Iranian Journal of Applied Linguistics (*IJAL*)

Vol. 22, No. 1, March 2019, 154-196

The Impact of Electronic vs. Human Observer Feedback on Improving Teaching of Translation Skills to Iranian EFL Students

Omid Tabatabaei *, Majid Fatahipour, Maryam Mohammadi Sarab
Department of English, Najafabad Branch, Islamic Azad University, Najafabad, Iran

Abstract

In the area of teaching translation, investigating the effectiveness of technological solutions in developing translation skills is both timely and in demand. As the first step, one can try to explore feedback generated by machines compared to humans. The purpose of this study was to examine the impact of electronic feedback provided by the designed translation software on the translation skills of Iranian EFL students compared to the feedback given by an expert human observer. To this end, 60 intermediate male and female students were selected as participants after the administration of Oxford Placement Test (OPT), followed by a translation skill test as pretest and posttest. The analysis of the pretest and posttest data revealed that after receiving the two types of feedback, both groups performed significantly better on posttest. Further analysis of the data, however, indicated that sustained electronic observer feedback was superior to human observer feedback in terms of developing translation skills. It can be discussed that taking advantage of an efficient electronic tool which enjoys the potential of providing some feedback would drive the focus away from the repetitive mistakes and frees up the time and focus on students' personal problems. The findings may have implications for translation education, training, and practice.

Keywords: Electronic observer feedback; human observer feedback; translation skills, translation software

Article Information:

Received: 30 November 2018 **Revised:** 9 February 2019 **Accepted:** 11 February 2019

Corresponding author: Department of English, Najafabad Branch, Islamic Azad University, Najafabad, Iran

1. Introduction

1.1 Background

The field of translation is now a technical as well as an academic activity in universities worldwide. If translators and translation students wish to promote in their profession, they need to become familiar with up-to-date technical tools. Having the potential of extending the knowledge and expertise of the users, they would be better enabled to flourish in the profession. However, the use of technology is not developed well or generally underused in various translation areas. For example, legal translators are only engaged with the fixed documentary sources; medical translators do not use translation memories; and technical translators refer to thesaurus (García-Izquierdo & Conde, 2012). On the one hand, it is obvious that the technical aspect of training the translators to improve the use of certain practical tools, such as manuals or thesauri are important. For example, making editorial feedback more automatic may increase our ability to make corrections and revisions that instantly can lower the errors often made in translation. On the other hand, an important aspect of any tool is flexibility, since change is constant and inevitable and any translation advisor constantly gets updated. The demands of students and users are not

just linguistic and an expert system cannot deal with those problems alone (Hassani Goodarzi & Rafe, 2012).

Two types of linguistic errors can be envisaged in translation. They are grouped into global and local errors. A global error is one which involves the overall structure of a sentence and a local error is one which affects a particular constituent such as omission of prepositions (Burt & Kiparsky, 1974). Identifying the linguistic errors may help teachers and material designers choose an appropriate pedagogical method. Thus, it is important to be recognized early on.

In many countries, there is a recurring scarcity of resources reported in literature regarding the integration of technology into translation. Even when it occurs, the technological advances rarely find their way into curricular practice or this transition is too unclear and still in many places, as it is reflected in the context of this study, no dedicated translation computer labs are available in which translation students could be trained more practically. Instead, students are generally advised to use computer as much as possible! This generality does not help practice much at all. Electronic skills have not been considered as part of translation competence for students or teachers of translation, and technology is not the linked with the prime purpose of translation instruction. Innovations of e-learning in Translator Training programs are still seen as unrelated to the type of

interactions needed within translator and interpreter training courses (Kearns, 2006).

Translation training and quality assessment should go parallel with the advancements in pedagogy and language teaching, where the classroom strategies are student-centered and the training should enable students to develop the skill of decision-making. With greater knowledge and awareness, translators can have more useful choices. Using online *translation software*, there will be less teacher input in the main-task phase; the students explore intended words or sentences, and the teacher monitors and responds to questions. Post-task feedback, provided by the teacher is the response given after reflecting on the translation process. It is derived from the results that depend greatly on the input that the students receive (Mitchell-Schuitevoerder, 2014).

The translation teachers should try to describe the actual translational decisions made by actual translators based on different socio-cultural variations, attitude, knowledge, and experience. Therefore, decision-making power is one of the requirements of professional translation. Knowledge of the source-language culture causes to produce a translation that is both readable and resistant. Translators must possess a commanding knowledge of the diverse cultural discourses to be able to offer an acceptable translation. The topic of translation can affect their socio-professional

profile, and their attitudes towards information resources. Accordingly, the current research suggests that the potential of artificial intelligence programs can be employed for distinguishing the quality of professional from non-professional work in translation that can represent a more accurate comparison of the translated texts done by some translators. This type of expert program can work with a checker program in which the translated text is rated at a certain level of professional quality whereby taking a score, which as a result, is closest to whatever the score is rated to manually marked score by the translation teacher.

2. Review of literature

2.1. Translation quality and reliability feedback

The positive effect of feedback on student learning is well-discussed in literature; however, in rare instances as Krish (2006) stated, research has aimed to uncover the power of feedback in an online environment in addition to a teaching tool. Thus, feedback has an important role in learning and provides the most productive understandings gained from specific opportunities in teaching to enhance student learning. Analysis of strong and weak points through the provision of feedback was made by the evaluators to help learners' practice in reflecting on identifying similar points. With the advent of computers, the traditional type of feedback called human observer gave its way to the more modern and recent computer-assisted feedback called electronic feedback. It should be noted that the second type cannot be

totally independent of human observance. However, a pure human type of feedback still abounds in translation classroom due to its facility and that it can be done in a simple classroom where no hi-tech devices are available to the teacher.

Research has indicated that learners welcome the feedback provided by a human observers (who are usually their teachers) and get more motivated, since they feel they can identify their own problems when the human observer is present. Waddington (2010) discussed the validity of different methods of evaluating student translations. At the beginning, students can earn initial points by the observer, although more accurate performance by students depend on the good guidance of the observer, having proper condition to present sufficient descriptions, even students' condition and their tension is not unimportant. Whatever the modality is, linguistic conditions and requirements should also be met all types of translations, as Tabatabaei and Fatahipour (2015) recommend.

2.2. Human observer feedback in the training process

Teachers can create a classroom environment that fosters and supports learning, by providing human observer feedback that is corrective, timely, and criteria-oriented, If teachers do not understand the learning objectives and processes, it would be difficult to make students understand what good

performance looks like. Feedback is invented in teaching to help students understand what is correct and what is incorrect(Hattie & Timperley, 2007). We know any feedback may be better than no feedback. However, effective feedback should also provide information about how close students come to meeting the criterion and teachers can provide elaboration in the form of practical examples (Shirbagi, 2007). Teachers can involve students in the feedback process by asking them to keep track of their performance as learning occurs during a unit or course (Dean, Ross Hubbell, Pitler, & Stone, 2012).

Despite all the benefits of the observer feedback, there are challenges that should be noticed. Correcting all the mistakes needs plenty of time that would make the trend impractical. If the students do not ask their questions in relation to the given text, such as facing an unknown word or the observer failing to express a particular and necessary point, monitoring the mistakes of all students would be rigid and tough. The observer might make a mistake in rating students or the scores might be totally displaced (Zieky & Perie, 2006).

2.3. Electronic feedback on the translation skills

Quality of observation demands a special set of knowledge and skills. Teacher feedback provided by a human observer often runs into challenges such as comparing the differences in students' performance whose gains are shown among scored assignments by a specified test, and through the use of

criteria for rating. Besides, it is too challenging for observers to produce accurate and meaningful feedback at a first attempt. For translation training, or any training, to be successful, certain essential activities need to be included to support students in gaining the required knowledge and skills. Higher-level understanding develops through effective questioning in the classroom. "Bias awareness training is needed to help observers to identify their preferences. They also need to understand their own personal tendencies to favor specific aspects of instruction, or to disfavor them, for particular reasons" (Archer, Cantrell, Holtzman, Joe, Tocci, & Wood, 2015, P. 45).

The employment of electronic feedback in translation does not have a long history. Benjamins, (2000, as cited in Tennent, 2005) examined the translation teaching methodology adopted by the various translators and discussed the underlying features of course content and structuring. Specific training was needed for translators to benefit from the opportunities offered by technological tools. The integration of technological tools had a good effect on translation, for instance translating with computers in a classroom, selecting texts at the right level for translation with certain technical criteria, or teaching how to translate text types which were in a translation memory data bank. Procedural knowledge has modeled activities, skills, and expertise that learners need within a systematic framework (Tennent, 2005).

One of the most important innovation in this e-learning model is the use of student self- and peer-assessment and tutor moderation as pillars of the assessment procedure. In a study, rank labels were used to rate the collected translations, and to discriminate between acceptable and unacceptable translations, which were then correlated with the scores assigned to outputs of machine translation systems (Zaidan & Callison-Burch, 2011).

Current machine translation equipment is still far from perfect because the output from these systems needs to be edited for some skipped errors. A way of increasing the productivity of the whole translation process is incorporating the human correction activities within the translation process itself. Typical solution to improving the quality level of the translations by a translation system requires manual post-editing. Thus, post-processing such as replacing the tags with their corresponding words will take place after the translated text by the user. System performance can be assessed by comparing sentence translations produced by the translation system as well (Barrachina, et al., 2009).

However, a simple error count is not advisable as a method of scoring a translation since it may not give credit for content richness and cannot measure the seriousness of the errors. Goff-Kfouri's (2005) argued that translation rubric is used to assess translation within a set of criteria that can be chosen to rate translation, based on general impression, and error counting.

As for teaching, rubrics used in related fields may be made applicable to translation. For example, Heaton (1990) has proposed an analytical or arithmetic grid for language teaching courses, which can be adopted for a translation scoring, too. The translation can be marked over fluency, grammar, terminology, general content, and the number of errors with x weighting for errors perceived to be more serious.

Translation Correction Criteria adopted from Heaton (1990, p. 110)

Error types	
Fluency /Flow	2x
Grammar	2x
Terminology	x
General Content	x

The test rubrics of time allocation, instructions, and test organization as well as grammar, terminology, and general content were used to identify and improve both accuracy and fluency of translators in the present study.

Two goals, adequacy and fluency, were the main criteria in machine translation evaluation. Statistical machine translation system are biased towards either longer or shorter output relevant for scoring the translation output. Instead of reporting human judgment of translation quality, researchers relied on automatic measures, BLEU score, which measures overlap with reference translations. Those researchers also provided empirical evidence that the estimated significance levels are accurate by comparing different texts (Koehn, 2004).

Now, the input could be controlled and engage considerable interaction between translated texts of translators through a designed program for receiving any electronic feedback. For high quality output, database, error management, and bug tracing are needed to be matched with the program task. One particularly important facility can be a complete software, which tells the computer what to do. Software generally consists of programs which were sets of instructions telling the computer to read some data, perform some calculations and comparisons, output some messages or digits, and so on (Hutchins & Somers, 1992). A discussion was reported by Newton (1992) that the translation system performs a translation task too boring for any human doing it to last for more than a few months, yet sufficiently constrained to allow an MT system to be devised which only makes mistakes when the input is ill-formed.

Nishida and Takamatsu (1990) designed an experimental system which analyzes error-correcting information given by post-editors and compares the intermediate expressions of the post-edited translation with those of the original translation, and somehow learn to identify the inappropriate or faulty parts to be corrected.

Finally, a teaching-related innovation in e-learning model of translation is employing the student self-assessment and peer-assessment subject to teacher moderation as pillars of the assessment procedure. Now, this can show a research gap to be addressed as to examine the interaction of three groups, two of which assigned as experimental groups going through a specific computerized feedback and another group as the control group.

3. Method

There are many ways to enhance *translation skills*. One is through assessing and evaluating translations through comparisons made between machine-generated and human-generated feedback. The current study is conducted to evaluate the usefulness of two different approaches to enhancing *translation skills* among undergraduate learners, namely *human observer feedback* and *electronic observer feedback*. Thus, the respective participants, instruments and procedures of the study ensue.

In the first part that follows, there is an elaboration on the general characteristics of the participants who took part in the study. Moreover, there are separate sections that follows up with explaining the details about instruments and procedures used in the study. The section on instruments also includes the materials which were adopted to gather the data. The next section follows with explaining the details of conducting the research. In other words, the procedures and the design through which the participants, instruments and materials have functioned. Finally, this section explains how data was gathered and prepared to be analyzed to set the scene for obtaining reliable results.

3.1. Participants

This study was conducted among undergraduate students of translation studies and language literature in a university in Tehran, Iran. To carry out the present research, intermediate students were recruited as participants. Adopting a quasi-experimental pretest-posttest design, the researcher recruited 60 participants from a pool of 100 students after they sat a language proficiency test, detailed in the next section. Then, 30 students were selected to receive solely human feedback and 30 students to receive electronic observer-mediated feedback. The participants were both males and females, with 22 participants who happened to be males and 38 participants who were females. The age of the participants ranged between 18 to 28 years old. The current study preplanned a regular treatment method

that occurred ina dedicated half an hour mini-session included during each normal session for the whole duration of the 16-session university course entitled "Translation of Journalistic Texts" which was provided for students in 90 minutes, once a week in two similar classrooms. The same teacher who was faculty member with over 12 years of teaching experience instructed both groups, thus method effect is minimized. Scores from only three of participants were disregarded during the process of intervention and data gathering because of their inconsistent participation or absence in the posttest. For this reason, the results of the study were reported based on 28 final participants in *human observer feedback* group and 29 participants in the *electronic observer feedback* group.

3.2. Instruments and materials

As part of their degree requirement, before being research participants, the students have registered a course module and their respective textbook was already assigned, as entitled *Translation of Journalistic Texts 1, which has been* authored by Gholamreza Tajvidi. The book was published by Payam-e- Noor publications, a local university publisher. Thus, the translation passages offered by the mentioned book were the limit of texts used in the classroom. Thus, course and material were already set as available, but students are serious and cooperative, on the plus side.

The following instruments were employed in order to collect the required data for the present study before and after the experiment itself, one is the well-used Oxford Placement Test (OPT) (See Appendix III) and the other is a researcher-developed *Translation Skill Test* to both suit their level, and their classroom schedule and textbook.

Both groups were found to be on intermediate level of language and translation skills and generally considered at the same footing with other less formal but important features such as the parameters of interest, breadth of knowledge, speed of action and success. The parameters which are necessary in experiencing the appropriate choice and decision for substituting words gradually were calculated by Goff Kfouri (2005) and Heaton's (1990) specified rubrics.

3.3. Oxford Placement Test (OPT)

To check for any significant difference between the language proficiency level of participants of the study, a version of OPT test (See Appendix for details), which is an English language examination (Provided by Oxford University Press and University of Cambridge Local Examinations Syndicate) was given to students. The test is especially designed to measure language learners' ability to communicate through English for everyday purposes. This is one of the major instruments in the present study because it can provide an index of ability which is independent and external to the translation construct and competence of students and translation scores. This

measurement instrument includes 60 multiple-choice test items. According to the score designation in the test, participants who score between 38 and 48 were assigned as intermediate learners. The participants were allowed to answer the questions in the time limit of 60 minutes with clear instructions provided for the students.

3.4. Translation skill test

In this study, the researcher designed *translation Skill Test*s to be used both as pretest and posttest. This test involved two sections: The first section consists of an English passage to be translated into Persian; the second section consists of a Persian passage to be translated into English. The passages are excerpts of recent news headlines. The test had been designed in a way to be like the tests they take at their class time in terms of design and topic, as much as possible (Appendix I).

Further, a computer program is designed in order to evaluate and ultimately help improving the students' translation skill by giving electronic feedback. The developed program (See Appendices for screen shots) were employed after the three consecutive pedagogical sessions and piloting that was provided for students for two hours a week. The different type of texts to translation that are prepared and recorded on the system along with their correct Persian translation, the students of each class were also tested by a

given translation text as a post-test. The students who were randomly chosen in each group had similar study major as English language translation undergraduate degree. Although the obtained scores of the students were considered individually, the obtained results of the groups are compared together and the impact of electronic and *human observer feedback* of this method is shown in results section.

3.5. Procedures

Since the translation test was designated, reliability, validity and feasibility were checked. For having an index of reliability, and also checking the feasibility of the test, the test was piloted before being used in the main study. To pilot the test, the same test was given to 16 comparable students of higher education institute in Tehran which was comparable to the university level designated as the target sample, at two-week intervals. The two sets of obtained scores were compared and a Cohen Kappa correlation of 0.82, which is considered acceptable, was obtained (Ubersax, 2016). As for the feasibility, the comments of students after the piloting sessions were collected and the test went under minor revision to make sure it does not have any problem with facets of the test rubric, namely, time allocation, instructions, and test organization. Thus, the test was shown to be practical and feasible in a similar environment as the environment in the main study. For the purpose of arguing for its validity, the test was checked through panel discussion with TEFL experts, especially as to its content and face

validity. The panel verified these aspects of validity of the test. The participants were asked to take the same test at the beginning (pretest) and end of the treatment (posttest) for making comparisons.

The rationale behind selecting these texts was their difficulty level which matched the proficiency level of participants of the present study. The difficulty level was ascertained by NASA Task Load Index (Hart, 1988 & 2006), which consists of six criteria on a 20 points scale. The 16 students who had been recruited in pilot study filled the difficulty scale. A difficulty index of 10.64 which stands for intermediate level text for translation (8 to 12 stands for intermediate level texts) was obtained (For further details, see Appendix IV).

The design of this study was quasi-experimental in which there were two groups: an experimental group and a control group. The control group in this study received their feedback from human observers while the experimental group received feedback from electronic sources. In the experimental group, the participants were provided with the specially designed software for developing *translation skills*. The software contained a database with a specified number of English and Persian sentences, and their equivalent translations. As the students typed the translation of a sentence, on the software, the mistranslated sections were highlighted in red and the student had to re-translate the erroneous part(s) of the passage. This

process was continued till the software accepted the translation completely. In the control group, however, the students were provided with some English or Persian short passages to be translated into the other language (See Appendix II). These materials were in form of printed hand-outs which were given to students to be translated. The participants translated the passages and handed them in to the teacher. Then, the teacher noted and commented on some problematic parts of the translated passages. Finally, the students re-translated the passages by taking teacher's feedback into account.

Thus, to take the procedure in a nutshell, the first session of the study started with administering the OPT to all participants of the study. The pretest which was *translation Skill Test* was administered to students of both groups, the next day at the same place and time of the day. Immediately after taking the test, the two groups were separated in two classrooms. The allocation was not arbitrary and participants in two predetermined classes were randomly assigned into two groups of experimental and control group. While the control group only received *human observer feedback*, the experimental group received electronic observation feedback. Finally, after the students of both groups finished the course materials and treatment, they came together and sat for a post-translation test.

It needs to be added that the students' performance in both pretest and posttest was evaluated by two raters based on Goff-Kfouri's Rubric (2005)

and also the criteria adopted from Heaton (1990) mentioned in the literature review section. To ascertain the reliability of the scores, the researcher used raters and measured inter-rater reliability. To this end, participants' translations of the passages were evaluated by two raters. Both raters were experienced translators and familiar with translation evaluation. The anonymized raters were assured of their privacy and confidentiality. They have a PhD degree in Applied Linguistics and Translation studies and over 10 years of experience in teaching and researching in those fields and practical experience in translating various texts for clients. Visual Studio has been used to design this software in the present study. It is a powerful technology used to design and build dynamic web pages. This designed translation software can be easily combined with other online translation software of translation machines.

As shown in the results, in the experimental group, the participants were provided with the specially designed software for developing *translation skills*. The software contained a database with a certain number of English and Persian paragraphs and their equivalent translations while it possessed the high capability to be encompassed any type and number of texts. In Implementation of translation comparison program, as the students were typing a translation text for the determined source texts in the study, on the software, the mistranslated sections were

highlighting in red along with their locations in green and the students could re-translate the erroneous part of the passage after comparison with the software. This process was continued till the software accepted the translation completely. The following pictures are related to designed program with C# software in some examples.

The text of figure 1 (Appendix II) is a correct translation of a given English text in which there was as prepared samples on the system. It was then compared with students' translated texts. The following paragraphs which were separated by hashtag were related to the translated texts by students.

Three separated translated texts were compared with the correct translation and the number of mistakes was computed for each text (See Figure 1, 2, and 3 in Appendix II). For instance, the number of differences between correct and incorrect translation was 1 in first line which means there was just one mistake in the first translated text by student. In fact, all of 30 translated texts can be compared to the correct translation. The performed translation of the given paragraph in figure 4 (Appendix II) was compared with main translation. All three lines turned red which means there were some errors on each line of the translated text. While the number of translation errors was found, the location of mistranslated parts was determined as well. The red lines were shown at the location of mistakes which were in 2 lines in this figure and the green lines were

shown to see what mistakes had the students. The white parts were the flawless sections. The students could re-translate their mistakes for several times to be understood their mistakes and compensated them. Some samples of compared translations in different modes of wrongful and low objection were brought in the figures above.

4. Results

Measuring the mean, standard deviation and normality of distribution for the scores are the information which were necessary for deciding what variables could be included with confidence in the primary analyses addressing the study's research questions. To start analyzing the questionnaire results, the researcher launched Smirnov- Kolmogorov to check the homogeneity of the participants of both groups. A normal distribution means two groups are significantly similar to one another. Since the results were satisfactory, the researcher compared means of both groups' test results, using t-test.

In order to measure the effects of the application of *electronic observer feedback* and *human observer feedback* on intermediate learners' *translation skills*, two paired samples t-test procedure were carried out to compare pretest and posttest language scores of the two groups. In order to see the difference between the *human observer feedback* and *electronic observer feedback* in terms of developing *translation skills*, an independent

samples t-test was carried out for comparing the pre- and posttest of the two groups.

Since measuring *translation Skill Test* of participants is to some extent subjective, two raters scored *translation Skill Test*s and the results taken from the two raters were compared through Pearson Correlation Coefficient. The *r value* for the two groups, before and after the treatment, is presented in Table 1.

Table 1

Correlation between the two raters, control and experimental groups, before and after treatment

Inter-rater	**Sig.**	**Pearson correlation**
Control, pretest	0.001	0.625**
Experimental, pretest	0.001	0.639**
Control, posttest	0.001	0.662**
Experimental, post test	0.001	0.936**

**. Correlation is significant at the 0.01 level (2-tailed).

As the correlation results indicate, the correlations have been significant between the two raters. This suggests that the given scores had been reliable scores which were given to students.

There was a need to check if mean scores have changed or not. Therefore, a paired t-test procedure was used to compare *translation skills* difference of experimental group before and after the receiving treatment. Descriptive statistics for the *translation skills* related to pretest and posttest among participants of experimental group is indicated in Table 2.

Table 2

Descriptive statistics for pretest and posttest translation skills scores (experimental group)

		Mean	**N**	**Std. Deviation**	**Std. Error Mean**
Translation skills	Pre-preliminary	14.3	28	.75864	.14088
	Post-preliminary	16.09	28	.63052	.11708

As the results of Table 2 indicate, *translation skills* were enhanced after the participants received electronic feedback. While *translation skills* in the experimental group' translation was around 14.3 before receiving electronic feedback, its size rose to 16.09 after participants received electronic feedback during intervention. The results of the t-test indicated that *translation skills* of experimental group participants after receiving electronic feedback were significantly enhanced. (p> 0.5; Sig. (2-tailed) = 0.01). More information is provided in Table 3.

Table 3

Paired samples t-test comparing translation skills before and after receiving electronic feedback

		Mean	**Std. Deviation**	**Std. Error Mean**	**95% Confidence Interval of the Diff.**				
					Lower	Upper	t	Df	Sig. (2-tailed)
Experimental	Translation skills	-.29207	.41850	.07771	-.45126	-.13288	-3.758	28	.001

As the t-test results indicate, the participants in the experimental group who received electronic feedback have significantly higher level of *translation skills* after receiving electronic feedback during intervention.

A paired samples t-test procedure was used to compare the pretest and post test results of *translation Skill Test* of the control group (*human observer feedback*). However, before comparing the results of pretest and post *translation Skill Test*, there was a need to check if mean scores of *translation skills* have changed or not. Descriptive statistics for participants' *translation skills* related to pretest and posttest comparison of the human observer (control) groups' *translation skills* is illustrated in Table 4.

Table 4

Descriptive statistics for pretest and posttest translation skills of control group (human observer)

Paired Samples Statistics

		Mean	**N**	**Std. Deviation**	**Std. Error Mean**
Translation	Pre-Human	14.63	29	.66579	.12363
	Post-Human	15.12	29	.76357	.14179

As the results of Table 4 indicate, the mean score for translation skills has changed after receiving *human observer feedback*. While their *translation skills* were around 14.63 before receiving *human observer feedback*, their *translation skills* were raised to 15.12 after intervention. To check the significance of *translation skills* difference which was resulted from *human observer feedback*, the mean *translation skills* scores from pretest and posttest were compared. The results of t-test indicated that students' *translation skills* have risen after receiving *human observer feedback*.

This means that the increase of *translation skills* among participants who received *human observer feedback* has been statistically significant (p> 0.5; Sig. (2-tailed) = 0.027). More information is provided in Table 5.

Table 5

Paired samples t-test comparing translation skills before and after receiving human observer feedback

Human-observer	**Mean**	**Std. Deviation**	**Std. Error Mean**	**95% Confidence Interval of the Difference**				
				Lower	Upper	T	Df	Sig. (2-tailed)
Pretest-posttest	.487	1.129	.20968	.05843	.91744	2.32	28	.027

As the t-test result indicates the participants in *human observer feedback* group have significantly different *translation skills* after receiving *human observer feedback*. The findings from the last two research questions indicate that both *human observer feedback* and electronic feedback enhance students' *translation skills*. Two independent t-test procedures were used to compare the translation test score between the two groups, before and after the intervention. Descriptive statistics for *translation skills* related to pretest comparison of the two groups' *translation Skill Test* scores is shown in Table 6.

Table 6

Descriptive statistics for pretest translation skills of the two groups

Group Statistics

	grouping	**Mean**	**N**	**Std. Deviation**	**Std. Error Mean**
Pretest Difference	Electronic	14.30	28	.758	.140
	Human	14.63	29	.665	.123

As Table 6 indicates, the two groups had different *translation skills* before receiving *human observer feedback* and electronic feedback. While

mean score of electronic feedback (experimental) group in pretest equaled 14.3, that of *human observer feedback* (control) group equaled 14.63. Though, mean translation score among the two groups during pretest was different before intervention, it needs to be statistically investigated if this difference is significant or not. To check the significance of *translation skills* score difference between the two groups, the means were compared (Table 7). The results of t-test indicated that, as we expected, the *translation Skill Test* score difference between the two groups, in the pretest was not significant (p> 0.5; Sig. = .11).

Table 7

Pretest comparison of the two groups' translation skills scores

Pretest	**F**	**Sig.**	**T**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**
Equal variances assumed	.582	.449	-1.594	56	.116	-.29207	.18318
Equal variances not assumed			-1.594	54.18	.117	-.29207	.18318

As the t-test result indicates the two groups were not significantly different before the treatment so that posttest *translation skills* score difference of both groups is indicated in Table 8. However, as the post test

result indicated, students' *Translation skills* score had changed after the treatment. While participants' pre-intervention means *translation skills'* score equaled 14.3 and 14.63 in electronic feedback group and *human observer feedback* groups, after intervention their mean *translation skills* score was raised to 16.09 and 15.12, respectively. The results are reported up to 2 or 3 decimals of accuracy.

Table 8

Descriptive statistics for posttest translation skills scores of the two groups

	Grouping	**Mean**	**N**	**Std. Deviation**	**Std. Error Mean**
	Electronic	16.09	28	.630	.117
	Human	15.12	29	.763	.141

Independent samples t-test was carried out to check if this difference is statistically significant or not. T-test results which for $p< 0.05$ equaled 0.012, have been shown in Table 9. The results of post-intervention indicate that electronic feedback raises participants' *translation skills* much more than traditional instruction does. In other word, based on the results,

participants' translation test scores among electronic feedback group was significantly higher (t-test for $p< 0.05$ is significant).

Table 9

Posttest comparison of the two groups' translation skills scores

		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Posttest	Equal variances assumed	.001	.972	2.594	56	.012	.48793	.18812
Difference	Equal variances not assumed			2.594	54.980	.012	.48793	.18812

5. Discussion

As shown by the analysis of the data, the important role of both electronic feedback and *human observer feedback* in developing translation skills among intermediate level participants can be argued and discussed. Moreover, the findings reveal that electronic feedback is more useful that *human observer feedback* in terms of developing translation skills. This means that utilizing electronic feedback is preferable to *human observer feedback* at least when the focus is on developing translation skills.

This point might be justified by Waddington (2010) who claim that effectiveness of human feedback depends on the follow-up of the observer and having a stable condition to present the sufficient descriptions. They believe that students' emotions and tension should not be disregarded (Waddington, 2010). Knowing that conducting this study in a different context, on different participants, different observers, and different context might reveal contrasting findings, it is still highly probable that machine-generated feedback can be positive regarded as lowering the tensions. On the other hand, it is possible that the observer personality traits turn out to produce more or less effective feedback.

The findings of the present study confirm the findings in Omar, Zaidan, and Callison Burch's (2010) study which reported that using a translation software can take the translation quality to near professional levels. They also added that the utilization of technology (for instance working with computers adds to a resourceful classroom, in selecting texts for translation with a certain degree, or regarding text type in a translation memory data bank) has a constructive effect on translation. Similarly, Koehn's (2003) findings support the use of electronic translation feedback.

Despite all this concurrence, heed should be observed, since not all of the studies previously are in line with the findings of this study.

Callison-Burch (2010) for instance reported that the manual (*human observer feedback*) evaluation of translation quality was not as expensive or as time consuming as generally thought (Chris Callison-Burch, 2010). Moreover, Callison-Burch (2010) add that human or manual translation evaluation has been indicated to enhance students' future translations quality.

6. Conclusion

To fulfill the aim of the current stud which is elaborating on role of *electronic observer feedback* in students' *translation skills*, the mean scores of pretest and post test results were compared in a translation test revealing that both *electronic observer feedback* and *human observer feedback* enhance participants' *translation skills*. However, further analysis of the data indicated that *electronic observer feedback* resulted in statistically higher *translation skills*, in comparison to *human observer feedback*. In other words, the change in electronic observer feedback was more significant than the change in *human observer feedback*. However, literature was not consistent in the role and the type of feedback in developing *translation skills* because there are many variables that might have interactions in achieving the results. For example, human observer's personality type might play an indispensable part in the efficiency of *human observer feedback*. Moreover, the amount of time allocated to the two types of feedback might have a decisive role. While some previous studies were in line with the

findings of this study (e.g. Koehn's, 2003), other studies had reported contradictory findings (e.g. Callison & Burch, 2010).

The researcher aimed at examining the function of two types of feedback on translation groups. The groups were designated as experimental groups scheduled with specific pedagogical plan. An intelligent system was designed to provide basic feedback and supervised to improve the students' *translation skill*. The obtained results were reported after the three consecutive pedagogical sessions and practice of different types of texts to be translated for four hours and to be recorded on the system along with their correct Persian translation. The students of each class were tested by a given translation text as a posttest. There is no claim in generalizability of results; however, they are consistent where possible. For example, student groups were chosen randomly and they were majoring in English language. Although the obtained scores of the students were considered individually, the obtained results of the groups could be compared together and the impact of electronic and *human observer feedback* was demonstrated.

In sum, the findings of this study suggest that receiving *electronic observer feedback* changes *Intermediate* learners' *translation skills* when their age is between 18 and 28. It was revealed that the factors relevant to the participants including context (general *translation skills*), proficiency, and age interact with the degree to which *electronic observer feedback* and

human observer feedback influence *translation skills*. For instance, advanced and elementary participants might be different from intermediate participants in terms of their ability to produce language related materials like portfolios and texts. Thus, similar effect of the participant groups was also attested in literature confirmed by Thomas (2002).

In all, the main purpose of this study was fulfilled, as the investigation revealed the possible usefulness of *electronic observer feedback* and *human observer feedback* on intermediate learners' *translation skills*. As it was shown, *electronic observer feedback* group outperformed the other group (*human observer feedback*) on *translation skills*. The results of this study are indicative of the fact that, under certain conditions *electronic observer feedback* appears to assist the development of *translation skills* even more than other types of feedback which is purely manual. This paves the ground for more explicit implementation and practice of *electronic observer feedback* in order to contribute and enhance *translation skills*.

A few pedagogical implications are as follows:

1. The curriculum developers and teachers of translation can incorporate *electronic observer feedback* as a task into developed materials and use it in classroom context in order to develop language students' *translation skills*.
2. Since *electronic observer feedback* is a process which lends itself well to the concept of autonomous learning, it is mostly suggested to those who like to

enhance their *translation skills* to practice through self-study approach in learning *translation skills*. This is suggested because the findings of the present study showed that *translation skills* can be enhanced in the absence of human observer.

3. The incorporation of technology into translation courses are highly recommended, as it is in other fields, because of its merits including its time saving nature as well as its effectiveness in enhancing *translation skills*.

The results and findings of the present study offer an untapped potential that may lead to further lines of research and inquiry into the role of technology in translation.

- First, Students' change of *translation skills* as a result of *electronic observer feedback* and *human observer feedback* might be due to quality factors. For instance, the quantity and the source of feedback (e.g. peer feedback) may result in findings which might differ from the findings of this study or might complement the findings.
- Second, the effects of long-term electronic and *human observer feedback* on learners' translation skill need more attention and exploration to inequality of short-term results and longer-term results.
- Third, using a set of qualitative methods in investigation of how *electronic observer feedback* results in higher *translation skills* can further this area of

research since the question that how learners perceive such feedbacks may not solely be revealed through quantitative methods.

- Fourth, for studies in the field of language teaching research, it is interesting to investigate learners' linguistic skill, perception, and attitude, in relation to and assuming the interactions amongst the components of language (e.g. grammar, vocabulary, reading, writing, etc.) and *electronic observer feedback*.
- Finally, the investigation of the effects of *electronic observer feedback* and *human observer feedback* on *translation skills* across gender, proficiency level, and age is still the area of interest for many researchers.

7. References

Archer, J., Cantrell, S., Holtzman, S. L., Joe, J. N., Tocci, C. M., & Wood, J. (2015), *Seeing it clearly: Improving observer training for better feedback and better teaching*. Retrieved from http://collegeready.gatesfoundation.org/teacher-supports/teacher development/measuringeffective-teaching/met-seeing-it-clearly-improving-observer training-for-better-feedback-andbetter-teaching.

Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., Lagarda, A., Ney, H., Tomas, J., Vidal, E., and Vilar, J.-M. (2009).

Statistical approaches to computer-assisted translation. *Computational Linguistics*, 35(1):3–28.

Burt, K.M. & C. Kiparsky. (1974). *The Gooficon*, Mass: Newbury House Publishers.

Dean, C. B., Ross Hubbell, E., Pitler, H., & Stone, B. (2012). *Classroom Instruction That Works*. Publisher: ASCD; 2nd edition (December 13, 2013). Published: Alexandria: ASCD, 2012. ISBN-13: 978-1416613626, ISBN-10: 1416613625

García-Izquierdo, I. & Conde, T. (2012). *Investigating Specialized Translators: Corpus and Documentary sources.* Universitat Jaume I & Universidad del País Vasco / Euskal Herriko Unibertsitatea (Spain). INVESTIGATING SPECIALIZED TRANSLATORS. Ibérica, núm. 23, 2012, pp. 131-156, ISSN 1139-7241

Goff-Kfouri, C. A. (2005). Testing and evaluation in the translation classroom. *Translation Journal, 9*(2), 75-99.

Fatahipour, M. and Tabatabaei, S. (2015) Excerpt Translation: Conditions and Requirements *Journal of Literary and Rhetorical Studies 3* (412), 52-65

Hart, S. 2006. "NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50* (9), 904 - 908. doi: 10.1177/154193120605000909.

Hart, S., and Staveland, L. 1988. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." *Advances in Psychology 52*, 139-183. doi: 10.1016/S0166-4115(08)62386-9.

Hassani, M., Goodarzi, M. H., & Rafe, V. (2012). Educational advisor system implemented by web-based fuzzy expert systems. *Journal of Software Engineering and Applications, 5*(07), 500. *Vol.5 No.7, DOI: 10.4236/jsea.2012.57058*

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research, Vol. 77, No. 1, pp. 81–112. DOI: 10.3102/003465430298487*

Heaton, J.B. (1990). *Classroom testing*. Longman, New York.

Hutchins, w. J. & Somers, H. L. (1992). An Introduction to Machine Translation.

Kearns, J. (2006). Curriculum Renewal in Translator Training: Vocational challenges in academic environments with reference to needs and situation analysis and skills transferability from the contemporary experience of Polish translator training culture, doctoral thesis, author John Keams, publisher: Dublin City University, 2006. School of Applied Language and Intercultural Studies, oai:doras.dcu.ie: 17625.

Koehn, P., Och, F. J., and Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology* - Volume *1*, NAACL '03, pages 48–54, Stroudsburg, PA, USA. Association for Computational Linguistics.

Koehn, P. (2004). Statistical Significance Tests for Machine Translation Evaluation. In EMNLP (pp. 388-395).

Krish, P. (2006). The power of feedback in an online learning environment. 3L; Language, Linguistics and Literature, *The Southeast Asian Journal of English Language Studies*, 12, 95-106.

Mitchell-Schuitevoerder, R., & Elizabeth, H. (2014). *A Project-Based Syllabus Design Innovative Pedagogy in Translation Studies* (Doctoral dissertation, Durham University) http://etheses.dur.ac.uk/10830/.

Newton, J. (1992). Computers in Translation: A Practical Appraisal.

Nishida, F. & Takamatsu, S. (1990). Automated Procedures for the Improvement of a Machine Translation System by Feedback from Postediting. DOI: 10.1007/BF00389819.

Shirbagi, N. (2007). Feedback in formative evaluation and its effects on a sample of Iranian primary students' achievement in science. *Pedagogika, 88*, 99–105.

Tennent, M. (2005). *Pedagogies for translation and interpreting: Training for the New Millennium* Amsterdam and Philadelphia: John Benjamins.

Thomas, W. (2002). A national study of school effectiveness for language minority students' long-term academic achievement. Santa Cruz: Center for Research on Education, Diversity an Excellence, University of California, Santa Cruz.

Ubersax J. (2016). *Kappa Coefficients*. Statistical Methods for Diagnostic Agreement. Available at http://john-uebersax.com/stat/kappa2.htm. Accessed July 16, 2016.

Waddington, C. (2010). Different Methods of Evaluating Student Translations: The Question of Validity (p. 313). Journal des traducteurs / *Meta: Translators' Journal, vol. 46,* n° 2, 2001, p. 311-325.

Zaidan, O. F. & Callison-Burch, C. (2011). Crowdsourcing Translation: Professional Quality from Non-Professionals. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 1220–1229, Portland, Oregon, June 19-24, 2011. ©2011 Association for Computational Linguistics.

Zieky, M. & Perie, M. (2006). A Primer on Setting Cut Scores on Tests of Educational Achievement. , Princeton, NJ: Educational Testing Service., the material on methods for setting cut scores has been excerpted from Passing Scores: A Manual for Setting Standards of Performance on Educational and

Occupational Tests by Samuel A. Livingston & Michael J. Zieky, originally published by ETS in 1982 and reprinted in 2004.

Notes on Contributors:

Omid Tabatabaei is an Associate Professor of English Language Teaching in the Department of English Language at Najafabad Islamic Azad University, Najafabad, Iran. He is presently the head of English department. He has published a number of articles in domestic and international journals and presented in various conferences. Moreover, He has authored various books in relation to the field of language teaching and assessment. His areas of interest are language assessment, teaching theories and skills, psycholinguistics, and research methodology.

Majid Fatahipour (PhD from UWE Bristol) is currently an Assistant Professor of Applied Linguistics in the Department of English at Parand Branch of the IAU, Parand, Iran. With an interest in always developing wider academic experience, he assumed variety of roles and did presentations in international conferences. In addition to a number of articles in peer-reviewed journals, he is constantly involved with some practical experience in translations besides academia, from certified for judiciary to multilingual websites such as Target. His research interest is

now keened on technology and legal translation in the area of translation studies.

Maryam Mohammadi Sarab is now a PhD student in TEFL at Najafabad University, Iran. She received her B.A. degree in Computer Software Technology Engineering form Gonbad Kavous Uiversity in 2011 and her MA in English language translation from Parand University in 2016. Since then, she has been researching English at different branches of interdisciplinary including English, teaching, and software. Her research interests are to combine the different methods together for presenting novel strategies in the fields of learning, teaching, and child L2 acquisition.