INTERNATIONAL JOURNAL OF
Practical and Pedagogical Issues in English Education
ISSN : 2980-9533 Homepage: www.ijpie.org 3(3), Sep. 2025

The Effect of Length Constraint on Task-Based L2 Narrative Retellings: A Study of EFL Learners' CALF Performance

Massoud Yaghoubi-Notash[*1], Shahabaddin Behtary[2], Sina Gobadi[3], Shiva Shahbazi Mojarrad[4]

[1] *English Language and Literature Department, University of Tabriz, Tabriz, Iran*

[2] *Department of English, Ard. C., Islamic Azad University, Ardabil, Iran*

[3] *Self-employed Researcher, Ardabil, Iran*

[4] *Freelance Researcher, Ardabil, Iran*

Corresponding Author's Email: masoud.yaghoubi@gmail.com

KEYWORDS

Length, Length Constraint, Narrative Retellings, Task, Variability

ARTICLE INFO

Article type: Research Article

Article history:
Received: 24 March 2025
Revised: 01 August 2025
Accepted: 07 August 2025
Published online: 07 August 2025

ABSTRACT

Task as a promising area of study has contributed to a wealth of the literature in the field of English as a Foreign Language (EFL) over the years. Length constraint as a task demand is one of the under-investigated areas of task performance. It refers to the restriction placed on the amount of spoken content, encouraging brevity and conciseness. For the purpose of the study, 23 EFL university students (10 male and 13 female) participated in the study. Adopting a within-subjects design, an oral reproduction text was chosen as an instrument in the form of a narrative retelling task to compare the participants' task-prompted complexity, accuracy, lexis and fluency across length-constrained and non-length-constraint performances. MANOVA results revealed that length constraint as the independent variable influenced EFL learners' narrative retellings when the four dependent variables of complexity, accuracy, lexis and fluency were taken together. However, length constraint was shown to have variable effect on the aspects of retelling performance. Positive significant differences were obtained regarding promoting accuracy and fluency, but not about complexity and lexis. Implications of the study are discussed in the light of theoretical and empirical claims in the literature.

Introduction

Ever since its phenomenal emergence on the scene of ESL/EFL, task has continued to stimulate research, inspire practice, and contribute to reflections on how to teach and learn second or foreign languages. Task as a realistic, process-based linguistic performance targets the pivotal point of 'naturalistic' acquisition as formulated by Krashen (as cited in Ellis et al, 2020). Shortly after being introduced, it broke free from merely pedagogic boundaries and turned into an autonomous area in its own right. Such a proliferation renders task a dynamic nature such that a hard and fast definition would be difficult to achieve. Quite many definitions and characterizations of task have been examined by Van den Branden (2006a). What underlies the entire set of definitions is the distinction regarding task between learning-oriented versus pedagogy-oriented accounts (Van den Branden, 2006b). On the other hand, a famous distinction is task as work-plan and task as activity. With the latter, the concept of learning process is implied which seems to be flawed since it would be impossible to define a concept in terms of an unpredictable outcome. Therefore, task as work-plan appears to be a stronger basis for characterizing task.

How to Cite: Yaghoubi-Notash, M., Behtary, S., Gobadi, S., Shahbazi Mojarrad, S. (2025). The Effect of Length Constraint on Task-Based L2 Narrative Retellings: A Study of EFL Learners' CALF Performance *International Journal of Practical and Pedagogical Issues in English Education*, 3(3), 21-37. DOI: 10.22034/ijpie.2025.513753.1094

Instead of providing a strict definition of a task, as is common with concepts in the human sciences, it may be more effective to describe it through its key characteristics. Ellis and Shintani (2014) provide four basic criteria for task-as-workplan, i.e. a) primary focus on meaning, b) gap in communication or knowledge, c) learner reliance on their linguistic and non-linguistic resources, and d) a clearly defined communicative outcome. Once the criteria are established, various task types can be identified, e.g. one-way vs. two-way, closed vs. open, convergent vs. divergent, and so forth.

Indeed, as soon as task is adopted for pedagogical practice, an important issue that emerges is determining the right sequence of tasks to ensure learning. Sequencing, as a generic problem in all approaches for any kind of input, offers a lot of challenges (Ellis, 2018). With primarily language-focused approaches, linguistic complexity and difficulty can be a convenient criterion for complexity, which in turn can be the basis for sequencing. However, when it comes to tasks, linguistic criteria are less than adequate since the type and nature of task-induced linguistic output are neither predictable nor controllable. What makes a difficult or complex task involves elements that are beyond simply linguistic in nature. As Ellis et al. (2020) put it, "tasks are conglomerates of variables and complexity is therefore influenced by the intersection of countless variables in ways that may make codification difficult if not impossible" (p. 14).

TBLT researchers, despite diverging perspectives on task, have little doubt that the issue of task-complexity "is a complex and multidimensional construct" (Awwad & Tavakoli, 2019, p. 3). A multitude of variables including task features, content, learner cognition, individual learner differences, etc. are involved which call for a theoretical framework rather than a simple categorization (see Robinson, 2007; Sasayama, 2016; Skehan, 2014; Vasylets et al., 2017). Among a few of the most proliferating theoretical frameworks is cognition hypothesis.

Following cognitive hypothesis, attentional capacities are limited; therefore, any activation of the resources for language production would need to act competitively (Ellis, 2024). In other words, a gain in terms of accuracy for instance, would amount to a sacrifice regarding fluency or complexity.

On the other hand, an idea that is explicitly or implicitly supported is that unlike the input which is more or less controllable, output or performance is not predictable. To put it differently, a vast range of factors "dynamically influence L2 performance. These factors can be cognitive, instructional and interactional in nature" (García-Ponce et al., 2022, p. 124). The present study raises the notion of constraint as a cognitively binding factor that shapes performance regarding yet another dimension of performance, which is length. The research has taken place in the Iranian context where EFL learners often face unique challenges in oral production due to limited authentic exposure to English and classroom instruction highlights the primacy of accuracy over fluency. The Iranian context provides ideal research setting where narrative as a typical genre of oral discourse can be examined for variability in L2.

Review of Literature

Length constraint regarding tasks and its influence in determining the oral performance in L2 has been intriguing enough in task-based language teaching (TBLT) research; however, such a variable has not been addressed independently. Robinson's Cognition Hypothesis and the Triadic Componential Framework are by far the most dominant frameworks in formulating tasks (Xing, 2019). Robinson (2001, p. 29) defines task complexity as "the result of attentional,

memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner". TC (Task Complexity) in Robinson's TCF (Triadic Componential Framework) is divided into resource-directing and resource-dispersing dimensions. The former refers to cognitive, conceptual requirements, including few/many elements, here-and-now/there-and-then, +/-reasoning demands, while the latter is associated with procedural and performance, learner-oriented requirements, including +/-planning, task structure, single/dual task, +/-prior knowledge.

According to Luo (2022):

> increasing task complexity in the resource-directing dimension leads to higher accuracy and complexity of L2 performance, but affects fluency, while increasing task complexity in the resource-dispersing dimension prevents learners from making full use of their existing L2 knowledge, thus affecting the quality of L2 performance (p. 152).

The three dimensions of performance or CAF which stands for complexity, accuracy and fluency continue to be explored and inform a wealth of literature in L2 studies.

Following Robinson's characterization (Robinson, 2016), complexity along the resource-directing dimension has been extensively studied. Kuiken and colleagues conducted several studies involving elements, and reported an increase in learners' linguistic accuracy, but no significant effect on syntactic or lexical complexity in complex tasks involving numerous elements (Kuiken et al., 2005; Kuiken & Vedder, 2007, 2012). Wang (2013) revealed that lexical complexity did increase in more complex tasks, but there was no significant change in syntactic complexity. Frear and Bitchener (2015) showed that more complex vocabulary and a greater number of adverbial clauses tended to be employed in tasks with more elements. Cho (2015), concentrating on two variables — elements and reasoning demands — demonstrated that fluency was higher in high-complexity tasks, but accuracy and syntactic complexity did not vary significantly. Similarly, Jin et al. (2020) indicated that attentional resources are allocated competitively between language accuracy and complexity. Findings revealed that the more difficult the task, the greater the proportion of attentional resources allocated to accuracy, whereas complexity received less attentional allocation.

The resource-dispersing dimension involves variables such asprior knowledge and planning (Divsar, 2017). Prior knowledge, also known as topic familiarity, was found by Adams and Nik (2014) to be more diverse in the vocabulary used by writers with less prior knowledge. Divsar (2022), in a very influential study focused on planning types (i.e., rehearsal, strategic, and online) and learners' writing, reported that rehearsal planning produced higher levels of complexity, accuracy, fluency, and lexical variety than the other types.

Aside from time and length constraints, the structure of the task may also result in L2 performance variability. Zohrabi and Hasanpour (2020), for instance, demonstrated that closed tasks — very much like length-constrained retellings —resulted in higher accuracy but lower syntactic complexity. On the other hand, open tasks led to greater lexical diversity at the cost of filtering out erroneous forms. This may indicate that task design acts as a mediating variable in attentional allocation, with constrained conditions tending to facilitate fluency and accuracy (Skehan, 2009), but less restrictive tasks may foster complexity (Robinson, 2001). Against this background, the present study's focus on length constraints is contextualized, as they may

function like other task-limiting variables, influencing priority-based cognitive access to certain performance aspects over others.

As already mentioned, the length of the task does not appear to be addressed frequently enough as a variable; therefore, the findings are neither conclusive nor exhaustive. Behtary and Yaghoubi-Notash (2009) investigated the effect of brevity demand as a task planning requirement on lexical density, lexical diversity, and grammatical accuracy. For the first two, awareness of length constraint as a planning factor resulted in significant changes in fluency and complexity, but not in grammatical accuracy. Divsar (2022) supported the privileged position of length of production in rehearsal planning as a subset of fluency. More recently, Skehan and Foster (2020) stated that limited attentional capacity forces learners to prioritize either fluency/accuracy or complexity/lexical diversity for shorter and longer tasks, respectively. Although not a very explicit finding, the theoretical account, which builds upon their earlier formulation of an exemplar-based system (Skehan, 2003), is insightful. Along the same lines, Jackson and Suethanapornkul (2023) noted that constrained in narrative tasks resulted in gains in accuracy at the cost of grammatical complexity, supporting the limited attention hypothesis in cognitive accounts of task-elicited performance.

Lambert and Kormos (2024), nevertheless, challenge this competing trade-off and suggest an alternative account: within a 'dynamic systems perspective,' length constraints interact unpredictably depending on learners' proficiency. They reported that advanced learners were more inclined to safeguard complexity under any constraint by actively using less demanding and more automatized structures; however, for beginners, fluency was achieved at the cost of less diversified production.

The role of working memory in mediating length-constrained performance is also duly noted in the literature. Saito and Hanzawa (2022) reported that high working memory learners produced more lexically diverse narratives under limits, whereas low working memory learners relied on ready-made, formulaic linguistic units. Michel and Kormos (2023) contribute to the literature by adding finer-grained dimensions, arguing that task features may govern the degree to which working memory is engaged. They showed, for instance, that constrained retellings, as opposed to original narratives, represented a lower burden on cognitive processing, as the provision of schemata played a facilitative role, thus liberating attentional resources for greater fluency.

De Jong and Perfetti (2021) showed that failures in producing fluent task-elicited utterances, such as increased pausing, tended to emerge in strictly constrained tasks, as learners' cognition was engaged in actively retrieving vocabulary. Such a finding lends support to the use of scaffolded constraints in striking a balance between delivery speed and utterance accuracy.

There are also correlational links between narrative structures and performance quality in CALF tasks. Tavakoli and Foster (2021) focused on how the rhetorical structure of the story can influence performance, arguing that shorter retellings resulted in grammatical avoidance but included a higher frequency of discourse markers to ensure cohesion in L2 production. Hsu (2023) supports these findings, observing that lexical diversity sharply decreased in retellings lasting less than a minute but was more stable when more time was allotted, suggesting a critical threshold for task performance condition.

Kormos and Trebits (2022) compared monologic and dialogic narration tasks and indicated that length constraints place greater burdens on monologues due to the absence of

collaborative planning. Their findings advocate for interactive retelling tasks to mitigate fluency losses.

Individual differences may also correlate with constraint-based variations in L2 narratives. Suzuki and Sunada (2023) detected improvements in fluency upon task repetition. This suggests that learners tend to align their production with constraint condition through intact retrieval of preplanned phrases for the ultimate goal of a high-speed delivery pace, without having to sacrifice accuracy. On the other hand, Sasayama and Malicka (2022) conducted a meta-analysis concluding that loose constraints benefitted low-proficiency learners more effectively, as more flexibility enabled them to retrieve lexical items and engage in self-repair.

Against this background, the present study addresses the following research question: Does length constraint affect EFL learners' narrative retellings in terms of a linear combination of complexity, accuracy, lexis and fluency?

Method

This study followed a within-subjects design in which the participants were not selected randomly, rather, they were chosen as an intact group. The study intended to investigate the effect of length constraint on the complexity, accuracy, lexis, and fluency of their oral reproductions.

Participants

The participants of this study were initially 23 junior English Language Teaching (ELT) students of Islamic Azad University in Ardabil. There were 10 male students and 13 female students participated in this study and the range of their age was between 19 to 26. They spoke two languages, Azeri Turkish and Persian, which were their mother tongue and official language, respectively. The participants have passed some basic courses in English namely: Pre-requisite English, Reading Comprehension, Grammar and Writing, Speaking and Listening, Study Skills, Advanced Grammar, Principles and Methods of Translation, Samples of Simple English Prose, Advanced Writing, Topic-based Conversation, Phonology, and Idiom Use in Translation. Table 1 presents the details of the participants.

Table 1

Profile of the Participants

Total participants	23
Gender distribution	10 Male, 13 Female
Age range	19–26 years
Languages spoken	Azeri Turkish (mother tongue), Persian (official language)
Field of study	English Language Teaching (ELT)
University	Islamic Azad University, Ardabil
Academic level	Junior students

Materials and Instruments

The text used in this study was a single-page story, number 57 from the book *Advanced Stories for Reproduction* (see Appendix). To select the text, first five books that were usually taught in Iranian universities for oral reproduction courses were selected. Afterwards, 10 units of those books were selected randomly and their difficulty levels were calculated using Flesch reading ease formula. After calculating the average Flesch reading ease score, the afore-mentioned text was chosen due to its score being close to that of the mean.

Flesch Reading Ease Score was developed by Rudolph Flesch in 1948 as the result of some efforts to improve the readability of newspaper. It varies between 0, for very difficult texts, and 100, for those which can easily be understood (Readable.com, 2025). The score for the above-mentioned story was calculated to be 84.8, indicating that the chosen text was easy to read and comprehend.

Another index to measure readability is Flesch-Kincaid grade level which was developed by Rudolf Flesch and J. Peter Kincaid. Despite the vague nature of understanding Flesch reading score, Flesch-Kincaid grade level provided a clear interpretation of the scores (Readable.com, 2025). To clarify, it is claimed that Flesch-Kincaid grade level is a conversion of Flesch reading ease score and its range is between 0, for very easy texts, and 18, for very difficult ones. The calculated index for the chosen text was 6.2.

The story was about a strange man who enters a village and catches the wild pigs with no weapon, using his wisdom, which amuses the villagers greatly. The story consisted of 349 words, 1377 characters excluding the spaces, and was organized in 6 paragraphs, 20 lines, and 18 sentences.

Procedures

The participants were divided into six groups and invited to the researchers' office. First of all, they were supposed to read the text for ten minutes. They were allowed to have four minutes as the first planning time. In their first oral reproduction, they were asked to reproduce the text as much in detail as possible for five minutes. Here there was no constraint for length. Then they were given the second four minutes of planning time. Next, they were supposed to reproduce the text as briefly as possible within five minutes. Here, the length constraint was imposed. The reproductions were all recorded by a mobile phone's voice recorder, to be exact, the brand of cell phone was i-phone 5S. One of participants had not pushed the record button correctly so his data was lost.

All the times used for reading the text, planning, and reproduction were calculated based on a plot study of five random cases from other students who were in the same level of language proficiency as the participants. The mentioned five students were asked to read, plan, and reproduce the text and the average amount of time that they needed were ten minutes for reading, four minutes for planning, and five minutes for reproduction. If the pilot test had not been carried out at all, no effect would have appeared because the data is gathered from the same participants twice. However, it was implemented painstakingly.

While reproducing the texts, the participants were not allowed to consult a dictionary or use any other resources. Having finished the recording of the data, all of the recordings were transcribed and checked twice. Five random recordings were checked for the accuracy of transcriptions by a colleague. The transcriptions even included the meaningless sounds produced

by the participants making sure everything is fully accurate to pave the way for any further study.

Complexity, accuracy, lexical diversity, and fluency (CALF) were calculated when all of the recorded files were checked and transcribed. To ensure inter-rater reliability, ten random cases were rated for CALF and also a colleague checked another random group of ten students ensuring that the given scores are accurate.

Data Collection

Complexity was measured by calculating the ratio of subordination, length of AS (Analysis of Speech) unit (Foster et al., 2000) and clause length as cited in Tavakoli et al. (2017). Foster et al. stated that "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" (p. 12).

To calculate accuracy, Foster and Wigglesworth (2016) proposed a more systematic approach to measuring clause-level accuracy, a weighted clause ratio (WCR) measure, which differentiates errors at different levels and distinguishes between the errors that impede communication, those that impair communication, and those that do not impair communication at all. Accuracy was measured within the mentioned framework. Percentage of correct use of verbs was used as a local measure of accuracy.

To demonstrate change in lexical diversity, measures of vocabulary diversity (D) (Malvern & Richards, 2002), and measure of textual lexical diversity (MTLD) were used.

The fluency measures in this study were mean length of run, speech rate, number of silent pauses clause internal and clause external, and a composite repair measure (Tavakoli et al., 2017).

Inter-rater reliability was calculated for all of the indexes except lexis. Five random cases were selected, transcribed and then scored by a colleague.

Having met the five general assumptions of parametric statistics, multicollinearity was checked by running correlation and checking the strength of correlation among the dependent variables. One-way repeated-measures multivariate analysis of variance (MANOVA) test was run to investigate if there were statistically significant differences in complexity, accuracy, lexis, and fluency between two reproductions due to the length constraint. The above-mentioned tests were run by means of latest SPSS Version (2023).

Data Analysis and Results

The descriptive statistics are presented in Table 2. Descriptive statistics calculated scores on two reproductions before and after imposing the length constraint.

Table 2

Descriptive Statistics

	N	Min.	Max	Mean	Standard deviation
Ratio of subordination 1	22	2.09	3.50	2.7950	.39286

Ratio of subordination 2	22	1.87	4.16	3.0936	.61467
Length of analysis of speech units 1	22	11.28	21.20	16.1614	2.46626
Length of analysis of speech units 2	22	12.12	28.66	19.7077	4.28757
Clause length 1	22	4.72	6.64	5.7700	.48258
Clause length 2	22	4.58	7.66	6.1950	.71723
Percentage of correct use of the verbs 1	22	12.12	59.45	36.5623	15.08358
Percentage of correct use of the verbs 2	22	16.00	83.33	42.6968	18.40099
Weighted clause ratio 1	22	.577	.886	.75755	.086045
Weighted clause ratio 2	22	.636	.933	.79923	.082751
Vocabulary diversity 1	22	22.871	61.604	44.02064	10.486035
Vocabulary diversity 2	22	25.207	60.890	43.34723	10.446934
Textual lexical diversity 1	22	24.143	57.145	39.77359	9.308824
Textual lexical diversity 2	22	24.241	148.120	48.19273	26.358131
Length of run 1	22	3.00	22.09	9.5250	5.30897
Length of run 2	22	3.54	44.60	11.5273	10.02564
Speech rate 1	22	37.86	108.26	70.5432	18.34080
Speech rate 2	22	51.00	116.73	82.1164	16.97664
Mean number of internal pauses 1	22	1.14	15.26	7.0673	3.53881
Mean number of internal pauses 2	22	.46	16.92	8.6036	5.13026
Mean number of external pauses 1	22	.00	11.80	4.1223	2.84205
Mean number of external pauses 2	22	.30	18.18	5.5564	4.30660
Mean number of repair measures 1	22	5.85	45.64	24.6614	11.43404
Mean number of repair measures 2	22	3.66	73.07	26.2107	16.07216

As cited in Tavakoli and Hunter (2018), previous literature showed that four measures: (a) ratio of subordination, (b) weighed clause ratio, (c) measures of vocabulary diversity, and (d) speech rate represent CALF consistently (Foster & Skehan, 1996; Michel et al., 2007; Tavakoli & Skehan, 2005). In this study the above-mentioned variables are used to limit the 12 dependent variables to 4.

Pre-requisites of (a) level of measurement, (b) random sampling, (c) normal distribution and (d) independence of observations were observed. Due to the fact that there is only one group participated in this study, checking homogeneity of variances is not required.

Using Kolmogorov-Smirnov and Shapiro-Wilk tests, the normality of scores was also checked and it proved to be normal. Total number of 22 cases were studied in the present research and four dependent variables were measured in two circumstances.

Checking the Histograms and Boxplots, no outliers were observed in the scores. Checking linearity and finding Mahalanobis score were not required as the groups are balanced. In addition, checking homogeneity of variance-covariance between groups was not necessary because the scores were obtained in two reproductions from just one group. Homogeneity of regression was not to be checked because a stepdown analysis was not performed in this research.

The presented results in Table 3 provide an answer to the research question checking if there is a difference in a linear combination of the four dependent variables. Due to the fact that the significance value of Wilk's Lambda is less than 0.05, it can be concluded that there is a difference in a linear combination of the four dependent variables across the two reproductions. The corresponding Partial Eta Squared index indicates that 80.6 percent of variance in the four dependent variables can be explained by the independent variable. Since a repeated measures design is used in the present study, the assumption of sphericity (Mauchly's test of sphericity) should be taken into account. Sphericity means that the variances of differences are not considerably different in each level of the independent variable. Owing to the fact that there are only two levels in the present study, this assumption cannot be tested. Then the Greenhouse-Geisser statistics are examined to check whether the differences in scores are statistically significant and could be explained by the independent variable. The results are presented in Table 4.

Table 3

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.806	18.643	4.000	18.000	.000	.806
Wilks' lambda	.194	18.643	4.000	18.000	.000	.806
Hotelling's trace	4.143	18.643	4.000	18.000	.000	.806
Roy's largest root	4.143	18.643	4.000	18.000	.000	.806

Table 4

Univariate Tests

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Reproductions	C	Sphericity Assumed	.981	1	.981	5.375	.031	.204
		Greenhouse-Geisser	.981	1.000	.981	5.375	.031	.204
		Huynh-Feldt	.981	1.000	.981	5.375	.031	.204
		Lower-bound	.981	1.000	.981	5.375	.031	.204

A	Sphericity Assumed	.019	1	.019	19.882	.000	.486
	Greenhouse-Geisser	.019	1.000	.019	19.882	.000	.486
	Huynh-Feldt	.019	1.000	.019	19.882	.000	.486
	Lower-bound	.019	1.000	.019	19.882	.000	.486
L	Sphericity Assumed	4.988	1	4.988	.189	.668	.009
	Greenhouse-Geisser	4.988	1.000	4.988	.189	.668	.009
	Huynh-Feldt	4.988	1.000	4.988	.189	.668	.009
	Lower-bound	4.988	1.000	4.988	.189	.668	.009
F	Sphericity Assumed	1473.324	1	1473.324	34.338	.000	.621
	Greenhouse-Geisser	1473.324	1.000	1473.324	34.338	.000	.621
	Huynh-Feldt	1473.324	1.000	1473.324	34.338	.000	.621
	Lower-bound	1473.324	1.000	1473.324	34.338	.000	.621

Note. Letters C, A, L, and F stand for complexity, accuracy, lexis, and fluency respectively.

To avoid committing Type I error, rejecting a null hypothesis that is actually true, the significance score of .05 should be divided by 4 which is the number of contrasts; and as a result, the alpha level in this study is .0125.

The significance value of 0.031 (>.0125) for complexity scores indicates that there is not a significant difference in complexity scores across two reproductions. The significance value was 0.000 (<.0125) for accuracy scores. Consequently, there is a considerable difference in accuracy scores across two reproductions. Considering the effect size, the corresponding Partial Eta Squared value illustrates that 48.6 percent of the variance in accuracy score can be explained by the independent variable. The percentage was calculated simply by multiplying the outcome by 100, shifting the decimal point two places to the right. For lexis, however, the significance value was 0.668 (>.0125) indicating that no significant difference is observed across the two reproductions. And the significance value of 0.000 (<.0125) for fluency shows that a significant difference was observed across two reproductions. Partial Eta Squared value of .621 in case of fluency illuminates that a considerable 62.1 percent of the variance in fluency score can be explained by the independent variable.

Discussion

This study aimed at investigating whether length constraint affects EFL learners' narrative retellings in terms of complexity, accuracy, lexis, and fluency (CALF). The findings indicated a significant positive effect of length constraint on learners' narrative L2 performance, particularly in accuracy and fluency; however, changes in complexity and lexis turned out to be non-significant. These results are in line with a number of earlier studies in the literature (e.g. Foster & Skehan, 1996; Mehnert, 1998; Ahmadian & Tavakoli, 2021), indicating that manipulating task features would influence cognitive demands and thereby result in L2 performance. The present findings also lend support to Skehan's (2009) trade-off hypothesis, suggesting that improvements in one performance dimension (e.g., accuracy) may come at the expense of others (e.g., complexity or lexis).

A notable finding was that learners produced more accurate and fluent retellings under length constraints, despite the increased cognitive load. This is in line with Foster and Skehan

(1996) and Skehan and Foster (1999), who stated that a higher level of task complexity would result in accuracy by pushing learners to rely on more controlled, error-free language. Similarly, Ahmadian and Tavakoli (2021) and Jackson and Suethanapornkul (2023) found that constrained tasks can result in gains regarding accuracy, as learners favor correctness over exploring the potential of language performance.

Interestingly enough, learners in this study produced longer yet more accurate sentences, contradicting the assumption that the demand to produce shorter utterances can improve accuracy. This, supported by Jackson and Suethanapornkul (2023), indicates that sentence length does not necessarily compromise accuracy, possibly due to learners' resort to preplanned or formulaic structures (Divsar, 2022; Michel & Kormos, 2023).

Fluency improvements under constrained condition tend to comply with Robinson's (2011) distinction between resource-directing and resource-depleting factors. Since length constraint primarily increases resource-depleting demands (e.g., faster processing), learners may compensate by retrieving high-frequency phrases, leading to smoother, more rapid speech (Suzuki & Sunada, 2023). This, however, fails to find support from Mehnert's (1998) argument that constraints essentially hinder fluency, though the present study suggests that repeated exposure (as in a second retelling) may moderate such effects.

No significant changes were observed in syntactic complexity. This tends to contradict Robinson's (2007) claim that higher cognitive demands result in more complex language. This discrepancy may grow out of learners' reliance on familiar structures from the provided text, limiting syntactic variety. Moreover, less planning time in the second retelling may have inhibited complex sentence production on the learners' part.

By the same token, lexical diversity remained uninfluenced by constraint, which is approved by Robinson (2001), who argues that resource-directing (rather than resource-depleting) factors trigger lexical variation. Since the input text remained identical across the two performances, it is highly likely that the learners may have re-processed and reorganized text-based lexical items rather than seeking novel expressions. This finding is contradicted by Behtary and Yaghoubi-Notash (2009), who reported lexical gains under constraints, but is confirmed by Awwad and Tavakoli's (2019) observation that intentional reasoning tasks (which require deeper processing) prompt lexically different production compared to simple retellings.

The results highlight the multidimensional nature of task complexity, suggesting that accuracy and fluency can improve simultaneously under constraints, while complexity and lexis may remain unchanged. This challenges a strict trade-off model, instead emphasizing adaptive resource allocation (Ellis, 2024; Skehan & Foster, 2020).

Conclusion

This study suggests that imposing a length constraint on narrative retelling tasks would increase pressure on learners' cognitive processing and working memory. Higher cognitive demand did not positively influence complexity nor accuracy. Accuracy and fluency scores involved positive changes in the second retelling imposing length constraint. The rise in two aspects of language performance confirmed the trade-off or competitiveness in literature. An in-depth analysis of data revealed that imposing length constraint does not improve lexis use. Overall, the present study showed that setting cognitive demands in the form of length constraint led to better L2 performance.

References

Adams, R., & Nik, N. M. (2014). Prior knowledge and second language task production in text chat. *System, 42,* 1-12. https://doi.org/10.1016/j.system.2013.10.011

Ahmadian, M. J., & Tavakoli, M. (2021). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research, 25*(3), 382-404. https://doi.org/10.1177/1362168819858249

Awwad, A. A., & Tavakoli, P. (2019). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. *International Review of Applied Linguistics in Language Teaching (IRAL), 60*(2), 169-196. DOI: 10.1515/iral.2018.0378

Behtary, S., & Yaghoubi-Notash, M. (2009). Explorations in L2 Speech Variability: A case for length constraint as task demand. *Rassegna Italiana di Linguistica Applicata, 41(3), 191-206.*

De Jong, N., & Perfetti, C. A. (2021). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61*(2), 533-568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

Divsar, H. (2020). The comparative effect of rehearsal, strategic, and online planning on the fluency, accuracy, complexity, lexical variety, and attentional allocation of EFL learners' argumentative writing. *Teaching English as a Second Language Quarterly (TESLQ), 41*(2), 75-111. DOI: 10.22099/TESL.2021.41004.3017

Divsar, H. (2022). The impact of planning time types on complexity, accuracy, fluency, and lexical diversity in EFL learners' writing. *Language Teaching Research, 26*(1), 3-24. https://doi.org/10.1177/1362168820906287

Ellis, R. (2018). *Reflections on task-based language teaching*. Multilingual Matters.

Ellis, R. (2024). Rod Ellis's essential bookshelf: Focus on form. *Language Teaching, 57,* 246–261 doi:10.1017/S026144482200012X

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (Eds.). (2020). Task-based language teaching: Theory and practice. Cambridge University Press. DOI: 10.1017/9781108643689

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.

Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language teaching research, 3*(3), 215-247.

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics, 36,* 98-116.

García-Ponce, E. E., Mora-Pablo, I., & Segovia-Hernández, J. G. (2022). Role of EFL learners' perceptions of task difficulty in complexity, accuracy and fluency: An exploratory case study. *Porta Linguarum, 37*, 123-142. DOI: https://doi.org/10.30827/portalin.vi37.15855

Hsu, H. (2023). Lexical diversity in constrained vs. unconstrained L2 narrative tasks. *TESOL Quarterly*, *57*(1), 78-102. https://doi.org/10.1002/tesq.312

Jackson, D. O., & Suethanapornkul, S. (2023). Time pressure and L2 oral performance: A trade-off between complexity and accuracy. *Language Learning, 73*(1), 45-72. https://doi.org/10.1111/lang.12512

Jin, T., Lu, X., & Ni, J. (2020). Attentional allocation in L2 task-based performance: A cognitive load perspective. *Applied Linguistics, 41*(5), 712-735. https://doi.org/10.1093/applin/amz046

Kormos, J., & Trebits, A. (2022). Task modality and pair formation method in L2 oral interaction. *Language Teaching Research, 26*(4), 567-589. https://doi.org/10.1177/1362168820906288

Kuiken, F., Mos, M., & Vedder, I. (2005). Cognitive Task Complexity and Second Language Writing Performance. *EUROSLA Yearbook, 5*, 195-222. DOI: 10.1075/eurosla.5.10kui

Kuiken, F., & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics in Language Teaching (IRAL), 45*(3), 261-284. DOI: 10.1515/iral.2007.012

Kuiken, F., & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143-170). John Benjamins. https://doi.org/10.1075/lllt.32.07kui

Lambert, C., & Kormos, J. (2024). Dynamic systems and task complexity: A longitudinal study of L2 performance. *Studies in Second Language Acquisition, 46*(1), 1-25. https://doi.org/10.1017/S027226312300045X

Luo, S. (2022). Task complexity and L2 performance: A meta-analysis. *Language Learning, 72*(1), 152-183. https://doi.org/10.1111/lang.12472

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language testing, 19*(1), 85-104.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in second language acquisition, 20*(1), 83-108.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching (IRAL), 45*(3), 241-259. DOI: 10.1515/iral.2007.011

Michel, M., & Kormos, J. (2023). Working memory and task conditions in L2 speech production. *Applied Psycholinguistics, 44*(2), 301-328. https://doi.org/10.1017/S0142716422000425

Readable.com (2025). *What is readability?* Retrieved July 30, 2025, from https://readable.com/readability/what-is-readability/

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework (p. 3). *Applied linguistics*, *22*(1), 27-57.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty *International Review of Applied Linguistics in Language Teaching (IRAL), 45*(3), 193-213. DOI: 10.1515/iral.2007.009

Robinson, P. (Ed.). (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (Vol. 2). John Benjamins Publishing.

Saito, K., & Hanzawa, K. (2022). Working memory and L2 oral fluency under time pressure. *Language Learning, 72*(3), 789-820. https://doi.org/10.1111/lang.12498

Sasayama, S. (2016). Is a 'Complex' Task Really Complex? Validating the Assumption of Cognitive Task Complexity. *The Modern Language Journal, 100*(1), 231-254. https://doi.org/10.1111/modl.12313

Sasayama, S., & Malicka, A. (2022). The role of task complexity in L2 performance: A meta-analysis. *TESOL Quarterly, 56*(2), 341-369. https://doi.org/10.1002/tesq.307

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics*, *30*(4), 510-532.

Skehan, P. (2014). Task-based performance and task-based instruction: Research contributions.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings (p. 4). *Language learning, 49*(1), 93-120.

Suzuki, Y., & Sunada, M. (2023). Fluency development through task repetition: The role of working memory. *Studies in Second Language Acquisition, 45*(1), 1-25. https://doi.org/10.1017/S0272263122000362

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in second language acquisition*, *15*, 165-179.

Tavakoli, P., & Foster, P. (2021). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 71*(1), 75-104. https://doi.org/10.1111/lang.12425

Tavakoli, P., & Hunter, A. M. (2018). Is fluency being 'neglected 'in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research, 22*(3), 330-349.

Tavakoli, P., Nakatsuhara, F., & Hunter, A. M. (2017). Validating speaking test rating scales through microanalysis of fluency using PRAAT.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In *Planning and task performance in a second language* (pp. 239-273). John Benjamins.

Van den Branden, K. (Ed.). (2006a). *Task-based language education: From theory to practice*. Cambridge University Press. https://doi.org/10.1017/cbo9780511667282

Van den Branden, K. (2006b). Introduction: Task-based language teaching in a nutshell. In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 1-16). Cambridge University Press. https://doi.org/10.1017/cbo9780511667282

Vasylets, O., Gilabert, R., & Manchón, M. M. (2017). The Effects of mode and task complexity on second language production. Language Learning, 67(2), 394-430. https://doi.org/10.1111/lang.12228

Wang, Z. (2013). The effects of task complexity on L2 lexical complexity. *Modern Language Journal*, *97*(1), 82-97. https://doi.org/10.1111/j.1540-4781.2013.01431.x

Wiggleworth, G. (2001). Influences on performance in task-based oral assessments. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 186-209). Longman. Xing, M. (2019). Task complexity and second language production: A meta-analysis. *System*, *84*, 1-12. https://doi.org/10.1016/j.system.2019.05.003

Zohrabi, M., & Hasanpour, S. (2020). The effect of open vs. closed tasks on Improving Iranian EFL learners' oral performance. *Teaching English as a Second Language Quarterly (TSOLQ)*, *38*(4), 205-238. https://doi.org/10.22099/jtls.2020.37075.2814

Massoud Yaghoubi-Notash received his B.A., M.A., and Ph.D. degrees in English Language and Literature from Islamic Azad University-Tabriz Branch, University of Tabriz, University of Tabriz, respectively. In 2007, he joined the Department of English, University of Tabriz, as an Assistant Professor in ELT. Since then, he has published research articles, books and presented articles in conferences. He has been supervising over 35 MA and a few PhD theses.

Shahabaddin Behtary received his B.A., and M.A. from Islamic Azad University-Ardabil Branch, and University of Tabriz, respectively. He is currently doing his doctoral studies as a PhD candidate at the English Department of the University of Tabriz. He is currently a full-time instructor at the Islamic Azad University- Ardabil Branch. Over the years, he has supervised theses, published papers, and presented articles in national and international conferences.

Sina Gobadi holds an MA in English Language Teaching (ELT). He is an English Tutor.

Shiva Shahbazi Mojarrad has an MA in English Language Teaching (ELT). She is a freelance researcher.

Appendix A

Story for Oral Reproduction: Indian Corns

It was a very wide river, with many great curves in it, and in one of these there lived a large number of wild pigs. Nobody could remember how they had got there, but they managed to live through floods, fires, ice and attacks by hunters.

Then one day a stranger came to the nearest village and asked where he could find the wild pigs. Somebody told him, and he went off. He had no weapons with him, and the village people wondered what he was going to do with the pigs.

When he came back a few months later and said that he had caught all the pigs, the villagers were still more surprised, but some of the men agreed to go with him when he asked for help in bringing the pigs out. They wanted to see whether he was telling the truth.

They soon discovered that he was. All the pigs were inside an enclosure which had a fence round it and a gate in one of its sides.

'How did you do it?' they asked the stranger.

'Well, it was quite easy really,' he answered. 'I began by putting out some Indian corn. At first, they would not touch it, but after a few weeks, some of the younger pigs began to run out of the bushes, take some of the corn quickly, and then run back. Soon all the pigs were eating the corn I put out. Then I began to build a fence round the corn. At first it was very low, but gradually I built it higher and higher without frightening the pigs away. When I saw that they were waiting for me to bring the corn each day instead of going and searching for their own food as they had done in the past, I built a gate in my fence and shut it one day while they were all eating inside the enclosure. I can catch any animal in the world in the same way if I can get it into the habit of depending on me for its food.'

INTERNATIONAL JOURNAL OF
Practical and Pedagogical Issues in English Education
ISSN : 2980-9533 Homepage: www.ijpie.org 3(3), Sep. 2025

Appendix B

Transcript of Students' Reproductions

Student A.

First Reproduction.

It was a very large river that there were a lot of curves in it in one of these one of these a large number of pigs live live on it live on it nobody knows nobody knows how they how they got there but they lived through the fire through the flu fires ice and also huntershe one one day a stranger ca came in there came in there and wanted people how he how how how how I could get the pigs get the pigs and they show them they show them he wents he wents there and after afterf a he wents of after e he wents of after after some month he came back he came back and said I I could I could caught all of them all of the pigs …and when he wanted people to help him to help him pe some people came c c some people ca came with him to sho to help him and also also know wether he was right or no but e whe when they're went the when they went there they they sho they say that he was truth he was truth and asked him how how how did you do that he said that he said that I began I began by I began by Indian corn I I put of the corn and show for them at first they didn't nitice for them but after the after the after the grasses and bushes run out they some some youngers of the pigs came and take it f take it from me and run out but after some days all of after some days most of them came and came came and e take from me take from me pi e the corns the corns then I started to then I started to make a make a hence arou hence around them at during they during they were eating during they were eating the corn corn and e then when and then e then then when they didn't go for search for searching the food for themselves as a past as a past I I I I I I made a I made a I I made a a gate gate e I put I put a gate in one of the in one of its sides in one of its sides and and then one of its sides and then I shot it I shot it I shot it| and I and I can get e I can get e all of the animals in the world if they dependence if they habit dependence if they… if they habits depends for me