



EFL Textbooks, L2 Contacts, and Teacher Self-Efficacy: Impact on Learners' Development of Oral Complexity, Accuracy, and Fluency

Manoochehr Jafarigohar^{1*}, Mehdi Karami², Zia Tajeddin³, Afsar Rouhi⁴

¹ Associate Professor of TEFL, Department of Teaching English as a Foreign Language (TEFL), Payame Noor University, Tehran, Iran

² PhD Candidate of TEFL, Payame Noor University (PNU), Tehran, Iran

³ Professor of Applied Linguistics, Department of English Language Teaching, Tarbiat Modares University, Tehran, Iran

⁴ Associate Professor of TEFL, Department of Teaching English as a Foreign Language (TEFL), Payame Noor University, Tehran, Iran

Received: 25 January, 2019

Accepted: 11 July, 2019

Abstract

Researchers have extensively compared different L2 learning contexts, such as EFL versus study-abroad, for their impacts on oral production; however, scant attention, if any, has been paid to comparing EFL settings in terms of input factors such as textbooks, amount of contacts in L2, and teachers. Accordingly, the effects of these factors on the oral production skills were investigated in this study. To this end, in a longitudinal study that spanned nearly three months, speech samples were elicited from three groups of Persian speaking advanced learners of English (N = 72) through oral narrative tasks and were scored for complexity, accuracy, and fluency (CAF). A one-way MANOVA was used to compare the means. The average number of subordinate clauses per AS-unit was used to measure grammatical complexity, "D" was a measure of lexical complexity, the percentage of error-free clauses was an index of accuracy, and the number of dysfluencies was calculated to be an indicator of fluency. After a period of time, the results provided strong evidence for the significantly different rates of progress among the learners of the three EFL settings on lexical complexity, accuracy, and fluency. Evaluation of course materials, amount of learners' contact in L2, and teachers' self-efficacy revealed that these different rates of progress might well be attributed to the characteristics of the speaking tasks in the textbooks. One important implication is that gains in a special dimension of oral production can be produced if EFL curriculum developers provide target learners with speaking tasks bearing particular features.

Keywords: Accuracy; Complexity; L2 contact; Fluency; Oral production; Teacher self-efficacy; Textbook

INTRODUCTION

Effective oral production in the second language (L2) requires learners to produce relatively

complex and accurate language, delivered in a fluent manner, and developing this skill is the primary and the most important goal of language teaching (Mora & Valls-Ferrer, 2012). Variabil-

*Corresponding Author's Email:
jafari@pnu.ac.ir



ity in oral production can be attributed to differences in many factors both from inside a dynamic system and from the outside world (Verspoor, Lowie, & van Dijk, 2008); however, L2 acquisition research has often related it to differences in learning contexts (Collentine & Freed, 2004). Thus, it can be predictable that different EFL settings and classrooms might play a key role in variable complexity, accuracy, and fluency (CAF) of oral proficiency among learners since the elements of input are probably different in the settings.

Despite this, it is curious that sources of input rarely if ever features in discussions of factors that influence L2 speech output (Verspoor et al., 2008). Ellis (2009) reviewed the studies on L2 learners' CAF of oral production; nevertheless, he has made no mention of the impact of any of input elements such as textbook, amount of learner contact, and teacher. Hence, exploring the nature of the learning context may be of fundamental importance for gaining an understanding of variation in EFL learners' CAF of oral performance. Thus, the present study aimed to investigate how this variation is caused by textbook, amount of learner contact in L2, and teacher self-efficacy.

Complexity, accuracy, and fluency or CAF have become widely available as dependent variables in L2 speaking research since the 1990s (Housen & Kuiken, 2009). The status of CAF has been justified both empirically and theoretically. Skehan (1998) noted that the three areas of CAF are effective indexes for measuring performance on a particular speaking task. Some factor analysis studies have also confirmed these categories as distinct areas of L2 oral production (Norris & Ortega, 2009). Complexity is characterized as "the ability to use a wide and varied range of sophisticated structures and vocabulary in the L2" (Housen, Kuiken, & Vedder, 2012, p. 1). Accuracy may best be defined as the ability to produce "error-free" (Ellis, 2003, p. 42) and "target-like" (Yuan & Ellis, 2003, p. 2) language. Fluency is the ability to produce lan-

guage in real time with "native-like rapidity" (Housen et al., 2012, p. 2) and without "undue pausing or hesitation" (Ellis & Barkhuizen, 2005, p. 139) or "reformulation" (Ellis, 2003, p. 342).

The present study sought to draw on the findings of studies on the learning context to explore the effects of different sources of input in EFL settings on CAF of oral proficiency. The comparison of the speech data from the London and Tehran (EFL) cohorts in the study by Tavakoli and Foster (2008) showed that the learning context has little or no effect on accuracy or fluency but a clear impact on lexical diversity. The study by Pérez-Vidal and Juan-Garau (2011) analyzed the effects of a period of study abroad (SA) in contrast with the impact of a period of formal instruction previously spent in an EFL context on L2 oral development. Comparing gains revealed a significant improvement as a result of SA in the fluency and accuracy of the participants. In contrast, grammatical complexity only showed a tendency toward significant improvement. Mora and Valls-Ferrer (2012) compared gains in oral production ability obtained after two terms of instruction in the EFL context to those gains obtained after spending the same period abroad. The results showed evidence of significant fluency gains, moderate gains in accuracy, and no gains in complexity occurring during a period of studying abroad, and lack of such gains during the EFL period. Dods (2017) sought to know whether exposure to the L2 through an SA experience can affect gains in complexity, accuracy, or fluency of oral performance. The results indicated greater gains in fluency and naturalness of speech through a higher speech rate, fewer dysfluencies, and greater use of native-like filled pauses and colloquial non-pause fillers. Changes in overall accuracy and complexity of speech, while evident, were less pronounced although.

Taken together, previous research suggests that SA periods yield more substantial gains in the learners' oral proficiency, especially fluency

and accuracy. However, there is still a need for comparing gains in CAF obtained from formal instruction at EFL settings, and the present study targeted this different, largely under-researched area. Thus, the main purpose of the present study was to understand how variability in EFL learners' oral CAF is caused by input factors such as textbook, amount of learner contact to L2, and teacher self-efficacy. To this end, two main questions were formulated in the present study:

1. Do EFL settings differ in terms of the impact that they have on the learners' CAF of oral production?
2. If there is a difference what are the sources of the difference? Textbook, amount of learner contact in L2, or teacher self-efficacy?

METHODS

The present study compared the effects of instruction in different EFL settings on the L2 learners' oral proficiency. A longitudinal design (Zedeck, 2014) was adopted consisting of EFL

textbooks, L2 contacts, and teacher self-efficacy as the independent variables and participants' CAF scores on oral narrative tasks as dependent variables. The study involved an evaluation of course materials, amount of learners' contact to L2, and teachers' self-efficacy in the three EFL settings where the participants had been instructed for years.

Participants

The required sample size for the study was calculated by G*Power (Faul, 2014) to be 72. Accordingly, 72 Persian speaking advanced learners of English, of which 37 were female and 35 male, were assigned to three groups. They had acquired English through 800 hours of formal instruction in three different private language schools, and each group had done it only via one course program from the beginning. Most of the participants ($n = 60$) in the three course programs (Table 1) were undergraduate students, majoring in different subjects. None of them had lived in an English speaking country previously. Their average age at testing was 17.5 years ($SD = 3.51$; range 16-30).

Table 1
Description of EFL Programs and Participants

EFL Program	Year of publication	Levels	Participants' Level	Male (n)	Female (n)
1. <i>New Headway</i>	(2015)	A1-C1	C1	12	12
2. <i>American English File</i>	(2014)	A1-C1	C1	11	13
3. <i>Top Notch & Summit</i>	(2012)	A1-C1	C1	12	12

Materials

Three picture-cued oral narrative tasks were used to elicit speech data from participants during the longitudinal period. In an oral narrative task, the speaker has a cartoon-strip story that can be visually depicted in 6-8 pictures, and the primary aim of the test taker is to produce a coherent story (Fulcher, 2003).

In EFL contexts, the amount of exposure to input and communication is tightly associated to the learners' oral performance (de Bot, Lowie, & Verspoor, 2005; Munoz, 2014); therefore,

Language Contact Profile (Freed, Dewey, Segalowitz, & Halter, 2004) or LCP, a socio-educational background questionnaire, was used to record the participants' amount of contact/interaction in English with native and non-native speakers, and frequency of exposure to English (reading, writing, speaking, listening).

In EFL programs, teacher qualifications affect the oral performance of learners. Therefore, it was decided to have the three teachers in the current study fill a teachers' sense of efficacy survey (Lee, 2009) to examine sets of qualifica-

tions they possessed related to English teaching confidence, personal teaching confidence, attitudes toward English, attitudes toward the current English education policy and practices, and English language proficiency.

In most EFL classroom contexts, up to 90 percent of classroom time is mediated by textbooks (Thornbury, 2014), and they usually determine the kind of methods being used (Richards, 2014). This might affect the learners' oral performance. Therefore, it was decided that all the speaking tasks in the textbooks used in the three EFL contexts should be evaluated using two frameworks for analysis of task characteristics. According to the framework proposed by Skehan (2001), speaking tasks vary in terms of the effects their characteristics have on CAF (Table 2). Tasks with familiar information will lead to greater fluency, since the easy access to information should make only limited demands on attention, allowing material to be assembled for speech more easily. Dialogic tasks are associated with greater complexity and such effects

are due to (1) collective reinterpretation of the task to make it more complex, and (2) scaffolded elaboration of partner's language. These tasks are also associated with greater accuracy because of (1) communication-driven push towards precision, (2) 'creation' of more time to focus on form, and (3) recycling of a partner's language, both with tendency to re-use correct language and to edit it (Skehan, 2001; Skehan & Foster, 1997). The lack of need to engage in large scale planning in tasks which contain clear inherent structure frees up attentional resources for on-line planning and higher fluency (Foster & Skehan, 1996; Skehan, 2001). Tasks also vary depending on the complexity of the decision that has to be made (reproducing the information vs. on-line computation). If tasks need transformations, this will lead to greater complexity as learners have to "wrestle with the need to bring the elements of the task into some sort of meaningful (and non-given) relationship with one another" (Skehan, 2001, p. 180).

Table 2

Summary of the Effects of Task Characteristics on Complexity, Accuracy, and Fluency (Skehan, 2001)

Task characteristic	Accuracy	Complexity	Fluency
Familiarity of information	No effect	No effect	Slightly greater
Dialogic vs. monologic tasks	Greater	Slightly greater	Lower
Degree of structure	No effect	No effect	Greater
Complexity of outcome	No effect	Greater	No effect
Transformations	No effect	Planned condition generates greater complexity	No effect

Skehan (2001) primarily draws on the cognitive load effects on CAF of oral production tasks; Ellis (2003) addresses the contextual factors as well. Therefore, his framework of task features (Table 3) was also reviewed to help us arrive at a working model for the analysis of speaking tasks in the course materials. A number of specific task features were proposed in this framework. The input to the task sometimes takes the form of a picture (contextual support) which must then be communicated verbally to the hearer. Tasks with no contextual support are linked to greater complexity and accuracy be-

cause they are more cognitively demanding. Tasks with contextual support are associated with greater fluency because they are less cognitively demanding (Ellis, 2003). The number of elements and features in a task that need to be manipulated by the speakers are linked to greater complexity because they are more cognitively demanding. They also affect fluency. For example, a story with four females interacting proves more difficult to narrate than a story with only one female and one male character (Ellis, 2003). Shared information tasks typically involve learners in decision-making and arguments over

a case. This argumentation requires conjunctions to mark the relationships between propositions and this benefits complexity of output (Newton & Kennedy, 1996). Split-information tasks, however, result in description. Tasks that pose a single demand favor complexity and fluency. A task that requires learners to describe a route on a map where the route to be taken is marked on the map involves a single task demand and contributes to greater fluency (Ellis, 2003). Open tasks (e.g., role play) are those where the participants know there is no pre-determined solution. They involve a rigid question-and-answer discourse structure and elicit more complex language. In open tasks learners are free to decide on the solution, and this will promote accuracy as well (Ellis, 2003). Closed tasks are those that

require learners to reach a single correct solution and are more associated with greater fluency (Skehan & Foster, 1997). Tasks which contain clear inherent structure (a pre-structured form), especially sequential structure, facilitate task performance by clarifying the macrostructure of the speech event. As a result, the lack of need to engage in large-scale planning frees up attentional resources for on-line planning and higher accuracy and fluency (Foster & Skehan, 1996; Skehan, 2001). Discourse mode is the specific rhetorical structure of an oral task that reflects its overall communicative function. A task with a narrative mode can elicit more complex language because it requires greater use of subordinating constructions than a descriptive commentary (Ellis, 2003).

Table 3
Task Design Features Affecting Learner Production (Ellis, 2003)

Design Variable	Fluency	Accuracy	Complexity
A. Input variables			
1. Contextual support	Tasks with contextual support	Tasks with no contextual support	Tasks with no contextual support
2. Number of elements	Tasks with few elements		Tasks with many elements
3. Topic	Tasks that generate conflict, tasks that are familiar		
B. Task conditions			
1. Shared vs. split information			Shared information tasks
2. Tasks demands	Tasks that pose a single demand		Tasks that pose a single demand
C. Task outcomes			
1. Closed vs. open tasks	Closed tasks	Open tasks	Open tasks with divergent goals
2. Inherent structure of the outcome	A clear inherent structure	A clear inherent structure together with opportunity for planning	
3. Discourse mode			Narrative task > descriptive task > Argument discussion Narrative > argument

Procedure

In response to research question one, a longitudinal study was conducted that spanned nearly

three months. During phase one of the study, participants in each EFL program (language institute) met the researcher (the first author) indi-

vidually, and the oral data were obtained through picture-cued oral narrative tasks. One minute was devoted to preparation to let the participants gather their thoughts about how they would narrate the story depicted in the pictures within 4 minutes. For accurate and quick transcription, the audio-recorded speech data were transcribed in CHAT format using the CLAN (Computerized Language Analysis) software of the CHILDES program (MacWhinney, 2000).

Preparation of the transcripts for coding and analysis began by segmenting each text into AS-units. An AS-unit is “a single speaker’s utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either” (Foster, Tonkyn, & Wigglesworth, 2000, p. 365). The examples in the Appendix taken from Foster et al. (2000) and Czwenar (2014) mark AS-unit and clause boundary clearly.

After that, the researchers graded the oral narrative monologs ($N = 72$) with an analytic grading rubric that included measures of quantifying speaking performance that tapped into different dimensions of CAF. As subordination measures have increasingly become the measure of choice in SLA research (Ellis & Barkhuizen, 2005; Norris & Ortega, 2009), the average number of subordinate clauses per AS-unit was used to allow grammatical complexity to be tapped into. Lexical complexity was measured by means of the “D” index that has been integrated within CLAN and is computable through the VocD program. D ranges between 10 and 100 and a higher value indicates a more diverse text (McKee, Malvern, & Richards, 2000). Accuracy was measured as percentage of error-free clauses, which is defined as “The number of error-free clauses divided by the total number of independent clauses, sub-clausal units and subordinate clauses multiplied by 100” (Ellis & Barkhuizen, 2005, p. 150). Comparing the groups’ accuracy performance included a total of 16 measures of morphological, syntactic, and morphosyntactic structures (e.g., subordinate-

conjunction accuracy). Fluency was measured by means of calculating number of dysfluencies. The total number of partial or complete functionless repetitions/reformulations, restarts, and repairs was divided by the total amount of time expressed in seconds and multiplied by 60 to calculate the number of dysfluencies (Ellis & Barkhuizen, 2005). It should be borne in mind that with fluency a reduction in values clearly represents an improvement. A sample of 10% of the total transcripts (i.e. 74/740 AS-units), randomly selected, was subjected to coding validation by an experienced SLA researcher. The inter-coder agreement, calculated as the percentage of identical scoring, proved to be greater than 95% for each measure.

A period of 45 days intervened between the initial oral narrative task and the second one in phase two. During the period, participants, who had been assigned to one of the three EFL programs, received 25 hours of classroom instruction offered by three different teachers and from three different textbooks. In a similar vein, after another 45 days, phase three included again 25 hours of classroom instruction with the textbooks and another oral narrative task.

In response to the second research question, we undertook evaluations of factors in the EFL settings that helped causing variability in oral production. Thus, Language Contact Profile or LCP (Freed et al., 2004) was used to record the participants’ contact with the target language in each EFL program. The scale had good internal consistency, with Cronbach’s alpha coefficient of .89. In addition, the three teachers in the current study filled in a teachers’ sense of efficacy survey (Lee, 2009) to help us know more about sets of qualifications they possessed. The questionnaire used a Likert scale ranging on a continuum from “Nothing/Not at all” to “A great deal” or “Strongly Disagree” to “Strongly Agree”. Furthermore, 409 speaking tasks from *New Headway*, 532 from *American English File*, and 658 from *Top Notch & Summit* were coded and categorized into the frameworks of task fea-

tures by Skehan (2001) and Ellis (2003). After identifying the features of the speaking tasks, the main orientation of each course material toward the development of oral proficiency could be identified (e.g., fluency-oriented or accuracy-oriented). A sample of 10% of the total corpus, randomly selected, was subjected to coding validation by an experienced SLA researcher. The inter-coder agreement, calculated as percentage of identical coding, proved to be greater than 88%.

Data Analysis

Quantitative analysis of the data included numerical scores obtained from the analysis of oral narratives. Given the nature of the research question one in the present study, as well as the 3-by-3 research design, a one-way between groups MANOVA was performed to compare the overall effect of three EFL contexts in three phases of the study on the average performance of L2 learners in oral narrative tasks and for the dependent variables—CAF. The alpha for achieving statistical significance was set at .05. Analysis of the data also included a one-way ANOVA for comparing the learners' contact with language in the EFL programs. Computing the scores of the teaching efficacy survey and calculating the mean percentage of speaking

tasks with CAF characteristics in the three course materials was also part of data analysis.

RESULTS

The Impact of EFL Settings on CAF

Before proceeding with the MANOVA results of oral narrative tasks, some preliminary assumption tests were conducted for CAF scores. A one-sample Kolmogorov-Smirnov was calculated to assess the normality of distribution of CAF scores. All scores were greater than .05 and normally distributed. Maximum Mahalanobis value was smaller than the critical value (18.47) for all scores suggesting multivariate normality and the presence of no multivariate outliers in our data file. Generating scatterplot matrices between each pair of the dependent variables and separately for each EFL group did not show any obvious evidence of non-linearity. The strength of correlations among scores was checked and most of the coefficients were moderate (.25 to .55). In our data file, Box's M significance value was .032 for phase one scores, and .018 for scores of phases two and three; therefore, the assumption of homogeneity of variance-covariance matrices has not been violated. No serious violations were noted. One-way between-groups MANOVAs were thus performed for CAF scores.

Table 4
Descriptive Statistics for CAF Scores

Index	EFL Program	n	Phase 1		Phase 2		Phase 3	
			M	SD	M	SD	M	SD
Mean number of clauses per AS-unit	New Headway	24	1.66	.32	1.86	.32	1.91	.32
	American English File	24	1.74	.31	1.89	.31	1.94	.31
	Top Notch & Summit	24	1.64	.36	1.84	.36	1.89	.36
D	New Headway	24	40.20	10.81	45.12	10.66	46.64	10.70
	American English File	24	33.16	13.32	34.66	13.32	35.16	13.32
	Top Notch & Summit	24	36.26	7.85	39.26	7.85	40.26	7.85
Percentage of error-free clauses	New Headway	24	62.63	19.97	66.63	19.97	68.63	19.97
	American English File	24	50.03	26.40	51.03	26.40	51.28	26.40
	Top Notch & Summit	24	51.21	18.58	53.21	18.58	54.21	18.58
Number of dysfluencies	New Headway	24	2.37	1.45	2.16	1.40	2.13	1.40
	American English File	24	1.97	1.25	1.17	1.20	1.15	1.18
	Top Notch & Summit	24	2.59	1.57	1.85	1.49	1.82	1.49

To answer the first research question and to probe the impact of EFL contexts on CAF of oral production, an inspection of descriptive statistics (Table 4) was conducted. The mean scores for the mean number of clauses per AS-unit, which was the index of grammatical complexity, indicated that the three EFL groups did not report very different levels of grammatical complexity. The mean scores for D, which was the index of lexical complexity, indicated that, particularly at phases two and three, the three groups became very different, and *New Headway* group obtained the highest mean scores. The mean scores for the percentage of error-free

clauses, which was the index of accuracy, indicated that, particularly at phases two and three, the three groups reported very different levels of accuracy, and the highest mean scores again belonged to the participants in *New Headway* group. An inspection of descriptive statistics for the mean number of dysfluencies, which was the index of fluency, indicated that, especially at phases two and three, the three course groups reported very different levels of fluency. *American English File* and *Top Notch & Summit* groups outperformed *New Headway* group in producing a lower number of dysfluencies (i.e., higher fluency).

Table 5
Tests of Between-Subjects Effects for CAF Scores

Dependent Variable	Index	Test period	Type III Sum of Squares	<i>f</i>	Mean Square	<i>F</i>	<i>Sig.</i>	Partial Eta Squared
Grammatical Complexity	mean number of clauses per AS-unit	Phase 1	.12		.06	.55	.57	.01
		Phase 2	.02		.01	.12	.87	.00
		Phase 3	.02		.01	.12	.87	.00
Lexical Complexity	D	Phase 1	597.44		298.72	2.51	.08	.06
		Phase 2	1318.24		659.12	5.60	.00	.14
		Phase 3	1587.12		793.56	6.73	.00	.16
Accuracy	percentage of error-free clauses	Phase 1	2327.52		1163.76	2.42	.09	.06
		Phase 2	3429.27		1714.63	3.56	.03	.09
		Phase 3	4144.29		2072.14	4.31	.01	.11
Fluency	number of dysfluencies	Phase 1	4.77		2.38	1.15	.32	.03
		Phase 2	12.22		6.11	3.23	.04	.08
		Phase 3	12.15		6.07	3.26	.04	.08

MANOVA results (Table 5) showed that the difference in grammatical complexity scores of oral narratives between the groups did not reach statistical significance neither at the end of phase one, $F(2, 69) = .55, p = .57$, nor at the end of phase two, $F(2, 69) = .12, p = .87$, and phase three, $F(2, 69) = .12, p = .87$. Partial eta squared values were respectively .01, .00, and .00, which according to the commonly used guidelines proposed by Cohen (1988) for interpretation of eta squared scores (.01 = small, .06 = moderate, .14 = large) suggest a quite small effect size. The difference in lexical complexity between the three groups did not reach statistical significance

at phase one, $F(2, 69) = 2.51, p = .08$, but at phase two, $F(2, 69) = 5.60, p = .00$, and phase three, $F(2, 69) = 6.73, p = .00$, it was highly significant. Partial eta squared value was .06 for phase one, but .14 and .16 for phases two and three which suggest large effect sizes. Difference in accuracy of oral production did not reach statistical significance at phase one, $F(2, 69) = 2.42, p = .09$, but at phase two, $F(2, 69) = 3.56, p = .03$, and phase three, $F(2, 69) = 4.31, p = .01$, it was highly significant. Partial eta squared value was .06 for phase one, but .09 and .11 for phase two and three, which suggests quite large effect sizes. As to the fluency of oral production,

difference in mean scores did not reach statistical significance at phase one, $F(2, 69) = 1.15, p = .32$, but at phase two, $F(2, 69) = 3.23, p = .04$, and phase three, $F(2, 69) = 3.26, p = .04$, it was highly significant. Partial eta squared values was .03 for phase one, and .08 for phases two and three, which suggests small effect size for phase one but quite large effect size for phases two and three.

MANOVA results show that over time no difference was noticed in grammatical complexity of oral production between EFL learners being instructed in three different EFL institutes. However, the groups became different in lexical complexity, accuracy, and fluency of oral production. These results left an unanswered question. At the end of the longitudinal period, why were participants instructed in *New Headway* program speaking more accurately and with a higher lexical diversity while learners in *American English File* and *Top Notch & Summit* groups were speaking more fluently? Only an evaluation of the input elements such as textbook, amount of learner L2 contact, and teacher self-efficacy in every EFL program would explain these results.

Table 6
Important LCP Findings

Sub-scales	EFL Contexts						$F(2, 69)$	Sig.	Eta squared
	<i>New Headway</i>		<i>American English File</i>		<i>Top Notch & Summit</i>				
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
days/week speaking	.83	.76	1.13	.85	.92	.65	.93	.39	.02
hours/day speaking	1.21	.41	1.33	.63	1.04	.20	2.48	.09	.06
days/week reading	1.31	1.55	1.06	1.34	1.42	1.18	.42	.65	.01
hours/day reading	1.56	.69	1.38	.72	1.50	.70	.43	.65	.01
days/week listening	2.54	2.12	1.96	1.68	2.13	1.80	.61	.54	.01
hours/day listening	2.17	1.16	1.96	.95	1.88	.68	.59	.55	.01
days/week writing	1.42	1.50	.88	1.15	1.17	1.55	.88	.41	.02
hours/day writing	1.45	.72	1.37	.71	1.41	.65	.08	.91	.00

The second element was teacher efficacy. Comparison of the teachers in the EFL groups was drawn between the raw scores of the teacher efficacy questionnaire. Examination of the most

Evaluation of Input Elements in EFL Programs

In response to the second research question, and to probe why EFL settings differed in terms of the impact they had on EFL learners' CAF of oral performance, the results of evaluation of three elements in *New Headway*, *American English File*, and *Top Notch & Summit* programs are presented here.

The first element in the settings was the amount of exposure to English as a foreign language. An inspection of median scores for different sub-scales of the LCP (Table 6) indicated that participant learners of the three EFL programs did not differ significantly from each other in any of the sub-scales. The median scores for different sub-scales of the LCP were entered into ANOVA (Table 6). There was not a statistically significant difference at the $p < .05$ level in the scores for the three EFL groups, and the actual difference in mean scores was quite small for all sub-scales. Moreover, post-hoc comparisons using the Tukey HSD test indicated that *American English File* group did not differ significantly from either *New Headway* or *Top Notch & Summit* in any of the sub-scales.

important English teaching efficacy factors such as English teaching confidence, personal teaching confidence, attitudes toward English, attitudes toward the current English education poli-

cy and practices, current level of speaking, sex, age, experience, attending in-service training programs, and use of English in English classes

revealed that teachers of the three EFL programs had no significant differences in their teaching efficacy and English proficiency (Table 7).

Table 7
Important Findings of Teacher Efficacy Questionnaire

Sub-scales	<i>New Headway</i> teacher	<i>American English File</i> teacher	<i>Top Notch & Summit</i> teacher
^a English Teaching Confidence	6.5	7	7.5
^b Personal Teaching Confidence	5	4.5	5
^b Attitudes toward English	4	5.5	5
^b Attitudes toward the current English education policy and practices	4	5	4
^b Current Level of Speaking	5.5	5.5	5
Sex	Male	Male	Male
Age	20s	30s	30s
Teaching experience (yrs)	8.5	8	11
Highest degree	MA in TEFL	MA in TEFL	MA in TEFL
Attending in-service training programs	Yes	Yes	Yes
English use in a period of English class	90-99%	90-99%	90-99%

^a max. score = 9; ^b max. score = 6

From these analyses it can be concluded that the three EFL programs were almost equivalent in terms of both the learners' contact to and practice of English and efficacy of the teachers. Thus, it might not have been any learner or teacher factor that had caused the differential gains in CAF of oral production among the three groups of participants.

The third element was the textbook. Analysis of speaking tasks in the course materials of the three EFL programs using Skehan's (2001) and Ellis's (2003) frameworks supported some generalizations. Table 8 displays the number and

percentage of tasks that have been classified as contributing to the complexity of oral production. The mean percentages of tasks with complexity characteristics were not very different in the course materials ($M = 40.21\%$, 40.78% , 40.23%). These results show that the amount of tasks which target complexity of oral production was almost equal in the three textbooks, and that no difference in grammatical complexity between the three groups of EFL learners reported in the MANOVA results (Table 5) might have been due to the fact that the three course materials put the same emphasis on complexity of oral production.

Table 8
Comparing Speaking Tasks in the Textbooks for Their Contribution to Complexity

	<i>New Headway</i>		<i>American English File</i>		<i>Top Notch & Summit</i>	
	<i>n/N</i>	%	<i>n/N</i>	%	<i>n/N</i>	%
Dialogic tasks	252/409	61.61	323/532	60.90	490/658	74.46
Tasks that need transformations	250/409	61.12	453/532	85.16	599/658	91.07
Tasks with no contextual support	139/409	33.82	182/532	34.24	154/658	23.40
Tasks with many elements	208/409	50.85	265/532	49.81	281/658	42.70
Shared information tasks	99/409	24.20	95/532	17.85	94/658	14.28
Tasks that pose a single demand	215/409	52.43	280/532	52.63	384/658	58.35

Open tasks	118/409	28.85	103/532	19.36	96/658	14.58
Narrative tasks	36/409	8.83	34/532	6.35	20/658	3.03
Mean percentage		40.21		40.78		40.23

Note: n/N = proportion of tasks with particular characteristics to total number of tasks in the course; % = percentage of tasks with particular characteristics in the course

Table 9 displays the number and percentage of tasks in each course material that have been classified as contributing to accuracy. The difference in the mean percentages of tasks with accuracy characteristics in the course materials was striking ($M = 44.15\%$, 37.55% , 43.15%). These results show that the proportion of dialogic tasks, tasks with clear inherent structure, tasks

with no contextual support, and open tasks which target accuracy of oral production was different in the three textbooks, and that the difference in accuracy between the three groups of EFL learners reported in the MANOVA results (Table 5) might have been due to the fact that *New Headway* contains more accuracy oriented speaking tasks.

Table 9

Comparing Speaking Tasks in the Textbooks for Their Contribution to Accuracy

	<i>New Headway</i>		<i>American English File</i>		<i>Top Notch & Summit</i>	
	n/N	%	n/N	%	n/N	%
Dialogic tasks	252/409	61.61	323/532	60.90	490/658	74.46
Tasks with clear inherent structure	214/409	52.32	190/532	35.71	396/658	60.18
Tasks with no contextual support	139/409	33.82	182/532	34.24	154/658	23.40
Open tasks	118/409	28.85	103/532	19.36	96/658	14.58
Mean percentage		44.15		37.55		43.15

The number and percentage of speaking tasks in each course material that have been classified as contributing to fluency are shown in Table 10. The difference in the mean percentages of tasks with fluency characteristics in the course materials was also great ($M = 56.34\%$, 59.33% , 71.70%). These results show that the amount of

tasks which target fluency of oral production was different in the three textbooks, and that the difference in fluency between the three groups of EFL learners reported in the MANOVA results might have been due to the fact that *American English File* and *Top Notch & Summit* put greater emphasis on fluency of oral production.

Table 10

Comparing Speaking Tasks in the Textbooks for Their Contribution to Fluency

	<i>New Headway</i>		<i>American English File</i>		<i>Top Notch & Summit</i>	
	n/N	%	n/N	%	n/N	%
Tasks with familiar information	355/409	62.34	463/532	87.03	608/658	92.40
Tasks with clear inherent structure	214/409	52.32	190/532	35.71	396/658	60.18
Tasks with contextual support	208/409	50.92	269/532	50.56	506/658	76.89
Tasks with few elements	200/409	48.89	266/532	50	378/658	57.44
Tasks that pose a single demand	215/409	52.43	280/532	52.63	384/658	58.35
Closed tasks	291/409	71.14	426/532	80.07	559/658	84.95
Mean percentage		56.34		59.33		71.70

The results of the task analysis were in close agreement with the MANOVA findings, and they suggest that the main cause of variability at least in accuracy and fluency of oral production among the learners of the three EFL programs was the course materials.

DISCUSSION

The main objective of the current study was to explore how and why different EFL programs (*New Headway*, *American English File*, and *Top Notch & Summit*) affect the development of L2 learners' oral performance differently. Two main questions were addressed: Do EFL settings differ in terms of the impact that they have on the learners' CAF of oral production? What are the sources of the difference: textbook, amount of learner contact in L2, or teacher self-efficacy?

In response to research question one, the effect of three EFL programs on CAF of oral production was investigated. No difference was observed in grammatical complexity, but the findings revealed that after a period of time the *New Headway* participants had a better performance in lexical complexity. Although detailed studies that focus on the relation between the contextual variables and variation in oral CAF are rare, some of the similar studies (Mora & Valls-Ferrer, 2012; Pérez-Vidal & Juan-Garau, 2011) found that over the years studying abroad, grammatical complexity only showed a tendency toward significant differences. Other studies showed that learners in the SA context used significantly more diverse vocabulary than learners in the EFL context (Tavakoli & Foster, 2008). The findings of the present study revealed that the learners in the *New Headway* program had a better performance in accuracy as well, and this is consistent with the results of some of the studies that found learning a language in SA contexts causes difference in accuracy (Pérez-Vidal & Juan-Garau, 2011). As to fluency, the findings of the present study revealed that *American English File* and *Top Notch & Summit* programs had better results. When Mora and Valls-Ferrer (2012) and Dods (2017) exposed their participants to SA context, their

fluency of oral production improved significantly. In response to the second research question, the findings of the longitudinal study were analyzed in light of the nature of tasks used in the course materials to elicit oral production since there are indications that characteristics of speaking tasks lead to selective improvements in particular areas of oral production—i.e., CAF (Skehan, 2001). *In close agreement with the results of oral narrative tasks*, the analysis of speaking tasks in the three course materials used in the three EFL programs revealed that mean percentage of speaking tasks with complexity characteristics was almost equal in the three textbooks (see Table 8), and consequently grammatical complexity appeared similar between the groups. The gained results of oral narrative tasks in the longitudinal study suggested that *New Headway* outperformed the other groups in lexical complexity. One explanation for this unexpected finding might be related to pre-task activities. Foster and Skehan (1996) found that providing participants with some guidance for pre-task planning was linked with greater complexity of language during the task. Thus, tasks which contain pre-task vocabulary activities seem to favor lexical complexity of oral production. Analysis of the speaking tasks in the textbooks also revealed that a better performance in accuracy on the part of *New Headway* learners can be presumably due to a considerable proportion of speaking tasks in the course with characteristics that encourage L2 learners to produce grammatically and lexically accurate language (see Table 9). Less proportion of dysfluencies (i.e., higher fluency) during production of oral narratives on the part of learners in *American English File* and *Top Notch & Summit* programs again can be attributed to the nature of tasks used in the course materials. The results of the task analysis revealed that most speaking tasks in these courses have mainly fluency features (see Table 10).

Based on the results of the present study, an intimate understanding of context-related variability in speaking was called for in the realms of research, theory, and practice in SLA. The find-

ings, although significant, have some limitations. The longitudinal period of this study spanned only two terms (3 months) which was not long enough to discover the actual impact of different EFL contexts on variability in oral production. Therefore, a longer longitudinal study of three different EFL programs, beginning immediately after the acceptance of beginner learners of English and ending right after they become advanced learners in the programs, may possibly provide more complete answers to the complex relationships between EFL program and variability in CAF of oral production.

CONCLUSION

The results of this study provide strong evidence for the impact of the L2 learning contextual and input elements on variability in oral production. The results suggest that some EFL programs have a positive impact on the learners' accuracy of oral production and some others do so for fluency. This varied contribution is more due the course materials used in these programs. There is a considerable proportion of speaking activities in one course which encourage accuracy of oral production (e.g., dialogic tasks), while in another there is a wealth of tasks with features that target an improvement in fluency (e.g., tasks with contextual support). The overall conclusion is that when learners in EFL contexts are similar in the amount of contact to L2 input and communication and when their teachers have similar qualifications, differences in their CAF of oral production might be attributed to the major source of input in the EFL context (i.e., course materials).

The findings of this study bear far-reaching implications for material developers, instructors, and SLA researchers and address a number of pedagogic issues related to EFL contexts and EFL course programs. One implication of this study is that task-based material developers should provide the learners with speaking tasks with particular characteristics if they aim at achieving considerable gains in a special dimension of oral performance. Indeed, for them this would mean that ELT materials should provide

learners with tasks with familiar information, tasks with clear inherent structure, tasks with contextual support, tasks with few elements, tasks that pose a single demand, and closed tasks if, for example, the main emphasis is on fluency. The greater the amount of practice of these tasks, the more fluent speaking performance occurs.

In addition, the findings of this study confirm the theoretical propositions of L2 variation (Larsen-Freeman & Long, 1991; Towell & Hawkins, 1994) which attributed variability in language learning to differences in learning environments. SLA theorists and researchers should begin to recognize course materials as one of the most important elements in EFL contexts that greatly affect learners' oral performance because they present learners with speaking tasks that have various demands and characteristics that differentially affect the aspects of L2 oral production (i.e., CAF).

References

- Collentine, J., & Freed, B. F. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition*, 26(2), 153-171. doi: 10.1017/S0272263104262015
- Czwenar, I. (2014). *Analysing spoken language for complexity, accuracy and fluency: Some methodological considerations*. Cham: Springer International Publishing.
- de Bot, K., Lowie, W., & Verspoor, M. (2005). *Second language acquisition: An advanced resource book*. London: Routledge.
- Dods, A. (2017). Changes in complexity, accuracy and fluency in a Japanese first year university student's oral production after study abroad. *Bulletin of Toyo Gakuen University*, 25, 279-295.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. (2009). The differential effects of three types of task planning on the fluency,

- complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30(4), 474-509. doi: 10.1093/applin/amp042
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Faul, F. (2014). G*Power (Version 3.1.9.2). Germany: University of Kiel.
- Foster, P., & Skehan, P. (1996). The Influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18(3), 299-323. doi: 10.1017/s0272263100015047
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354-375. doi: 10.1093/applin/21.3.354
- Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, 26(2). doi: 10.1017/s027226310426209x
- Fulcher, G. (2003). *Testing second language speaking*. London: Longman.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473. doi: 10.1093/applin/amp048
- Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency* (pp. 1-20). Amsterdam: John Benjamins.
- Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London: Longman.
- Latham-Koenig, C., & Oxenden, C. (2014). *American English file* (2nd ed.). New York, NY: Oxford University Press.
- Lee, J.-A. (2009). *Teachers' sense of efficacy in teaching English, perceived English language proficiency, and attitudes toward the English language: A case of Korean public elementary school teachers*. (Doctoral dissertation), The Ohio State University. Retrieved from <https://etd.ohiolink.edu/>
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). New York, NY: Psychology Press.
- McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15(3), 323-338. doi: 10.1093/lc/15.3.323
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46(4), 610-641. doi: 10.1002/tesq.34
- Munoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35(4), 463-482. doi: 10.1093/applin/amu024
- Newton, J., & Kennedy, G. (1996). Effects of communication tasks on the grammatical relations marked by second language learners. *System*, 24(3), 309-322. doi: 10.1016/0346-251x(96)00024-3
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555-578. doi: 10.1093/applin/amp044
- Pérez-Vidal, C., & Juan-Garau, M. (2011). The effect of context and input conditions on oral and written development: A Study abroad perspective. *International Review of Applied Linguistics in Language Teaching*, 49(2), 157-185. doi: 10.1515/iral.2011.008
- Richards, J. C. (2014). *The ELT textbook*. New York, NY: Palgrave Macmillan.
- Saslow, J. M., & Ascher, A. (2012). *Summit 2* (2nd ed.). White Plains NY: Pearson Education.
- Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*,

- 18, 268-286. doi: 10.1017/s0267190500003585
- Skehan, P. (2001). Tasks and language performance assessment. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 167-185). London: Longman.
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1(3), 185-211. doi: 10.1177/136216889700100302
- Soars, L., Soars, J., Hancock, P., & Williamson, M. (2015). *New Headway English course* (4th ed.). Oxford: Oxford University Press.
- Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58(2), 479-473. doi: 10.1111/j.1467-9922.2008.00446.x
- Thornbury, S. (2014). English language teaching textbooks: Content, consumption, production. *ELT Journal*, 69(1), 100-102. doi: 10.1093/elt/ccu066
- Towell, R., & Hawkins, R. (1994). *Approaches to second language acquisition*. Clevedon: Multilingual Matters.
- Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214-231. doi: 10.1111/j.1540-4781.2008.00715.x
- Yuan, F., & Ellis, R. (2003). The Effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24(1), 1-27. doi: 10.1093/applin/24.1.1
- Zedek, S. (2014). *APA dictionary of statistics and research methods*. Washington, DC: American Psychological Association.

Biodata

Dr Manoochehr Jafarigohar is an associate professor of TEFL at Payame Noor University, Tehran, Iran. He teaches research and second language acquisition at post-graduate level. His research interests include CALL, research methodology, and second language acquisition. He has authored numerous textbooks and papers and has presented in conferences worldwide.

Email: jafari@pnu.ac.ir

Mr Mehdi Karami is a PhD candidate of Teaching English as Foreign Language (TEFL) at Payame Noor University, Tehran, Iran. He is currently teaching English in private language schools and state schools of Hamadan, Iran. His research interests include task-based language teaching (TBLT), computer assisted language learning (CALL), and second language teacher education.

Email: karami_m58@yahoo.com

Dr Zia Tajeddin is a professor of Applied Linguistics at Tarbiat Modares University, Iran, where he teaches doctoral courses in Interlanguage Pragmatics and Second Language Teacher Education. He is the co-editor of *Applied Pragmatics* (John Benjamins) and editor of *Journal of Second Language Teacher Education*, and sits on the editorial/review boards of journals such as *RELC Journal* and *TESL-EJ*. His research interests center on L2 pragmatic instruction and assessment, classroom discourse analysis, sociocultural theory, teacher identity and cognition, and EIL/ELF. He has presented papers in international conferences and published papers in many journals, including *The Language Learning Journal*, *Journal of Intercultural Communication Research*, *RELC Journal*, *Australian Journal of Teacher Education*, *Language and Intercultural Communication*, *The Asia-Pacific Education Researcher*, *TESL-EJ*, and *TESL Canada Journal*.

Email: zia_tajeddin@yahoo.com

Dr Afsar Rouhi is an associate professor of TEFL at Payame Noor University, Tehran, Iran. He teaches at post-graduate level, and his current research interests include form-focused instruction, individual differences, and task-based instruction.

Email: afsar.rouhi@gmail.com