



## RESEARCH REPORT

# Longitudinal development of second language utterance fluency, cognitive fluency, and their relationship

Jimin Kahng 

Department of Modern Languages, University of Mississippi, Oxford, MS, USA  
Email: [jkahng@olemiss.edu](mailto:jkahng@olemiss.edu)

(Received 19 July 2022; Revised 01 November 2023; Accepted 09 November 2023)

### Abstract

The development of L2 utterance fluency has been extensively researched, whereas that of cognitive fluency has rarely been examined. This study investigated the longitudinal development of L2 utterance and cognitive fluency and their relationship. Thirty-one Chinese learners of English completed speaking tasks and a set of tasks for cognitive fluency before and after 5 months' study abroad. The results showed that participants made a significant improvement in mean syllable duration, end-clause pause frequency, and the speed of syntactic encoding and articulation but not in mid-clause pause frequency or lexical retrieval speed. Mixed-effects modeling confirmed a significant relationship between syntactic encoding speed and mean syllable duration and mid-clause pausing. Furthermore, the significant relationships were maintained over time. The findings highlight (a) the differences between mid-clause and end-clause pausing in terms of their developmental patterns and relationship with cognitive fluency and (b) a significant role of syntactic encoding speed in L2 utterance fluency.

Whereas fluency can be defined various ways (e.g., Fillmore, 1979; Freed, 2000), a narrow sense of fluency—how smoothly oral speech is delivered (Lennon, 1990)—has recently been examined extensively based on its significance in understanding of second language (L2) proficiency (Iwashita et al., 2008). Segalowitz (2010) further distinguishes three facets of fluency—utterance, cognitive, and perceived fluency. Utterance fluency (UF) refers to the temporal and repair features of speech, whereas cognitive fluency (CF) refers to the underlying cognitive processes responsible for fluent speech production. The current study investigates the development of utterance and cognitive fluency, and its interrelationship in L2.

### *L2 utterance fluency and its development*

Compared with L1 speech, L2 speech tends to be slower and have more pauses and repairs (e.g., Kahng, 2014; Riazantseva, 2001). Importantly, L2 speech often has more

pauses within clauses (Kahng, 2014; Tavakoli, 2011) and within analysis of speech (AS) units<sup>1</sup> (De Jong, 2016) than L1 speech does.

L2 UF has been found to exhibit significant correlations with L2 oral proficiency. Among various linguistic measures (e.g., grammatical accuracy, complexity, pronunciation), speech rate exhibited the strongest influence on distinguishing a range of L2 speaking proficiency levels (Iwashita et al., 2008). Speech rate, articulation rate, and mean length of run in Ginther et al. (2010) and mean syllable duration and mid-clause silent pause ratio in Kahng (2014) were moderately to strongly correlated with speaking scores.

In longitudinal studies, O'Brien et al. (2007) found that the L1-English L2-Spanish learners improved speech rate and mean length of run without fillers after a semester of study abroad. In Mora and Valls-Ferrer (2012), Catalan-Spanish bilingual learners of English showed gains in speech rate, mean length of run, and pause frequency and duration after a 3-month stay abroad. In their 2-year-long study with L1-English L2-Spanish learners, Huensch and Tracy-Ventura (2017) reported gains in mean syllable duration appeared quickly and were maintained after return from 9 months' study abroad, whereas gains in pause frequency appeared later and were sensitive to attrition after return home.

### *L2 cognitive fluency and its relation to utterance fluency*

Although L2 CF and its relation to UF has been underexamined, a couple of proposals (Kormos, 2006; Skehan et al., 2016) have been made regarding the connection between different UF measures and speech production stages. According to Levelt (1989), speech production consists of three main stages. First, a preverbal message is generated using world knowledge (conceptualization). Second, this message is put into words through lexical, grammatical, morphophonological, and phonetic encoding (formulation). Third, the generated utterance is articulated (articulation). During speech planning and after speech articulation, one's own speech is monitored. Kormos (2006) proposed that end-clause pausing can reflect L2 speaker's conceptualization, mid-clause pausing relates to formulation, repairs may reflect speech monitoring, and speed measures can involve all dimensions of L2 speech production. The proposal has been used by Saito et al. (2018) in interpreting their cross-sectional findings on UF correlates of different levels of perceived fluency. Based on the distinctive length of residence profile of their fluency groups, they argued that L2 fluency development could initially be observed in end-clause pause frequency (relating to conceptualization), followed by mid-clause pause frequency (relating to formulation), and articulation rate.

On the other hand, in examining UF–CF relationship, only a handful of studies have measured CF and related them to UF measures. In Segalowitz and Freed (2004), L1-English L2-Spanish speakers completed an oral proficiency interview, a semantic classification task, and an attention control test before and after a semester either at a home university (AH) or a study-abroad (SA) setting. They found both groups improved in the speed and efficiency of L2 lexical access but only the SA group made gains in UF (e.g., speech rate, mean length of run without fillers). As for UF–CF relationship, they found significant correlations between the speed and efficiency of

<sup>1</sup>An AS-unit is defined as “a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either” (Foster et al., 2000, p. 365).

lexical access and mean length of run without fillers in pretest but did not find significant relationships between pretest CF and UF gains or between CF gains and UF gains.

De Jong and her colleagues examined different aspects of CF and their relationship with UF. In De Jong et al. (2013), mean syllable duration has been identified as a strong indicator, whereas pause duration as the weakest indicator of L2 linguistic knowledge and cognitive skills. In particular, mean syllable duration was strongly correlated with sentence-building speed and moderately correlated with lexical-retrieval speed. Overall, the speed of sentence building and lexical retrieval (formulation) significantly correlated with the number of silent pauses, filled pauses, corrections, and repetitions. De Jong and Mora (2019) focused on the relationship between articulation and UF and found that articulatory skills explained 19% and 27% of silent pause rate and silent pause duration, respectively, but were not related to articulation rate. In examining the influence of conceptualization difficulty on UF, Felker et al. (2019) found that abandoning and regenerating a speech plan leads to dysfluencies for both L1 and L2 speakers; however, L2 speakers needed more additional time to regenerate a speech plan than L1 speakers did.

More recently, based on the association of UF in L1 and L2 (De Jong et al., 2015; Duran-Karaoz & Tavakoli; 2020), Kahng (2020) investigated to what extent different measures of L2 UF can be explained by L2-specific CF and/or the corresponding L1 UF measures. The results suggest that mid-clause pause frequency and mean syllable duration are indicative of L2-specific cognitive measures, whereas silent pause duration is indicative of L1 UF.

### Current study

Taken together, L2 fluency has been widely investigated from an UF perspective, whereas its underlying cognitive processes have been largely underexamined. A few recent studies on L2 CF discussed above have started revealing its complex relationships with L2 UF, suggesting mean syllable duration and mid-clause pause frequency be potential correlates of L2 CF. However, the seminal research by Segalowitz and Freed (2004) appears to be the only study that investigated the development of CF along with that of UF. We need more empirical research to examine the proposals on UF–CF relationship (Kormos, 2006; Skehan et al., 2016) and the development of CF, which can contribute to the development of L2 speech-production model. Therefore, the current study investigated the development of UF and CF and their relationship with the following research questions:

RQ1. Which aspects of L2 utterance and cognitive fluency demonstrate changes after 5 months' study abroad?

RQ2. Which aspects of L2 utterance and cognitive fluency have significant relationships? Are the relationships maintained over time?

Regarding UF, although it has been extensively researched, the selection of UF measures has often not been comparable across studies. In particular, pause location has rarely been examined in the longitudinal development of UF. As discussed earlier, pauses between clauses and within clauses have been proposed to involve different stages of speech production (e.g., Kormos, 2006; Skehan et al., 2016). Considering that pause location has been suggested to be an important indicator of development (Kahng,

2014; Saito et al., 2018) and be closely related to CF (Kahng, 2020; Kormos, 2006; Skehan et al., 2016), the current study categorized pauses into mid-clause versus end-clause pauses and examined their relationship with CF measures.

Furthermore, as the oft-cited UF measures in the previous studies such as speech rate and mean length of run involve multiple fluency features (e.g., speed and frequency and/or duration of pauses), even if gains are observed it is difficult to identify to which aspect of UF the gains are attributable. Therefore, in the current study UF measures were purposefully selected based on the theoretical distinction between the three aspects of UF—speed, breakdown, and repair fluency (Skehan, 2003).

In examining L2 CF, the speed of lexical retrieval and syntactic encoding was measured for formulation, and articulation speed was measured for articulation. In addition, the scores of elicited imitation tasks (EITs) were included in the analysis to investigate participants' changes in overall oral proficiency before and after 5 months' study abroad and to address potential effects of participants' initial proficiency level on their fluency development.

## Method

### *Participants*

The current study is part of a larger project wherein forty-four Chinese learners of English participated and received \$50 per session. Informed consent was obtained from all participants. This study focuses on the data of 31 learners (17 males, 14 females) who completed speaking tasks and CF tasks before and after 5 months' study abroad while taking undergraduate or graduate courses at a university in the United States. Their ages ranged 21–46 ( $M_{\text{age}}=28$ ,  $SD_{\text{age}}=6$ ). They had lived in the USA less than 6 months ( $M_{\text{LOR}}=2$  months,  $SD_{\text{LOR}}=1$  month) at the onset of the study. They started to learn English around the age of 11 ( $M_{\text{AO}}=11.1$ ,  $SD_{\text{AO}}=2.0$ ). Based on the grammar and vocabulary sections of DIALANG, a diagnostic test developed by Lancaster University, the majority of them were intermediate learners (3 A2, 26 Bs, and 2 C1s), according to the common European framework.

### *Speaking tasks*

#### *Materials*

Two types of questions were used as prompts—one addressing personal preference from a given category (e.g., important time, people, places) and the other asking personal choice between two options (e.g., living in a big or small city). For each type, six comparable prompts on daily life were developed to avoid any practice effects of using the same prompts across sessions. In each session, one of six prompts from each type was randomly selected for each participant. Participants answered in total four different prompts across two sessions.

#### *Procedure*

In each time, participants answered two questions. For each question, they had 15 s to prepare for their answer and talked for about a minute. Their speech was recorded using Praat (Boersma & Weenink, 2018), with a Blue Snowball USB microphone (frequency response 40 Hz–18 KHz) at a 44 KHz sampling rate (16-bit resolution, 1 channel).

*Utterance fluency measures*

All speech samples were transcribed to include information regarding silent and filled pauses, repetitions, corrections, and clause boundaries (Foster et al., 2000). Silent pauses (>250ms; De Jong & Bosker, 2013) and filled pauses were identified and their length was measured in milliseconds (ms) using Praat (Boersma & Weenink, 2018). Following Skehan (2003), speed, breakdown, and repair fluency were measured. For speed fluency, mean syllable duration was calculated by dividing speech time excluding pause time by total number of syllables (De Jong et al., 2013). For breakdown fluency, based on the identified clause boundaries, pauses were categorized into mid-clause or end-clause pauses. The number of mid-clause silent and filled pauses and of end-clause silent and filled pauses per 100 syllables and mean duration of silent pauses were calculated (Kahng, 2020). For repair fluency, the number of repetitions and corrections per 100 syllables were calculated.

*Tasks for cognitive fluency*

The CF measures were adopted to represent subprocesses of Levelt's speech production model (1989), focusing on formulation and articulation (De Jong et al., 2013). For formulation, the speed of lexical retrieval and syntactic encoding were measured, and for articulation the speed of articulation was measured. The details are described below.

*Lexical retrieval speed*

*Materials.* Forty pictures of objects (e.g., strawberry, horse) were selected from Snodgrass and Vanderwart (1980). Half of the pictures were randomly selected for a picture-naming task in Time 1 and the other half for Time 2. An advanced Chinese learner of English was consulted to ensure the familiarity of the items to Chinese participants and changes were made when necessary. There was no significant difference in the number of syllables of the words used in Time 1 and Time 2,  $t(38) = 0.44$ ,  $p = .67$ , 95% CI: -0.36–0.56).

*Procedure and measure.* Using PsychoPy (Peirce et al., 2019), pictures were presented on the screen and participants named each of them as fast and accurately as possible. Following De Jong et al. (2013), after a fixation cross was presented for 1,500 ms, the target picture was presented for 2,000 ms, which was followed by a blank screen for 500 ms. The pictures were presented in a random order for each participant. There was a practice session with a few pictures not included in the actual tasks to ensure familiarity with the task. The recording procedure was same as for the speaking tasks. The reaction time (RT) between the presentation of the picture and the beginning of the correct response was measured using Praat (Boersma & Weenink, 2018). After inspection of the data, RTs below the minimum of 50 ms and RTs higher than 3 *SD* above the grand mean were identified as outliers (i.e., 0.8% of the data) and replaced via multiple imputation by chained equations (van Buuren & Groothuis-Oudshoorn, 2011).

*Syntactic encoding speed*

*Materials.* Participants saw the beginning of a sentence (e.g., *I expect ...*) and selected an option (e.g., *A. them... B. go...*) that best followed the beginning (Hulstijn et al., 2009). As the task is designed to assess syntactic processing speed rather than grammatical accuracy, basic syntactic structures were targeted so that intermediate learners

were able to answer them. The items covered syntactic structures including word order in declarative and interrogative sentences, subject-verb agreement, and different types of phrase structure. Forty items were developed for two sessions. Two comparable syntactic encoding tasks were devised, each of which had 20 items covering the same syntactic structures. All the items were pilot tested by English speakers.

*Procedure and measure.* Using PsychoPy (Peirce et al., 2019), participants were first presented with the beginning of a sentence and on the next screen two possible options followed. Participants selected an option as fast and accurately as possible. They were told that the options would not complete the sentence but one of them would best follow the beginning of the sentence. There was a practice session with a few items not included in the actual experiment. The interval between the presentation of the two options and participant's keyboard response was automatically measured, and the RTs for correct responses were used for analysis. After inspection of the data, RTs below the minimum of 50 ms and RTs higher than 3 *SD* above the grand mean were identified as outliers (i.e., 1.7% of the data) and replaced via multiple imputation by chained equations (van Buuren & Groothuis-Oudshoorn, 2011).

#### *Articulation speed*

*Materials.* The materials were the same as for the lexical retrieval speed.

*Procedure and measures.* Participants completed the picture-naming task once more, but this time they waited until a cue was given before naming a picture. This delay was to provide them with time for lexical retrieval and phonetic encoding so that the duration of response of the delayed picture-naming task mainly reflects articulation rather than formulation (De Jong et al., 2013).

After a fixation cross was presented for 500 ms, the target picture was presented for 2,000 ms, which was followed by a short beep (De Jong et al., 2013). Participants named the picture right after they heard the beep. The picture remained on the screen for another 1,000 ms. Pictures were presented in a random order for each participant.

The duration of response between the beginning and the end of their correct response was measured using Praat (Boersma & Weenink, 2018). After inspection of the data, durations below the minimum of 50 ms and higher than 3 *SD* above the grand mean were identified as outliers (i.e., 3.1% of the data) and replaced via multiple imputation by chained equations (van Buuren & Groothuis-Oudshoorn, 2011).

#### ***Task for speaking proficiency: Elicited imitation task (EIT)***

The EIT consisted of 29 sentences with a wide range of syllables ( $M_{\text{syllable}} = 13$ , range = 8–20 syllables) to achieve a high discriminability. The EIT was individually administered using PsychoPy (Peirce et al., 2019) with a laptop and a headset in a quiet room. First, participants heard stimulus sentences in a random order, repeated each of the sentences, and pressed a spacebar to move on to the next sentence. Participants' responses were transcribed and coded for the percentage of accurately repeated words. Randomly selected 30% of the data was transcribed by an additional coder, and interrater reliability was high (Cronbach's  $\alpha = 0.97$ ). There was a significant positive correlation between the EIT scores and the composite scores of the grammar

and vocabulary sections of DIALANG,  $r = .51$ ,  $p < .05$  (see Kahng & Otonya, 2021, for the instrument and analysis).

### Analysis

The recordings of the speaking tasks and picture-naming tasks were first transcribed, annotated, and measured by a research assistant. Next, the accuracy of the transcription, annotation, and measures was checked by another research assistant and corrections were made when necessary.

The two research questions were examined using two sets of linear mixed-effects modeling using the *lme4* (Bates et al., 2015), *parameters* (Lüdtke, et al., 2020), and *MuMIn* (Bartoń, 2022) packages in R (R Core Team, 2020). The data satisfied the assumptions in linear mixed models including linearity, normality of the residuals, homoscedasticity, and multicollinearity. When fitting models, once significant random intercepts to each model have been established, each set of fixed variables (e.g., time, CF, and Time  $\times$  CF) was incorporated into the baseline model. Improvement in model fit was identified by a lower Akaike information criterion (AIC) demonstrated by new models along with a statistically significant  $\chi^2$  change between the new and the baseline model.

## Results

### *Utterance fluency and cognitive fluency in Time 1 and Time 2*

In the speaking tasks, on average, participants produced 161 syllables (SD = 61) and spent 65 s (SD = 22) in Time 1 and produced 344 syllables (SD = 139) and spent 132 s (SD = 51) in Time 2 per speech sample. Table 1 shows means, standard deviations, medians, and 95% CIs of UF and CF measures and EIT scores in Time 1 and Time 2. Before examining the effect of time using mixed-effects modeling, 95% CIs roughly show that there was little to slight overlap, especially for mean syllable duration, articulation speed, and EIT.

Table 2 demonstrates the results of the final model on the effects of time, with the included random intercepts for each of the UF measures, and participants' initial oral proficiency level (EIT scores in Time1) as a covariate to address its potential effect. The results show that between Time 1 and Time 2, there was a significant decrease in mean syllable duration, the number of end-clause pauses, syntactic encoding speed, and articulation speed. In addition, mean syllable duration, the number of mid-clause pauses, repetitions, and corrections, and all three CF measures demonstrated significant associations with oral proficiency.

### *Relationship between utterance fluency and cognitive fluency across time*

The UF–CF relationship across time was examined using another set of mixed-effects modeling with UF measures as dependent variables and time, CF measures, and interaction terms between time and CF measures as potential independent variables. The results in Table 3 show that the majority of the final models did not include the interaction of time and CF.

Across time, syntactic encoding speed was significantly associated with mean syllable duration, mean silent pause duration, the number of mid-clause pauses, and

**Table 1.** Measures of utterance fluency, cognitive fluency, and EIT in Time 1 and Time 2

	Time 1				Time 2			
	<i>Mean</i>	<i>SD</i>	<i>Median</i>	<i>95% CI</i>	<i>Mean</i>	<i>SD</i>	<i>Median</i>	<i>95% CI</i>
<i>Utterance fluency</i>								
Mean syllable duration <sup>a</sup>	310	48	308	293–328	290	42	286	275–306
Mean silent pause duration <sup>a</sup>	589	136	559	539–638	579	148	524	524–633
Number of <sup>b</sup>								
Mid-clause silent pauses	8.25	4.53	7.00	6.59–9.92	8.32	4.34	7.46	6.73–9.91
End-clause silent pauses	8.68	3.02	8.49	7.58–9.79	7.67	3.06	6.76	6.55–8.80
Mid-clause filled pauses	3.49	2.83	2.58	2.45–4.53	3.75	2.76	2.94	2.74–4.76
End-clause filled pauses	3.30	2.20	2.86	2.49–4.11	2.52	1.79	2.38	1.86–3.18
Repetitions	1.89	1.77	1.25	1.24–2.54	1.97	1.60	1.65	1.38–2.55
Corrections	1.21	0.96	1.17	0.86–1.56	1.37	0.95	1.18	1.02–1.72
<i>Cognitive fluency</i>								
Syntactic encoding speed <sup>a</sup>	1,946	922	1,718	1,618–2,289	1,750	911	1,472	1,420–2,081
Lexical retrieval speed <sup>a</sup>	1,283	354	1,204	1,154–1,413	1,266	482	1,235	1,089–1,442
Articulation speed <sup>a</sup>	584	131	583	536–632	493	152	491	437–548
<i>EIT</i>								
Accuracy (0–100%)	74.7	12.5	79.5	69.9–79.4	87.4	10.3	88.7	83.5–91.3

<sup>a</sup>in milliseconds.<sup>b</sup>per 100 syllables.



**Table 2.** Changes in utterance and cognitive fluency between Time 1 and Time 2

Random effects		Fixed effects				
Mean syllable duration	By-participant intercepts, SD = 0.72	$R^2_{Fixed} = .130; R^2_{Fixed+Random} = .628$				
			$\beta$	SE	t(114)	p
		Intercept	< .001	0.14	0.04	.971
		Time	-0.18	0.06	-3.07	.003
Proficiency	-0.32	0.14	-2.24	.027		
Mean silent pause duration	By-participant intercepts, SD = 0.72	$R^2_{Fixed} = .039; R^2_{Fixed+Random} = .533$				
			$\beta$	SE	t(114)	p
		Intercept	< .001	0.15	0.03	.978
		Time	-0.04	0.06	-0.56	.579
Proficiency	-0.20	0.15	-1.36	.176		
Number of mid-clause silent pauses	By-participant intercepts, SD = 0.70	$R^2_{Fixed} = .200; R^2_{Fixed+Random} = .669$				
			$\beta$	SE	t(114)	p
		Intercept	< .001	0.14	0.05	.961
		Time	0.03	0.05	0.56	.577
Proficiency	-0.46	0.14	-3.30	.001		
Number of end-clause silent pauses	By-participant intercepts, SD = 0.75	$R^2_{Fixed} = .054; R^2_{Fixed+Random} = .601$				
			$\beta$	SE	t(114)	p
		Intercept	< .001	0.15	0.03	.972
		Time	-0.18	0.06	-2.97	.004
Proficiency	-0.16	0.15	-1.05	.296		
Number of mid-clause filled pauses	By-participant intercepts, SD=0.64	$R^2_{Fixed} = .163; R^2_{Fixed+Random} = .554$				
			$\beta$	SE	t(114)	p
		Intercept	<.001	0.13	0.04	.968
		Time	0.07	0.06	1.15	.253
Proficiency	-0.41	0.13	-3.08	.003		
Number of end-clause filled pauses	By-participant intercepts, SD = 0.75	$R^2_{Fixed} = .037; R^2_{Fixed+Random} = .583$				
			$\beta$	SE	t(114)	p
		Intercept	<.001	0.15	0.04	.965
		Time	-0.18	0.06	-2.99	.003
Proficiency	-0.07	0.15	-0.49	.626		
Number of repetitions	By-participant intercepts, SD = 0.53	$R^2_{Fixed} = .179; R^2_{Fixed+Random} = .454$				
			$\beta$	SE	t(114)	p
		Intercept	<.001	0.12	0.03	.976
		Time	0.06	0.07	0.93	.354
Proficiency	-0.43	0.12	-3.58	<.001		
Number of corrections	By-participant intercepts, SD = 0.22	$R^2_{Fixed} = .046; R^2_{Fixed+Random} = .094$				
			$\beta$	SE	t(114)	p
		Intercept	<.001	0.10	0.01	.988
		Time	0.08	0.09	0.87	.385
Proficiency	-0.20	0.10	-2.11	.037		
Syntactic encoding speed	By-participant intercepts, SD = 0.46 By-item intercepts, SD = 0.21	$R^2_{Fixed} = .165; R^2_{Fixed+Random} = .413$				
			$\beta$	SE	t(999)	p
		Intercept	<.001	0.10	<.001	.998
		Time	-0.08	0.03	-2.37	.018
Proficiency	-0.40	0.09	-4.35	<.001		

(Continued)

Table 2. (Continued)

	Random effects	Fixed effects				
Lexical retrieval speed	By-item intercepts, SD = 0.31	$R^2_{\text{Fixed}} = .022; R^2_{\text{Fixed+Random}} = .119$				
			$\beta$	SE	t(789)	p
		Intercept	0.04	0.06	0.64	.525
		Time	-0.05	0.06	-0.78	.437
Proficiency	-0.14	0.03	-4.23	<.001		
Articulation speed	By-item intercepts, SD = 0.51	$R^2_{\text{Fixed}} = .097; R^2_{\text{Fixed+Random}} = .359$				
			$\beta$	SE	t(872)	p
		Intercept	<.001	0.09	0.01	.988
		Time	-0.30	0.09	-3.44	<.001
Proficiency	0.10	0.03	3.53	<.001		

end-clause filled pauses. In addition, lexical retrieval speed had a significant relationship with mean syllable duration.

On the other hand, the number of mid-clause pauses had significant interaction of time and CF measures. The number of mid-clause silent pauses had a significant interaction of time and articulation speed, and the number of mid-clause filled pauses had an interaction of time and lexical retrieval and of time and articulation speed. The follow-up analysis showed there were some changes in the direction of relationships between the variables over time (for the number of mid-clause silent pauses,  $\beta_{AS} = -0.06$  in Time 1 and  $\beta_{AS} = 0.21$  in Time 2; for the number of mid-clause filled pauses,  $\beta_{LR} = -0.13$ ,  $\beta_{AS} = -0.17$  in Time 1 and  $\beta_{LR} = 0.20$ ,  $\beta_{AS} = -0.18$  in Time 2); however, none of the relationships were statistically significant (see Table S2 in Supplementary Material).

In addition, considering the small sample size, post hoc power analyses were conducted using *simr* (Green & MacLeod, 2016). The simulation results demonstrate that some analyses achieved lower power than others. For example, the power for the effect of syntactic-encoding speed on mean syllable duration and the number of mid-clause-filled pauses was strong, whereas the power for the interaction effects on the number of mid-clause-filled pauses was less than 70% (see Table S3 in Supplementary Material).

## Discussion

### Changes in L2 utterance and cognitive fluency

The results of mixed-effects modeling showed that participants exhibited a significant improvement in mean syllable duration, the number of end-clause silent and filled pauses for UF, and the speed of syntactic encoding and articulation for CF even after proficiency was addressed in the model.

Based on previous studies on the development of L2 UF and its relationship with proficiency (Huensch & Tracy-Ventura, 2017; Kahng, 2014), gains found in mean syllable duration are expected. On the other hand, a lack of changes in mid-clause pause frequency and an improvement in end-clause pause frequency may appear somewhat unexpected. Considering that most previous studies on the relationship between L2 UF and proficiency were cross-sectional including participants with a range of proficiency levels, the current findings may suggest that improvement in mid-clause pausing can

**Table 3.** Contribution of CF measures and time to UF measures

Random effects		Fixed effects				
Mean syllable duration	By-participants intercepts, <i>SD</i> = 0.76	$R^2_{Fixed} = .138; R^2_{Fixed+Random} = .664$				
			$\beta$	<i>SE</i>	<i>t</i> (116)	<i>p</i>
		Intercept	0.006	0.15	0.04	.967
		Time	−0.08	0.09	−0.89	.374
		SE	0.37	0.11	3.45	<.001
		LR	−0.19	0.09	−2.09	.039
		AS	0.17	0.13	1.32	.190
Mean silent pause duration	By-participants intercepts, <i>SD</i> = 0.71	$R^2_{Fixed} = .104; R^2_{Fixed+Random} = .571$				
			$\beta$	<i>SE</i>	<i>t</i> (116)	<i>p</i>
		Intercept	<.001	0.14	0.02	.980
		Time	0.15	0.09	1.58	.116
		SE	0.25	0.11	2.20	.030
		LR	−0.01	0.10	−0.13	.898
		AS	0.31	0.13	2.33	.021
Number of mid-clause silent pauses	By-participants intercepts, <i>SD</i> = 0.72	$R^2_{Fixed} = .134; R^2_{Fixed+Random} = .660$				
			$\beta$	<i>SE</i>	<i>t</i> (113)	<i>p</i>
		Intercept	0.09	0.14	0.61	.544
		Time	0.15	0.08	1.78	.077
		SE	0.27	0.10	2.65	.009
		LR	0.07	0.09	0.81	.420
		AS	0.21	0.12	1.70	.092
		Time:SE	−0.02	0.08	−0.26	.792
		Time:LR	−0.04	0.08	−0.48	.633
Time:AS	0.16	0.07	2.27	.025		
Number of end-clause silent pauses	By-participants intercepts, <i>SD</i> = 0.74	$R^2_{Fixed} = .041; R^2_{Fixed+Random} = .585$				
			$\beta$	<i>SE</i>	<i>t</i> (116)	<i>p</i>
		Intercept	<.001	0.15	0.05	.962
		Time	−0.23	0.09	−2.46	.015
		SE	0.04	0.11	0.36	.716
		LR	−0.03	0.10	−0.27	.786
		AS	−0.11	0.13	−0.85	.400
Number of mid-clause filled pauses	By-participants intercepts, <i>SD</i> = 0.70	$R^2_{Fixed} = .214; R^2_{Fixed+Random} = .646$				
			$\beta$	<i>SE</i>	<i>t</i> (113)	<i>p</i>
		Intercept	0.08	0.14	0.53	.598
		Time	<.001	0.09	0.05	.960
		SE	0.40	0.11	3.73	<.001
		LR	−0.09	0.10	−0.89	.375
		AS	−0.19	0.13	−1.46	.146
		Time:SE	−0.09	0.08	−1.05	.297
		Time:LR	0.20	0.08	2.45	.016
Time:AS	0.16	0.07	2.14	.035		
Number of end-clause filled pauses	By-participants intercepts, <i>SD</i> = 0.73	$R^2_{Fixed} = .089; R^2_{Fixed+Random} = .605$				
			$\beta$	<i>SE</i>	<i>t</i> (116)	<i>p</i>
		Intercept	<.001	0.14	0.05	0.964
		Time	−0.22	0.09	−2.46	0.015
		SE	0.22	0.11	2.04	0.044
		LR	−0.13	0.10	−1.33	0.185
		AS	−0.13	0.13	−0.98	0.327

Note. SE = syntactic encoding speed, LR = lexical retrieval speed, and AS = articulation speed; no variables had significant effects on the number of repetitions or corrections (see Table S1 in Supplementary Material).

take longer (e.g., more than 5 months) than those in end-clause pausing. Saito et al. (2018) also suggest that the frequency of end-clause pauses reaches native-like levels before that of mid-clause pauses and is the first aspect of UF that L2 learners develop. In addition, the lack of changes in the number of mid-clause pauses seem to align with a lack of changes in lexical-retrieval speed, which has been associated with mid-clause and mid-AS pausing (e.g., De Jong, 2016, Kahng, 2014, 2020).

With respect to CF, the current study is one of the first to investigate the development of the syntactic-encoding speed and articulation speed, and the findings captured L2 speakers' improvement after 5-months' study abroad. For lexical retrieval speed, in Segalowitz and Freed (2004) speed and efficiency of L2 lexical access improved after a semester of study abroad. On the other hand, in this study statistical changes were not observed, which might partly be related to the materials. Although the picture-naming tasks used oft-cited standardized pictures (Snodgrass & Vanderwart, 1980), some of the words such as animals may not have been ideal for some Chinese undergraduate/graduate student participants, who grew up in large cities and have been learning English for academic purpose. Even if they had known the words, their mental representation of those words might not have been as robust as that for other words they used more often.

### *Relationship between L2 utterance and cognitive fluency over time*

The examination of effects of time and CF on UF revealed a significant relationship between CF measures, especially syntactic-encoding speed, with mean syllable duration and the number of mid-clause silent and filled pauses. The findings confirm a significant relationship of mean syllable duration and mid-clause pause frequency with CF, as has been found in previous studies (De Jong et al., 2013; Kahng, 2020), consistent with the proposals by Kormos (2006) and Skehan et al. (2016). This study also replicated the interesting findings in De Jong and Mora (2019), in which articulation speed was significantly correlated with mean silent pause duration but not with mean syllable duration (inverse articulation rate). With respect to end-clause pause frequency, its lack of a significant association with CF measures seems to reflect the absence of a conceptualization measure in the current study.

Regarding the UF–CF relationships over time, syntactic-encoding speed exhibited a robust relationship with mean syllable duration and mid-clause pausing across time, without an interaction of CF and time. For mean syllable duration, when CF measures were included in the model (Table 3), the significant effect of time identified earlier (Table 2) disappeared, highlighting its tight association with CF across time. On the other hand, the speed of lexical retrieval and articulation speed had an interaction effect of time on mid-clause pausing, suggesting potentially more complex relationships. However, the interaction effects had low statistical power and call for further investigation.

### **Conclusion**

This study is one of the first attempts to examine the longitudinal development of L2 utterance and cognitive fluency and sheds insights on their developmental patterns and relationships. In particular, the findings emphasize the differences between mid-clause and end-clause pausing in their development and relationships with CF. Another novel finding pertains to the robust relationship between syntactic-encoding speed and mean

syllable duration and mid-clause pause frequency over time. Previous studies have shown the significant relationship between CF and mean syllable duration and mid-clause pausing, whereas this study is the first to show the significant relationships were maintained over time.

The current study has two primary limitations that need to be acknowledged, and further research is required for replication and elaboration on this topic. The first issue relates to the tasks used to measure UF and CF. In measuring UF, a personal narrative was used because it has ecological validity in that it resembles a real-world speech task and involves all stages of speech production including conceptualization, unlike a picture-narrative task wherein general content is given. Yet, whether the current findings are applicable to other types of speaking tasks is an empirical question that requires further investigation. In measuring CF, the current study did not include a measure of conceptualization. The influence of conceptualization on L2 fluency has rarely been examined (cf. Felker et al., 2019) and the conceptual link between conceptualization and end-clause pausing needs to be validated in the future.

Second, this study is based on 31 participants' 5-month study-abroad experience, and the findings should be interpreted with caution. For future research, tracking a larger number of participants' changes in UF and CF at multiple points over a longer period will allow us to probe remaining questions such as whether the speed of lexical retrieval and articulation has a more dynamic relationship with mid-clause pausing than syntactic-encoding speed does and whether mid-clause pausing improves after end-clause pausing as suggested by Saito et al. (2018) and the current study.

Longitudinal development of CF in connection with UF is an underexplored area in L2 research and offers a rich line of inquiry with theoretical and practical implications for researchers and educators in SLA and language assessment.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/S0272263123000591>.

**Acknowledgments.** This research was supported by funding from the Language Learning Early Career Research Grant Program. I am grateful to all the participants and to the various research assistants, especially to Ben Arakawa, Weiwei An, and Mandy Bozart, who helped with data collection and coding. Finally, I would like to acknowledge the constructive comments provided by the reviewers and the handling editor.

**Competing interest.** The author declares none.

## References

- Bartoń, K. (2022). *Package 'MuMIn'* [Computer software]. R Foundation for Statistical Computing. <https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67, 1–48.
- Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (Version 6.0.43) [Computer program]. Pratt. <http://www.praat.org/>
- De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, 54, 113–132.
- De Jong, N. H., & Bosker, H. R. (2013, August 21–23). *Choosing a threshold for silent pauses to measure second language fluency* [Paper presentation]. Sixth Workshop on Disfluency in Spontaneous Speech, Stockholm, Sweden.
- De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36, 223–243.

- De Jong, N. H., & Mora, J. C. (2019). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, 41, 227–239.
- De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34, 893–916.
- Duran Karaoz, Z., & Tavakoli, P. (2020). Predicting L2 fluency from L1 fluency behaviour: The case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition*, 42, 671–695.
- Felker, E. R., Klockmann, H. E., & de Jong, N. H. (2019). How conceptualizing influences fluency in first and second language speech production. *Applied Psycholinguistics*, 40, 111–136.
- Fillmore, C. (1979). On fluency. In C. Fillmore, D. Kempler, & W. S.-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). Academic Press.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375.
- Freed, B. F. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243–265). University of Michigan Press.
- Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, 27, 379–399.
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498.
- Huensch, A., & Tracy-Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal*, 101, 275–293.
- Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, 30, 555–582.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29, 24–49.
- Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, 64, 809–854.
- Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, 41, 457–480.
- Kahng, J., & Otonya, M. (2021). What elicited imitation can show us: Quantitative and qualitative analyses of longitudinal changes. *TESOL Quarterly*, 55, 284–295.
- Kormos, J. (2006). *Speech production and second language acquisition*. Erlbaum.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–417.
- Levelt, W. (1989). *Speaking: From intention to articulation*. MIT Press.
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., & Makowski, D. (2020). Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, 5, Article 2445. <https://doi.org/10.21105/joss.02445>
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity informal instruction and study abroad learning contexts. *TESOL Quarterly*, 46, 610–641.
- O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, 29, 557–582.
- Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203.
- R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>
- Rizantseva, A. (2001). Second language proficiency and pausing. *Studies in Second Language Acquisition*, 23, 497–526.
- Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*, 39, 593–617.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173–200.
- Skehan, P. (2003). Task based instruction. *Language Teaching*, 36, 1–14.
- Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching*, 54, 97–111.

- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology*, 6, 174–215.
- Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, 65, 71–79.
- van Buuren S., & Groothuis-Oudshoorn K. (2011). “MICE: Multivariate Imputation by Chained Equations in R.” *Journal of Statistical Software*, 45(3), 1–67.

---

**Cite this article:** Kahng, J. (2024). Longitudinal development of second language utterance fluency, cognitive fluency, and their relationship. *Studies in Second Language Acquisition*, 46: 535–549. <https://doi.org/10.1017/S0272263123000591>