Applied Psycholinguistics (2020), **41**, 457–480
doi:10.1017/S0142716420000065

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency

Jimin Kahng*

University of Mississippi
*Corresponding author. Email: jkahng@olemiss.edu

(Received 24 November 2018; revised 08 January 2020; accepted 21 January 2020)

Abstract

Given that utterance fluency in a second language (L2) is associated with not only L2 cognitive fluency but also utterance fluency in the first language (L1), the study examined to what extent different measures of L2 utterance fluency can be explained by L2-specific cognitive fluency and/or the corresponding L1 utterance fluency measures. Utterance fluency measures on speed, breakdown, and repair phenomena and cognitive fluency measures including speed of lexical retrieval, syntactic encoding, and articulation were collected in the L1 and the L2 from 44 Chinese learners of English. The results show that most L2 utterance fluency measures are accounted for by the combination of L2-specific cognitive fluency measures and the equivalent L1 utterance fluency measure, whereas the number of mid-clause silent pauses and corrections, and mean syllable duration are largely explained by L2-specific cognitive measures, suggesting that they reflect L2-specific knowledge and cognitive skills. In contrast, mean silent pause duration and the number of filled pauses are mainly explained by the corresponding L1 utterance fluency measures.

Keywords: cognitive fluency; L1 fluency; L2 fluency

Speaking is a skill under time pressure. As nicely illustrated in de Jong (2014), when people participate in a conversation in a second language (L2), L2 speakers might very well follow the gist of the conversation and know what they want to say. However, by the time they have figured out how to use the appropriate words, structures, and sounds, the conversation may have often already changed topics. Compared to their first language (L1), people typically not only have less knowledge of their second language but also are considerably less fluent using the L2 knowledge they have (Segalowitz, 2010). Along with accuracy and complexity, fluency constitutes a crucial aspect of understanding L2 performance and proficiency (e.g., Bosker, Pinget, Quené, Sanders, & de Jong, 2013; Cucchiarini, Strik, & Boves, 2002; Housen, Kuiken, & Vedder, 2012; Iwashita, Brown, McNamara, & O'Hagan, 2008; Skehan, 1998). In particular, Iwashita et al. (2008) showed that fluency had a stronger impact on distinguishing overall levels of L2 speaking proficiency than any other linguistic features of spoken L2 production such as grammatical accuracy and complexity, vocabulary, or pronunciation.

Although the importance of fluency in L2 acquisition and education has been acknowledged by both researchers and practitioners, how to define the term has been an issue in the field partly due to its polysemous nature. Lennon (1990) points out that there are mainly two different ways the term is used. In the broad sense, it refers to global speaking proficiency, whereas in the narrow sense it relates to how easily and smoothly speech is delivered. According to Segalowitz (2010), even this narrow sense of fluency is a multidimensional construct and a distinction should be made among the three dimensions of fluency—utterance, cognitive, and perceived fluency. Utterance fluency refers to the temporal, pausing, and repair characteristics of utterances. Cognitive fluency is about the speaker's capacity to utilize the underlying cognitive processes responsible for fluent speech production. In contrast, perceived fluency has to do with how listeners make inferences about the speaker's cognitive fluency based on their utterances.

The purpose of the current study is to investigate the underlying mechanism of L2 utterance fluency and to understand what makes L2 speakers fluent. Based on the Segalowitz's (2010, 2016) framework of fluency, L2 utterance fluency is dependent upon L2 cognitive fluency. In addition, previous L2 fluency studies have demonstrated that certain aspects of L2 fluency are associated with one's individual speaking style, or L1 utterance fluency (e.g., de Jong, Groenhout, Schoonen, & Hulstijn, 2015; Derwing, Munro, & Thomson, 2007). Therefore, this study examined the two main factors proposed in the literature to modulate L2 utterance fluency—L2 cognitive fluency and L1 utterance fluency—and their relative contributions to different aspects of L2 utterance fluency. In what follows, previous studies on the relationship between utterance and cognitive fluency, and the L1–L2 relationship in utterance fluency are discussed in turn.

Understanding utterance fluency vis-à-vis cognitive fluency

Utterance fluency can be objectively measured by temporal variables in speech samples, and it has a few different aspects such as speed fluency, breakdown fluency (pause and hesitation phenomena), and repair fluency (Skehan, 2003, 2009; Tavakoli & Skehan, 2005). To identify reliable oral production features of L2 fluency, previous studies have employed a few different approaches: comparing speech from fluent and nonfluent speakers (e.g., Ejzenberg, 2000; Kahng, 2014; Riazantseva, 2001; Riggenbach, 1991; Tavakoli, 2011), investigating the longitudinal development of fluency (e.g., Derwing et al., 2007; Freed, 1995, 2000; Lennon, 1990; Mora & Valls-Ferrer, 2012; Towell, Hawkins, & Bazergui, 1996; Wood, 2010), and relating utterance fluency to perceived fluency by correlating fluency ratings with temporal variables (e.g., Bosker et al., 2013; Cucchiarini et al., 2002; Derwing, Rossiter, Munro, & Thomson, 2004; Kormos & Dénes, 2004; Rossiter, 2009; Suzuki & Kormos, 2019).

The main findings were that speed and pause measures, and to a lesser degree, repair measures have been found to be associated with L2 oral fluency development and perceived fluency (e.g., Bosker et al., 2013; Kormos & Dénes, 2004;

Lennon, 1990; Saito, Ilkan, Magne, Tran, & Suzuki, 2018; Segalowitz & Freed, 2004; Towell et al., 1996). In addition, Bosker, Quené, Sanders, and de Jong (2014) showed that fluency ratings of native and nonnative speech were similarly influenced by the number and length of silent pauses, and by speed manipulations. Kahng (2014) demonstrated that one of the biggest differences between L1 and L2 utterance fluency was found in the number of pauses within a clause and further revealed that listeners were also sensitive to pause location and rated speech samples with pauses within a clause to be less fluent than those with pauses between clauses (Kahng, 2018). More recently, Suzuki and Kormos (2019) also showed that perceived fluency was most strongly correlated with the frequency of mid-clause pauses. In line with these findings, in a cross-sectional study, Saito et al. (2018) found that the number of final-clause pauses distinguished low- from mid-level fluency performance; that of mid-clause pauses further distinguished mid- from high-level performance; and articulation rate further distinguished high from nativelike performance.

What has been underexamined in L2 fluency research is what enables L2 speakers to produce fluent speech, or the cognitive processes responsible for fluent speech production. In understanding cognitive fluency, it is essential to consider what types of knowledge and processes are involved in speech production. According to Levelt (1989, 1999), speech production consists of three major 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). In addition, during speech planning and after speech articulation, one's own speech is self-perceived and monitored (monitoring). These stages are hypothesized to operate simultaneously; however, for less fluent speakers, processes particularly involved in formulation, including lexical retrieval and grammatical encoding, may not be fully automatized, thus resulting in a breakdown in parallel processing and the slowing down of speech (Kormos, 2006; Segalowitz, 2010).

Some previous L2 fluency studies attempted to indirectly explore cognitive fluency through utterance fluency without separate measures of cognitive fluency. For instance, Saito et al. (2018) interpreted their cross-sectional findings on the development of different aspects of utterance fluency based on Kormos's (2006) proposal on the association between different utterance fluency measures and cognitive processes of speech production—the number of final-clause pauses reflects conceptualization; the number of mid-clause pauses reflects formulation (lexical, syntactic, phonological encoding); repair fluency measures reflect monitoring; and speed fluency measures involve all dimensions of speech production, indicating overall automatization of speech production. Based on the proposal, they argued that the developmental pattern of utterance fluency found in their cross-sectional data (i.e., improvement in final-clause pauses > mid-clause pauses > articulation rate) seems to suggest development in conceptualization, formulation, and overall speech production stages, respectively. In a similar manner, Skehan, Foster, and Shum (2016) related different dimensions of utterance fluency to stages of speech production. They proposed a distinction between discourse-level fluency (e.g., end-clause pauses) and clause-level fluency (e.g., mid-clause pauses and repairs) by linking the former to conceptualization and the latter to the formulation and articulation stages of Levelt's speech production model.

However, only a few studies have thus far investigated cognitive fluency with a separate measure and related it to utterance fluency. Kahng (2014) investigated L2 cognitive fluency using stimulated recall and found that the lower proficiency learners reported to have thought about a lot more issues regarding L2 declarative knowledge on grammar and vocabulary at the time of speaking than the higher proficiency learners. This finding was corroborated by the fact that silent pause rate within a clause exhibited one of the strongest correlations with L2 speaking scores and supported the claim in previous studies that pauses within clauses reflect processing difficulties such as lexical retrieval in speech production (e.g., Kircher, Brammer, Levelt, Bartels, & McGuire, 2004; Pawley & Syder, 2000).

There are relatively few studies that examined cognitive fluency by measuring subprocesses of speech production in relation to utterance fluency. Segalowitz and Freed (2004) measured cognitive fluency using a semantic classification task and an attention control test and related the results to gains in utterance fluency. They measured cognitive fluency in the L1 and the L2, and the cognitive fluency measures were corrected by partialling out L1 measures in order to capture L2-specific cognitive processing skills. They found correlations between mean length of run and lexical access speed and efficiency. In their large-scale study, de Jong, Steinel, Florijn, Schoonen, and Hulstijn (2013) aimed to identify utterance fluency measures indicative of cognitive fluency. They measured linguistic knowledge (e.g., vocabulary, grammar, and pronunciation) and processing skills (e.g., speed of morphosyntactic processing, lexical selection, and articulation). Results showed that mean syllable duration was most strongly related to linguistic knowledge and skills, explaining 50% of the variance, whereas pause length was the weakest, explaining only 5% of the variance (number of silent pauses, 22%; filled pauses, 18%; corrections, 25%; and repetitions, 12%). The findings suggest that mean syllable duration is a strong indicator of cognitive fluency but silent pause length is not. More recently, focusing on the articulation stage, de Jong and Mora (2019) examined to what extent L1 and L2 utterance fluency can be explained by individual differences in articulatory skills. Articulatory skills were measured by delayed picture naming tasks in the L1 and the L2 and a diadochokinetic production task (DDK; i.e., producing /pa/, /ta/, /ka/, /pa.ta/, and /pa.ta.ka/ as fast as one can for 5 s). The results showed that the articulatory skills explained only 10% and 7% of variance for silent pause rate and silent pause duration, respectively, in the L1, and 19% and 27% for silent pause rate and silent pause duration, respectively, in the L2, but were not related to articulation rate.

To synthesize findings of previous studies, L2 utterance fluency has been widely researched due to its theoretical and pedagogical importance, and speed, pause, and repair measures have been found to be associated with L2 oral fluency development and perceived fluency. In particular, pure speed measures excluding pause time (i.e., mean syllable duration and articulation rate) and pause rate within a clause seem to be indicative of L2 cognitive fluency. However, L2 cognitive fluency and its relationship with utterance fluency have been underexamined and require more empirical evidence. Although de Jong et al.'s (2013) seminal study provided much insight to capture a comprehensive picture of the relationship between L2 utterance and cognitive fluency, a limitation of the study is that L1 base measures for cognitive fluency were not controlled for. Therefore, the current study examined whether their findings can be replicated when L2-specific cognitive measures were used by addressing individuals' L1 cognitive fluency (Segalowitz, 2010, 2016) with different L1–L2 pairs and settings. Furthermore, based on the potential importance of pause location proposed in the previous literature (de Jong, 2016; Kahng, 2014, 2018; Kormos, 2006; Skehan et al., 2016), the present study categorized pauses into mid-clause and end-clause pauses and examined their relationships with cognitive fluency separately.

L1–L2 relationship in utterance fluency

L1 speakers are often assumed to talk fluently by default (Riggenbach, 1991); however, individual differences in terms of temporal aspects of speech have been documented (e.g., Goldman-Eisler, 1968; Shriberg, 1994). Some previous studies on L2 fluency have further shown that certain aspects of L2 utterance fluency are correlated with those of L1 utterance fluency (e.g., de Jong et al., 2015; Huensch & Tracy-Ventura, 2017; Peltonen, 2018). For instance, de Jong et al. (2015) examined to what extent L2 utterance fluency measures are reliable indicators of L2 proficiency given that fluency is also influenced by personality or speaking style. They measured Dutch L2 proficiency and utterance fluency in the L1 and the L2 of Turkish and English native speakers. The results showed that all the utterance fluency measures of speed, pause, and repair phenomena in the L1 and the L2 were moderately to strongly correlated ($r = .60–.76$), except for mean syllable duration ($r = .37$). In addition, except for silent pause duration, for most measures of utterance fluency, both the corrected (for L1 measures) and the uncorrected utterance fluency measures significantly predicted L2 proficiency; number of silent pauses, filled pauses, repetitions, and corrections each explained 10%–19% of the variance. For mean syllable duration, the corrected measure had a stronger predictive power of L2 proficiency than the uncorrected measure did (41% and 30% of variance explained, respectively).

It should be noted that not all studies on L1–L2 utterance fluency exhibited a high level of correlations across the measures. In their longitudinal study, Derwing, Munro, Thomson, and Rossiter (2009) investigated whether cross-linguistic differences influence the L1–L2 relationship in utterance fluency with Slavic and Mandarin speakers of English. They did find significant L1–L2 correlations with respect to speech rate, number of pauses, and pruned syllables per second. However, the correlations were higher for Slavic speakers than for Mandarin speakers, and the L1–L2 correlations for Mandarin speakers disappeared about 1 year after their arrival to Canada, possibly due to the lack of improvement of the Mandarin group as a whole. Huensch and Tracy-Ventura (2017) also examined the L1–L2 relationship of utterance fluency in their longitudinal study with English L1 Spanish and French majors. Only two of the seven temporal measures of fluency (i.e., mean syllable duration and the number of silent pauses per second) for both groups showed significant L1–L2 correlation before and after 5 months residing abroad. Furthermore, the L1–L2 relationship changed over time and was modulated by cross-linguistic differences and proficiency. They argued that one explanation for their less strong L1–L2 association might be related to participants' experience of residing abroad where the target language is spoken. Participants in de Jong et al. and Derwing et al. had already lived

https://doi.org/10.1017/S0142716420000065 Published online by Cambridge University Press

abroad for an average of 4.5 years and 5 months, respectively, whereas participants in Huensch and Tracy-Ventura had never resided abroad when they were tested.

To summarize, previous studies on the L1–L2 relationship in utterance fluency suggest that certain aspects of utterance fluency are correlated; however, findings are mixed in terms of number and strength of significant correlations and are claimed to be influenced by cross-linguistic differences and proficiency.

The present study

Taken together, L2 utterance fluency has been widely researched; however, the underlying cognitive processes responsible for fluent utterance (i.e., cognitive fluency) and its relationship with utterance fluency has been underresearched (cf. de Jong et al., 2013). In addition, although previous studies have shown that certain aspects of L2 utterance fluency are associated with one's speaking style, or L1 utterance fluency, there have been mixed results in the nature and strength of L1–L2 association, suggesting influence of mitigating factors such as cross-linguistic differences and speaker's L2 proficiency level. Furthermore, these two main factors of L2 utterance fluency—L2 cognitive fluency and L1 utterance fluency—have not been examined simultaneously, thus making it difficult to answer which aspects of L2 utterance fluency reflect L2 cognitive fluency and/or L1 utterance fluency. Therefore, this study attempted to address the gap in the literature and investigated the relative contribution of L2 cognitive fluency and L1 utterance fluency to different measures of L2 utterance fluency with the following research questions.

1. Which of the utterance fluency measures (i.e., speed, breakdown, and repair fluency) show correlation between the L1 and the L2?
2. Which of the L2-specific cognitive fluency measures are correlated with L2 utterance fluency measures?
3. To what extent can different measures of L2 utterance fluency be predicted from L2-specific cognitive fluency and/or the corresponding L1 utterance fluency measures? Based on the analyses, which of the utterance fluency measures can be considered to be more indicative of L2 cognitive fluency than L1 utterance fluency and vice versa?

Before answering the third question, which is the main research question of the study, findings of the first two questions will provide basic information regarding the L1–L2 relationship in utterance fluency and the relationship between cognitive fluency and utterance fluency in L2. They will also demonstrate to what extent the current study replicates previous findings with a different L1–L2 pair and context.

To answer the research questions, utterance fluency measures on speed, breakdown, and repair phenomena and cognitive fluency measures including speed of lexical retrieval, syntactic encoding, and articulation were collected in the L1 and the L2 from Chinese learners of English. Cognitive fluency was measured in both the L1 and the L2 in order to control for L1 base measures and capture L2-specific cognitive fluency. Following Segalowitz's (2010, 2016) suggestion, L2-specific cognitive fluency was operationalized by partialling out L1 measures from L2 cognitive fluency measures.

Finally, any cross-linguistic differences in utterance fluency between English and Mandarin need to be noted because when comparing utterance fluency in the L1 and the L2, some differences might be attributed to differences between the languages themselves rather than individual differences. Pellegrino, Coupé, and Marsico (2011) conducted a cross-linguistic analysis on speech information rate and showed that the syllabic rate in Mandarin was significantly slower than English (5.18 vs. 6.19 syl/s). Mandarin exhibited the slowest syllabic rate among the eight languages analyzed (the other seven included Vietnamese, German, English, Italian, French, Spanish, and Japanese). In terms of pauses, based on the analysis of 267 spoken monologues in Mandarin, Yuan, Xu, Lai, and Lieberman (2016) showed that about 27 silent pauses were used every 100 words in Mandarin speech. For filled pauses, American English and other modern Germanic languages have two common filled pauses: *uh* and *um* (Clark & Fox Tree, 2002; Wieling et al., 2016), whereas Yuan et al. (2016) identified two basic filled pauses in Mandarin: *e* and *en*; however, they found that the use of the nasal-final filled pause (*en*) was higher in female than in male speakers, as was found in the studies of Germanic languages (*uh* vs. *um*).

Method

Participants

Forty-four Chinese speakers of English (24 males, 20 females) participated in the study. They voluntarily participated in the study and received $50 for their participation. Their age ranged from 20 to 43 ($M_{age} = 27$; $SD_{age} = 6$). Their length of residence in an English-speaking country was less than 6 months ($M_{LOR} = 1.9$ months; $SD_{LOR} = 1.8$ months). They started to learn English around the age of 10 ($M_{AO} = 10.6$; $SD_{AO} = 2.4$). Based on the results of the grammar and vocabulary sections of DIALANG, an online diagnostic language test developed by Lancaster University, most of them were identified to be intermediate learners of English (3 A2s; 38 Bs; 3 C1s), according to the common European framework (Council of Europe, 2001).

Speaking tasks

Materials

Participants answered two questions on familiar topics of personal relevance both in English and in Mandarin. The questions resembled the two independent speaking tasks of iBT TOEFL (Educational Testing Service, 2018). The first question asked them to express their personal preference from a given category (e.g., important time, people, and places) and the other question asked them to make and support their personal choice between two options (e.g., living in a big or small city, and eating out or at home). For each type of question, one of six such prompts was randomly selected and presented to each participant. The two sets of six prompts were on daily life so that the questions were comparable and participants were familiar with the topic and were able to talk naturally without much difficulty.

Procedure

Each participant answered two questions (i.e., one on personal preference and the other on choice between two options) first in English and next in Mandarin,

https://doi.org/10.1017/S0142716420000065 Published online by Cambridge University Press

answering four different questions in total. For each question, participants had 15 s to prepare for their answer and were asked to talk for about a minute. Participants' 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 recordings were transcribed including information regarding silent and filled pauses, repetitions, and corrections. Silent and filled pauses were identified and their length was measured in milliseconds (ms) by listening to each speech sample and examining the waveform and spectrogram using the Annotate to TextGrid (silences) function in Praat (Boersma & Weenink, 2018). Silences of 250 ms or longer were counted as silent pauses (de Jong & Bosker, 2013; Kahng, 2012). Pauses were further categorized into mid-clause pauses and final-clause pauses depending on their location (Foster, Tonkyn, & Wigglesworth, 2000). Following Skehan's (2003) taxonomy of utterance fluency, its three different dimensions (speed, breakdown, and repair) were measured. For speed fluency, mean syllable duration (ms) was calculated by dividing speech time excluding pause time by total number of syllables (de Jong et al., 2013).[1] For breakdown fluency, along with mean length of silent pauses, the number of silent and filled pauses in the middle and at the end of clauses per 100 syllables were calculated.[2] For repair fluency, the number of repetitions and self-corrections per 100 syllables were calculated. All the utterance fluency measures were aggregated across the two speaking tasks.

Tasks for cognitive fluency

A battery of cognitive fluency tasks was used in order to measure linguistic knowledge and cognitive processes involved in speech production (de Jong et al., 2013). To provide an overview of the test battery, for linguistic knowledge, L2 vocabulary and grammar knowledge were assessed. Based on the potential link between formulaic language and fluent speech processing and production (e.g., Bolibaugh & Foster, 2013; Bybee, 2002; Siyanova-Chanturia & Pellicer-Sánchez, 2018; Wood, 2010), a separate measure of phrasal vocabulary size was included in the test battery (Martinez, 2011; Martinez & Schmitt, 2012). In examining cognitive processes, measures were adopted to represent subprocesses of Levelt's speech production model (1989, 1999), focusing on the formulation and articulation stages (de Jong et al., 2013). For the formulation stage, speed of lexical retrieval and syntactic encoding were measured, and for the articulation stage, speed of articulation was measured using a delayed picture naming task. Processes involved in the conceptualization stage were not included for measurement because while planning what to say next, world knowledge utilized in the stage is not assumed to be organized in language-specific terms (de Bot, 1992; Levelt, 1989, 1999); therefore, L2-specific fluency issues are not expected (Segalowitz, 2010). Following Segalowitz's (2010, 2016) suggestion, all the tasks assessing subprocesses of speech production were conducted in both the L1 and the L2 in order to control for the baseline L1 cognitive process and obtain L2-specific cognitive processing skills by partialling out L1

https://doi.org/10.1017/S0142716420000065 Published online by Cambridge University Press

measures. In the following sections, detailed descriptions of each of the tasks and measures are provided.

L2 vocabulary and grammar knowledge

Materials and procedure. To measure participants' L2 vocabulary knowledge, two separate tests were used—the vocabulary section of DIALANG and the Phrase Vocabulary Size Test (PVST: Martinez, 2011). PVST is a test to estimate how many phrasal expressions (e.g., *have to*, *go away*) an English learner knows receptively and contains 50 multiple-choice questions, 10 at each of the first 5 thousand-levels of word frequency based on the PHRASE List (Martinez & Schmitt, 2012). To measure L2 grammar knowledge, the grammar section of DIALANG was used. The DIALANG tests were administered individually using a laptop, and the test scores generated by the system were recorded by the researcher. The participants took a paper-and-pencil version of PVST, and for each correct response, 1 point was awarded.

Lexical retrieval speed: Formulation stage

Materials. Forty pictures were selected from the list of pictures by Snodgrass and Vanderwart (1980). Half of the pictures were randomly selected for an L1 picture naming task and the other half were used for an L2 task. A Chinese speaker 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 in the list of words in Mandarin and in English, t (38) $= 1.80$, $p = .08$.

Procedure and measure. The task was administered by the PsychoPy software package (Peirce et al., 2019). Participants were asked to name each of the pictures presented on the screen as fast and accurately as possible. Following de Jong et al. (2013), after a fixation cross was presented for 1500 ms, the target picture was presented for 2000 ms, which was followed by a blank screen for 500 ms. The pictures were presented in a random order for each participant. Before the actual experiment, each participant completed a practice session with a few pictures that were not included in the actual tasks, to ensure familiarity with the task. Participants' production was recorded with a Blue Snowball USB microphone (frequency response 40 Hz–18 KHz) at a 44 KHz sampling rate (16-bit resolution; 1 channel). The reaction time between the presentation of the picture and the beginning of the correct response was measured using Praat (Boersma & Weenink, 2018).

Syntactic encoding speed: Formulation stage

Materials. In the syntactic encoding task, participants were presented with the beginning of a sentence (e.g., *I expect* . . .) and asked to select an option (e.g., A. *them* . . . B. *go* . . .) that best followed the beginning of the sentence (Hulstijn, Van Gelderen, & Schoonen, 2009; Lim & Godfroid, 2015). The task required them to process a sentence fragment and to choose what best fits into the syntactic structure they created. Following Lim and Godfroid (2015), the sentence fragments and the options were kept as short as possible (ranging from one to three words) in order to limit higher level semantic analysis and focus on participants' syntactic

https://doi.org/10.1017/S0142716420000065 Published online by Cambridge University Press

processing skills. Twenty items containing the beginning of a sentence and two corresponding options were developed for each of the English and Mandarin versions. As the task was designed to assess the speed of syntactic processing rather than grammatical accuracy, basic syntactic structures were targeted so that intermediate learners were able to answer them. The syntactic structures covered by the items included word order in declarative and interrogative sentences, subject–verb agreement, and different types of phrase structure. All the items were pilot-tested by English and Chinese speakers for the corresponding language version.

Procedure and measure. The syntactic encoding task was administered using the PsychoPy software package (Peirce et al., 2019). Participants were first presented with the beginning of a sentence on the first screen, and on the next screen two possible options followed. Participants were asked to select an option as fast and as accurately as possible and were also told that the options would not complete the sentence but one of them would best follow the beginning of the sentence. Before the actual experiment, each participant completed a practice session with a few items that were not included in the actual experiment, to ensure familiarity with the task. The period between the presentation of the two options and each participant's correct keyboard response was automatically measured and the reaction times for correct responses were used for analysis.

Speed of articulation (pronunciation duration): Articulation stage

Materials. The materials were the same as the ones used for the lexical retrieval measure.

Procedure and measures. Participants completed the picture naming task one more time. However, this time they were instructed to prepare their response but wait until a cue was given before naming a picture. Following de Jong et al. (2013), after a fixation cross was presented for 500 ms, the target picture was presented for 2000 ms, which was followed by a short beep. Participants were asked to name the picture right after they heard the beep. The picture remained on the screen for another 1000 ms. Pictures were presented in a random order for each participant. Before the actual experiment, each participant completed a practice session with a few pictures that were not included in the actual tasks, to ensure familiarity with the task. Participants' production was recorded with a Blue Snowball USB microphone (frequency response 40 Hz–18 KHz) at a 44 KHz sampling rate (16-bit resolution; 1 channel). The duration of response between the beginning and the end of their correct response was measured using Praat (Boersma & Weenink, 2018).

Analysis

All speech recordings (i.e., spontaneous speaking tasks, picture naming tasks, and delayed picture naming tasks) were transcribed, annotated, and measured by two Mandarin–English bilingual research assistants for Mandarin speech and two native English-speaking research assistants for English speech. Once the samples were transcribed, annotated, and measured by the first research assistant, their accuracy was checked by a second research assistant and corrections were made when necessary. The research assistants and the author worked closely with each other through weekly or biweekly meetings to ensure the English and Mandarin recordings were transcribed, annotated, and measured as similarly and accurately as possible.

For an overview of statistical analyses used in the study, first of all, differences in fluency between the L1 and the L2 were examined using multivariate analyses of variance (MANOVAs) and follow-up univariate analyses of variance (ANOVAs) with Bonferroni correction after testing assumptions of parametric tests. The variables that violated the assumptions of parametric tests (i.e., mean silent pause duration, number of mid-clause silent pauses, filled pauses, repetitions, and corrections) were log-transformed. All the transformed data improved in terms of normality after the transformation.

In examining the first two research questions of the L1–L2 relationship in utterance fluency and the relationship between L2 utterance fluency and L2-specific cognitive fluency, Pearson correlations were conducted. In particular, for the second research question, in order to capture L2-specific cognitive fluency, the residuals from linear regression models predicting L2 cognitive fluency from L1 cognitive fluency were calculated and used as the corrected L2-specific measures of cognitive fluency.

Finally, to examine the third question of the predictive power of L2 cognitive fluency and L1 utterance fluency for L2 utterance fluency, stepwise multiple regressions were conducted. Before carrying out the regression analyses, assumptions of linear regression including normality, linearity, homoscedasticity, and multicollinearity were tested for all the models. Based on the examination of P-P plots, scatterplots of the residuals, and VIF values, the regression models and variables satisfied the assumptions.[3]

Results

In what follows, language differences in the measures of utterance fluency and cognitive fluency (i.e., results of MANOVAs and ANOVAs), and the results of the correlation analyses and the regression analyses are reported in turn.

Language differences in utterance fluency and cognitive fluency

To summarize the mean utterances produced by the participants, they produced 226 syllables ($SD = 98$) in L1 Mandarin and 155 syllables ($SD = 67$) in L2 English per speech sample. Table 1 shows descriptive statistics and language differences in the measures of utterance fluency and cognitive fluency. Results of two MANOVAs, using Pillai's trace, showed that participants' L1 and L2 performances were significantly different for the measures of utterance fluency, $V = .90$, $F\,(8, 14) = 16.25$, $p < .001$, and for cognitive fluency, $V = .85$, $F\,(3, 39) = 71.34$, $p < .001$. Follow-up univariate ANOVAs on the eight dependent variables in utterance fluency revealed significant language differences, after Bonferroni correction, in mean syllable duration, number of mid-clause and final-clause pauses, and repetitions with large effect sizes ($\eta_p^2 = .01 =$ small, $.06 =$ medium, $.14 =$ large; Cohen, 1988). In contrast, there was no significant difference in mean silent pause duration or number of corrections between their L1 and L2 speech. For cognitive

https://doi.org/10.1017/S0142716420000065 Published online by Cambridge University Press

Table 1. Descriptive statistics and language differences in utterance fluency and cognitive fluency

	L1		L2					
	M	*SD*	*M*	*SD*	*F*	*df*	*p*	η_p^2
Utterance fluency								
Mean syllable duration (ms)	203	28	317	49	135.42	1	<.001*	.87
Mean silent pause duration (ms)$_t$	575	139	580	162	0.33	1	.570	.02
Number of								
Mid-clause silent pauses/100 syllables$_t$	3.81	1.98	10.55	5.28	69.44	1	<.001*	.77
Final-clause silent pauses/100 syllables	5.61	1.45	8.36	2.20	47.54	1	<.001*	.69
Mid-clause filled pauses/100 syllables$_t$	1.24	1.32	4.21	3.48	28.42	1	<.001*	.58
Final-clause filled pauses/100 syllables$_t$	1.59	1.07	3.08	1.85	20.85	1	<.001*	.50
Repetitions/100 syllables$_t$	0.45	0.50	1.74	1.35	29.60	1	<.001*	.59
Corrections/100 syllables$_t$	1.02	0.72	1.00	0.81	0.29	1	.60	.01
Cognitive fluency								
Syntactic encoding speed (ms)	1062	207	2062	598	184.86	1	<.001*	.82
Lexical retrieval speed (ms)	1093	138	1332	157	76.80	1	<.001*	.65
Pronunciation duration (ms)	570	100	588	62	1.40	1	.24	.03
L2 grammar knowledge (max = 6)			3.82	0.97				
L2 vocabulary knowledge (max = 6)			3.87	0.77				
L2 phrasal vocabulary size (max = 50)			37.46	4.99				

Note: The subscript t next to a variable indicates that the variable was log-transformed for inferential statistics. An asterisk indicates significant difference after Bonferroni correction.

fluency, follow-up univariate ANOVAs on the three dependent variables showed that after Bonferroni correction, there was a significant L1–L2 difference in the speed of syntactic encoding and lexical retrieval, whereas there was no significant language difference in pronunciation duration.

Relating L1 and L2 fluency

In order to examine which fluency measures in the L1 and the L2 are correlated to what extent, Pearson correlations were conducted for utterance fluency and cognitive fluency as shown in Tables 2, 3, 4, and 5. The tables show the L1–L2 correlations for the utterance (Tables 2 and 3) and cognitive fluency measures (Tables 4 and 5). Tables 2 and 4 show the results between the measures in L1 Mandarin, and Tables 3 and 5 show the results in L2 English. The diagonal in Tables 3 and 5 shows the correlations between the L1 and the L2 for each measure. The tables show that in terms of the intercollinearity of measures within the L1 and the L2, the majority of correlations are low (r = .25) to moderate (r = .4; Plonsky & Oswald, 2014). In Tables 2 and 3, the correlations between the number of mid-clause silent pauses and mid-clause filled pauses is relatively higher than

https://doi.org/10.1017/S0142716420000065 Published online by Cambridge University Press

Table 2. Pearson correlations between utterance fluency measures within L1 Mandarin

	1	2	3	4	5	6	7	8
1. Mean syllable duration	1							
2. Mean silent pause duration	.19	1						
3. Number of mid-clause silent pauses	.51**	.04	1					
4. Number of final-clause silent pauses	.30*	.44**	.34*	1				
5. Number of mid-clause filled pauses	.45**	−.27	.55**	.29	1			
6. Number of final-clause filled pauses	.25	−.10	.13	.32*	.61**	1		
7. Number of repetitions	.31	.09	.12	.45*	.46*	.46*	1	
8. Number of corrections	.21	−.25	.52**	.07	.53**	.15	.57*	1

*$p < .05$. **$p < .01$.

Table 3. Pearson correlations between utterance fluency measures within L2 English and with the relation between the L1 and the L2 on the diagonal (in bold)

	1	2	3	4	5	6	7	8
1. Mean syllable duration	**.43****							
2. Mean silent pause duration	−.01	**.57****						
3. Number of mid-clause silent pauses	.41**	.35*	**.34***					
4. Number of final-clause silent pauses	.39**	.17	.52**	**.22**				
5. Number of mid-clause filled pauses	.50**	−.05	.61**	.19	**.36***			
6. Number of final-clause filled pauses	.36*	−.08	.12	.46**	.31*	**.46****		
7. Number of repetitions	.10	.11	.59**	.15	.35*	−.12	**.04**	
8. Number of corrections	−.01	.20	.46**	.35*	.24	−.10	.50**	**.41***

*$p < .05$. **$p < .01$.

Table 4. Pearson correlations between cognitive fluency measures within L1 Mandarin

	1	2	3
1. Syntactic encoding speed	1		
2. Lexical retrieval speed	.56**	1	
3. Pronunciation duration	.05	.12	1

*$p < .05$. **$p < .01$.

other measures in both the L1 ($r = .55$) and the L2 ($r = .61$), which seems to be in line with a previous finding that in both L1 and L2 speech, filled pauses often accompany silent pauses (Kahng, 2014). The diagonal in Table 3 shows that the L1–L2 correlation of mean silent pause duration was stronger than other measures

https://doi.org/10.1017/S0142716420000065 Published online by Cambridge University Press

Table 5. Pearson correlations between cognitive fluency measures within L2 English and with the relation between the L1 and the L2 on the diagonal (in bold)

	1	2	3
1. Syntactic encoding speed	**.64****		
2. Lexical retrieval speed	.42**	**.34***	
3. Pronunciation duration	.11	−.19	**.52****

*$p < .05$. **$p < .01$.

($r = .57$), whereas those of mean syllable duration, the number of silent pauses, repetitions, and corrections were weak to moderate. The diagonal in Table 5 reveals moderate L1–L2 correlations in syntactic encoding speed and pronunciation duration and a low correlation in the speed of lexical retrieval.

Relating L2 utterance fluency with L2-specific cognitive fluency

In examining which measures of L2 utterance fluency are related to L2-specific cognitive fluency to which extent, Pearson correlations were carried out. L2 cognitive fluency measures were corrected by partialling out L1 measures in order to gauge L2-specific cognitive processing skills.

As Table 6 shows, the number of mid-clause silent pauses exhibited correlations with the majority of cognitive measures (i.e., speed of syntactic encoding and lexical retrieval, vocabulary knowledge, and phrasal vocabulary size). Mean syllable duration was significantly correlated with speed of syntactic encoding and lexical retrieval. The number of mid-clause filled pauses was also correlated with speed of lexical retrieval and phrasal vocabulary size. The number of repetitions showed positive correlation with syntactic encoding speed and negative correlation with phrasal vocabulary size. The number of corrections was positively correlated with syntactic encoding speed and the number of final-clause silent pauses was negatively correlated with vocabulary knowledge. However, neither mean silent pause duration nor the number of final-clause filled pauses was significantly correlated with any of the L2-specific cognitive measures. Among the cognitive measures, pronunciation duration and grammar knowledge did not demonstrate correlation with any of the L2 utterance fluency measures.

Predicting L2 utterance fluency

In order to examine which of the L2-specific cognitive fluency and/or L1 utterance fluency measures have a predictive power for different L2 utterance fluency measures, stepwise multiple regressions were performed with eight L2 utterance fluency measures as outcome variables (see Table 1). As predictor variables, the corresponding L1 utterance fluency measure of each dependent variable (e.g., L1 mean syllable duration for L2 mean syllable duration) and a set of six cognitive measures (see Table 1) were included in the model.

Table 7 shows the results of stepwise multiple regressions including total R^2, changes in R^2, standardized coefficients, and F values when predictors were added

https://doi.org/10.1017/S0142716420000065 Published online by Cambridge University Press

Table 6. Pearson correlations between L2 utterance fluency and L2-specific cognitive fluency

	Mean syllable duration	Mean silent pause duration	Number of mid- clause silent pauses	Number of final-clause silent pauses	Number of mid-clause filled pauses	Number of final-clause filled pauses	Number of repetitions	Number of corrections
Syntactic encoding speed	.39*	.07	.47**	.18	.28	.08	.31*	.54**
Lexical retrieval speed	.50**	.22	.30*	.20	.32*	.07	.02	.05
Pronunciation duration	.04	–.03	–.13	–.08	–.13	–.03	.06	–.06
Grammar knowledge	–.02	–.02	–.14	–.04	–.02	.05	–.12	–.14
Vocabulary knowledge	–.13	–.12	–.41**	–.33*	–.25	–.18	–.17	–.19
Phrasal vocabulary size	–.26	.02	–.32*	–.19	–.36**	–.11	–.33*	–.18

*$p < .05$. **$p < .01$.

https://doi.org/10.1017/S0142716420000065 Published online by Cambridge University Press

Table 7. Results of stepwise multiple regressions

Outcome variables	Predictors	R^2	ΔR^2	β	ΔF	*df*	*p*
L2 mean syllable duration	L2-specific lexical retrieval speed	.24	.24	.42	12.22	1, 39	.001
	L2-specific lexical retrieval speed + L2 phrasal vocabulary	.40	.16	−.42	9.97	1, 38	.003
	L2-specific lexical retrieval speed + L2 phrasal vocabulary + L1 mean syllable duration	.47	.07	.29	5.43	1, 37	.025
L2 mean silent pause duration	L1 mean silent pause duration	.33	.33	.58	18.91	1, 39	<.001
	L1 mean silent pause duration + L2-specific lexical retrieval speed	.40	.07	.26	4.17	1, 38	.048
L2 number of mid-clause silent pauses	L2-specific syntactic encoding speed	.23	.23	.48	11.46	1, 38	.002
L2 number of mid-clause filled pauses	L2 phrasal vocabulary	.15	.15	−.51	6.60	1, 39	.014
	L2 phrasal vocabulary + L1 number of mid-clause filled pauses	.35	.20	.47	12.04	1, 38	.001
L2 number of final-clause filled pauses	L1 number of final-clause filled pauses	.23	.23	.47	11.33	1, 39	.002
L2 number of repetitions	L2 phrasal vocabulary	.10	.10	−.33	4.45	1, 39	.041
	L2 phrasal vocabulary + L1 number of repetitions	.21	.11	.32	4.93	1, 38	.032
L2 number of corrections	L2-specific syntactic encoding speed	.30	.30	.55	16.58	1, 39	<.001

to the model. There is some variability in the amount of variance of L2 utterance fluency explained by corrected L2-specific cognitive fluency and L1 utterance fluency, ranging from 21% (repetitions) to 47% (mean syllable duration).

Most L2 utterance fluency measures turned out to have the combination of one or more L2-specific cognitive fluency measures and the equivalent L1 utterance fluency measure as significant predictors. In contrast, it is noteworthy that the number of mid-clause silent pauses and the number of corrections had an L2-specific cognitive fluency measure only in the model (i.e., speed of syntactic encoding), which explained 23% and 30% of the variance, respectively. In a similar manner, for L2 mean syllable duration, L2-specific cognitive fluency measures (i.e., lexical retrieval speed and phrasal vocabulary size) explained most of its variance (40%) and the equivalent L1 measure explained an additional 7% of the variance. In contrast, the corresponding L1 utterance fluency explained the majority of variance of mean silent pause duration (33%) and the number of filled pauses in the L2, 20% for the mid-clause and 23% for the final-clause filled pauses, respectively. One more thing to note is that none of the predicting variables significantly explained the L2 number of final-clause silent pauses.

Discussion

Which of the utterance fluency measures (i.e., speed, breakdown, and repair fluency) show correlation between the L1 and the L2?

The participants were more fluent in their L1 compared to their L2 for most utterance fluency measures, including mean syllable duration, number of mid-clause and final-clause pauses, and repetitions with large effect sizes. The findings are corroborated by the fact that most participants were intermediate learners of English; however, it is noteworthy that mean silent pause duration and the number of corrections in L1 and L2 speech were not significantly different.

Correlation analyses showed that for the L1–L2 relationship in utterance fluency, mean silent pause duration exhibited the strongest L1–L2 correlation ($r = .57$), whereas those of mean syllable duration, the number of mid-clause pauses, and corrections were weak to moderate. Overall, the findings are in line with previous studies on L1–L2 utterance fluency. De Jong et al. (2015) also found that mean silent pause duration has the strongest L1–L2 correlation and mean syllable duration has a weak L1–L2 correlation with Turkish and English learners of Dutch. The findings suggest that L2 mean silent pause duration seems to largely reflect an individual's speaking style, transferred from their first language.

Another thing to note is that, as found in Huensch and Tracy-Ventura (2017), the L1–L2 correlations in the current study tend to be less strong than those in de Jong et al. (2015). The weaker L1–L2 association might be related to participants' short length of residence at the time of testing (Huensch & Tracy-Ventura, 2017). Participants in de Jong et al. had already lived abroad for an average of 4.5 years, whereas Huensch and Tracy-Ventura's participants had never resided abroad and the length of residence of the current participants was also minimal ($M = 1.9$ months).

Which of the L2-specific cognitive fluency measures are correlated with L2 utterance fluency measures?

In order to capture L2-specific cognitive fluency measures, L2 cognitive measures were corrected by partialling out L1 cognitive measures. The correlation analyses between L2-specific cognitive fluency and L2 utterance fluency revealed that the number of mid-clause silent pauses, in particular, had significant correlations with the majority of L2 cognitive fluency measures such as speed of syntactic encoding and lexical retrieval, L2 vocabulary knowledge, and L2 phrasal vocabulary size, suggesting that the measure is a strong indicator of L2 knowledge and cognitive skills. Mean syllable duration was also significantly correlated with speed of syntactic encoding and lexical retrieval. However, mean silent pause duration was not significantly correlated with any of the L2-specific cognitive measures. Among the cognitive processing measures, syntactic encoding speed and lexical retrieval speed demonstrated association with L2 utterance fluency measures, whereas pronunciation duration did not demonstrate correlation with any of the L2 utterance fluency measures.

The findings are in general compatible with de Jong et al. (2013) although their L2 cognitive measures were not corrected for L1 baseline measures. They also found that mean syllable duration was correlated with all of the cognitive measures, whereas mean silent pause duration was not associated with most of them. For the cognitive processing measures as well, sentence construction speed and lexical retrieval speed exhibited tight association with L2 utterance fluency measures but pronunciation duration did not, as found in the present study. However, the current finding on mid-clause silent pauses could not be compared as the measure was not included in de Jong et al. (2013). Taken together, the findings seem to suggest that L2 utterance fluency measures tend to be more strongly associated with the formulation than with the articulation stage of speech production.

In addition, the association found between phrasal vocabulary size and L2 utterance fluency is noteworthy. Phrasal vocabulary size was negatively correlated with the number of mid-clause silent and filled pauses, and repetitions. Although the importance of formulaic language on fluent speech has been pointed out in the literature (e.g., Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Bybee, 2002; Siyanova-Chanturia, & Pellicer-Sánchez, 2018; Wood, 2010), this is one of the first studies that demonstrates a significant correlation between receptive phrasal vocabulary size and L2 utterance fluency.

To what extent can different measures of L2 utterance fluency be predicted from L2-specific cognitive fluency and/or the corresponding L1 utterance fluency measures?

To examine which of the L2-specific cognitive fluency and/or L1 utterance fluency measures have a predictive power for different L2 utterance fluency measures, stepwise multiple regressions were conducted. Whereas most L2 utterance fluency measures were predicted by the combination of L2-specific cognitive fluency measures and the equivalent L1 utterance fluency measure, the number of mid-clause silent pauses had an L2 cognitive fluency measure only (i.e., syntactic encoding speed) as a predictor in the model. Along with its weak L1–L2 correlation and significant correlations with the speed of the formulation stage (syntactic encoding and lexical retrieval) in the L2, the results suggest that the number of mid-clause silent pauses in the L2 seems to mainly reflect L2-specific knowledge and skills.

The findings on the number of mid-clause silent pauses are consistent with previous studies in L1 and L2 speech production. Research on L1 speech production has shown that pauses at clause boundaries are associated with more general long-term planning of the following clause, whereas pauses within clauses tend to occur before unpredictable and infrequent words and be related to speech planning and in particular lexical retrieval (Kircher et al., 2004; Levelt, 1983; Maclay & Osgood, 1959). For instance, using functional magnetic resonance imaging, Kircher et al. (2004) found that pauses within clauses are associated with an activation of the left temporal lobe, which is argued to suggest that pauses within clauses are a correlate of lexical retrieval.

Previous studies have also shown that L2 speech often has pauses in the middle of clauses, whereas L1 speech has pauses at syntactic boundaries (Deschamps, 1980; Kahng, 2014; Tavakoli, 2011). The silent pause rate within a clause has also been shown to be significantly correlated with L2 speaking proficiency (Kahng, 2014). In Saito et al. (2018), the number of final-clause pauses only distinguished low- from mid-/high-level fluency performance, whereas that of mid-clause pauses was able to further distinguish among low-, mid-, and high-level fluency performance, suggesting that the number of mid-clause pauses is an important indicator of L2 fluency development. The current findings also seem to support proposals by Kormos (2006) and Skehan et al. (2016) in which mid-clause pauses are argued to be related to the formulation stage of speech production where lexical, syntactic, and phonological encoding occurs. The number of mid-clause pauses in an L2 in the current study was predicted by L2-specific syntactic encoding speed and significantly correlated with the L2-specific speed of syntactic encoding and lexical retrieval.

In terms of L2 mean syllable duration, the two cognitive measures in particular—lexical retrieval speed and phrasal vocabulary size—explained most of its variance (40%) and the equivalent L1 measure explained an additional 7% of the variance. The results are very much in line with the studies by de Jong and her colleagues. In de Jong et al. (2013), 50% of variance of L2 mean syllable duration was explained by L2 linguistic knowledge and processing skills altogether. In addition, de Jong et al. (2015) demonstrated that the predictive power of mean syllable duration for L2 proficiency increased from 30% to 40% when the measure was corrected for L1 behavior by partialling out L1 mean syllable duration.

In contrast, the current study showed that for mean silent pause duration and the number of filled pauses, each of the equivalent L1 utterance fluency measures has a stronger predictive power than L2 cognitive fluency measures, suggesting they reflect more language-general fluency rather than L2-specific fluency. The findings are also compatible with de Jong et al. (2015) in which both mean silent pause duration and the number of filled pauses per second exhibited strong L1–L2 correlations ($r = .65–.73$).

Another interesting finding is that none of the cognitive measures and the corresponding L1 utterance fluency measures significantly predicted the number of final-clause silent pauses in the L2. This contrasts with the finding that 23% of the variance of the number of mid-clause silent pauses was predicted by an

L2 cognitive fluency measure. The findings seem to confirm that silent pauses in different locations (i.e., within a clause vs. between clauses) involve different cognitive processing and are fundamentally different as discussed earlier. Whereas mid-clause pauses may be related to the formulation stage, final-clause pauses may be associated with the conceptualization stage, or more general long-term planning of the following clause (Kircher et al., 2004; Kormos, 2006; Skehan et al., 2016). The current findings on the mid-clause pauses seem to support the hypothesis. However, the association between final-clause pauses and conceptualization has not been tested in the current study, and thus requires further research.

Conclusion

L2 utterance fluency is dependent upon L2 cognitive fluency and has also been shown to be related to L1 utterance fluency. Nevertheless, these two main factors of L2 utterance fluency have not been examined simultaneously, thus making it difficult to understand the relative contribution of L2-specific cognitive fluency and L1 utterance fluency to different measures of L2 utterance fluency. To address the gap in the literature, the present study examined to what extent different measures of L2 utterance fluency can be predicted from L2-specific cognitive fluency and/or the corresponding L1 utterance fluency measures and identified which aspects of L2 utterance fluency are indicative of L2 cognitive fluency or L1 utterance fluency.

The results showed that most L2 utterance fluency measures were predicted by the combination of L2 cognitive fluency measures and the equivalent L1 utterance fluency measure, whereas the number of mid-clause silent pauses and mean syllable duration in L2 were largely predicted by L2 cognitive measures, suggesting that they reflect L2-specific knowledge and processing skills. However, mean silent pause duration and the number of filled pauses in the L2 were largely predicted by the corresponding L1 utterance fluency measures.

The current study aimed to contribute to the understanding of what constitutes L2 utterance fluency in relation to both L1 utterance fluency and different aspects of L2-specific cognitive fluency of the speech production process. The findings may help researchers and test developers in selecting fluency measures for different purposes. For instance, when L2 utterance fluency measures are collected to reflect L2 cognitive skills or proficiency, the number of mid-clause silent pauses and the mean syllable duration are likely to serve the purpose better than the mean silent pause duration.

In addition, one of the pedagogical implications of the study relates to the close relationship between L2 utterance fluency measures and L2 receptive phrasal vocabulary size. Learning and utilizing formulaic language has been suggested to be a way to improve L2 fluency (e.g., Boers et al., 2006), and its potential usefulness has been confirmed by the significant negative correlations between receptive phrasal vocabulary size and the number of mid-clause silent and filled pauses, and repetitions in the current study.

This study is one of the first attempts to examine the relationship between utterance fluency and cognitive fluency in the L1 and the L2 in a comprehensive manner and the findings need to be replicated with a larger sample size in different languages. Although the study included different aspects and measures of utterance and cognitive fluency in both the L1 and the L2, there were measures that were not included in the analysis. For instance, whether and/or what kind of cognitive tasks to use in measuring processing involved in conceptualization (e.g., Felker, Klockmann, & de Jong, 2019) and L1 linguistic knowledge (e.g., grammar and vocabulary) could be addressed in future research. Another important avenue of future research is whether and how the relationship between utterance fluency and cognitive fluency in the L1 and the L2 changes over time with more experience in the L2.

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.

Notes

1. As a pure measure of speed fluency, articulation rate (i.e., number of syllables produced per minute excluding pause time) is another popular choice. An advantage of mean syllable duration (i.e., inverse articulation rate) is that the direction of the measure is compatible with the rest of the utterance fluency measures. In other words, like other utterance fluency measures, higher numbers represent more disfluency and lower numbers represent more fluency, facilitating interpretation of the findings. In addition, the measure was selected over articulation rate in order to increase direct comparability with recently published relevant studies such as de Jong et al. (2013, 2015) and Huensch and Tracy-Ventura (2017), where the same measure was used.

2. Length of silent pauses was not computed separately for mid- and final-clauses based on previous findings. For instance, in Kahng (2014), there was no significant difference between native and nonnative speakers in terms of mean length of silent pauses in different locations. In addition, the length of silent pauses was not correlated with L2 proficiency.

3. Power analysis was conducted using the G*Power software (Faul, Erdfelder, Buchner, & Lang, 2009). The results of the post hoc power analyses showed that the mean level of achieved power was over 0.81. With the maximum number of predictors (7), the power was over 0.74.

References

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, H. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, **10**, 245–261.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0.43. Retrieved September 8, 2018, from http://www.praat.org/

Bolibaugh, C., & Foster, P. (2013). Memory-based aptitude for nativelike selection: The role of phonological short-term memory. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 205–230). Amsterdam: Benjamins.

Bosker, H. R., Pinget, A., Quené, H., Sanders, T., & de Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs, *Language Testing*, **30**, 159–175.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). The perception of fluency in native and nonnative speech. *Language Learning*, **64**, 579–614.

Bybee, J. (2002). Phonological evidence of exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, **24**, 215–221.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, **84**, 73–111.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. New York: Cambridge University Press.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, **111**, 2862–2873.

de Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, **13**, 1–24.

de Jong, N. H. (2014). Teaching speaking. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1071–1077). London: Wiley-Blackwell.

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). *Choosing a threshold for silent pauses to measure second language fluency*. Paper presented at the 6th Workshop on Disfluency in Spontaneous Speech, Stockholm.

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. doi:10.1017/S0272263117000389

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.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, **29**, 359–380.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, **31**, 533–557.

Derwing, T. M., Rossiter, M., Munro, M., & Thomson, R. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, **54**, 655–679.

Deschamps, A. (1980). The syntactic distribution of pauses in English spoken as a second language by French students. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 271–285). The Hague, Netherlands: Mouton.

Educational Testing Service. (2018). *TOEFL test prep planner*. Retrieved from https://www.ets.org/s/toefl/pdf/toefl_student_test_prep_planner.pdf

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 287–314). Ann Arbor, MI: University of Michigan Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, **41**, 1149–1160.

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. doi:10.1017/S0142716418000474

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, **21**, 354–375.

Freed, B. F. (1995). Do students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Amsterdam: Benjamins.

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). Ann Arbor, MI: University of Michigan Press.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency. Investigating complexity, accuracy and fluency in SLA* (pp. 1–20). Amsterdam: Benjamins.

Huensch, A., & Tracy-Ventura, N. (2017). Understanding L2 fluency behavior: The effects of individual differences in L1 fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, **38**, 755–785.

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. (2012). *How long should a pause be? Effects of cut-off points of pause length on analyzing L2 utterance fluency*. Poster presented at the Fluent Speech Workshop, Utrecht, the Netherlands.

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. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, **39**, 569–591.

Kircher, T. T. J., Brammer, M. J., Levelt, W., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: Engagement of left temporal cortex during pauses in speech, *NeuroImage*, **21**, 84–90.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, **32**, 145–164.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, **40**, 387–417.

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, **14**, 41–104.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford: Oxford University Press.

Lim, H., & Godfroid, A. (2015). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, **36**, 1247–1282.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, **15**, 19–44.

Martinez, R. (2011). *The development of a corpus informed list of formulaic sequences for language pedagogy*. Unpublished doctoral dissertation, University of Nottingham, UK.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, **33**, 299–320.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity informal instruction and study abroad learning contexts. *TESOL Quarterly*, **46**, 610–641.

Pawley, A., & Syder, F. (2000). The one clause at a time hypothesis. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 163–191). Ann Arbor, MI: University of Michigan Press.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., . . . Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, **51**, 195–203.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, **87**, 539–558.

Peltonen, P. (2018). Exploring connections between first and second language fluency: A mixed methods approach. *Modern Language Journal*, **102**, 676–692.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, **64**, 878–912.

Riazantseva, A. (2001). Second language proficiency and pausing. *Studies in Second Language Acquisition*, **23**, 497–526.

Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes*, **14**, 423–441.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, **65**, 395–412.

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*. New York: Routledge.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, **54**, 79–95.

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.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies*. Unpublished doctoral dissertation, University of California, Berkeley.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2018). *Understanding formulaic language: A second language acquisition perspective*. London: Routledge.

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

Skehan, P. (2003). Task based instruction. *Language Teaching*, **36**, 1–14.

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

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.

Suzuki, S., & Kormos, J. (2019). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*. Advanced online publication. doi:10.1017/S0272263119000421

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, **65**, 71–79.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–276). Amsterdam: Benjamins.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, **17**, 84–119.

Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., & Liberman, M. (2016). Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change*, **6**, 199–234.

Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. London: Continuum.

Yuan, J., Xu, X., Lai, W., & Lieberman, M. (2016). *Pauses and pause fillers in Mandarin monologue speech: The effects of sex and proficiency*. Paper presented at Speech Prosody, Boston.

Cite this article: Kahng, J. (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics* **41**, 457–480. https://doi.org/10.1017/S0142716420000065