Hyland & Hyland, 2006).

Lai's (2010) study focused on investigating problems and benefits of technologies in L2 writing. It compared an AWE tool with peer evaluation, and the results showed that the participants preferred peer evaluation to AWE due to several aspects including social learning, feedback strategies, computer anxiety, and cultural impact. However, in general, the effectiveness of these two modes of writing evaluation was confirmed from the learners' points of view.

Besides, another survey results, based on interviews and observations, indicated students' positive perception of using laptop programs for a variety of learning purposes such as improving their writing. "More than 50% of students responded that they like writing with the tool, it is easy to use, they revise their writing more, and their confidence in writing has increased when using MY Access!" (Zheng, Warschauer, & Farkas, 2013, p. 290). Students' interview also revealed other reasons for using laptop programs, including writing at their own pace, having autonomous learning, and being motivated to write better. Their instructors also mentioned that the burden of evaluating students' essays declined because of using AWE tools, although they were not

sure about the accuracy of AWE scoring. Regardless of teachers and students' negative attitudes toward employing AWE programs which were merely in low socio-economic status classrooms since they were hardly able to make use of computer-generated feedback, many teachers were willing to use such programs in their classrooms to facilitate teaching and classroom management as well as to motivate learners (Zheng et al., 2013). In line with this study are the results of another survey indicating that AWE tools could facilitate classroom management and raise EFL learners' motivation to write and revise (Grimes & Warschauer, 2010).

According to interview analyses of Li et al. (2015) concerning the use of an AWE tool for EFL learners' writing, the instructors' views on the potential of the AWE tools in assisting their learners with accuracy (i.e., grammar and mechanics) were significantly positive; however, the quality of the AWE feedback on content was not completely satisfactory. On the other hand, their findings showed that the learners' attitudes towards the AWE feedback on both accuracy and content were more positive. That is, AWE corrective feedback along with setting scores was important for them to be encouraged to engage in their writing assignments. Their results also suggested that students' use of AWE tool depends largely on instructors' pedagogical perspectives and in-class practices. Nevertheless, AWE feedback potential in L2 writing cannot be neglected or denied.

Blending Computer-generated and Teacher feedback

Computer-generated feedback is required to be integrated with teacher feedback and considered as a supplement to the teacher's feedback; it cannot effectively be employed alone or applied instead of teacher feedback as a replacement in classrooms (Burstein & Marcu, 2003; Burstein, Chodorow, & Leacock, 2004; Chen & Cheng, 2008; Kellogg, Whiteford, & Quinlan, 2010).

Researchers also mention that ecological validity can be obtained by examining different methods of integrating AWE feedback with teacher feedback as well as with classroom writing instruction (Stevenson & Phakiti, 2014). Highlighting the pedagogical value of integrating AWE into classroom writing instruction, Stevenson and Phakiti introduce several practical ways of such an integration.

Cheng (2017) expresses that his study's limitation is that the unique role of the teacher feedback is not taken into account, although it cannot be totally replaced with OAF because each bears its own pros and cons. With this in mind, he calls for future studies to explore how one can properly blend both teacher feedback and OAF for writing improvement. Li et al. (2014) asserted that teachers take advantage of AWE scores for the purpose of pre-evaluating (or diagnostic purposes) and a classroom formative assessment. Therefore, there would be a come and go relation between both computer-based and teacher feedback for the evaluation of EFL learners' writing.

Another study argues that using an AWE tool, EFL learners would be able to identify local grammatical errors in their writing so that the teacher would only focus on EFL learners' global errors, then this would reduce the teacher's task and time by half (Liao, 2016). Liao states that his participants' positive performance on writing was due to repetitive practices, gap noticing as well as applying an AWE tool under the integrated process and structural pedagogy. However, it should be taken into account that firstly the teacher ought to scaffold EFL learners how to use an AWE tool and make them aware of the rationale behind using AWE. Then, independently producing written texts on AWE tool for several rounds, EFL learners will gradually develop autonomy.

To sum up, as feedback on writing is regarded of paramount importance in improving EFL learners' writing, on one hand, with the pervasiveness of

technology and growing access to the Internet, EFL learners can have further access to the computer-generated feedback; on the other hand, researchers hardly reach a consensus whether to employ teacher-only feedback or computer-generated feedback or both of them. Applying both teacher- and computer-generated feedback simultaneously and integratively to foster EFL learners' writing remains a less explored realm particularly in the Iranian context. Therefore, further research in this regard is required to fill this gap and demonstrate further proof in terms of the effect of the integration of both sources of feedback on EFL learners' essay writing.

The importance of the present study is highlighted due to several drawbacks of prior studies. First, according to Li et al. (2015), it is arguably said that computer-generated feedback cannot offer accurate corrections because technology-based corrections do not bear the capability of reporting cognitive processes involved in EFL learners' dynamic writing process. Hence, it is essentially the teachers' role to guide students to use computer-generated feedback in meaningful ways concerning why errors occur, how to make corrections, and how to use AWE feedback. Second, based on Li et al.'s view (2015), the corrective feedback given by AWE might not be the mere reason for improving the accuracy of EFL learners' writing; therefore, there may be other factors contributing to learners' writing accomplishment. Third, numerous researchers (Cheng, 2017; Dikli & Bleyle, 2014; Huang, 2014; Kim, 2014; Li et al., 2014; Liao, 2016; Stevenson & Phakiti, 2014) emphasize that AWE feedback should be integrated with teacher feedback because it cannot potentially be employed on their own or even it may be problematic if teacher feedback is replaced merely with AWE. Fourth, few researchers have applied an OAF tool as a computer-generated feedback provider which is a 'web-based' program (such as Cheng, 2017; Li et al., 2015; Wang, Shang, & Briody, 2013), and the tools used in their studies have cost them a good

amount of money. On the contrary, the OAF tool preferred in this study is easily available whenever one has access to the Internet, and the teachers will not take the difficulty of purchasing it since it is free of charge.

This study addresses the following questions:

1. Does blending both OAF and teacher feedback significantly influence EFL learners' essay writing ability?
2. How do EFL learners perceive the use of integrated OAF and teacher feedback for essay writing?

Consequently, the following null hypothesis in accordance with the first question was formulated:

H_0. There is no significant difference between the essay writing of EFL learners who use both OAF and teacher feedback and that of those EFL learners who use teacher-only feedback.

Method

Research Design

This study used a mixed-method research design to investigate the effect of both OAF and teacher feedback on EFL learners' essay writing ability. In addition, it aimed to explore EFL learners' views on the use of blended feedback for their essay writing. These two purposes could be best served through a mixed-method design as the first and second research questions are quantitative and qualitative by nature, respectively. This allows us to corroborate the quantitative findings through qualitative ones and gain a more complementary view of the issue under study. The independent variable in this study was the integration of OAF and teacher feedback. As the name suggests, OAF was produced by an online automatic system, and, along with teacher feedback, was accessed by EFL learners in the experimental group but not by those in the control group. The dependent variable was the holistic

scores of EFL learners' two-paragraph essay writing. The quantitative and qualitative methods included an analysis of EFL learners' essay writing scores, a questionnaire survey, and a semi-structured interview with the experimental group.

Participants

Initially, the study began with 100 students selected from seven language institutes including 25 males and 75 females; however, there was participants' attrition due to a few reasons, including a low or high level of proficiency, reluctance, and their lack of free time to take part in the main phase of the study. Participants' level of proficiency was determined through a placement test in the first session of the study, and those not being placed in the relevant range were dismissed. In fact, 40 students (15 males and 25 females) no longer participated for this reason. Moreover, 30 students (5 males and 25 females) left the research within one to three sessions upon the launch of the experiment due to their unwillingness, time shortage, and a few other personal problems. Hence, the data were collected from the remaining 30 students (5 males and 25 females) who successfully went through the whole experiment. The above-mentioned sample size is enough since it is generally stated that, at 5% confidence level, a sample size between 30 and 500 is sufficient for many studies (Altunışık, Coşkun, Bayraktaroğlu & Yıldırım, 2004, as cited in Delice, 2010).

Therefore, the remaining 30 Iranian EFL learners, including 5 males and 25 females, from the intermediate level of proficiency were asked to voluntarily accept the invitation to participate in this study. Their ages ranged from 20 to 30 years and their level of proficiency was intermediate. Additionally, the average age of the participants was (M = 25.77). The institute and the participants were selected based on convenience sampling.

Prior to the study, all participants signed a form containing the purpose of the study, and delivered their informed consent. One group with 15 students was randomly assigned to the experimental group while the rest of the students were assigned to the control group. Both experimental and control groups belonged to the same level of proficiency (i.e., homogenous groups) based on the results of a placement test and a first writing test. It is worth noting that since students were from seven different language institutes, from the second session (pre-test writing), the experimental and control group were asked to attend the research site, a language institute, on odd days and even days, respectively, and as a part of their extracurricular activities.

Instruments

Placement test. An Oxford Placement Test, designed by Oxford University Press and University of Cambridge Local Examinations Syndicate (UCLES, 2001), was used to homogenize the learners in terms of their level of proficiency, and control the possible effect of proficiency. This test includes 40 multiple-choice questions to be answered within 20 minutes. It tested learners' knowledge in terms of vocabulary and grammar which are essential for writing skill. The accepted scores were in the range of 20 to 34, and the average score of the participants was calculated as well (M = 26.53), and their maximum score was (Max = 33).

Writing tests. Four writing tests were administered to the participants in both groups based on two-paragraph problem-solution essay writing (i.e., the same genre) with topics selected from *Academic Writing: From Paragraph to Essay* by Zemach and Rumisek (2013). These writing tests were used to diagnose students' problems, and errors upon which the relevant feedback was provided in the next step based on both computer-based and teacher feedback. For all writing tests, the writing genre, scoring rubrics (elaborated in the next

section under Online Automated Feedback Tool), and situations, were all the same. Writing tests were scored twice by the teacher as well as the OAF tool to ensure the inter-rater reliability of the scores. The percentage agreement in both cases was found to be above 90 percent, indicating an excellent degree of consistency.

Online Automated Feedback tool. An online web-based automatic tool, as an OAF tool, named Paper-Rater (http://www.paperrater.com/free_paper_grader) was adopted to evaluate students' essay writing and provide computer-generated feedback. Prior to the main phase of the study, a pilot study was carried out to assure the effectiveness of the selected OAF tool. This proposed OAF tool was designed to provide feedback on EFL learners' essay writing in terms of spelling, grammar, word choice, style of writing including usage of transitional phrases, sentence length and variability, sentence voices, sentence beginnings, usage of academic vocabulary, and finally it gave a holistic score ranging from 1 to 100 as well as from A to D. To add more, the scores, ranging from 90 to 100, the grade "A" was given and the rest was as follows, respectively: (80-89) = "B", (70-79) = "C", (60-69) = "D", and finally scores less than 60 were regarded as 'unaccepted writing'. The feedback provided by this program was of both direct and indirect types, that is, with and without explicit correcting the errors.

Questionnaire. A five-item questionnaire was used to probe into EFL learners' perceptions associated with integrating computer-based feedback and teacher feedback, being a restructured model developed by Cheng (2017) (see Appendix 1). Responses were either rated on a five-point Likert scale (Strongly agree = 5, Agree = 4, Neutral = 3, Disagree = 2, Strongly disagree = 1) or expressed in the open-ended written comments. All the items of the

questionnaire were examined for internal reliability via Cronbach's Alpha (.765) suggesting acceptable reliability (Clark & Watson, 1995).

Semi-structured Interview. On the basis of Cheng's (2017) model, a four-item interview was restructured for the purpose of the triangulation of the questionnaire results, and conducted with all of the participants in the experimental group (see Appendix 2). The interview was transcribed in order to be codified and analyzed. The interview was written down in order to be codified and analyzed. It was called a semi-structured interview since it was a less rigid one in which the researchers used a written list of questions as a guide, while still having the freedom to ask for more information related to the questions in order to clarify them (Mackey & Gass, 2016).

Procedure

The study was conducted in seven sessions, each for 60 minutes. The participants went through the phases respectively as outlined in Table 1.

Table 1.

Research Phases

Session 1	Administrating the placement test in order to homogenize the participants in terms of their level of proficiency + familiarizing them with a sample of problem-solution paragraphs
Session 2	Administrating the first writing test (pre-test writing) + assigning the two classes to experimental and control groups
Session 3	Introducing the OAF tool to the experimental group + administering the integrated OAF tool and teacher feedback to the experimental group on EFL learners' first writing + the control group received teacher-only feedback on their first writing
Session 4	Administering the integrated OAF tool and teacher feedback to the experimental group on EFL learners' second writing + the control group received teacher-only feedback on their second writing

Session 5	Administering the integrated OAF tool and teacher feedback to the experimental group on EFL learners' third writing + the control group received teacher-only feedback on their third writing
Session 6	Administering the final writing test to both groups + administering the Questionnaire
Session 7	Conducting the Interview

Prior to the main phase, a pilot study was carried out with 10 EFL learners with characteristics similar to those of the participants of the study to ensure the effectiveness of the selected OAF tool and the other instruments. Then, the main study was conducted in seven sessions, each for 60 minutes. As expressed earlier, both the institute and participants of this study were chosen on the basis of convenience sampling, being from the intermediate classes. They were selected from intermediate level of proficiency classes since essay writing is mainly taught for this level of proficiency in most English institutes and they have knowledge of basic English grammatical structures and vocabulary that are essential for essay writing. Prior to the study, all students were informed about the general purpose of this project and signed a form to participate in it.

Administering the placement test. The first session was allocated to administrating the placement test to determine the participants' level of proficiency. The placement test was the Oxford Placement Test, designed by Oxford University Press and University of Cambridge Local Examinations Syndicate (UCLES, 2001), and including 40 multiple-choice questions for which students were given 20 minutes to answer them. Since the students were from seven different institutes, the administration of the placement test to all of them in the same situation, at the same time and place, was impractical for the researchers. However, the procedure of how to take the test was explained clearly in the same way by the same researcher in each institute. After this test, the papers were evaluated and scored in order to ensure that the

participants were at the same level of proficiency. If their scores were placed within range (20-34) out of 40, they were considered to be at an intermediate level of English proficiency based on the test manual. After the placement test, the participants' number was reduced to 60 because the result of the test showed that 40 participants (15 males and 25 females) were in a lower or higher level of proficiency. Additionally, 30 students (5 males and 25 females) left the research within one to three sessions upon the launch of the experiment owing to reasons such as unwillingness, lack of time, and other personal problems. Hence, the data were collected from the 30 students (5 males and 25 females) who successfully went through the whole experiment.

Reading sample provision. In the second part of the first session, the teacher (i.e., one of the researchers) also provided students with a reading sample containing a two-paragraph problem-solution text (retrieved from Zemach & Rumisek, 2013, p. 51) in order to make them familiar with the genre, topic, as well as the structure of the writing which was the main purpose of the research. To explain, a reading passage composed of two paragraphs was read in the class and the teacher tried to give them an understanding of how to write a two-paragraph problem-solution essay writing. Briefly, as its name suggested, a two-paragraph problem-solution essay writing was supposed to be written in only two paragraphs. It first described and discussed a problem in the first paragraph, and then proposed one or more solutions to that problem in the second one. The teacher also drew students' attention to the usage of linking phrases, transitional words, conditional sentences, topic sentence, supporting and concluding sentences in the provided reading text, and reminded students to make use of them in their writing in the following sessions.

Administering the first writing test. In the second session, the first writing test as a pre-test was given to the participants. It was a two-paragraph

problem-solution composition based on the same genre that each participant was required to write and type. All the students took this pre-test in the same condition, i.e., at the same time and place. They were given the teacher's instructions, five minutes to think about the topic, and then their 25-minute composition process started with the sign of the teacher.

To begin with, the teacher had students think of their own problem, or problem among their friends, acquaintances or around the world, for example, air pollution, traffic, overcrowded classrooms, a messy roommate, talkative friends, a latecomer person, lack of money, etc. Then, the teacher asked them to write the first paragraph describing one of their problems and explain why this issue is a problem. Then, they were expected to write the second paragraph introducing one or more solutions to that problem and explain how such solutions could remedy the problem.

After 25 minutes, the teacher gave the stop sign to the students and the papers were collected to be scored by an OAF tool. The two classes of participants were then randomly assigned into experimental and control groups. It was also decided that, for the remaining sessions, the experimental group was supposed to attend the class on odd days and the control group was expected to attend it on even days. The experimental and control groups were instructed by the same teacher using the same materials except that the experimental group also received OAF.

Introducing the OAF tool. In the third session, the teacher introduced an OAF tool to the experimental group, in particular the one employed in this study (i.e., Paper-Rater), and explained how to use it which was necessary to be taught to EFL learners for the purpose of the present research. This introduction part was done at the very beginning of the study in the experimental class in order to eliminate the possible interference of low skills of students in using the OAF tool.

To this end, the experimental group was provided with a two-minute video clip the role of which was introducing the OAF tool (i.e., Paper-Rater). In the video file, first, they were asked to type the address http://www.paperrater.com/free_paper_grader on their browser. Second, they copied their writing (or uploaded their writing file) into the existing empty box. Third, the students were expected to choose their "grade", "type of writing" (which was 'essay' for the purpose of the study), and then, put a tick in the square to accept the terms of service. Finally, they clicked on "get report" to receive feedback on their writing automatically after a few seconds of processing.

Administering the blended OAF tool and teacher feedback. For the third session of the experimental group, the students' first writing from their first writing test was assessed by the teacher as well as through the OAF tool (as a diagnostic tool) in order to identify problems, mistakes, and errors that students might have with their writing. For the next step of the third session, the teacher provided feedback (including both direct and indirect feedback) based on which the students were required to correct their current composition and write another piece of writing while using the OAF tool to help them improve their writing. This process of writing and providing feedback on the basis of the blended two sources of feedback continued in the subsequent sessions for the second and third times. That is, for the following sessions, the experimental group was asked to write their next two-paragraph problem-solution writing, considering the feedback provided by the teacher and the OAF tool on their previous writing. It is worth mentioning that the feedback provided by the teacher was based on the OAF tool rubrics with the purpose of complementing the computer-generated feedback.

On the other hand, in the third session for the control group, the students were provided only with teacher feedback on their first writing without using

an OAF tool. In other words, the control group was not allowed to apply the OAF tool. Throughout sessions, they only received feedback from the teacher based on which they were asked to write their next writing for the following sessions.

Administering the final writing test. In the last session of the experiment, both experimental and control groups wrote the final two-paragraph problem-solution essay writing within 30 minutes, and submitted it to the teacher for holistic automated scoring via the OAF tool. To this end, the two groups sat the 30-minute post-test, writing on the same genre and also in the same situations as the pre-test phase.

Administering the questionnaire and interview. Finally, as the second part of the sixth session, the researcher administered a questionnaire to all the participants of the experimental group, and they also took part in a semi-structured interview. The questionnaire, which was distributed among the students of the experimental groups to fill out, lasted between five to ten minutes, i.e., the participants gave back their answers to the questionnaire within ten minutes. In the next session, the teacher-researcher interviewed them based on the pre-planned questions and jotted their responses down for future analysis.

Data Analysis

In this study, the data analysis was carried out through SPSS software (Version 23). During the research, four sets of scores were collected. These quantitative data to answer the first research question were analyzed using independent-samples t-test and repeated-measures ANOVA. The independent-samples t-test aimed to determine whether there were significant differences between students' writing scores in the experimental group and those in the control group, whereas repeated-measures ANOVA was used to

show if there were significant differences in the scores across four different essay writings within the experimental group. Then, the information resulting from the questionnaire was analyzed based on the percentage and frequency of the items to answer the second qualitative research question. After that, the qualitative data from the interview was transcribed verbatim, summarized and thematically categorized. The interview data were utilized to cross-validate the results obtained via the questionnaire.

Results

The answers to the research questions are discussed in the rest of this section. In order to answer the first question, the descriptive results from the four writing scores of the two experimental and control groups indicated the participants' essay writing ability (see Table 2).

Table 2.
Group Statistics for Four Writing Scores

	Group	N	Mean (SD)
Writing Score 1	Exp.	15	71.13 (5.514)
	Cont.	15	68.27 (5.106)
Writing Score 2	Exp.	15	76.60 (4.940)
	Cont.	15	73.47 (5.730)
Writing Score 3	Exp.	15	85.33 (4.981)
	Cont.	15	77.00 (4.123)
Writing Score 4	Exp.	15	89.73 (3.882)
	Cont.	15	80.33 (6.137)

The raw data were analyzed for the sake of comparison between the two groups and to determine whether the means of the two groups were significantly different from one another or not. The sample was assumed to be normally distributed since regarding Kolmogorov-Smirnov Z-test statistics,

the p-values were not statistically significant i.e., ($\rho = .20 > .05$). Moreover, 5% Trimmed Mean equals the Mean (i.e., both are 69.70). Additionally, the values for scores followed a normal curve. At first glance, as shown in Table 2, we see that the experimental group (i.e., Exp.) for Writing Score 4 attained an average of $M_e = 89.73$, whereas the control group (i.e., Cont.) obtained an average of $M_c = 80.33$. However, the standard deviations for average are not identical (3.882 and 6.137, respectively). Levene's test results are shown in Table 3.

Table 3.

Levene's test results of Writing Score 1, 2, 3, and 4 between the two groups

	Levene's Test for Equality of Variances	
	F	Sig.
Writing Score 1	.014	.906
Writing Score 2	.183	.672
Writing Score 3	3.953	.057
Writing Score 4	1.826	.187

Levene's test results suggested that the equal variances assumption was held between the two experimental and control groups for all scores due to their Sig. > .05. Therefore, to examine if there were significant differences in the scores between the two experimental and control groups, independent samples t-tests were run. The results of the tests are shown in Table 4.

Table 4.

Independent-Samples t-Test for Writing Score 1, 2, 3, and 4

		t-test for Equality of Means						
							95% Confidence Interval of the Difference	
		T	df	Sig. 2-tailed	Mean Dif.	Std. Error Dif.	Lower	Upper
Writing Score 1	Equal variances assumed	1.477	28	.151	2.867	1.940	-1.108	6.841
	Equal variances not assumed	1.477	27.835	.151	2.867	1.940	-1.109	6.842
Writing Score 2	Equal variances assumed	1.604	28	.120	3.133	1.953	-.868	7.135
	Equal variances not assumed	1.604	27.404	.120	3.133	1.953	-.872	7.139
Writing Score 3	Equal variances assumed	4.991	28	.000	8.333	1.670	4.913	11.753
	Equal variances not assumed	4.991	27.056	.000	8.333	1.670	4.908	11.759
Writing Score 4	Equal variances assumed	5.013	28	.000	9.400	1.875	5.559	13.241
	Equal variances not assumed	5.013	23.655	.000	9.400	1.875	5.527	13.273

For Writing Score 1, as Table 4 shows, Sig. (2-tailed) = .151 is greater than .05. As a result, the assumption that there was not any difference between the experimental and control groups at the beginning of the study and before the experiment was confirmed. That is, both groups were the same in terms of their writing ability prior to the treatment. For Writing Score 2, Sig. (2-tailed) = .120 is greater than .05. Hence, there was no significant difference in the means of Writing Score 2 for the EFL learners' second essay writing between experimental and control groups. Consequently, both groups were still the same in terms of their writing ability.

The experimental groups, later for Writing Score 3, outperformed the control group in terms of the mean scores; that is, ($M_e = 85.33$, $M_c = 77.00$, $p < .05$), and Sig. (2-tailed) = .000 is less than .05. To this effect, there was a significant difference in the means of Writing Score 3 for the participants' third essay writing between the two groups. In fact, the experimental group receiving two sources of feedback performed better than the control group in the third essay writing task.

For Writing Score 4, Table 4 provides the answer to the first research question of the study. As Table 4 shows, the significance level of Sig. = .000 reveals that it is very unlikely that the observed mean difference of 9.400 was due to chance. Particularly, we can state that there was a .000 (Sig. = .000 < .05) probability that the observed difference in Writing Score 4 of the two groups could be attributed to chance and that the null hypothesis can be rejected. In other words, the integrated OAF and teacher feedback had a significant effect on EFL learners' writing ability. It is concluded that the experimental group receiving two sources of feedback performed better in their essay writing ability ($M_e = 89.73$) when compared with the control group's average ($M_c = 80.33$). In other words, the null hypothesis of no difference was rejected, and the study had statistical evidence in support of the idea that experimental and control groups' performance in their writing ability were significantly different after the treatment.

To further examine if there were significant changes in learners' writing scores across their four essay writing tasks (i.e., during four sessions) for the experimental group, a repeated-measures ANOVA was also performed for this group. Looking at Table 2, the reader can notice a consistent increase in the mean scores of the experimental group from Writing Score 1 to Writing Score 4 while also noting that the variances are approximately equal. According to the results of Mauchly's test for equality of variances, the level

of significance was greater than .05, namely Sig. = .366. Therefore, it could be concluded that the null hypothesis is not rejected which is an indication of equal variances. Furthermore, the results of the tests of within-subjects effects indicated the level of significance (i.e. F = 63.55, Sig. = .000 < .05); i.e. the means of the four writing scores were significantly different. That is, the statistical evidence showed that essay-writing scores of the experimental group did significantly change with the passage of time over four writing tasks. Table 5 presents the tests of significance for all possible pairs of the four writing scores of the experimental group.

Table 5.
Pairwise Comparisons for All Possible Combinations

(I) Session	(J) Session	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	-5.467*	1.612	.004	-8.924	-2.009
	3	-14.200*	1.722	.000	-17.892	-10.508
	4	-18.600*	1.712	.000	-22.272	-14.928
2	1	5.467*	1.612	.004	2.009	8.924
	3	-8.733*	1.395	.000	-11.726	-5.740
	4	-13.133*	.956	.000	-15.183	-11.084
3	1	14.200*	1.722	.000	10.508	17.892
	2	8.733*	1.395	.000	5.740	11.726
	4	-4.400*	1.397	.007	-7.395	-1.405
4	1	18.600*	1.712	.000	14.928	22.272
	2	13.133*	.956	.000	11.084	15.183
	3	4.400*	1.397	.007	1.405	7.395

As Table 5 shows, the significance levels for all comparisons showed a value of .00, (Sig.= .00 < .05), which provided evidence that the mean scores for these four writing scores of the experimental group were significantly

different. As a result, the increase in the means of essay writing scores of the experimental group over four sessions (71.13, 76.60, 85.33, and 89.73, respectively) demonstrated a statistically significant increase. Therefore, the statistical evidence revealed that the EFL learners' essay writing scores in the experimental group did improve as the sessions progressed. All in all, the results suggested that the use of integrated OAF and teacher feedback could improve the EFL learners' essay writing ability.

The second question addressed how EFL learners perceive the use of integrated OAF and teacher feedback for essay writing. To answer this question, the data were collected from both the questionnaire survey and interview with all participants of the experimental group. Table 6 presents the results of participants' responses to the fixed-response questions in the administered questionnaire.

Table 6.

Results of EFL Learners' Responses to the Fixed-response Questions (Part 1)

Question	Response	Number of respondents	Percentage of respondents
1- How much do you agree that blending computer-generated and teacher feedback could help you identify the strengths and weaknesses in your essay writing ability?	*Strongly agree*	11	73.3
	Agree	4	26.7
	Neutral	0	0.0
	Disagree	0	0.0
	Strongly disagree	0	0.0
2- How much do you agree that blending computer-generated and teacher feedback could help you improve your essay writing ability?	*Strongly agree*	12	80.0
	Agree	3	20.0
	Neutral	0	0.0
	Disagree	0	0.0
	Strongly disagree	0	0.0

Question	Response	Number of respondents	Percentage of respondents
3- To what extent are you willing to continue receiving feedback from both computer and teacher on your essay writing in near future?	*To very large extent*	11	73.3
	To a large extent	3	20.0
	To some extent	1	6.7
	To a small extent	0	0.0
	Not at all	0	0.0

As Table 6 shows, 100% of the participants strongly agreed or agreed that blending OAF and teacher feedback could help them identify the strengths and weaknesses in their essay writing ability. A similar proportion (100%) strongly agreed or agreed that integrating OAF and teacher feedback could help them improve their essay writing ability. Nearly 93% of the participants said that they were to a large or very large extent willing to continue receiving feedback from both OAF and teacher on their essay writing in near future, and only 7% expressed they were willing to some extent to do so. The results indicated that most of the participants perceived the benefits of the integrated OAF and teacher feedback and they were positive towards receiving feedback from both sources of feedback on their essay writing.

Table 7 presents the results of the participants' responses to the open-ended items of the administered questionnaire.

Table 7**.**

Results of EFL Learners' Responses to the Open-ended Questions (Part 2)

Question	Category	No. of responses	Percentage
4- What do you like **most** about blending computer-generated and teacher feedback?	- Simultaneous indication of areas for writing ability improvement	8	53.3
	- Simultaneous feedback as well as provision of suggestions and examples	3	20.0
	- Preparing learners for standardized exams	1	6.7
	- Motivation encouragement	1	6.7
	- Scoring method	2	13.3
5- What do you like **least** about blending computer-generated and teacher feedback?	- Slow feedback time and boredom	2	13.3
	- Lack of sufficient provision of suggestions and examples	2	13.3
	- Scoring method	3	20.0
	- Nil (Nothing)	8	53.3

Table 7 shows that among the participants' responses, eight (53%) pointed out the simultaneous indication of areas for writing ability improvement, for example, "*I like its view to all aspects [of my writing] such as vocabulary, grammar, sentences, etc.*" or "*[I like] categories and suggestions which are very useful in self-study in order to improve my writing.*" Three responses (20%) were about the simultaneous feedback as well as provision of suggestions and examples, for instance, "*They [two sources of feedback] can find problems and give me tips and advice about my writing which is really invisible for me.*" One of the responses (7%) was about preparing learners for standardized exams, as the respondent stated, "*It makes me sure about my strengths and weaknesses, also because I am going to pass TOEFL exam, I need to make my writing ability better through both computer*

and teacher feedback." Another response (7%) was about the motivation encouragement, for example, "*That was so helpful. […] Generally, after those writings, I am interested in learning English more, now.*" Finally, two responses (13%) were about the scoring method, as the two participants shared, "*I love all things in this method, especially its grade.*", and "*[I like] multiple and various feedback from two different sources with its ways of rating.*"

On the other hand, when asked about what participants liked least about the integrated feedback (i.e., the fifth item), eight responses (53%) expressed nothing in particular showing that over half of them were satisfied with the integrated feedback. As an example, consider this participant's idea, "*I saw no weaknesses in this tool. So, in relation to the above question, these two types of feedback are very useful for my writing.*" Three responses (20%) were about the scoring method, for example, "*Sometimes computer-based analysis and its scores might not be accurate because of technology failure.*" Two responses (13%) were about the lack of sufficient provision of suggestions and examples, as the two respondents expressed, "[…] *it must suggest some vocabulary to improve this ability to everyone, not by randomly introducing vocabulary*", and "*It could have given me a few ranges of vocabulary that I can use instead of my own vocabulary.*" Finally, two responses (13%) were about the slow feedback time and boredom, for instance, "*Each feedback can be given on two days one after another in order not to get students bored.*", and "*Two feedbacks take more time and it is boring.*"

The results suggested that participants perceived the strengths of the integrated feedback mainly in terms of feedback content, motivation encouragement, and scoring method. Moreover, the results also indicated that the participants perceived the weaknesses of the integrated feedback mainly in terms of its time-consuming process, the accuracy of scoring method, and

lack of examples. However, nearly 53% of the participants noted that there was nothing in particular that they liked least about the integrated feedback. To add more, the scoring method was either liked or not liked by the participants.

As explained earlier, 15 experimental EFL learners participated in the semi-structured interview to collect their views on the use of the integrated OAF and teacher feedback for essay writing. The results showed that they all agreed that the use of the two sources of feedback, namely OAF and teacher feedback could help them improve their essay writing ability. An interviewee mentioned, "*Well, it certainly helped my essay writing and when a human and technology are combined together, it can correct my problems on every aspect of my writing.*"

All interviewees also noted that the integrated OAF and teacher feedback could differ from teacher-only feedback in four ways: 1) The integrated feedback is more comprehensive as well as more accurate than the teacher-only feedback. The majority of the interviewees (74%) mentioned this difference, for example, "*Computer-teacher feedback points include all parts of the writing skill such as vocabulary, grammar, sentence structures, etc. But teacher-only feedback just mentions minor points.*" or "*From my point of view, mostly a teacher is not able enough to be as exact as a computer system.*" 2) The integrated feedback is more reliable than teacher-only feedback. 3) Teacher-only feedback is less satisfying, for instance, "*It's clear that teacher-only feedback is not reliable as both are, but when it comes to blending teacher and technology to give feedback, it will be more satisfying than the only one.*" 4) The integrated feedback is more convenient than teacher-only feedback. An interviewee, for example, mentioned, "*I like both of them, but I personally prefer the integrated computer-teacher feedback because it is more complete and convenient to use.*"

When interviewees were asked what they liked most about the integrated feedback, five satisfactory features were noted. Eight (out of 15) interviewees liked the feature of identifying mistakes/errors accurately as well as comprehensively, for example, "*What I like most is being so careful and detailed in its every part.*" or "*[…] it identifies all of my writing problems exactly.*" The rest of the interviewees commented that they were in favor of features like provision of suggestions and examples, being motivating, speed of feedback provision, and ease of use. With respect to these features, some of their statements were as follows. Four interviewees mentioned that they were in favor of the provision of suggestions and examples, like "*What I like most are tips and examples on the length of sentences used in the text.*" Four of them responded that it was good to motivate them in learning English writing, for instance, "*It is a big help for writing essays. I didn't know this kind of thing existed and now that I got to know it, I will be using it a lot in the future.*" while three said that the speed of feedback provision was satisfactory, "*What I like most is its speed in answering.*" Finally, three of them expressed the ease of use as a liked feature, for instance, "*What I like most is that it is easier to use and identify all of my writing problems exactly.*"

When asked what they liked least about the integrated feedback, seven interviewees commented nothing in particular. However, the rest of them stated that there were several features that may have an adverse effect on their writing, such as lack of sufficient provision of suggestions and examples, slow feedback time and boredom, scoring method deficiency, unreliable feedback due to teacher tiredness or technology failure, lack of coincidence of two sources of feedback, and difficulty in understanding the provided feedback.

With respect to the last item of the interview, six interviewees responded that the integrated feedback can be promoted through educating experts, for example, "*In my opinion human being can promote computer technology on*

giving feedback on educational aspects and educate experts to give out the best results." Five of them suggested nothing in particular, while the rest (four out of 15) suggested providing sufficient suggestions and examples for each area, coincident feedback, and an online teacher could enhance the integrated feedback. Their ideas can be found in the following comments, respectively.

- I hope that it could be more intelligent in suggesting words to improve the errors, such as in 'bad phrases' section instead of saying 'replace these words', it could introduce replacement words.

- Enhancing the coincidence of the two sources of feedback

- An online teacher who can interact with students online will be a great option.

Taken together, it was confirmed that the use of the integrated feedback could significantly improve EFL learners' essay writing ability. The main reasons for such improvement included raising EFL learners' awareness of the essential areas for the improvement of essay writing ability, identifying mistakes/errors then providing suggestions and examples on their essay writing, providing comprehensive as well as accurate feedback via two sources of feedback, and increasing learners' motivation. While the majority of participants thought that the integrated feedback generated by the OAF tool as well as the teacher was beneficial, two questioned whether it could provide sufficient suggestions and examples for their essay writing mistakes/errors. Another concern of the participants was whether the integrated feedback could take less time so that it could be less boring to receive feedback on their essay writing. Three of them were also concerned about its scoring method deficiency in terms of their essay writing. Additionally, a few of them stated that the integrated feedback was unreliable due to teacher tiredness or technology failure. The interview data also indicated that the benefits of the

integration of OAF and teacher feedback is more important and valuable than its disadvantages. However, integrated feedback could not meet all EFL learners' needs.

Discussion

The purpose of this study was two-fold. Firstly, it was designed to examine whether blending both OAF and teacher feedback had a significant effect on Iranian EFL learners' essay writing ability or not. Secondly, it aimed at exploring Iranian EFL learners' perceptions on the use of blended OAF and teacher feedback to foster their essay writing.

Participants' four essay writing scores, obtained during the research sessions, were analyzed in the two groups (i.e., experimental and control) based on the blended feedback. The results of the statistical analyses demonstrated that there was no significant difference between the groups in the means of writing scores up to session four (i.e. up to Writing Score 2). To this end, both groups were the same in terms of their writing ability. However, a significant difference was revealed in the means of Writing Score 3 and Writing Score 4 for the EFL learners' third and fourth essay writing between the two groups. That is, in the fifth and sixth sessions the two groups were not the same in terms of their writing ability, and the experimental group receiving two sources of feedback outperformed the control group in their third and fourth essay writing. Therefore, it can be claimed that using the integrated OAF and teacher feedback did positively affect EFL learners' essay writing improvement.

Regarding whether there was any significant difference within the experimental group's writing scores as a consequence of receiving the integrated feedback or not, the statistical evidence showed that the experimental group outperformed the control group in their essay writing; this

implied that the use of integrated OAF and teacher feedback could improve the EFL learners' essay writing ability.

Additionally, the questionnaire findings showed revealed the participants' positive attitudes towards the usefulness of the integrated OAF and teacher feedback. On the other hand, the interview data were also used to cross-validate the questionnaire results. As a result, several categories that emerged from the interview analyses revealed that the integrated OAF and teacher feedback could benefit EFL learners through specifying the areas to improve essay writing ability, identifying mistakes/errors, providing suggestions and examples on their essay writing, providing comprehensive as well as accurate feedback through two sources of feedback, and increasing learners' motivation. Other identified categories showed the limitations of the integrated OAF and teacher feedback mainly in four areas: lack of sufficient provision of suggestions and examples, slow feedback time and boredom, scoring method deficiency, and unreliable feedback due to teacher tiredness or technology failure. The qualitative data indicated that the attractions of the integrated OAF and teacher feedback outweighed its drawbacks; however, the integrated feedback could not meet all EFL learners' needs.

The findings of this study provided an answer for Cheng's (2017) second study limitation which focused on the use of OAF only. The results of his study proved that, on one hand, teacher feedback played a crucial role that could be neither replicated nor replaced by OAF. The reason is that both OAF and teacher feedback have their own pros and cons, and no single feedback source is able to meet all EFL learners' writing needs. On the other hand, designing an effective approach to blending both teacher feedback and OAF for writing was worth investigating. As Cheng indicated, his results, on the basis of using only OAF, had several drawbacks including incapability of understanding human language, lack of feedback provision on the subject

agreement, main idea, supporting and concluding sentences, and the absence of the fitness of the learner's writing content. These all could be compensated by integrating teacher feedback and OAF in the present study.

It is worth mentioning that few previous studies practically had blended two sources of feedback, namely OAF and teacher feedback. They merely investigated the effects of these two sources of feedback separately. Yet, the findings of their studies concluded that the computer-generated feedback is required to be integrated with teacher feedback and it cannot efficiently be applied separately; that is, they cannot be employed interchangeably in classrooms (Chen & Cheng, 2008; Kellogg et al., 2010; Stevenson & Phakiti, 2014). The reason is that both OAF and teacher feedback have their own pros and cons, and no single feedback source is able to meet all EFL learners' writing needs (Cheng, 2017).

In another study, Li et al. (2014) concluded that the teacher takes advantage of computer-generated evaluation for diagnostic purposes and there is a come and go relation between both computer-generated and teacher feedback for the evaluation of EFL learners' writing, which is in line with the qualitative findings of the present study; that is, using the integrated feedback for specifying the essential areas to improve writing ability, identifying learners' mistakes/errors then providing suggestions and examples on their essay writing, providing comprehensive as well as accurate feedback through two sources of feedback.

Regarding the EFL learners' attitude, the findings of the present study are also in accordance with Chen and Cheng's (2008) views asserting that AWE feedback can be perceived favorably when it is followed by the teacher and peers feedback to revise their writing. Their study also revealed that AWE feedback could not replace teacher and peer feedback since it caused frustration to EFL learners and limited their learning of writing. However, it

is said that AWE can be exploited as a supplement to teacher and peer feedback (Chen & Cheng, 2008). In line with the present findings are the results of another study (Liao, 2016) stating that his participants' positive performance on writing was due to repetitive practices, gap noticing as well as applying an AWE tool under the integrated process (i.e. the blended computer-generated and teacher feedback). Therefore, we could come to this result that the overall findings of the present study support the findings of previous studies proving that the blended feedback can potentially raise EFL learners' motivation to be aware of their writing problems and address them appropriately (Cheng, 2017; Dikli & Bleyle, 2014; Kim, 2014).

Conclusion

The findings of the present study bore several major educational and pedagogical implications in the area of English essay writing for teachers, students, and other stakeholders such as test-developers and curriculum developers. The results also have several implications for practitioners, administrators, and researchers. In educational contexts, stakeholders are expected to integrate the two sources of feedback (i.e., OAF and teacher feedback) and consider those factors that students find most important and beneficial for their L2 essay writing learning. With this in mind, the EFL teachers are required to be trained on how to integrate an OAF tool appropriately in their writing classes. Stakeholders employing the integrated feedback should consider how they could best train their instructors and their students to work with this type of feedback. Institutions can also develop and support further research and training on the integrated feedback and offer ongoing support for its use and the technology and pedagogical systems that support it.

The effectiveness of the blended feedback can be determined by teachers' attitudes and their skills in terms of using AWE tools, as well as EFL learners' characteristics and goals for learning writing. At the instructor level, an understanding of the implications of the integrated feedback can help instructors choose the sources that suit their feedback philosophy. For instance, an understanding of the implications of the integrated feedback can assist teachers and university administrators make decisions about the types of technological support and tools they should offer to their EFL students. Instructors and the administrators keen on applying the integrated feedback on their learners' writing will be able to make informed choices about investing their time and funds. Using the blended feedback in their classrooms, the teachers can reduce their workload and time, identifying local grammatical errors via OAF tool, and then the teacher feedback addresses global errors only. The blended feedback can also be helpful for instructors' populated classrooms and their learners' high expectations since the teacher can employ OAF to provide extensive feedback in a short time and complete it with the teacher feedback.

At the student level, perceiving the positive effect of the blended feedback, they can identify mistakes/errors more accurately and comprehensively. As a result, they can improve their attitudes towards using blended feedback, increase their motivation to learn, and autonomous learning.

Test-developers, additionally, can make use of the rubrics and holistic scores of the integrated feedback and develop writing tests in line with the results of this study. The findings of the present research also suggest when the traditional approach of feedback provision (i.e., teacher feedback) is accompanied by technology-based feedback (i.e., computer-generated feedback), the stakeholders and institutions will have opportunities to create a

more convenient and motivating L2 learning environment. To this effect, they are also required to equip the educational context with the latest technologies such as computers, and the Internet.

Based on the limitations and delimitations of the present study, some suggestions can be put forward for future studies. Further research is required to move towards investigating the advantages and disadvantages of the integrative OAF and teacher feedback in order to obtain a more comprehensive and in-depth understanding of its pros and cons. The next recommendation for further research is that other writing genres as well as EFL learners at different levels of language proficiency can be addressed. Moreover, the researchers also recommend focusing more on the effects that the integrated OAF and teacher feedback has on the motivation of EFL learners. Further research can also integrate peer feedback to the integrated OAF and teacher feedback to investigate the effect of the three sources of feedback simultaneously.

References

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater V. 2. *The Journal of Technology, Learning and Assessment, 4*(3), 1-31.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. *Ai Magazine, 25*(3), 27-27.

Burstein, J., & Marcu, D. (2003). Automated evaluation of discourse structure in student essays (chap. 13). In M. D. Shermis, & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 209-229). Mahwah, New Jersey: Lawerence Erlbaum.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*(3), 267-296.

Chapelle, C. A. (2001). *Computer applications in second language acquisition*. London: Cambridge University Press.

Chen, C.-F. E., & Cheng, W.-Y. E. C. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94-112.

Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education, 34*, 18-27.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319.

Conrad, S. M., & Goldstein, L. M. (1999). ESL student revision after teacher-written comments: Text, contexts, and individuals. *Journal of Second Language Writing, 8*(2), 147-179.

Cunningham, K. J. (2018). *Modes of feedback in ESL writing: Implications of shifting from text to screencast.* (Ph.D. Dissertation), Iowa State University, ProQuest LLC.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18*(1), 7-24.

Delice, A. (2010). The sampling issues in quantitative research. *Educational Sciences: Theory and Practice, 10*(4), 2001-2018.

Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing, 22*, 1-17.

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition. The Cambridge applied linguistics series*. New York: Cambridge University Press.

Ellis, R. (2008). A typology of written corrective feedback types. *ELT Journal, 63*(2), 97-107.

Ene, E., & Upton, T. A. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System, 46*, 80-95.

Ericsson, P. F., & Haswell, R. H. (2006). *Machine scoring of student essays: Truth and consequences*. Utah: Utah State University Press.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*(3), 161-184.

Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29*(1), 33-53.

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315-339.

Fotoohnejad, S. (2018). *Using an automated writing evaluation program (Criterion) for EFL learners' writing improvement: A process-writing approach.* (MA Thesis), Vali-e-Asr University of Rafsanjan, Iran.

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment, 8*(6), 1-42.

Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing, 8*(1), 5-16.

Hearst, M. A. (2000). The debate on automated essay grading. *IEEE Intelligent Systems and their Applications, 15*(5), 22-37.

Hoon, T. B. (2006). Online automated essay assessment: Potentials for writing development. Retrieved from http://ausweb.scu.edu.au/aw06/papers/refereed/tan3/paper.html

Huang, S.-J. (2014). Automated versus human scoring: A case study in an EFL context. *Electronic Journal of Foreign Language Teaching, 11*(1), 149-164.

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing, 7*(3), 255-286.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83-101.

Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research, 42*(2), 173-196.

Keshavarz, M. H. (2017). *Contrastive analysis, error analysis, and interlanguage* (2nd ed.). Tehran: Rahnama.

Kim, J. E. (2014). The effectiveness of automated essay scoring in an EFL college classroom. *Multimedia-Assisted Language Learning, 17*(3), 11-36.

Klobucar, A., Elliot, N., Deess, P., Rudniy, O., & Joshi, K. (2013). Automated scoring in context: Rapid assessment for placed students. *Assessing Writing, 18*(1), 62-84.

Krishnamurthy, S. (2005). A demonstration of the futility of using Microsoft Word's spelling and grammar check. Retrieved from http://faculty.washington.edu/sandeep/check/

Lai, Y. h. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology, 41*(3), 432-454.

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1-18.

Li, Z., Link, S., Ma, H., Yang, H., & Hegelheimer, V. (2014). The role of automated writing evaluation holistic scores in the ESL classroom. *System, 44*, 66-78.

Liao, H.-C. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System, 62*, 77-92.

Mackey, A., & Gass, S. M. (2016). *Second language research: Methodology and design*: Routledge.

Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Great Britain: Pearson.

Robinson, P. (2001). *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51-65.

Wang, Y.-J., Shang, H.-F., & Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning, 26*(3), 234-257.

Whithaus, C. (2006). Always already: Automated essay scoring and grammar checkers in college writing courses. In R. H. Haswell, & P. F. Ericsson

(Eds.), *Machine scoring of sudent essays: Truth and consequences* (pp. 166-176). Logan, UT: Utah State University Press.

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, *50*(3), 483-503.

Yeh, Y.-L., Liou, H.-C., & Yu, Y.-T. (2007). The influence of automatic essay evaluation and bilingual concordancing on EFL students. *English Teaching & Learning*, *31*(1), 117-160.

Zemach, D. E., & Rumisek, L. A. (2013). *Academic Writing from Paragraph to Essay*. Oxford: Macmillan.

Zheng, B., Warschauer, M., & Farkas, G. (2013). Digital writing and diversity: The effects of school laptop programs on literacy processes and outcomes. *Journal of Educational Computing Research*, *48*(3), 267-299.

Appendix 1: Questionnaire

In the Name of God

Questionnaire

Name: …… Age: ……..…..

Dear participant,

Please answer questions No. 1 to 3 based on (Strongly agree = 5, Agree = 4, Neutral = 3, Disagree = 2, Strongly disagree = 1), and questions No. 4 to 5 in written comments.

1- How much do you agree that blending computer-generated and teacher feedback could help you identify the strengths and weaknesses in your essay writing ability? ① ② ③ ④ ⑤

2- How much do you agree that blending computer-generated and teacher feedback could help you improve your essay writing ability? ① ② ③ ④ ⑤

3- To what extent are you willing to continue receiving feedback from both computer and teacher on your essay writing in near future? ① ② ③ ④ ⑤

4- What do you like **most** about blending computer-generated and teacher feedback?

5- What do you like **least** about blending computer-generated and teacher feedback?

Thank you very much for your cooperation!

Appendix 2: Interview

1- Do you think that blending computer-generated and teacher feedback could help you improve your essay writing?

2- What is your opinion about the main differences between the blended computer-teacher feedback and teacher-only feedback?

3- What do you like **most** and **least** about the blended computer-teacher feedback?

4- Do you have any suggestions to enhance the blended computer-teacher feedback?