International Journal of Research Studies in Language Learning
2015 July, Volume 4 Number 3, 107-123

The effect of dynamic assessment on complexity, accuracy, and fluency in EFL learners' oral production

Ebrahimi, Elham ✉
Najafabad Branch, Islamic Azad University, Najafabad, Iran (ebrahimi_elham@rocketmail.com)

Received: 1 October 2014 **Revised**: 28 November 2014 **Accepted**: 30 November 2014
Available Online: 13 January 2015 **DOI**: 10.5861/ijrsll.2015.982

ISSN: 2243-7754
Online ISSN: 2243-7762

OPEN ACCESS

Abstract

The present study aimed to investigate the effect of dynamic assessment on complexity, accuracy and fluency (CAF) in EFL learners' oral production. Participants were 44 intermediate; they were all female and aged from 11 to 15 in a language institute in Isfahan. They were divided into two experimental and control groups: The experimental group received the treatment (dynamic assessment/DA), and the control group received regular instruction under institute's normal situation. Findings revealed that the implementation of DA promoted more complex and accurate oral productions while it did not have any effect on the fluency of learners' oral productions. In addition, results manifested strong positive correlations between measures of CAF and learners' oral proficiency scores. Finally, the factor analysis revealed that there were three main factors (CAF) underlying the six measures used in the present study. The results of this study have a number of implications for EFL practitioners and teachers as well.

Keywords: dynamic assessment; complexity; accuracy; fluency and EFL practitioners

The effect of dynamic assessment on complexity, accuracy, and fluency in EFL learners' oral production

1. Introduction

In the last two decades, a great number of examinations of different applications of dynamic assessment (DA) have been begun, which is rooted in both Vygotsky's (1978) learning theory, conceptualization of a zone of proximal development (ZPD) and Feuerstein's (1979) theory of mediated learning experiences. Vygotsky (1978) believes that the early development of understanding occurs while having interaction with other people and more important and higher level achievement occurs when a child learns by working with a more educated and informed guide. Furthermore, students can develop the mental functioning required in social interaction within ZPD (Brown, 2004). The term 'dynamic assessment' includes a range of methods and materials to assess this opportunity for teaching and learning dynamically, which use two models; pretest–treatment–posttest and successive hinting. But, static kind of achievement assessed by typical and usual tests.

DA methods focus on how learning is occurring not measuring static experience knowledge. In addition, these two models consider the conditions of changing the performance not just performance itself. Also, the models examine the interaction of the assessor and the examinee, the influence of this interaction on performance, and the responsiveness of the examinee to the interaction (Lidz, 1991). In other words, as Vygotsky believes that the unification of assessment and instruction can lead development. Moreover, Vygotsky's sociocultural theory in assessment shows that DA suggests new and various views of assessment in the language classroom as well. As we frequently hear that a teacher acknowledges the drawbacks of any specific assessment method which some top students take final test weakly, however, s/he does tasks well in the class.

Although a number of studies have been done on DA, but little has been done in relation to the effect of DA on aspects of performance, such as complexity, accuracy and fluency (CAF) in students' oral production. To do so, the aim of this study is to apply dynamic assessment procedures to identify the present learners' abilities, their potential future abilities and the development of learners' oral proficiency in terms of CAF.

2. Literature Review

2.1 Theoretical framework of DA

According to all researches which were done in DA area, Intervention, Interactionist, Sandwich and Cake formats have been discussed. Lantolf and Poehner (2004) intend the terms interactionist and interventionist as two main and general types of mediation. Intervention is as a part of assessment that according to Sternberg and Grigorenko (2002, p. 30), instructor can monitor the learners in a case of providing an opportunity, in order to achieve a view of their changes. These authors believe that however all DA procedures cannot be helpful but some of can be as evaluations instruments which will be applied to suggest support which are graded from standardized hints to dialogic interaction for the learners and also examine their reactions so that mediator can make predictions about their levels of ability of learning. Reporting of their reactions and readiness to the teacher, parents, and other decision-makers can predict their real levels of knowledge and also development that may occur during the course; it is worth mentioning that this procedure was the main procedure applied in the present study. Vygotsky himself imagined the relationship between the examiner and examinee as "cooperation" (Vygotsky, 1998, p. 201) rather than intervention, clearly inferring a dialogic interaction, with both participants discussing and sharing in the responsibility for development.

2.2 Studies of DA

This section has introduced briefly some studies which have been done in DA area to illustrate the point of present study. It has been begun by reviewing two dynamic-like assessment studies in a L2 context. First, Schneider and Ganschow (2000) suggest that awareness of metalinguistic strategies could be especially helpful for learners with dyslexia by using interaction. And second, Grigorenko, Sternberg, and Ehrman (2000) developed a formal testing instrument to measure learners' ability to deal with novel problems.

As it mentioned earlier, intervention model is one of the general model of DA. Carlson and Wiedl have proposed two intervention techniques: They support interrupting test administration as essential to provide feedback and elicit verbalization rather than introducing a separate intervention phase. Another study with the same ground, interventionist approach to DA, has been carried by Kozulin and Garb. Due to the large numbers of adult immigrants to Israel who are included their research, Kozulin and his colleagues have relied upon an interventionist format in which teaching is sandwiched between a static pretest and post-test. They define the difference between the learners' pre- and post-test scores and made this difference as low, intermediate and high, and instructional recommendations for each group.

The Second general approach to DA is interactionist. In this regard, the first attempt to apply this approach in a L2 context was done by Antón (2003) to achieve a view on the participants' language abilities in order to more accurately place them in the L2 Spanish program that this work has a lot in common with interventionist approach to obtain the validity of results of assessment. Another field of studies is interactionist L2 DA in an ESL writing program by Aljaafreh and Lantolf (1994). They reported on their collaboration with ESL learners for a writing intensive class. The mediator monitored students to find out the level of correctness of usages of tense, modal verbs, prepositions and articles. Another study of this approach has been conducted by Poehner (2005) which was cake format and interactionist of DA for L2 French learners' oral proficiency. It could provide the view on learners' problems while doing the tasks and quantity and quality of interactions between learners and mediator to solve their problems.

The aim of the next study which has been done by Abdolrezapour, Tavakoli and Ketabi (2013), was to gain a picture of students' emotions in an EFL context by using dynamic assessment models to the development of learners' emotional intelligence (EI). Finding of this study showed that there is the potential of emotional of DA for increasing one's emotional intelligence and it leads to promoting L2 development and promoting learners' knowledge about emotional characteristics.

2.3 Aspects of Oral Production: Complexity, Accuracy and Fluency

The constructions of complexity, accuracy, and fluency (CAF) have been applied some studies on the acquisition and using second language; they are used for doing language tasks by native speakers or first language learners, most frequently used as dependent variables to examine the differences between independent variables such as acquisitional level or task features such as DA. These three notions measures can also be used to describe performance by native speakers or first language learners.

One related study that has been conducted by Ahmadian and Tavakoli (2011) which is also described since it was specifically drawn upon for the description statistics and measurement of the variables of the present study. This study was hypothesized that the trade-off involves accuracy and fluency with both careful online planning and task repetition enhancing complexity in oral production. It was confirmed that the simultaneous use of these two implementation variables leads to an exponential increase in learners' complexity of oral language production.

Complexity - According to review of CAF researches, complexity has different meanings; so it the most problematic construct of the CAF triad. Complexity refers to complexity of both tasks (objective difficulty) and language performance (subjective difficulty) (Skehan, 1998; Robinson, 2007). Also, Skehan defines complex as

'challenging language'. Some authors identify complex with 'acquired late'; Ellis defines complexity as 'the capacity to use more advanced language', while Skehan speaks of 'more advanced language, leading to complexity'.

Accuracy - Accuracy is perhaps the simplest coherent construct, referring to the degree of conformity to correct and error free norms. Wolfe-Quintero et al. (1998, p. 33) defined accuracy measures as the comparison with target like use. It's worth mentioning that there is a different between accuracy and comprehensibility (e.g., Cumming & Mellow, 1996). For example, a text is not accurate but is understandable and also is maybe not accurate but more 'more developed'.

Fluency - Fluency can be defined as 'the ability to produce speech at normal rate and without pausing or hesitation' (Skehan, 2005; Ellis & Barkhuizen, 2005). These definitions infer to native speakers' behavior. Accorcing to the other ideas, fluency usually refers to a person's general language proficiency specially in both speech and writing, as characterized by perceptions of ease, and 'smoothness' (Lennon, 1990; Chambers, 1997; Guillot, 1999; Freed, 2000; Koponen & Riggenbach, 2000; Hilton, 2008).

The present study will be intended to investigate the impact of dynamic assessment on CAF. As a few studies on the topic have been mentioned above, they showed the importance of CAF and applying DA separately. The present study, more specifically, intends to integrate assessment into instruction to investigate the effect of DA on aspects of CAF in learners' oral production. To do so, the research questions to be addressed are:

- Does DA have any effect on the complexity of Iranian EFL learners' oral productions?
- Does DA have any effect on the accuracy of Iranian EFL learners' oral productions?
- Does DA have any effect on the fluency of Iranian EFL learners' oral productions?
- Are there any relationships between the complexity, accuracy and fluency of Iranian EFL learners' oral productions and their oral proficiency?
- Do the six measures used in this study truly represent the three aspects of complexity, accuracy, and fluency?

3. Method

The present study used quasi-experimental (because of limitation of language institute) design where the participants were assigned into two groups (four classes), experimental group (n=24) and control group (n=20). The experimental group was taught by applying dynamic assessment materials, called dynamic assessment group (henceforth DA-group) and control group was taught under institute's normal procedures called none dynamic assessment group (henceforth non-DA-group), who undergoes normal procedures that error corrections are in only one step for them. The treatment for experimental group involves interventionist and interactionist DA approaches. Two groups received a complete course of instruction that were 16 sessions in 8 weeks and each session took 110 minutes. However, about 20 to 30 minutes of 10 sessions were devoted to dynamic assessment to DA-group. The instructor of two groups was the present researcher.

To make sure of the homogeneity of the two groups in terms of their level of proficiency prior to the experiment, all participants took an oral placement test which was conducted by the researcher, according to guidelines given by Farhady, Ja'farpur and Birjandi (2010, p. 216) which have been explained in procedure section. Results revealed that participants in two groups had a range of scores between 67 and 82. Finally, at the end of the course, participants took an oral exam as a final exam. Learners' responses in final exam provided the data of present research. It's worth mentioning that the classes were equipped with camera to record participants' oral activities and responses, both during the course and for final exam.

This study has one independent variable that is the implementation of DA in a natural classroom context,

and the dependent variable is the participants' oral proficiency at three levels of complexity, accuracy and fluency.

3.1 Participants

Participants were forty four EFL learners who were all female and aged from 11 to 15. They enrolled in a language institute studying "Hey There! 2B" (Morales, Myers, & Jackson, 2009). They were all in four classes. Learners were native speakers of Farsi and they had taken English courses for about two years. Their level of English proficiency was considered as intermediate level learners according to the language center's standards and the placement tests that they had taken, that is, all participants were currently taking the intermediate EFL course offered by the institute. None of these participants had additional exposure to the English language, apart from the regular TV programs and the Internet nor had ever been to a country where English is spoken as a native language.

3.2 Instruments

For the purpose of the present study, different materials were prepared and used. These materials were divided into two categories; instructional materials and assessment materials which will be explained in detail below.

Instructional Materials

- Storybook: Hey There! 2B has a storybook which is compatible with the learners' level; its name was 'E.T.' (the extra-terrestrial). The book consists of 11 sequential chapters of 2 or 3 pages that the first 6 chapters are covered in Hey There! 2A and the rest of them are covered in Hey There! 2B term. Some sessions learners had to be ready to narrate one chapter of their storybook which was assigned its previous session. In this study, one session was devoted to storybook for DA-group to receive dynamic assessment that was chapter 7 and the title was 'E.T. Makes a Transmitter' which included simple past tense verbs and 4 new words.

- Movie: 5 episodes of a movie were provided for this term which its name was 'Extra' that each episode took about 5 minutes. It was shown learners every multiple sessions of 5 and they had to give a report after watching that, while the movie was shown again silently, so that learners could remember sentences. In present study, third episode which was called 'Animal Protest' was chosen to be under dynamic assessment condition for DA-group.

- Picture-cued story: There are 8 pictures on pages 64 and 65 In Hey There! 2B that are related to each other. Learners were asked to make a story of those pictures in one session which was considered as another dynamic assessment session for DA-group.

- Audio parts: In Hey There! 2B, there are four units, 5 to 8, that each unit has one listening part which is included a conversation or a story. Each session that learners had this part, they listened to the audio track. After the first time, they were asked a general question and then it was played for the second time; after that they had to give a summary. In order to carry out the dynamic assessment for DA-group, the listening part of unit 7, was played that was a short story about some girls and boys who have a picnic but one of them while climbing the tree injured and other help her.

- Pictures of objects: There was a big picture on page 98 and 99 in unit 8, that learners were asked to describe the picture that were included different subjects while doing different actions. They had to use simple present and present continues tense.

- Text as a reading comprehension: In Hey There! 2B, there are four units, 5 to 8, that each unit has one reading part. Learners have to talk about reading parts as a summary of their books, as a report and

then they were expected to make questions from text sentence by sentence. Reading of unit one was selected for implementing dynamic assessment for DA-group.

- News event: The first fifteen minutes of all odd sessions learners had to narrate the news which they read in newspaper, magazine and internet or heard from television or anyone else. To do dynamic assessment for DA-group, session thirteen was considered.

- Telling anecdote: Learners were given a topic to tell their personal anecdote three times during the term; in order to check them whether they used grammar points correctly or not. To make sure of providing the same condition for all learners, they were given topic in the class to get ready after some minutes to think about it not in advance.

- Pair work: Students can learn through interaction with their teacher, but also through interaction with their classmates. Several studies have elucidated the impact that peer mediation can have on a student's learning. Moreover would help the teacher in running activities and would enhance the effectiveness of the intervention. In Some sessions learners were asked to practice different parts of book that concluded ask and answer each other the questions related to the grammar point of unit and making the conversation the same as unit's conversation which was as a model. This time of the class, instructor went around the class and checked their oral productions. One session was devoted to run the dynamic assessment for DA-group.

- Group work: Within small group work, students often take responsibility for assisting each other; and also group work is a common configuration in foreign language classrooms because it allows all learners the opportunity to speak (Curtain & Dahlberg, 2010). Learners did such activities in pair work, but in groups of three. One session was carried out dynamic assessment for DA-group through group work. In this part, CAF were observed to gain their development by learners' proficiency level and participations as well as pair work.

Assessment Materials

- Oral interview: In this study two interviews were done. The first one was a non-dynamic pre-test as a placement test to make sure the homogeneity of the learners and the second one was a non-dynamic post-test as a final exam to gain differences between non-DA group and DA group after treatment. The best way to test one's abilities to produce the language is through an interview test (Farhady, Ja'farpour, & Birjandi, 2010). In this study, interview contained some questions which assessed their oral productions in terms of CAF. To do so, answers assessed analytically that their oral productions were recorded and assessed by two raters. The purpose of discrete-point or analytic marking is to identify learners' strengths and weaknesses of various components of oral production. To do so, that teacher knows what to focus on as they continue to improve their spoken English. For example, comprehensibility was scored by some items; such as, pronunciation, including consonants, vowels, intonation, rhythm, and sentence and word stress. Also, fluency was scored by smoothness and flow. For instance, accuracy is related to correctness of answer grammatically which its grammar point was taught during the course; so structure, tense of verbs and word ordering were assessed separately.

- Oral narrative tasks: As mentioned in instructional materials, learners were asked to narrate the content of audiotapes, episodes of movie, storybook and picture-cued story in order to induce learners to produce units of language that are not influenced by interactional variables (Yuan & Ellis, 2003). One session for implementing each material was devoted to run dynamic assessment for DA-group from all sessions.

3.3 Procedure

The dynamically assessed tasks in this study provided the opportunity of performing speaking activities by applying approaches of DA. This study applied interactionist, intervention and sandwich models of DA in DA group. Interactionist DA is referred to Vygotsky' s preference for cooperative dialoging in which assistance emerged from the interaction between the examiner and the learner, simply relationship between them with both participants negotiating and sharing in the responsibility for development. This model was used in two ways; first, by answering learners' problems and second, by asking a question in order to know the development of learners that it showed whether applying interventionist is useful or not.

Interventionist DA focus on development of learner individually and in a group by successive hinting for treatment, which in this study was a series of 8 hints that has been proposed by Lantolf and Poehner (2006). The mediating moves were ordered from most implicit (pause) to most explicit (explanation). These hints were considered a part of assessment materials as well. Interventionist model concluded these steps which were explained by an example that occurred in the class while DA-group was receiving dynamic assessment:

- Pause
- Repeat the whole phrase questioningly
- Repeat just the part of the sentence with the error
- Teacher points out that there is something wrong with the sentence
- Teacher points out the incorrect word
- Teacher asks either/or question
- Teacher identifies the correct answer
- Teacher explains why

This mediation inventory enabled the instructor to be highly systematic in interactions with her students. As they worked through the activities, her initial response to learner difficulties (in the case under consideration, when learners failed to correctly mark sentences) was to pause. For some students, this sent a clear message that something was amiss with their performance and they attempted to work through the difficulty, often with a positive outcome. For other students, the pause either resulted in an inappropriate response or it failed to trigger any response at all. When this occurred, the instructor moved to the next prompt in which she would repeat the utterance produced by the student with rising intonation as a way of signaling that there was something wrong but without indicating the nature of the problem (e.g. lexical, syntactic, morphological) nor precisely where in the construction the problem was located. For instance, in narrating storybook that they were asked to narrate by simple past tense, if a student said 'they don't hear their mom on the stairs', the instructor repeated 'they don't hear their mom on the stairs'?; If the student responded with the appropriate form of didn't, the interaction concluded, and the activity continued with another student. If the prompt failed, however, the instructor would move to the next level of mediation (3) in an attempt to narrow the student's focus to the portion of the utterance where the error had occurred. In the example at hand, this was 'don't' produced with rising intonation; and this protocol continued to get a correct response from student. It's worth mentioning that the most mediation was concluded before step (5).

The last model which was applied in this research is sandwich model that follows traditional experimental research, pretest-treatment-posttest. This model was used for two groups, DA group and non-DA group. An oral test was conducted by the researcher (instructor), based on Farhady, Ja'farpour and Birjandi (2010, p. 216) as a proficiency test instead of pretest.

Depending on the purpose and defined criteria of a test, scoring may be done holistically or analytically. The former related to an overall impression according to which the interviewee either receives "excellent", "good", "fair" or "poor" or "pass/fail". The latter was applied for this study, on the other hand, rates the interviewee's performance separately on scales that pertain to accent, structure, vocabulary, fluency, and comprehension. These ratings were then weighed and added up to determine a final score for all participants. The highest and lowest scores have been brought here; the results are average scores which two raters assessed and showed no statistically significant difference among participants.

Table 1

The Highest and Lowest Scores in Oral Proficiency Test

Scale	Rating	Weighted
The Highest:		
Accent	5	1
Structure	5	30
Vocabulary	5	20
Fluency	5	10
Comprehension	5, 6	21
Total score		82
The Lowest:		
Accent	3	2
Structure	4	24
Vocabulary	3, 4	16
Fluency	4	8
Comprehension	4	17
Total score		67

Scoring on the basis of detailed operational statements is more reliable and provides a useful profile of the interviewee's relative strengths and weaknesses. Moreover, an interviewer may decide to do away with accent scale as a whole because it doesn't cover test objectives or simply because there is really no agreement as to what constitutes good pronunciation. Under such condition, a specific and precise scoring criterion is required. The next step in this study was carrying out the treatment. Treatment in non-DA group was traditional and ordinary instruction, teaching with traditional error correction, only by explicit feedbacks; such as teacher correction or another student correction, by asking who can correct this sentence?; without any hints, prompts, and leading questions; but in DA-group successive hinting was used in 10 sessions of 20 to 30 minutes through they were receiving dynamic assessment, as mentioned above. At last, final exam was considered as the posttest. That was oral exam which learners were asked six similar questions that questions are appropriate to their levels and materials which taught. Their responses were recorded to measure the CAF. Learners did not get any feedback or help while they were answering.

4. Results

As said before, this study was designed to investigate the effect of dynamic assessment on the students' oral productions in terms of complexity, accuracy, and fluency. Moreover, this research aimed to consider if there was any relationship between the three aspects of complexity, accuracy and fluency of EFL learners' oral productions and their overall oral proficiency scores. Finally, this research aimed to consider if the six measures used in the present research truly represent the three distinct components of complexity, accuracy, and fluency. To achieve these aims, the quantitative data were collected. The research questions, previously formulated in introduction section, are reiterated and answered in order. This section explains the data treatment for each of the instruments employed in this study and presents the responses from each of the instruments. Results of analysis will be reported for each research question. In each part the data will be displayed, the hypotheses of the study will be tested, and the results will be stated. Table 2 below shows the descriptive statistics of the study.

Table 2

Descriptive Statistics

Factors	*N*	*Mean*	Min	Max	Range	*SD*	Variance	Sum
Oral Proficiency	44	74.14	67.00	82.00	15.00	4.46	19.93	3262.0
Complexity No of Verb Forms	44	7.59	5.00	11.00	6.00	1.93	3.74	334.00
Complexity Clauses/Subordination	44	1.23	1.01	1.59	.58	.18	.03	54.12
Accuracy Error-Free Clauses	44	76.08	61.80	96.44	34.64	10.68	114.01	3347.7
Accuracy Use of Target-like Verbs	44	70.28	54.00	92.38	38.38	10.54	111.19	3092.4
Fluency Syllables/Min	44	54.24	40.95	64.45	23.50	6.29	39.61	2386.0
Fluency Meaningful Syllables/ Min	44	50.75	37.33	61.12	23.79	6.42	41.27	2232.4

4.1 Analysis of the research questions

The effect of DA on the complexity of learners' oral productions - The first research question addressed the effect of DA on Iranian EFL learners' oral production in terms of complexity. In response to this question, hypothesis one was formulated which will be examined in this section.

Hypothesis 1: DA does not have any effect on the complexity of Iranian EFL learners' oral productions.

In order to investigate the first hypothesis, independent samples t-tests were conducted between the two experimental and control groups. The minimum alpha for confirmation of the research hypothesis was .05. The mean scores and standard deviations of the two groups with respect to the two complexity measures (total number of different grammatical verbs and ratio of clauses to amount of subordination) are presented in Table 3 Additionally, t-tests results are demonstrated in Table 4.

Table 3

Descriptive of independent samples t-tests for complexity measures

Factors	Group	*N*	*Mean*	Std. Deviation	Std. Error Mean
Complexity	Control Group	20	6.8000	2.04167	.45653
No of Verb Forms	Experimental Group	24	8.2500	1.59483	.32554
Complexity	Control Group	20	1.1260	.13721	.03068
Clauses/Subordination	Experimental Group	24	1.3167	.16584	.03385

Table 4

Independent samples t-tests for complexity measures

		F	*Sig.*	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Complexity No of Verb Forms	Equal variances assumed	.673	.416	-2.645	42	.011	-1.45000	.54821
	Equal variances not assumed			-2.586	35.626	.014	-1.45000	.56071
Complexity Clauses/Subordination	Equal variances assumed	4.363	.043	-4.101	42	.000	-.19067	.04649
	Equal variances not assumed			-4.173	42.000	.000	-.19067	.04569

The results of the T-tests, illustrated in Table 4, showed that there were statistically significant differences ($p < .05$) between the control and experimental groups, regarding the two complexity measures. As illustrated in Table 3, the mean score of the first complexity measure i.e., total number of different grammatical verbs ($M = 8.25$) was statistically higher in the experimental group. In addition, Table 3 reported that the mean score of the second complexity measure i.e., ratio of clauses to amount of subordination ($M = 1.32$) was also statistically higher in the experimental group (see Figure 1). Therefore, the obtained results showed that both measures of complexity were greater in the DA group, indicating that DA resulted in more complex oral production. Thus, based on these results the first null hypothesis predicting that DA does not have any effect on the complexity of Iranian EFL learners' oral productions is rejected.

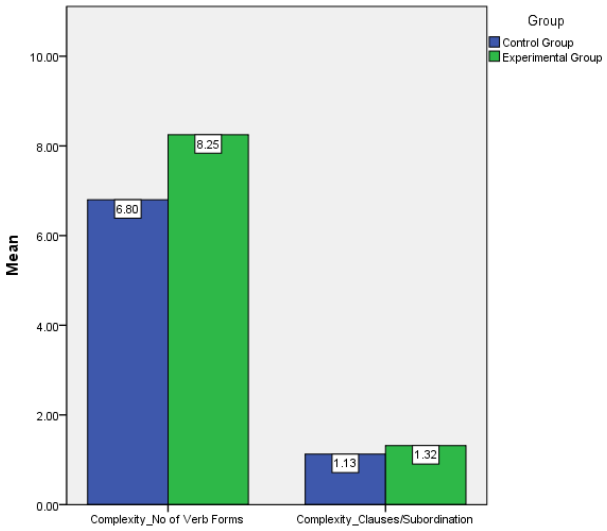

Figure 1. Complexity scores of the experimental and control groups

The effect of DA on the accuracy of learners' oral productions - The second research question tried to investigate if DA has any effect on the accuracy of Iranian EFL learners' oral productions. Subsequently, the second null hypothesis was made in reply to this question.

Hypothesis 2: DA does not have any effect on the accuracy of Iranian EFL learners' oral productions.

In order to examine the second hypothesis, two independent samples t-tests were carried out on each accuracy measure in order to determine for which measures differences reached significance. At first the descriptive data for the two accuracy measures are displayed in Table 5. Similarly, summary of the results from the t-tests is displayed in Table 6.

Table 5

Descriptive of independent samples t-tests for accuracy measures

Factors	Group	*N*	*Mean*	Std. Deviation	Std. Error Mean
Accuracy	Control Group	20	72.4105	10.30653	2.30461
Error-Free Clauses	Experimental Group	24	79.1458	10.19193	2.08042
Accuracy	Control Group	20	66.4585	9.57234	2.14044
Use of Target-like Verbs	Experimental Group	24	73.4708	10.43360	2.12975

Table 6

Independent samples t-tests for accuracy measures

		F	*Sig.*	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Accuracy Error-Free Clauses	Equal variances assumed	.122	.729	-2.172	42	.036	-6.73533	3.10150
	Equal variances not assumed			-2.169	40.414	.036	-6.73533	3.10473
Accuracy Use of Target-like Verbs	Equal variances assumed	.431	.515	-2.304	42	.026	-7.01233	3.04374
	Equal variances not assumed			-2.322	41.578	.025	-7.01233	3.01949

The results of the t-tests, illustrated in Table 6, showed that there was a statistically significant difference ($p < .05$) between the experimental and control groups regarding both accuracy measures. As shown in Table 5, the mean score for percentage of error-free clauses in the experimental group ($M = 79.14$) was statistically greater than the mean score in the control group ($M = 72.41$; t (42) = -2.17, $p = .36$). Likewise, as for the other dependent variables i.e., percentage of target-like use of correct verbs, the mean score of the experimental group

(*M* = 73.47) was statistically higher than the mean score of the control group (*M* = 66.45; *t* (42) = -2.30, *p* = .26). In order to show the differences more clearly findings are illustrated in Figure 2. Therefore, regarding the effect of DA on the accuracy of learners' oral productions, t-test results indicated that learners produced more accurate oral productions in the DA group in comparison to the learners in the control group. Thus, the second null hypothesis stating that DA does not have any effect on the accuracy of Iranian EFL learners' oral productions is rejected.

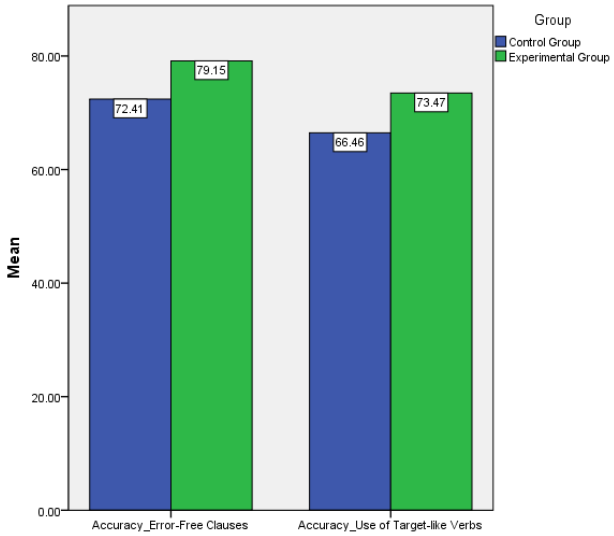

Figure 2. Accuracy scores of the experimental and control groups

The effect of DA on the fluency of learners' oral productions - The third research question addressed the effect of DA on the fluency of Iranian EFL learners' oral productions. As mentioned in chapter one hypothesis three was formulated in response to this research question.

Hypothesis 3: DA does not have any effect on the fluency of Iranian EFL learners' oral productions.

In order to examine the effect DA on fluency of oral productions, two independent sample t-tests were conducted between the experimental and control groups for both measures of fluency. Table 7 reported the descriptive data of fluency scores with regard to each of the control and experimental groups. Additionally, t-test results are demonstrated in Table 8.

Table 7

Descriptive of independent samples t-tests for fluency measures

Factors	Group	*N*	*Mean*	Std. Deviation	Std. Error Mean
Fluency	Control Group	20	53.8650	6.58136	1.47164
Syllables Per Minute	Experimental Group	24	54.5500	6.16843	1.25913
Fluency	Control Group	20	50.2820	6.94110	1.55208
Meaningful Syllables Per Minute	Experimental Group	24	51.1375	6.08433	1.24196

Table 8

Independent samples t-tests for fluency measures

		F	*Sig.*	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Fluency Syllables Per Minute	Equal variances assumed	.028	.868	-.356	42	.724	-.68500	1.92515
	Equal variances not assumed			-.354	39.509	.725	-.68500	1.93678
Fluency Meaningful Syllables Per Minute	Equal variances assumed	.085	.772	-.436	42	.665	-.85550	1.96372
	Equal variances not assumed			-.430	38.188	.669	-.85550	1.98781

Table 7 showed that the fluency scores were not equivalent for the two groups. The mean scores of both fluency measures (i.e., number of all syllables produced per minute and number of meaningful syllables per minute) were slightly higher in the experimental group than in the control group (see Figure 3). Although the mean score were slightly different among the two groups, t-test results in Table 8 revealed that the differences were not statistically significant for both fluency measures ($P > .05$). Consequently, there was no statistically significant difference between the fluency of learners' oral productions in the DA and non-DA group. Based on these results, therefore, the third null hypothesis stating that DA does not have any effect on the fluency of Iranian EFL learners' oral productions is confirmed.

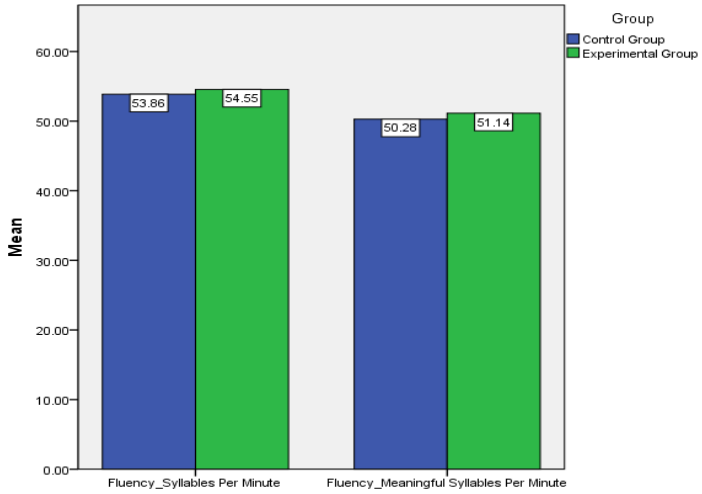

Figure3. Fluency scores of the experimental and control groups

The relationship between CAF and learners' oral proficiency - The forth research question addressed the relationship between learners' oral proficiency and the three aspects of complexity, accuracy and fluency of their L2 oral productions. In response to this question, hypothesis four was formulated which will be examined in this section.

Hypothesis 4: There are not any relationships between the complexity, accuracy and fluency of Iranian EFL learners' oral productions and their oral proficiency.

Pearson product-moment correlation coefficient was run in order to investigate the correlation among oral proficiency scores and any of the three aspects of fluency, complexity, and accuracy of learners' oral productions. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity. The results of the correlation coefficients are presented in Table 9.

The results of the correlation coefficient in Table 9 revealed that there were positive significant correlations between learners' oral proficiency scores and all of the six measures of CAF ($p < .01$). One thing to consider is the size of the value of the correlation coefficient which can range from –1.00 to 1.00. This value will indicate the strength of the relationship between the two variables. Cohen (1988, pp. 79–81) suggested the following guidelines: small $r = .10$ to .29; medium $r = .30$ to .49; large $r = .50$ to 1.0. Therefore, as presented in Table 9, there were large correlations between CAF measures and oral proficiency scores (r is above .5), indicating that all of the six measures were strongly correlated with oral proficiency scores for both groups.

Based on the observed results, it can be concluded that there were strong positive correlations between measures of complexity, accuracy and fluency of Iranian EFL learners and their L2 oral proficiency. As a result, the forth null hypothesis, as there are not any relationships between measures of complexity, accuracy and fluency of Iranian EFL learners' oral productions and their oral proficiency is rejected.

Table 9

Pearson product-moment correlation coefficient

		Oral Proficiency
Complexity	Pearson Correlation	.707**
No of Verb Forms	Sig. (2-tailed)	.000
	Sum of Squares and Cross-products	262.455
	Covariance	6.104
	N	44
Complexity	Pearson Correlation	.663**
Clauses/Subordination	Sig. (2-tailed)	.000
	Sum of Squares and Cross-products	22.860
	Covariance	.532
	N	44
Accuracy	Pearson Correlation	.773**
Error-Free Clauses	Sig. (2-tailed)	.000
	Sum of Squares and Cross-products	1585.354
	Covariance	36.869
	N	44
Accuracy	Pearson Correlation	.841**
Use of Target-like Verbs	Sig. (2-tailed)	.000
	Sum of Squares and Cross-products	1703.080
	Covariance	39.607
	N	44
Fluency	Pearson Correlation	.633**
Syllables Per Minute	Sig. (2-tailed)	.000
	Sum of Squares and Cross-products	765.048
	Covariance	17.792
	N	44
Fluency	Pearson Correlation	.732**
Meaningful Syllables Per Minute	Sig. (2-tailed)	.000
	Sum of Squares and Cross-products	903.228
	Covariance	21.005
	N	44

The underlying construct of complexity, accuracy, and fluency - The last research question addressed the extent to which the six measures used in this study truly determine the three aspects complexity, accuracy, and fluency. In response to this question, hypothesis five was formulated.

Hypothesis 5: The six measures used in this study do not truly represent the three aspects of complexity, accuracy, and fluency.

In order to examine whether the six measures in these three sets of variables truly represent three distinct aspects of language, a principal components factor analysis (stipulating a three-factor solution) was carried out. Prior to the analysis, the suitability of the data for factor analysis was investigated. Inspection of the correlation matrices revealed the presence of coefficients of .70 and above. The Kaiser-Meyer-Oklin value was .835, exceeding the recommended value of .60 (Kaiser, 1974). Barlett's Test of Sphericity reached statistical significance (p = .000), supporting the factorability of each correlation matrix (see Table 10).

Table 10

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.835
Bartlett's Test of Sphericity	Approx. Chi-Square	248.479
	df	15
	Sig.	.000

After checking the suitability of the data for factor analysis, the scores of the six measures were subjected to

principal components analysis. The percentage of variance explained by this three factor solution is shown in table 11. Besides, the results obtained from factor analysis are presented in Table 12, which contains factor loadings.

Table 11

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	4.539	75.658	75.658	4.539	75.658	75.658	3.864
2	.780	13.003	88.661	.780	13.003	88.661	3.296
3	.271	4.518	93.179	.271	4.518	93.179	3.766
4	.187	3.112	96.291				
5	.141	2.355	98.646				
6	.081	1.354	100.000				

Note. Extraction Method: Principal Component Analysis.
a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table 12

Results from factor analysis

	Component		
	1	2	3
Complexity Clauses/Subordination	**.967**	-.067	-.041
Complexity No of Verb Forms	**.854**	.130	-.013
Fluency Syllables Per Minute	.068	**.988**	.101
Fluency Meaningful Syllables Per Min	-.056	**.866**	-.201
Accuracy Error-Free Clauses	.004	.023	**-.966**
Accuracy Use of Target-like Verbs	.329	.089	**-.628**

Note. Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
Rotation converged in 9 iterations.

Table 11 showed that, 93.18 percent of the variance was explained for the two-factor solution. In addition, as Table 11 presented, factor one was made up of two high loadings on measures of the ratio of clauses to amount of subordination (.967) and total number of different grammatical verbs forms (.854). These measures refer to different aspects of complexity and therefore confirm the existence of a general complexity factor. Moreover, Table 12 revealed that the two fluency measures loaded highly on the second factor: number of all syllables produced per minute (.988) and number of meaningful syllables per minute (.866). This suggested that both variables were related to one underlying construct that could be taken to represent fluency.

Finally, results obtained from factor analysis indicated that the third factor is made up of two loaded measures: percentage of error-free clauses (.966) and percentage of target-like use of correct verbs (.628). These two measures loaded highly on this factor and could be taken to define that factor as signifying accuracy. In conclusion, the factor analysis revealed that there were three main factors (complexity, accuracy, and fluency) underlying the six measures used in the present study. Therefore, hypothesis five predicting that the six measures used in this study do not truly represent the three aspects of complexity, accuracy, and fluency is rejected.

5. Discussion and Conclusion

To conclude, analyzing the results obtained from the experimental and the control groups with regard to the effect of DA on the complexity, accuracy, and fluency of oral performance, deductions could be made that DA promoted more complex and accurate oral productions while it did not have any effect on the fluency of learners' oral productions. Therefore, hypothesis 1 and 2 were rejected, and hypothesis 3 was confirmed. Considering the

relationship between measures of complexity, accuracy and fluency of Iranian EFL learners' oral productions and their oral proficiency, results manifested strong positive correlations between these measures and learners' oral proficiency scores. Unquestionably, then, the forth null hypothesis was rejected. Finally, regarding the last research question, the factor analysis revealed that there were three main factors underlying the six measures used in the present study. Therefore, hypothesis five predicting that the six measures used in this study do not truly represent the three aspects of complexity, accuracy, and fluency was rejected.

The three major hypotheses of the present study were that implementation of DA do not have any effect on CAF. The findings did not confirm the all three hypotheses by showing that the differences between DA group which received dynamic assessment and non-DA group which was taught under normal context classroom. The obtained results showed that both measures of complexity and both measures of accuracy were greater in the DA group, indicating that DA resulted in more complex and more accurate oral production. But there was no statistically significant difference between the fluency of learners' oral productions in the DA and non-DA group.

The question may arise as to why there were not any differences between two groups' participants' fluency but there were significant differences between their complexity and accuracy. A plausible interpretation is that although in DA group learners received dynamic assessment but they had to have more concentration on their oral productions than non-DA group so that they could focus on grammatical and length of their oral productions to produce responses which they were asked while receiving eight steps to reach the correct form of response. Moreover, learners were expected to produce the most accurate and complex responses so they ignored the speed of their production and how to improve it simultaneously with their production' accuracy and complexity. Therefore, one of the contributions of this study is that it ponders the next researchers to pay enough attention to learners' fluency simultaneously with their production' accuracy and complexity.

Skehan (2009) believed that complexity, accuracy, and fluency are important aspects of second language performance; according to his findings, accuracy and complexity are simultaneously raised the same as the present study's result. The fourth hypothesis stated that there is no relationship between the complexity, accuracy and fluency of Iranian EFL learners' oral production and oral proficiency. But obtained results revealed that there were strong positive correlations between measures of complexity, accuracy and fluency of Iranian EFL learners and their L2 oral proficiency. The reason that is given is that these three aspects of oral production considered all aspects of improving the oral proficiency and there are not any others aspects. CAF measures are a good starting point for defining language performance.

Finally, the fifth hypothesis formulated for whether CAF in control and experimental group truly represent the three aspects of complexity, accuracy, and fluency or not. A significant progress in the field would thus be the identification of a limited set of standardized measures to be used across studies. It is also important to remember that one has to make choices and that measures are necessarily partial. It is therefore not advisable to group too many notions under the same term which, instead of providing a wider picture, just makes it more blurred. Despite of these limitations for measuring CAF, the findings of present study were obtained exactly by measuring all these three aspects through six different ways. These findings revealed that six measures used in this study (Complexity: number of verb forms and clauses/Subordination, Accuracy: error-free clauses and use of target-like verbs, Fluency: all syllables/min and meaningful Syllables/min) in control and experimental group truly represent the three aspects of complexity, accuracy, and fluency that led to reject the fifth null hypothesis. To do so, some studies, such as the present one need to measure all of them obviously.

In this study participants took an oral test that was conducted by the researcher (instructor), based on Farhady, Ja'farpour and Birjandi (2010, p. 216) as a proficiency test to make sure that there is not any significant differences among them. This research's findings showed that the learner who got the minimum scores in oral proficiency exam did not get the minimum point in all aspects of oral production in final exam somewhat similar to Kozulin and Garb (2002) and Sternberg and Grigorenko (2002). So, this study showed both regression and progression of learners, the same as findings of Abdolrezapour, Tavakoli, and Ketabi (2013). In this regard,

comparing final scores of DA group with non-DA group conforms to Vygotsky's idea in that implementation of DA includes both evolutionary and revolutionary changes, regressions, gaps, zigzags, and conflicts.

In an attempt to remain true to both applied linguistic theory and research on effective language teaching, DA in the present study was situated within a meaningful and authentic context. The materials for the course were based on natural condition of the institute which was appropriate for learners' levels. This research's findings added a new dimension to the available empirical literature of dynamic assessment to second language performance and second language production; previous attempts on using dynamic assessment paid little attention to the details of aspects of production which they are complexity, accuracy and fluency in oral production of learners and also have not paid any attention to the improvement of learners' oral production in terms of three aspects of complexity, accuracy and fluency by integrating it to DA procedures. In this regard, present findings lay a path between CAF and assessing learner's production and establish a solid basis for the integration of details of important aspects of oral production within teaching and assessment tasks and using them to improve one's performance.

The last point is that however this research did not consider this model, but self-monitoring mechanism and covert repair could be the background of DA in this study. Assessment was to give eight hints to learner to reach the correct form of learner oral production. So the learner monitored her speech after each hint. The major contribution that this study makes to the existing literature is the discovery that running DA procedures positively impacts the EFL learners' accuracy, complexity, and has no significant effect on fluency in their oral production. So, learners in DA group produced more accurate and complex as compared to the participants in non-DA group. It was also confirmed that there is a strong relationship between the complexity, accuracy and fluency of EFL learners' oral production and oral proficiency.

5.1 Implications

These findings have some implications for language pedagogy in EFL context. First, EFL teachers can benefit from the new treatment and using DA instead of statistic assessment or testing which is done as usual error correction. Thus it does suggest the usefulness of DA implementation in language learning programs. This technique provides the teacher with a clearer idea of students' learning while also promoting development for students within the class. In addition, teachers could be able either modify or select more appropriate methods, the material of teaching and the type of assessment by monitoring learners and their improvements during the instruction. As Ball and Forzani (2009) argue, teaching is unnatural. Teachers must be taught the importance of providing mediation to their students attuned to the ZPD of those students. Second, the present study has important implications for learners as well. Since learners have to be more accurate about oral productions of themselves and their classmates in order to correct the wrong productions, they could be more qualified for oral production while monitoring themselves and their classmates.

6. References

Abdolrezapour, P., Tavakoli, M., & Ketabi, S. (2012). Emotionalized dynamic assessment as a key to enhancing learners emotion in an L2 context. *Proceedia- Social and Behavioral Sciences, 70,* 323-330.

Abdolrezapour, P., Tavakoli, M., & Ketabi, S. (2013). Qualitative analysis of mediational strategies in emotionalized dynamic assessment of L2 reading comprehension. *International Journal of Research Studies in Language Learning, 6*(1), 1-12.

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task. *Language Teaching Research*, *15*(1), 35-59. http://dx.doi.org/10.1177/1362168810383329

Allal, L., & Pelgrims, A. (2000). Assessment of or in the zone of proximal development. *Learning and Instruction*, *10*(2), 137-152. http://dx.doi.org/10.1016/S0959-4752(99)00025-0

Anton, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, *42*(3), 576-598. http://dx.doi.org/10.1111/j.1944-9720.2009.01030.x

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., & Cohen, A. D. (1998). *Interfaces between second language acquisition and language testing research*. Cambridge: Cambridge University Press.

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

Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing, 13*(3), 257–279. http://dx.doi.org/10.1177/026553229601300303

Berk, L. E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education*. Washington, DC: National Association for the Education of Young Children.

Brown, A. L., & Ferrara, R. A. (1985). Diagnosing zones of proximal development. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (pp. 273-305). New York: CUP.

Campione, J. C. (1989). Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities*, *22*(3), 151-165. http://dx.doi.org/10.1177/002221948902200303

Campione, J. C., & Brown, A. L. (1990). Guided learning and transfer: Implications for approaches to assessment. In N. Frederiksen, R. Glaser, A. Lesgold, & M.G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 141-172). Hillsdale: New Jersey.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Davin, K. J. (2011). *Group dynamic assessment in an early foreign language learning program: Tracking movement through the zone of proximal development.* Unpublished Doctoral dissertation, University of Pittsburgh.

Farhady, H., Ja'farpur, A., & Birjandi, P. (2010). *Testing language skills: From theory to practice* (pp. 186-189). Tehran: Samt.

Holzman, L. (2009). *Vygotsky at work and play*. London: Routledge.

Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31-36. http://dx.doi.org/10.1007/BF02291575

Larsen-freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Journal of University of Michigan, 30*(4), 583-585.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin: Springer Publishing. http://dx.doi.org/10.1007/978-0-387-75775-9

Poehner, M. E., & Lantolf, J. P. (2005). *Dynamic assessment in the language classroom language teaching research.* New York: University Park.

Riggenbach, H. (2000). *Perspectives on fluency*. Ann Arbor: The University of Michigan Press.

Robinson, P. (2007). *Task complexity*. *Theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty*. *IRAL, 45*, 193–213. http://dx.doi.org/10.1515/iral.2007.009

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language, 50*(4), 696–735. http://dx.doi.org/10.2307/412243

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Skehan, P. (2009). Modeling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532. http://dx.doi.org/10.1093/applin/amp047

Skehan, P., & Foster, P. (2005). 'Strategic and on-line planning: The influence of surprise information and task time on second language performance'. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 1-12). John Benjamins. http://dx.doi.org/10.1075/lllt.11.12ske

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*, 439–73. http://dx.doi.org/10.1111/j.1467-9922.2008.00446.x

Xiaoxiao, L., & Yan, L. (2010). A case study of dynamic assessment in EFL process writing. *Chinese Journal of Applied Linguistics, 33*(1), 6-12.