ISSN 1798-4769
Journal of Language Teaching and Research, Vol. 6, No. 5, pp. 1063-1074, September 2015
DOI: http://dx.doi.org/10.17507/jltr.0605.20

The Impact of Dynamic Corrective Feedback in Developing Speaking Ability of Iranian Intermediate EFL Learners

Esmaeil Bagheridoust
Islamic Azad University, South Tehran Branch, Iran

Ali Mohammadi Kotlar
Islamic Azad University, South Tehran Branch, Iran

Abstract—**Speaking ability is one of the major skills of learning English, and during the process of learning and speaking committing errors is inevitable, but it should be treated properly and in a systematic way. The purpose of the present study was to see to what extent does the implication of the principles of a newly developed strategy of error correction_ dynamic corrective feedback (DCF) affect the overall oral development of Iranian intermediate EFL learners, and to what extent does it affect its major components, specifically accuracy, fluency, and complexity. Therefore in study quasi-experimental study that used a pretest, treatment, post-test, there was one control group (N= 26), and one experimental group (N=28) with both male and female young adult learners. The SPSS software was used to compute and analyze the amount of the treatments impact, and the independent t-test and multivariate ANOVA (MANOVA) built up the core statistical analyses of the study. The findings of this study indicated that DCF has been mostly successful, since it benefits from strong theoretical principles derived from DWCF in the area of writing error correction. The retrieved principles imply that provided corrective feedback should be meaningful, timely, constant, and manageable and it must reflect the individual learners most immediate needs based on what they produce (Evans et al. , 2010).**

Index Terms—**dynamic corrective feedback (DCF), corrective feedback (CF), DWCF, speaking ability, accuracy, fluency, complexity**

I. Introduction

Making errors when learning a new foreign/second language is a natural part of it and it is inevitable. The errors can be of various kinds, for example in pronunciation and syntax, or word choice errors. Feedback is needed to avoid fossilization. As errors cannot be self-corrected, teachers' reaction toward error in the form of corrective feedback is essential. If errors are not corrected, various aspects of a learner's inter-language may become fossilized and he/she will not be able to "progress to fully mature linguistic competence" (Tomasello & Herron, 1988, p. 237).

Almost all EFL teachers agree on the importance of provision of corrective feedback, but there might be disagreements on determining the type of corrective feedback that should be provided at different levels of proficiency. According to Lyster and Ranta (1997) teachers react to learners' errors in one or more of six different ways; i.e. explicit correction, recast, clarification request, metalinguistic feedback, elicitation, and repetition. Translation and multiple feedback are two other spoken feedback types which are added to Lyster and Ranta's (1997) list by other researchers.

The efficacy of the corrective feedback in second language acquisition has been discussed and challenged for long, through a great number of theoretical and empirical studies, but there is still a need for further researches. Russell and Spada (2006) also confirm that "much more work needs to be done" (p. 156), and reaffirm that "to stablish clear patterns across studies" (P. 156), "similar variables in a consistent manner" need to be investigated.

Regarding the efficacy of one type of corrective feedback over the others or even inefficacy of it Guénette (2007) recommends that we should investigate on comparable groups over time. Moreover she encourages the effort for designing proper EC strategies based on "the students' proficiency levels and developmental readiness" (p. 51), at the same time she insists on paying attention to external variables i.e. classroom context and student differences, when trying to develop and design such new types of corrective feedback (Guénette 2007).

Considering the importance of EC, and the actual practices of EC in today's classes especially in the area of speaking, it seems that the conventional methods are mostly un-focused, un-systematic, and are done with less care about its effect on learners' speaking development. At the same time, providing corrective feedback for the speaking as a productive skill seems to be difficult to deal with. From the view point that any oral production disappear soon after it is being produced, providing proper and meaningful corrective feedback due to the lack of time, lack of control over it, and lack of consistency of the provided feedback in any area of difficulty, short after an error is committed it becomes very difficult if not impossible. There is also no guaranty that learners can learn properly from the provided feedback, since when it is done it is likely to be forgotten or ignored by the learners in the future.

Having all these findings, debates, and concerns in mind, this study tried to follow the issue of error correction in a more specific area of learning, speaking ability, to have more focus on it and deal with it from the CF's perspective. Therefore, the main purpose of the study was to see, to what extent does the implication of the principles of dynamic corrective feedback affect the overall oral development of Iranian intermediate EFL learners, and to what extent does it affect its major components, specifically complexity, accuracy, and fluency. To address the stated problems, and the purpose of the study, the following research question was raised:

Q1: Does dynamic corrective feedback affect the overall oral proficiency of Iranian intermediate EFL learners?

Along with the main question, the following sub-questions were also raised.

Q2: Does dynamic corrective feedback affect the accuracy development of Iranian intermediate EFL learners' speaking?

Q3: Does dynamic corrective feedback affect the fluency development of Iranian intermediate EFL learners' speaking?

Q4: Does dynamic corrective feedback affect the complexity development of Iranian intermediate EFL learners' speaking?

Accordingly, to probe the above research questions, four relevant null hypotheses were respectively made.

II. Theoretical Backgrounds

A. Complexity, Accuracy, and Fluency (CAF)

As many language learners try to achieve native-like speaking ability, Skehan (1996) describes that this achievement is in connection with developing three major areas of performance which are complexity, accuracy, and fluency. Relevantly Skehan (1992, 1996 as was stated in in Skehan & Foster, 1999) describes these three competitive traits that try to use the attentional resources for themselves, as follows:

Fluency: the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems.

Accuracy: the ability to avoid error in performance, possibly reflecting higher levels of control in the language, as well as a conservative orientation, that is, avoidance of challenging structures that might provoke error complexity/range the capacity to use more advanced language, with the possibility that such language may not be controlled so effectively. This may also involve a greater willingness to take risks, and use fewer controlled language subsystems. This area is also taken to correlate with a greater likelihood of restructuring, that is, change and development in the interlanguage system. (P. 4)

It's worth mentioning that along with the above definition, these traits of complexity, accuracy, and fluency, have been operationally defined and measured differently in different studies, although the core idea is similar. Accordingly, they have been measured through different language domains differently, with the use of various tools, scales, and frameworks.

B. DWCF

Evans and his colleagues (Evans, Hartshorn, McCollum, & Wolfersberger, 2010) designed a method of error correction basically for their higher level students to enable them to have higher accuracy in their writings. Evans (Evans et al., 2010) used dynamic written corrective feedback (DWCF) as a new trend of error correction. They believed that contextual variables played a crucial role when designing the new method. They (Evans et al., 2010) also considered the issue that 'how contextual factors affect our learning, teaching and research' (Ferris, 2004; Guénette, 2007) as most of the newer studies do.

DWCF is based on the two central principal characteristics as mentioned by Evans et al., (2010) "Feedback reflects what the individual learner needs most as demonstrated by what the learner produces; and tasks and feedback are manageable, meaningful, timely, and constant for both the learner and teacher" (P. 452). In other words, DWCF is meaningful when learners understand the provided feedback and know how they are expected to utilize it. DWCF is timely when the learners receive at immediate proper time intervals. Feedback is constant when it is provided to the learners at regular, frequent intervals over an extended period of weeks or months (Evans, Hartshorn, & Strong-Krause, 2011, p. 232). So DeKeyser's concept of skill acquisition theory (2001, 2007, as stated in Evans et al., 2011) affirms that, in order for the learners to get a meaningful level of automatic L2 production, writing practices, application of the feedback need to be timely and constant. Feedback is called manageable when the teachers and students have a list of errors to provide quality feedback and students have enough time to process and apply the feedback they receive (Hartshorn et al, 2010).

C. DCF

The main idea of DCF roots back in successful results of a newly developed method of error correction in the area of writing_ dynamic written corrective feedback (DWCF) exemplarily in the works of Evans et al., 2010; Evans et al., 2011; Hartshorn et al., 2010.

Although (as mentioned earlier) these studies focused on the implementation of the Dynamic Corrective Feedback in the area of writing (DWCF), but the current study focuses on the DCF on oral production, hence the main principles and premises of dynamicity in error correction is derived from the aforesaid studies. Since it was not enough and applicable to apply all those principles and still keep it intact, just some minor modifications were applied to those premises to enable the implementation of DWCF in the area of speaking, that is here termed as DCF.

DCF is neither just written, nor just oral but also it is an amalgamation of both. DCF implies its techniques and strategies in a way that makes the issue of providing corrective feedback for the oral production more flexible and systematic. DCF uses the two central principal characteristics of the DWCF as mentioned by Evans et al., (2010, P.452) “Feedback reflects what the individual learner needs most as demonstrated by what the learner produces; and tasks and feedback are manageable, meaningful, timely, and constant for both the learner and teacher”.

In this study, the first key characteristics of the DWCF as mentioned above, is still the same and it is kept intact, but some modifications are applied around the second key principle. Therefore, the DCF in this study keeps the premise of being manageable, meaningful, timely, and constant (Evans et al., 2010), and adds a fifth premise that DCF should provide better alternatives in order to enrich the students production with more to the point, and varied technical alternatives.

Consequently, in term of theoretical principles, the two types of dynamic corrective feedback_ DWCF in previous studies (Evans et al., 2010; Evans et al., 2011; Hartshorn et al., 2010), and DCF (the present study) are quite similar, but the differences are much more tangible in term of their implication and actual practices. Since the current study tried to provide DCF on the students’ oral production, and from the viewpoint that oral production is disappeared and forgotten soon after it is being produced, the theoretical premises and principles of the DCF that were explained earlier seem more crucial.

Accordingly in this study, it is tried to make the corrective feedback more dynamic in its actual practice, by the use of tally sheet and error list, and the provided columns in them, along with the systematic and principled based techniques of providing corrective, and also with the regular revision of the errors, plus some other techniques of cooperation. Therefore, the premises are being timely, constant, manageable, and meaningful (Evans et al., 2010), plus the very last one, that asserts that DCF should provide alternatives (i.e. grammatical, lexical) to become more meaningful. Moreover, the students in DCF receive the feedback in different forms and provided differently such as self-correction, peer-correction, or teacher correction, but we should bear in mind that more important than who corrects the errors, is how to treat them; in order to take advantage of it and avoid committing them in similar occasions.

III. Research Methodology

A. Participants

Before the homogenization, there were 82 available students in the target English language institute in Tehran, who had already passed their previous course-book, Top Notch 3 (by Saslow & Ascher, 2006), and enrolled for the next semester_ Summit 1 (by Saslow & Ascher, 2006). Then 52 students whose test score fell between one standard deviation (SD=5.62) above and below the mean (M=54.36) were selected. They were both men and female young adult EFL learners ranging in age from 16-29, with the average age of 19. The number of males exceeded the females; consequently, after the random placement of the students in control and experimental groups, there were 30 males and 24 females in all. In other words, there were six classes of control and experimental with about 8-10 students in each.

Two experienced and trained examiners scored and rated the papers for the parts that could not be objectively scored. Therefore, during the PET test, the two raters scored the writing and the speaking sections, using PET’s analytic scoring rubric. Their inter-rater consistency were then calculated.

One of the authors of the study was the only teacher who participated, and taught in both groups of control and experimental. Although it was difficult to have all classes with the same teacher during six days of the week, but since most of the students had experienced his classes at least for one semester in lower levels, with an acceptable level of satisfaction (as the feedback the manager of the institute received indicated it), there was a great hope to have their full cooperation.

B. Instruments

1. Preliminary English Test (PET) and the coursebooks:

A version of Preliminary English Test (2003) was used to homogenize the students and to make sure the students were all in the same level of proficiency. The speaking section of the test was also used for the pretest in the next step. Parallel to the pretest, another version of PET’s speaking section was used in the posttest for the sake of posttest measurement and analysis. The PET’s scoring rubric was used for the scoring of the PET’s relevant sections, during the homogenization phase, and a digital voice recorder was used during the pretest and the posttest to audio record the participants’ voices, for further measurement in a later time.

2. Error list and tally sheet:

An error list, which was specifically designed for this research, was a piece of an A4 paper given to the experimental group in the beginning of the treatment section of the class, with a table drawn on it. The table contains some columns, including a list of *committed errors*, the *corrected form* of the error, and *alternatives*. The error list is a kind of draft that, every student initially, takes notes in it and completes the parts individually while they are having conversations, and then, during the intervals, and also at the end of the class, all members cooperatively and collaboratively complete the parts.

A tally sheet which is a more complete form of the error list, has some more columns such as the error type, and the general category of the error types. An error log is a column that he who has committed the error checks it for him-self to keep track of his/her repeated errors and its number until that error is totally removed. Similar to the error list, the tally sheet is kept cooperatively in the class, but their difference is that the tally sheet is like a final draft. In other words, the tally sheet is more complete, logs all the errors of oneself and others which does not belong to only one session, but to the whole semester.

C. Procedure

1. Homogenization:

In this study, first a version of PET test was administered to homogenize 76 available students who had already passed their previous course-book Top Notch 3, and enrolled for the subsequent semester, which was Summit 1 in that institute.

It is worth mentioning that due to the difficulty and limitation for the administration of the speaking section to all the students, the process of homogenization took place in two phases. First, all sections of the PET test except for the speaking section were administered to all the students before the beginning of the semester. Then 58 students who had met the criteria of getting a score that falls between one standard deviation (SD=5.62) below and above the mean (M=54.36) were randomly placed in two groups of control and experimental. It was only after the beginning of the semester that the speaking section was administered to both control and experimental groups, but this time their full scores of all sections of the test, including the speaking scores were calculated for the second homogenization phase. As a result, four more students who were outliers due to their really strong or weak performances in the speaking test, and their scores was out of the range of one SD (5.62) above or below the mean (M=73.48, for the full PET score) were omitted. Therefore in the study phase we had 54 students in all, that were placed in two groups of control (N= 26) and experimental (N= 28) with the same teacher teaching in all six classes, and with 8-10 students in each.

For the rating purposes, two experienced teachers corrected and rated the papers, and their inter rater consistency was calculated that was at a good level of agreement (see Table 18).

2. *Pretest:*

During the administration of the PET test, as it was supposed to have the speaking section of the test performances also used for the pretest, all the participants' voices in this section were audio-recorded for further analyses and measurements regarding the complexity, accuracy, and fluency, along with their overall oral proficiency through the whole speaking test.

3. The study context:

Except for test and homogenization sessions, the study lasted for 14 sessions for each class, and they were held for two sessions of 90 minutes per week, and the treatment groups were all placed during six working days of the week, parallel to the timetable of the institute for their regular classes. There were only two study classes held in each day_ one experimental and one control class, therefore the only teacher would just go to one experimental and one control group each day. As the main purpose of the study in this phase was to intervene in the way of speaking practices and applying the innovative method of DCF on the experimental group, to compare its results with the results of the control group using conventional methods, the last 35 minutes of each session in both groups was devoted to the treatment practices. The other 55 minutes in the beginning of each class was to cover the book, present the materials of each unit, follow the normal schedule of each class, and above all to prepare and keep the speaking topics for more practice and extended free talks, for the treatment section of that session.

4. Instructional methods:

It is worth mentioning that in this study all the measurement of the errors and the focus of error correction of the speaking, was based on the grammatical and lexical errors, therefore, other types of speaking deficiencies such as pronunciation, or intonation were not focused, except for the overall speaking scores.

Since this paper sought to apply the techniques of a new approach regarding error correction in speaking (DCF), it was tried to have the topics and the speaking practices of the course-book, kept for the last 35 minutes of each session. Although, the basis of the speaking parts were taken from their course-books, but we did not limit our-selves to it and sometimes we went far beyond the initial topics. When necessary, some of the parallel units of the "Let's talk 3" speaking book, were also used as a supplementary book to support and enrich the topics of the main course-book, in order to let the conversation go and make it more interesting.

The instructional method in the control group was similar to that of the experimental group. The only differences were in the way of error correction, or the way errors were treated. In both groups the students errors were corrected when they were necessary to be corrected, but the way of correction, the person who sought for corrected the errors, and also the amount of focus and attention on the errors (even after the correction has happened), totally differed.

In the control group, the errors were just corrected by the teacher or sometimes by peers, and no further attention or focus was given to the errors unless the students asked for, but in the experimental group, the errors were sought, welcomed and treated in a systematic way.

A typical error treatment in the experimental group is as follows: In the beginning, the students were given an error list with some provided columns to jot the errors down while they were having conversation. Students were taught to participate in class or group conversations, but at the same time actively seek for the committed errors and put them down, while the others were talking. For the sake of fluency, they were not allowed to interrupt the conversation unless the errors were global, or needed to be helped and corrected on the spot. Every 2-4 minutes of time intervals (depending on the conversation and the number of committed errors), they were made to cooperatively recall the errors, correct and write the true form of it in the error list, and to think of any alternatives or options for saying the same idea in some other words (or with an alternative grammatical/lexical form when possible).

The teacher, who was following the same steps and was working cooperatively with the students played some more important roles too. He managed, handled, facilitated, and monitored the conversations and discussions, but he only corrected the errors that were not noticed or not corrected by peers, or the ones that needed more help. Sometimes he signaled the students for poor performances that needed a better grammatical or vocabulary alternatives, by slowly knocking his pen or his index finger on the desk; and also by shaking his pen or finger slowly to the left and right as a NO sign for a committed error that needed immediate help from the peers. He would also try to help the students with more alternatives through grammar or vocabulary, and to introduce the contingencies and possible grammatical mistakes mostly in review sections.

It was no matter who had committed the errors, but they were supposed to not to repeat it, and mark the errors of their own to track it and consequently, have more attention on it.

At the end of the class the error list (first draft) that was completed and revised in that session, and was limited to that session's error, was kept to be reviewed and transmitted to a more complete form of it (tally sheet), in the beginning of the next speaking practice before the new error correction cycle be started.

Since it is not easy to have control on everything and think of all possibilities on the spot, the further reviews in the beginning of the following sessions' speaking played a crucial role. The new given list (Tally sheet) was more complete than the first draft list (Error list). The students were asked to transmit and rewrite the content of the former list in the latter one, and to mark the error type, plus keeping track of one's own recurred errors or error types (logging the errors), before it is totally removed.

5. Posttest:

At the end of the term and after 14 sessions of the study, another version of PET's speaking test with a different topic administered to both groups. Similar to the pretest, the voices during the posttest were audio recorded and then transcribed for the analytic scoring and measurement of the complexity, fluency, and accuracy traits. Besides, the overall oral proficiency was scored by the same raters in the test session, and their inter-rater reliability was calculated afterwards.

D. Measuring the Speaking Components

For the measurement of the speaking components namely, accuracy, fluency, and complexity, that is a threefold procedure, first, the voices were transcribed and then, a different scale for each individual trait was used to measure the amount of the complexity, accuracy and fluency separate from one another. To measure the complexity, which refers to the ability to use a more advanced language, following Foster and Skehan (1996), the proportion of clauses to C-units were calculated. A C-unit as stated by Foster and Skehan, (1999) is defined as "a simple clause, or an independent subclausal unit, together with subordinated clauses associated with them" (p. 106). It is measured by dividing the total number of clauses by the total number of c-units, and since every C-unit has at least one clause, the minimum score for it is 1:00 (Skehan & Foster, 1999). A clause is respectively considered as "either a simple independent finite clause or a dependent finite or non-finite clause" (Foster & Skehan, 1999, P. 228).

For the measurement of the accuracy, Foster and Skehan's (1996) formula was used, therefore the percent of the number of error-free clauses by the total number of clauses in each transcription was calculated. To do this first the number of clauses and then the number of error-free clauses (the clauses with no errors) were counted and calculated. For the sake of fluency calculation Mochizuki and Ortega's trend (2008) was followed, therefore the mean number of words per minute was the base of calculation of the fluency measurement.

For the analysis of the acquired data, the SPSS software was used to compute and analyze the amount of the treatments impact. The independent t-test and multivariate ANOVA (MANOVA) built up the core statistical analyses of the study, which is presented in the results section.

IV. RESULTS

A. PET General Language Proficiency Test

An independent t-test was run to compare the experimental and control groups' mean scores on PET general language proficiency test in order to prove that the two groups enjoyed the same level of general language proficiency prior to the main study. As displayed in Table 1 the mean scores for experimental and control groups on proficiency test were 54.36 and 54.35 respectively.

TABLE 1.
DESCRIPTIVE STATISTICS OF PET BY GROUPS

Group	N	Mean	Std. Deviation	Std. Error Mean
Experimental	28	54.36	2.556	.483
Control	26	54.35	2.399	.470

The results of the independent t-test (t (52) = .016, P > .05, r = .002 it represents a weak effect size) indicate that there was not any significant difference between experimental and control groups on proficiency test. Thus, it can be concluded that the two groups enjoyed the same level of general language proficiency prior to the main study.

TABLE 2.
INDEPENDENT T-TEST; PET BY GROUPS

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.005	.945	.016	52	.987	.011	.676	-1.345	1.367
Equal variances not assumed			.016	51.992	.987	.011	.674	-1.342	1.364

It should be noted that the assumption of homogeneity of variances was met (Levene’s F = .005, P > .05). That is why the first row of Table 2, i.e. “Equal variances assumed” was reported.

B. *Pretest of Speaking*

As displayed in Table 3 the mean scores for experimental and control groups on Pretest of Speaking were 19.07 and 19.19 respectively.

TABLE 3.
DESCRIPTIVE STATISTICS; PRETEST OF SPEAKING BY GROUPS

Group	N	Mean	Std. Deviation	Std. Error Mean
Experimental	28	19.07	1.152	.218
Control	26	19.19	1.266	.248

The results of the independent t-test (t (52) = .367, P > .05, r = .051 it represents a weak effect size) indicate that there was not any significant difference between experimental and control groups on Pretest of Speaking. Thus, it can be concluded that the two groups enjoyed the same speaking ability level prior to the main study.

TABLE 4.
INDEPENDENT T-TEST; PRETEST OF SPEAKING BY GROUPS

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.145	.705	.367	52	.715	.121	.329	-.539	.781
Equal variances not assumed			.366	50.564	.716	.121	.330	-.542	.784

It should be noted that the assumption of homogeneity of variances was met (Levene’s F = .145, P > .05). That is why the first row of Table 4, i.e. “Equal variances assumed” was reported.

C. *Pretests of Speaking Accuracy, Fluency and, Complexity*

A multivariate ANOVA (MANOVA) was run to probe any significant difference between the experimental and control groups on the pretests of speaking accuracy, fluency and complexity in order to prove that the two groups were also homogeneous in terms of their ability in using components of speaking. Before reporting the main results it should be noted that the assumption of homogeneity of variances – as tested through the Levene’s F-values – was met. As displayed in Table 5 the probabilities associated with the Levene’s F-values were all higher than .05. Thus, the assumption of homogeneity of variances was met.

TABLE 5.
HOMOGENEITY OF VARIANCES; PRETESTS OF COMPONENTS OF SPEAKING

	F	df1	df2	Sig.
Accuracy Pre	.177	1	52	.675
Fluency Pre	.001	1	52	.973
Complexity Pre	.008	1	52	.927

Based on the results displayed in Table 6 it can be concluded that there were not any significant differences between the experimental and control groups on the pretests of speaking accuracy, fluency and complexity (F (3, 50) = .29, P > .05, Partial η^2 = .018 it represents a weak effect size). Thus, it can be concluded that the two groups were homogeneous in terms of their ability in using the components of pretests of speaking.

TABLE 6.
MULTIVARIATE TESTS; PRETESTS OF COMPONENTS OF SPEAKING BY GROUPS

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.999	16349.627	3	50	.000	.999
	Wilks' Lambda	.001	16349.627	3	50	.000	.999
	Hotelling's Trace	980.978	16349.627	3	50	.000	.999
	Roy's Largest Root	980.978	16349.627	3	50	.000	.999
Group	Pillai's Trace	.018	.299	3	50	.826	.018
	Wilks' Lambda	.982	.299	3	50	.826	.018
	Hotelling's Trace	.018	.299	3	50	.826	.018
	Roy's Largest Root	.018	.299	3	50	.826	.018

Note. It should be mentioned that the SPSS produces four F-values. If the assumptions of normality and homogeneity of variances are met – as is the case in this study – the first F-value, i.e. Pillai's Trace should be reported. For a complete discussion of these statistics please refer to Filed (2009).

The F-value of .29 indicated that there were not any significant differences between the means of the two groups on the pretests of components of speaking as a total score. What follows is the comparison of the two groups on each test separately. Based on the results displayed in Table 7 and Table 8 it can be concluded that;

A: There was not any significant differences between the experimental and control groups on the pretest of speaking accuracy (F (1, 52) = .019, P > .05, Partial η^2 = .000 it represents a weak effect size). As displayed in Table 8 the means for the experimental and control groups on the pretest of speaking accuracy were 49.03 and 48.69.

TABLE 7.
TESTS OF BETWEEN-SUBJECTS EFFECTS; PRETESTS OF COMPONENTS OF SPEAKING BY GROUPS

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	Accuracy Pre	1.590	1	1.590	.019	.891	.000
	Fluency Pre	6.258	1	6.258	.058	.811	.001
	Complexity Pre	.000	1	.000	.001	.988	.000
Error	Accuracy Pre	4358.503	52	83.817			
	Fluency Pre	5657.390	52	108.796			
	Complexity Pre	.830	52	.016			
Total	Accuracy Pre	133329.000	54				
	Fluency Pre	250353.000	54				
	Complexity Pre	96.059	54				

B: There was not any significant differences between the experimental and control groups on the pretest of speaking fluency (F (1, 52) = .058, P > .05, Partial η^2 = .001 it represents a weak effect size). As displayed in Table 8 the means for the experimental and control groups on the pretest of speaking fluency were 67.64 and 66.96.

TABLE 8.
DESCRIPTIVE STATISTICS; PRETESTS OF COMPONENTS OF SPEAKING BY GROUPS

Dependent Variable	Group	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Accuracy Pre	Experimental	49.036	1.730	45.564	52.508
	Control	48.692	1.795	45.089	52.295
Fluency Pre	Experimental	67.643	1.971	63.687	71.598
	Control	66.962	2.046	62.857	71.066
Complexity Pre	Experimental	1.328	.024	1.280	1.376
	Control	1.328	.025	1.278	1.377

C: There was not any significant differences between the experimental and control groups on the pretest of speaking complexity (F (1, 52) = .001, P > .05, Partial η^2 = .000 it represents a weak effect size). As displayed in Table 8 the means for the experimental and control groups on the pretest of speaking accuracy were 1.32 and 1.32.

D. Examining Research Questions

1. Examining research question 1:

As displayed in Table 9 the mean scores for experimental and control groups on Posttest of Speaking were 20.61 and 19.73 respectively.

TABLE 9.
DESCRIPTIVE STATISTICS; POSTTEST OF SPEAKING BY GROUPS

Group	N	Mean	Std. Deviation	Std. Error Mean
Experimental	28	20.61	1.571	.297
Control	26	19.73	1.430	.280

The results of the independent t-test (t (52) = 2.13, P < .05, r = .28 it represents an almost moderate effect size) indicate that there was a significant difference between experimental and control groups on Posttest of Speaking. Thus, it can be concluded that the first null-hypothesis as dynamic corrective feedback does not affect the overall oral proficiency of Iranian intermediate EFL learners is rejected.

TABLE 10.
INDEPENDENT T-TEST; POSTTEST OF SPEAKING BY GROUPS

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.284	.596	2.138	52	.037	.876	.410	.054	1.699
Equal variances not assumed			2.146	51.982	.037	.876	.408	.057	1.696

It should be noted that the assumption of homogeneity of variances was met (Levene’s F = .284, P > .05). That is why the first row of Table 10, i.e. “Equal variances assumed” was reported.

2. Examining research questions 2, 3, and 4:

A multivariate ANOVA (MANOVA) was run to probe any significant difference between the experimental and control groups on the posttests of speaking accuracy, fluency and complexity in order to probe the effect of dynamic corrective feedback on the development of the Iranian intermediate EFL learners’ speaking accuracy, fluency and complexity. Before reporting the main results it should be noted that the assumption of homogeneity of variances – as tested through the Levene’s F-values– was met. As displayed in Table 11 the probabilities associated with the Levene’s F-values were all higher than .05. Thus, the assumption of homogeneity of variances was met.

TABLE 11.
HOMOGENEITY OF VARIANCES; POSTTESTS OF COMPONENTS OF SPEAKING

	F	df1	df2	Sig.
Accuracy Post	.790	1	52	.378
Fluency Post	.456	1	52	.503
Complexity Post	.410	1	52	.525

Based on the results displayed in Table 6 it can be concluded that there were not any significant differences between the experimental and control groups on the pretests of speaking accuracy, fluency and complexity (F (3, 50) = 236.59, P < .05, Partial η^2 = .934 it represents a large effect size). Thus, it can be concluded that there were significant differences between the experimental and control groups’ means on the components of the posttests of speaking.

TABLE 12.
MULTIVARIATE TESTS; POSTTESTS OF COMPONENTS OF SPEAKING BY GROUPS

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.993	2282.662	3	50	.000	.993
	Wilks' Lambda	.007	2282.662	3	50	.000	.993
	Hotelling's Trace	136.960	2282.662	3	50	.000	.993
	Roy's Largest Root	136.960	2282.662	3	50	.000	.993
Group	Pillai's Trace	.934	236.591	3	50	.000	.934
	Wilks' Lambda	.066	236.591	3	50	.000	.934
	Hotelling's Trace	14.195	236.591	3	50	.000	.934
	Roy's Largest Root	14.195	236.591	3	50	.000	.934

Note. It should be mentioned that the SPSS produces four F-values. If the assumptions of normality and homogeneity of variances are met – as is the case in this study – the first F-value, i.e. Pillai’s Trace should be reported. For a complete discussion of these statistics please refer to Filed (2009).

The F-value of 236.59 indicated that there were significant differences between the means of the two groups on the posttests of components of speaking as a total score. What follows is the comparison of the two groups on each test separately. Based on the results displayed in Table 13 and Table 14 it can be concluded that:

A: There was a significant differences between the experimental and control groups on the posttest of speaking accuracy (F (1, 52) = 61.941, P < .05, Partial η^2 = .544 it represents a large effect size). As displayed in Table 14 the experimental group (M = 73.10) outperformed the control group (M = 54.84) on the posttest of speaking accuracy. Thus, the second null-hypothesis as dynamic corrective feedback does not affect the accuracy development of Iranian intermediate EFL learners’ speaking is rejected.

TABLE 13.
TESTS OF BETWEEN-SUBJECTS EFFECTS; POSTTESTS OF COMPONENTS OF SPEAKING BY GROUPS

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Group	Accuracy Post	4495.585	1	4495.585	61.941	.000	.544
	Fluency Post	95.342	1	95.342	.853	.360	.016
	Complexity Post	.159	1	.159	7.457	.009	.125
Error	Accuracy Post	3774.063	52	72.578			
	Fluency Post	5812.973	52	111.788			
	Complexity Post	1.112	52	.021			
Total	Accuracy Post	231635.000	54				
	Fluency Post	265631.000	54				
	Complexity Post	113.880	54				

B: There was not any significant differences between the experimental and control groups on the posttest of speaking fluency (F (1, 52) = .853, P > .05, Partial η^2 = .016 it represents a weak effect size). As displayed in Table 14 the means for the experimental and control groups on the posttest of speaking fluency were 68.07 and 70.73. Thus, the third null-hypothesis as dynamic corrective feedback does not affect the fluency development of Iranian intermediate EFL learners' speaking is supported.

TABLE 14.
DESCRIPTIVE STATISTICS; POSTTESTS OF COMPONENTS OF SPEAKING BY GROUPS

Dependent Variable	Group	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Accuracy Post	Experimental	73.107	1.610	69.876	76.338
	Control	54.846	1.671	51.494	58.199
Fluency Post	Experimental	68.071	1.998	64.062	72.081
	Control	70.731	2.074	66.570	74.892
Complexity Post	Experimental	1.496	.028	1.441	1.552
	Control	1.388	.029	1.330	1.445

C: There was a significant differences between the experimental and control groups on the posttest of speaking complexity (F (1, 52) = 7.457, P < .05, Partial η^2 = .125 it represents an almost large effect size). As displayed in Table 14 the experimental group (M = 1.49) outperformed the control group (M = 1.38) on the posttest of speaking complexity. Thus, the fourth null-hypothesis as dynamic corrective feedback does not affect the complexity development of Iranian intermediate EFL learners' speaking is rejected.

E. Validity

A factor analysis through the varimax rotation is carried out to probe the construct validity of the tests administered in this study. Before commenting on the results of the factor analysis, it should be mentioned that the present sample size was adequate for running the factor analysis (KMO = .82 > .60) and the correlation matrix was appropriate for the analysis (χ = 644.98, P < .05) (Table 15).

TABLE 15.
KMO AND BARTLETT'S TEST OF SPHERICITY

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.825
Bartlett's Test of Sphericity	Approx. Chi-Square	644.980
	df	21
	Sig.	.000

The SPSS has extracted only one factor which account for 73.18 percent of the total variance.

TABLE 16
TOTAL VARIANCE EXPLAINED

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.241	74.866	74.866	5.123	73.187	73.187
2	.994	14.200	89.066			
3	.597	8.527	97.594			
4	.091	1.301	98.894			
5	.050	.712	99.606			
6	.018	.261	99.867			
7	.009	.133	100.000			
Extraction Method: Principal Axis Factoring.						

TABLE 17.
COMPONENTS MATRIX

	Factor
	1
Accuracy Pre	.993
Fluency Pre	.991
Complexity Pre	.978
Complexity Post	.934
Fluency Post	.913
Accuracy Post	.698
PET	.093

Finally, as displayed in Table 17 the components of the pretests and posttests of speaking and the PET test load on the only extracted factor. The rather low factor loading of the PET on this factor indicates that the extracted factor can be labeled as "speaking proficiency" factor.

F. Inter-rater Reliability

The Pearson correlations between the two raters are displayed here in Table 18. Based on these results it can be concluded that:

A: There is a significant agreement between the two raters who rated the students' writings (r (74) = .86, P < .05, it represents a large effect size).

B: There is a significant agreement between the two raters who rated the students' pretest of speaking (r (52) = .61, P < .05, it represents a large effect size).

TABLE 18.
INTER-RATER RELIABILITY

		Writing_Rater1	Sp_PreT_Rater1	Sp_PostT_Rater1
Writing_Rater2	Pearson Correlation	.862**		
	Sig. (2-tailed)	.000		
	N	76		
Sp_PreT_Rater2	Pearson Correlation		.612**	
	Sig. (2-tailed)		.000	
	N		54	
Sp_PostT_Rater2	Pearson Correlation			.697**
	Sig. (2-tailed)			.000
	N			54

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

*. Correlation is significant at the 0.05 level (2-tailed).

C: There is a significant agreement between the two raters who rated the students' posttest of speaking (r (52) = .69, P < .05, it represents a large effect size).

V. DISCUSSION

In previous section, after the calculation of the homogeneity of the participants in both control and experimental groups prior to the treatment, a series of analyses were run on their scores in pretest and posttest in order to find proper answer to the study's research questions.

The results of the data analysis indicated that DCF affected the overall oral proficiency of the Iranian intermediate EFL learners in the experimental group significantly greater than in the control group. The story was somewhat the same regarding the sub-questions, it means the effect of DCF on experimental group has been greater in accuracy and complexity traits compared to the results of the control group, and the experimental group out-performed the control group. The only trait that was not affected positively in the experimental group, or the effects seem to be less than development of the participants in the control group, was the fluency development. It may be concluded that as much as the students tried to be more accurate and produce less erroneous clauses and focused on the quality of their production, they lost their concentration on their quantity and speed of their production that resulted in lower fluency.

The reason for having the overall development of the experimental group (M= 20.61) greater than that of the control group (M= 19.73), but this difference was not as much as those for accuracy and complexity needs to be considered. Since the results of the independent t-test (t (52) = 2.13, P < .05, r = .28 represented an almost moderate effect size of DCF on posttest of overall speaking, while this effect on accuracy and complexity (as mentioned earlier) was of high effect size, it can be resulted from two factors that should be considered. On one hand, the overall speaking is like an umbrella term regarding its components; accuracy, fluency, and complexity. Therefore, the variance in each of these traits results in a change in the overall performance (overall speaking) due to the overlap between the content of the overall speaking with its components' content. On the other hand, since each of these four traits was measured separately and with a different scale (i.e. the use of body language or eye contact), it is not necessarily expected to have the same results of the speaking components reflexed in the overall speaking performance. In other words, one's performance might be poor on the individual components of speaking, but not as much poor (or even well) as on the overall speaking performance; or respectively the other way around.

To conclude, the findings of this study is in line with the previous studies who implemented the DWCF for their students writing in term of accuracy development, i.e. N.W. Evans et al. (2011), although this development in accuracy was in some of them of less effect, but none of them were ineffective or with negative effect, i.e. (Evans et al., 2010; Evans et al., 2011; Hartshorn et al., 2010).

VI. CONCLUSION

The findings of this study indicated that DCF which is a systematic and innovative method of error correction in the area of speaking correction has been mostly successful, since it benefits from a strong theoretical principles derived from DWCF in the area of writing error correction. The principles imply that the provided corrective feedback should be meaningful, timely, constant, and manageable and it must reflect the individual learners most immediate needs based on what they produce (Evans et al., 2010). It should be noted that based on the findings described earlier, DCF was more effective in the area of accuracy and complexity, and quite effective on overall speaking of the participants, but its impact on the fluency of the learners was not as much positive as in the control group, since the control group outperformed the experimental group quite significantly. Further, this study and its positive findings is just a beginning for the DCF in the area of oral production, and it needs further investigations, modification, and practices to be more reliable and practical.

However, although this study just focused on a limited number, context, and level of participants (54 Iranian intermediate EFL learners), and there is still a long path to go, the findings could be implemented in smaller contexts, and for personal or even educational uses. As the last word, it should be noted that after further investigations and proves, and modifications of the presented technique of DCF in larger scales, it is highly potential to be used globally by teachers, teacher trainers, and also syllabus designers.

REFERENCES

[1] DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge, England: Cambridge University Press.

[2] DeKeyser, R., 2007. Skill acquisition theory. In: VanPatten, B., Wiliams, J. (Eds.), *Theories in Second Language Acquisition.* Lawrence Erlbaum, Mahwah, NJ, pp. 97e113

[3] Evans, N. W., Hartshorn, K. J., McCollum, R. M., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research, 14*(4), 445-463. doi: 10.1177/1362168810375367.

[4] Evans, N. W., James Hartshorn, K., & Strong-Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System, 39*(2), 229-239.

[5] Evans, N. W., James Hartshorn, K., & Strong-Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System, 39*(2), 229-239.

[6] Evans, Norman W., James Hartshorn, K., & Strong-Krause, Diane. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System, 39*(2), 229-239. doi: 10.1016/j.system.2011.04.012.

[7] Ferris, D. R. (2004). The “grammar correction” debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime…?). *Journal of Second Language Writing, 13*, 49-62.

[8] Foster P., and P. Skehan. (1996). “The Influence of Planning on Performance in Task-Based Learning.” *Studies in Second Language Acquisition*, 18, 3: 299–324.

[9] Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research, 3*(3), 215–247. doi:10.1177/136216889900300303

[10] Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing, 16*, 40-3.

[11] Hartshorn, K. James, Evans, Norman W., Merrill, Paul F., Sudweeks, Richard R., Strong-Krause, Diane, & Anderson, Neil J. (2010). Effects of Dynamic Corrective Feedback on ESL Writing Accuracy. *TESOL Quarterly, 44*(1), 84-109. doi: 10.5054/tq.2010.213781

[12] Hartshorn, K. James, Evans, Norman W., Merrill, Paul F., Sudweeks, Richard R., Strong-Krause, Diane, & Anderson, Neil J. (2010). Effects of Dynamic Corrective Feedback on ESL Writing Accuracy. *TESOL Quarterly, 44*(1), 84-109. doi: 10.5054/tq.2010.213781.

[13] Lyster, R., Ranta, L. (1997) Corrective feedback and learner uptake: negotiation of form in communicative classrooms. *Studies in Second Language Acquisition 19*: 37-66.

[14] Mochizuki, N., and L.Ortega. 2008. “Balancing Communication and Grammar in Beginning Level Foreign Language Classroom: A Study of Guided Planning and Relativization.” *Language Teaching Research*, 12: 11–37.

[15] Quintana, J. (2003). PET: practice tests: with explanatory key. Oxford University Press.

[16] Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133-164). Amsterdam: John Benjamins.

[17] Saslow, J., & Asher, A. (2006). Summit 1. New York: Pearson Education.

[18] Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17,* 38–62.

[19] Skehan, P., & Foster, P. (1999). The Influence of Task Structure and Processing Conditions on Narrative Retellings. *Language Learning, 49*(1), 93–120. doi:10.1111/1467-9922.00071.

[20] Tomasello, M., and C. Herron. (1988). “Down the Garden Path: Inducing and Correcting Over generalization Errors in the Foreign Language Classroom.” *Applied Psycholinguistics*, 9: 237–46.

Esmaeil Bagheridoust is an assistant professor at Islamic Azad University, South Tehran Branch.

During the last 15 years, he carried out a number of research projects, composed and translated a couple of books and articles, and took part in a number of conferences and seminars inside and outside the country. He took the chance to be the secondary research assistant at the University of Ottawa (2005) and the visiting scholar of the University of Texas at Austin (2006-8). He proposed a model of language testing to implement the ACTFL ILR-based assessment system of PDAT in the Middle Eastern Department of UT.

Ali Mohammadi Kotlar was born in Tehran, Iran in 1983. He received his MA in TEFL/TESOL from Islamic Azad University, South Tehran branch in 2013, and holds a BA in English language and literature from Islamic Azad University of Roudehen (2008).

He has worked as a Teacher and Supervisor at many different institutes and schools in Tehran for more than 8 years. He is interested in Language Teaching theories and practices, Language Testing, Teacher Education, and Sociocultural Theory; and he has been recently doing some researches on Scaffolding, Dynamic Assessment, and Corrective Feedback.