Konin Language Studies
Faculty of Humanities and Social Sciences, State University of Applied Sciences in Konin, Poland
KSJ 8 (4). 2020. 349-368
http://ksj.pwsz.konin.edu.pl
doi: 10.30438/ksj.2020.8.4.1

Second language speaking performance: The role of L2 working memory capacity

Ayşegül Nergis
Istanbul University, Turkey
https://orcid.org/0000-0002-0498-2705
aysegulnergis@gmail.com

Abstract

The present study aimed to investigate to what extent advanced L2 learners with low and high working memory (WM) capacity differ with respect to their L2 speaking. 52 English for Academic Purposes (EAP) students enrolled in a freshmen course with a focus on academic speaking skills participated in this study. In order to assess participants' working memory (WM) capacity, Daneman's (1991) speaking span test was used. Speech samples collected through an oral argumentation task were analyzed in terms of acoustic measures of oral fluency, accuracy, and lexical and structural complexity. Multivariate tests showed that L2 WM capacity can explain variances in lexical complexity; but, WM was not significantly associated with any of the oral fluency, accuracy or syntactic complexity measures. Theoretical implications of the findings were discussed.

Keywords: working memory; oral fluency; complexity; speaking in L2; speaking span

1. Introduction

L2 speaking as a complex concept is associated with many performance-related phenomena (such as temporal fluency, articulation rate, grammatical accuracy, grammatical, lexical and syntactic complexity) and cognitive factors (e.g., working memory, attention, L1 speaking style) and recently has gained a wider recognition in SLA research. It is widely recognized that L2 speakers, especially those with limited proficiency, have difficulties in allocating their attention on both form and meaning of linguistic items. This situation forces them to make a choice between these two dimensions while speaking: they either focus on meaning or they focus on form (Anderson, 1995; VanPatten, 1990) and their speech becomes slower, less accurate or complex. Indeed, L2 speaking is a difficult skill to develop because there are so many different linguistic components (i.e., grammar rules, vocabulary knowledge, FS, etc.) and cognitive processing limitations involved (i.e., working memory capacity, attention, etc.) during oral performance. Gatbonton and Segalowitz (2005) and Segalowitz (2010) argued that research on L2 speaking performance should consider taking the role of abovementioned between-subjects and within-subject effects into consideration. This paper aims to contribute to our understanding of the effect of learner differences in working memory capacity on speaking performance in English as a second language.

2. Aspects of L2 oral performance

L2 oral performance is often examined by breaking aspects of speech into three subcomponents: accuracy, fluency and complexity (Skehan, 1996, 1998; Skehan & Foster, 1997; Vercellotti, 2017). Skehan (1996, 1998) describes fluency as the learner's capacity to produce speech with fewer hesitations, repetitions, self-corrections and pauses. Complex speech is characterized by elaboration within the limits of the currently possessed linguistic proficiency, a considerable amount of subordination (embedding) and lexical variety in speech (Yuan & Ellis, 2003), and accuracy refers to the minimum level of errors that do not impede communication and correspondence to the target language norms.

These three components of speech have not been consistently defined by previous researchers. For example, fluency has been highly associated with accuracy of speech (Ahmadian & Tavakoli, 2011), easy and smooth flow of speech, and a minimum amount of pauses and hesitations in speech (Mizera, 2006). Some researchers even associated fluency with good command of a language and overall L2 proficiency (Fulcher, 2003; Wood, 2012). According to Brumfit (1984) and Ur (1981) language teachers have an intuitive understanding of these terms and tend to classify classroom activities as "fluency" or "accuracy." Fulcher (2003) further argues that in terms of language testing, accuracy and fluency are related to, but not dependent on each other in that they are seen as being at opposite ends of a continuum where in one end there is accurate and dysfluent speech and in the other end is the inaccurate but fluent speech. However, together they constitute the construct definition of L2 speaking (Fulcher, 2003, p. 27).

EAP teachers aim to improve oral accuracy in their classrooms, but they are also aware that their students occasionally make errors when speaking. Some of these errors are ignored if they do not interfere with the communicative message. But even in EAP classrooms, where students are widely accepted as advanced L2 learners, some serious errors can impede the intended message. Fulcher (2003) listed the common types of errors that are penalized or ignored in testing situations according to their gravity (the degree of interference with communication). According to Fulcher (2003), the most serious error that L2 speakers make while speaking in English is word order and omission of words. This type of errors decrease rapidly as the learner improves beyond the stage of beginner. The second type of errors is the misuse of pronouns and relative clauses. This information is conducive to the operational definition of accuracy of speech. In previous studies, accuracy has been associated with correct use of certain structures in speech. For example, Yuan and Ellis (2003) measured accuracy as the percentage of accurately used verb forms and the percentage of clauses that did not contain any error. Ahmadian and Tavakoli (2011) operationalized accuracy of speech as the number of error-free clauses and correct verb forms in speech. These measures seem as appropriate to be used with advanced speakers of English because they include correct use of complex structures.

Researchers associate complexity of speech with the existence of subordination in utterances and a variety of vocabulary used in speech. Previous studies mostly dealt with the relationship between task planning conditions and complexity of speech. According to Foster and Skehan (1996), pre-task planning results in greater complexity of language production. Mehnert (1998) found that complexity in speech is observed only when planning time is longer than 10 minutes. In the one-minute planning condition, the participants produced accurate speech which was less complex than 10-minute planning condition. Yuan and Ellis (2003) argue that accuracy and complexity of speech is enhanced in online planning conditions. In online planning conditions, the participants have to allocate their attentional resources into the ongoing task and have to prioritize form over meaning. However, this results in a decrease in fluency. According to Ahmadian and Tavakoli (2011), complexity is also enhanced when learners are given a chance to repeat a task. Previous research operationalized oral complexity in inconsistent ways. For example, Vercellotti (2017) used distinct measures for grammatical and lexical complexity. Thai and Boers (2016), on the other hand, used only syntactic complexity measures.

Fluency is considered to be an aspect that is difficult to operationalize by SLA researchers because it is not directly dependent on vocabulary vault and there is not a “fluency store” in the speaker’s mind (Lennon, 1990, p. 391). Segalowitz (2010) suggested that a comprehensive definition of L2 fluency should also refer to the perspective of cognitive sciences. According to Segalowitz (2010), there are three aspects of fluency that can be distinctly observed and examined: *cognitive fluency, utterance fluency*, and *perceived fluency*. *Cognitive fluency* refers to the speaker's ability to control the underlying cognitive systems that feed into production of utterances, which majorly concerns the monitoring processes. This aspect includes both lexical retrieval and information access that will support the message but also tailoring the utterance in accordance with the intended message and articulation processes. All these processes should be conducted almost simultaneously and with great efficiency to aid smoothness of speech. *Utterance fluency*, on the other hand, deals with the features of utterances such as the amount of pauses, hesitations, repetitions, and repairs in speech. However, utterance fluency is not about how the utterance is perceived by a listener, so it is the actual performance of a speaker as measured in terms of temporal variables (such as the number of syllables or words uttered in a minute). *Perceived fluency* is related to the inferences of listeners about speakers' utterances. It refers to the conditions where a listener (who is most of the time an examiner) is supposed to listen to a person or a recorded speech sample (in an online or an offline task) and judge the speaker as fluent or dysfluent according to impressions he/she draws from the speech sample. L2 oral fluency, as a temporal quality, has been associated with several different components of speech. Although research on L2 oral fluency is scarce, researchers investigating oral fluency as a performance variable share similar views on the types of temporal variables to be measured (Wood, 2012, p. 12). Today, it is commonly agreed that fluent speech is diagnosed by longer fluent runs, increased speech rate, and fewer hesitations and pauses (Kormos, 2006; Lennon, 1990; Segalowitz, 2010; Wood, 2012). In accordance with this perspective on fluency, Skehan (2003) and Tavakoli and Skehan (2005) proposed that construct definition of fluency contains these three aspects: *speed fluency, repair fluency*, and *breakdown fluency.* While speed fluency refers to articulation speed, repair fluency concerns hesitations and corrections, and breakdown fluency deals with silent (unfilled) pauses and filled pauses such as *hmm, uhh*.

As can be seen from works cited above, the three aspects of speech are interrelated, but still need to be distinguished from each other because they require different type of processing on the learner's part (Skehan, 1996). As human beings have a limited capacity to process information, they are unable to attend to all aspects of a task. This occurs when speakers have difficulty in allocating attentional resources and have to prioritize one aspect over the others (Anderson, 1995; Skehan, 1996; VanPatten, 1990). This so-called deficiency of language learners calls for a better understanding of the mechanisms underlying difficulties they have during oral performance.

3. Working memory and L2

Working memory (WM) is a central system "responsible for the processing and temporary storage of information in the performance of complex cognitive tasks" (Daneman, 1991, p. 446). WM has a limited range; in other words, there is a limit to the number of items that can be maintained during immediate consciousness. According to Wen (2014), it is widely accepted among SLA researchers that WM plays an important role in the language learning and processing and its role is comparably more important in L2 acquisition. He explains that this is due to the premise that L1 acquisition occurs automatically in early years of life, while L2 learning mostly occurs later in life and therefore requires greater controlled language processing and cognitive demand on the part of the learner (McLaughlin, 1995).

A myriad of studies examined the interaction between WM and L2 learning. To summarize some of the key works, McLaughlin (1995) argued that cognitive demand is much higher in SLA in contrast to L1 acquisition because it requires more effort and control for processing language. On the other hand, Miyake and Friedman (1998) put forward that WM, in fact, corresponds to language aptitude. They reported that for advanced learners L1 and L2, WM relies on similar cognitive resources. O'Brien, Segalowitz, Collentine, and Freed (2006, 2007) investigated the role of WM in L2 proficiency and showed that WM played an important role in narrative development at earlier stages of L2 learners and in grammatical competence at later stages of language learning. Juffs and Harrington (2011) also showed that WM has strong links to L2 processing, vocabulary development, and L2 proficiency. As for the speaking skill, Daneman (1991) asserts that, in contrast to what most people think, speaking is not an automatic and inherently effortless task in that it requires the speaker to temporarily store what to say next while the person is still in the action of producing speech. She also asserts that WM, in fact, may be the system that distinguishes skilled and unskilled speakers.

In the recent years, there have been only a few studies looking into the role of WM in SLA. To summarize some earlier studies, two decades ago, Atkins and Baddeley (1998) examined the relationship between WM and L2 vocabulary learning and found that participants' WM (phonological memory span) was conducive to their L2 lexical development. Wen (2014) emphasizes that this study is one of the few studies that provided empirical evidence about the relationship between WM and L2 vocabulary acquisition. Fortkamp (1999) found that L2 WM capacity was correlated with L2 oral fluency and L1 WM capacity with L1 oral fluency. He explained that this could be the outcome of cross-linguistic difference of WM capacity. On the other hand, Mizera (2006) found no correlation between L2 speaking and WM capacity. He concluded that different groups of language learners may need different levels of WM use. Kormos and Safar (2008) looked into the relationship between working memory capacity and English language competence with bilingual students and found that English test scores of the participants correlated with their WM scores. Gilabert and Munoz (2010) followed a different path and compared L1 WM capacity with L2 oral performance measures and found that with a group of 59 undergraduate students, WM scores correlated with L2 oral fluency, accuracy, and lexical complexity scores. In a recent study, Wen (2016) found that Chinese university students' executive memory scores highly associated with lexical measures of their L2 oral production, but not with their L2 oral fluency or accuracy scores. In another study, Georgiadou and Roehr-Brackin (2017) found that WM and the number of pauses in speech were in negative correlation, indicating an association between WM and fluency; however, they did not find any correlation between speed and WM.

As can be seen from works cited above, there have been many attempts to explore the role of WM in L2 performance; however, there is still much to discover about their association, at least for advanced learners. In order to address this issue, in the present study, the role of working memory capacity for L2 speech production of advanced EAP students was investigated. Previous research also investigated the role of WM in L2 oral performance; however, measures used in these studies did not directly address the skill they were investigating. For example, Cho (2018) investigated the role of WM in L2 oral performance by using a reading span measure. In the same vein, Georgiadou and Roehr-Brackin (2017) investigated working memory in relation to fluency in L2 speech and used a backward digit span test and a listening span test (LST). Kormos and Safar (2008) also used a backward digit span test to measure working memory capacity. As Wen (2016) asserted, executive function of WM also plays a more important role as L2 proficiency increases, suggesting that speaking span tests that deal with this specific function of WM (such as Daneman's speaking span test) give a better estimate of cognitive processes involved in advanced learners' L2 production. Therefore, in this study, Daneman's speaking span test was used to measure L2 WM capacity.

4. Methodology

4.1. Participants

Fifty two freshmen students enrolled in an academic speaking course in an English-medium university in Istanbul participated in this study. According to the university exam bureau, the participants met the proficiency level above or near the minimum requirement of TOEFL IBT 80 or IELTS academic module 6,5. The age range was 18-24, with an average of 19.2 years and they were from a variety of disciplines (Social Sciences, Humanity, Law, Nature Sciences, and Engineering). The medium of instruction in the university was English; in other words, the context was an English medium instruction (EMI). Students enrolled in this academic speaking course were prepared for their academic studies in their departments. In order to pass the course, they completed individual, pair, and group tasks (such as presentations, discussions, reports, watching videos, or responding to academic forums via audio recordings of their speech) in and outside the class.

Of the 52 students, 25 were female and 27 were male. The first language of the participants was Turkish. Students who reported to have a speech impairment (that could adversely affect the results) or speaking anxiety did not participate in the study. Participants were informed before data collection that participation was on voluntary basis and their response would not affect their grades. They signed a standard consent form prepared by the university's research and development unit.

4.2. Data collection procedures

4.2.1. Speaking span test

Daneman's (1991) speaking span test was used to test the participants' working memory capacity in English as a second language. It was chosen over other measures available because it is a measure that directly deals with speaking. The test includes 100 common English words organized in a set of two to six words, consisting of five words for each sub-set (see Appendix 2). The number of words increased progressively from two to six; so the load on working memory also increased during the course of the test. Each word was displayed on a computer screen for one second. After each set was completed, a blank screen appeared. Participants generated (by speaking aloud) a grammatically, syntactically, and semantically correct sentence by using the each given word in the displayed set. Each participant was given 60 seconds to produce sentences for each set. Their speech was recorded via an audio-recorded for scoring procedures.

In the speaking span test, typically there is no restriction on the sentence length and complexity. Participants exercised with three practice trials before conducting the test. To prevent the participants to rely on a fixed syntactic pattern, they were discouraged to repeat the same sentence patterns (such as "I saw the machine" and "I saw the compass"); since such strategies would tax the processing component of working memory (Mizera, 2006). All the words included in the test were seven letter words with two syllables. The maximum score of this test was 100 and a participant's speaking score was the number of grammatically acceptable sentences they could produce with the given words. Lenient scoring was used, as opposed to strict scoring which only accepts exact form and presentation of the target words. Adopting scoring procedures by Weissheimer (2007) and Tavares (2008), half credit (0.5) was given to words recalled in a different order and in slight deviations from the original forms given in the set (e.g., “machines” instead of “machine”). In this way, individual differences in performance in terms of working memory were more effectively captured.

After each participant’s WM score was calculated, the median value of the sample was also calculated to determine low- and high-WM groups. Speakers with WM scores below the median were assigned to the Low-WM Group (*N* = 23) and the ones who scores at or above the median were assigned to the High-WM Group (*N* = 29).

4.2.2. Oral argumentation task

Oral data were obtained from an argumentation task. According to Hulstijn, Schoonen, De Jong, Steinel, and Florijn (2012), complex, formal and argumentative speech elicitation tasks are suitable for B2 and above levels of learners. This type of task was also preferred because in order for individual differences in working memory capacity to emerge, the task performed has to be rather complex (Fortkamp, 1999, 2000). Previous research working on similar groups of learners also used this type of task for data elicitation (e.g., Zalbidea, 2017).

In this task, participants were given a list of world problems and asked to evaluate their significance and rank them according to their perceived importance (see Appendix 1). Participants were familiar with the task type and content as it was the focus of the content of the EAP course they were enrolled in. Task content and prompts were also reviewed by a small group of experienced EAP instructors for the relevance of the task to the sample in terms of their content familiarity and lexical and grammatical competence to accomplish the task.

Participants were given 30 seconds to plan their speech and 3 minutes to complete the task. A digital chronometer showing the remaining time was present in front of them. They first completed the oral argumentation task and then the speaking span test in a silent room and they were informed that their speech would be recorded. Both tasks were piloted with similar groups before the experiment.

4.3. Data analysis

All speech data were transcribed by the researcher and then coded by two coders who were experienced EAP instructors (one coder was a PhD candidate in ELT and the other held an MA in ELT). Inter-coder reliability analyses for the acoustic measures were done through calculating Cohen’s *kappa* (*k*) for each measure. Analyses showed high levels of agreement.

4.3.1. L2 oral fluency

Five acoustic measures were calculated for each speech sample. To automatically calculate the number of syllables, length and the number of pauses and phonation time in speech, a PRAAT script was used (De Jong & Wempe, 2009). The cut-off point for pauses was determined as >250 ms as was also used in previous studies (Bosker Pinget, Quené, Sanders, & De Jong, 2013; De Jong & Perfetti, 2011; Nergis, 2018).

Speed fluency was represented by speech rate (SR). For breakdown fluency, three measures were used: the number of silent pauses per phonation time (NSP), the number of filled pauses per phonation time (NFP), and the mean length of silent pauses (MLSP). For the aspect of repair fluency, one measure was used: the number of dysfluencies (repetitions and self-corrections) per phonation time (ND). A log transformation was performed on the scores to approximate the data to the normal distribution.

Table 1 Acoustic measures of fluency

Aspects	Acoustic measures	Calculation
Speed	Speech rate (SR)	Number of syllables/phonation time (Kormos & Dénes, 2004)
Breakdown	Number of silent pauses (NSP)	Number of silent pauses/phonation time (Bosker et al., 2013)
	Number of filled pauses (NFP)	Number of filled pauses/phonation time (Kang, 2012)
	Mean length of silent pauses (MLSP)	Sum of length of silent pauses/number of silent pauses (Bosker et al., 2013)
Repair	Number of dysfluencies (ND)	Number of self-corrections and repetitions/phonation time (Suzuki & Kormos, 2019)

4.3.2. L2 oral accuracy

Two measures were used for accuracy: Error-free AS-units (EFAS) and Errors per AS-unit (EAS). According to Foster, Tonkyn, and Wigglesworth (2000), an Analysis-of-Speech unit (AS-unit) is “a speaker’s utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either” (p. 365).

Table 2 Acoustic measures of accuracy

Acoustic measures	Calculation
Error-free AS-units (EFAS)	Percentage of error-free AS-units (Foster et al., 2000)
Errors per AS-unit (EAS)	Mean number of errors per AS-unit (Foster et al., 2000)

4.3.3. L2 oral complexity

Syntactic complexity of the participants' speech was assessed through two measures: Mean number of clauses (CAS) and Mean length of AS-unit (MLAS). For lexical complexity, two measures were used: Guiraud's index (GI) and Mean segmental type-token ratio (MSTTR). Operational definitions are given in Table 3 below.

Table 3 Acoustic measures of complexity

Aspects	Acoustic Measures	Calculation
Syntactic complexity	Mean number of clauses (CAS)	Mean number of clauses per AS-unit (Tavakoli & Foster, 2008)
	Mean Length of AS-unit (MLAS)	Mean number of words per AS-unit (Norris & Ortega, 2009)
Lexical complexity	Guiraud's index (GI)	Number of types divided by the square root of tokens (Gilabert, 2007; Mora, 2012)
	Mean segmental type-token ratio (MSTTR)	Number of different words divided by the total number of words in every 40-words (Yuan & Ellis, 2003)

5. Findings

In order to determine how low-WM and high-WM groups differ in terms of L2 oral performance, a series of multivariate analysis of variances were conducted (for fluency, accuracy and complexity measures; see Table 4 for descriptive statistics). Assumptions of multivariate normality, homogeneity of variances and multicollinearity were satisfied.

Table 4 Descriptives for dependent variables

	Low WM		High WM	
	M	*SD*	*M*	*SD*
Fluency				
Speech rate (SR)	1.66	.05	1.68	0.4
Number of silent pauses (NSP)	.68	.04	.69	0.3
Number of filled pauses (NFP)	.57	.02	.58	.03
Mean length of silent pauses (MLSP)	2.3	.04	2.4	.03
Number of dysfluencies (ND)	0.21	.04	.021	.03
Accuracy				
Error-free AS-units (EFAS)	0.85	.08	.086	.09
Errors per AS-unit (EAS)	.53	.07	.52	.06
Complexity				
Mean number of clauses (CAS)	.58	.05	.57	.04
Mean length of AS-unit (MLAS)	8.1	.09	7.9	.07
Guiraud's index (GI)	6.90	1.21	7.11	1.32
Mean segmental type-token ratio (MSTTR)	.82	.12	.85	.11

Table 5 Multivariate effects for the interaction between oral fluency and L2 WM capacity

Effect	Value	*F*	Hypothesis *df*	Error *df*	*p*
Wilk's Λ	.096	.486	5	47	.838

Note. *$p < .05$; **$p < .01$; ***$p < .001$ ($N = 52$)

Table 6 Univariate effects on oral fluency scores

Dependent Variable	*df*	*SS*	*F*	*p*	η^2
Speech rate (SR)	1	.001	.111	.741	.003
Number of silent pauses (NSP)	1	.140	1.272	.266	.032
Number of filled pauses (NFP)	1	.003	.372	.546	.010
Mean length of silent pauses (MLSP)	1	.009	.421	.520	.011
Number of dysfluencies (ND)	1	.011	.437	.512	.011

Note. *$p < .05$; **$p < .01$; ***$p < .001$ ($N = 52$)

As can be seen in Table 5, the results of the MANOVA showed no significant differences in oral fluency scores according to their working memory capacity ($F(1, 51) = .486$, $p > .05$, Wilk's $\Lambda = .851$, partial $\eta 2 = .096$). There was not a statistically significant difference in oral accuracy based on working memory ($F(1,51) = 1.751$, , $p > .05$, Wilk's $\Lambda = .933$, partial $\eta 2 = .067$).

Table 7 Multivariate effects for the interaction between oral accuracy and L2 WM capacity level

Effect	Value	F	Hypothesis df	Error df	p
Wilk's Λ	.067	1.751	2	49	.184

Note. *$p < .05$; **$p < .01$; ***$p < .001$ ($N = 52$)

Table 8 Univariate effects on oral accuracy scores

Dependent Variable	df	SS	F	p	η^2
Error-free AS-units (EFAS)	1	.002	.592	.445	.012
Errors per AS-unit (EAS)	1	.011	3.394	.071	.064

Note. *$p < .05$; **$p < .01$; ***$p < .001$ ($N = 52$)

Table 9 Multivariate effects for the interaction between oral complexity and L2 WM

Effect	Value	F	Hypothesis df	Error df	p
Wilk's Λ	.891	1.962	3	48	.132

Note. *$p < .05$; **$p < .01$; ***$p < .001$ ($N = 52$)

Table 10 Univariate effects on oral complexity scores

Dependent Variable	df	SS	F	p	η^2
Mean number of clauses (CAS)	1	.022	3.828	.096	.021
Mean Length of AS-unit (MLAS)	1	.003	.037	.849	.001
Guiraud's Index (GI)	1	20.199	4.074	.049	.075
Mean Segmental TTR (MSTTR)	1	4491.872	4.739	.034	.087

Note. *$p < .05$; **$p < .01$; ***$p < .001$ ($N = 52$)

The results of the MANOVA showed that working memory capacity did not hold a statistically significant difference on oral complexity scores (*F* (1,51) = 1.962, $p > .05$, Wilk's Λ = .891, partial $\eta 2$ = .109). However, significant univarite effects were observed for MSTTR (*F*(1,51) = 4.739, , $p < .05$, partial $\eta 2$ = .087) and GI (*F*(1,51) = 4.074, , $p < .05$, partial $\eta 2$ = .075), both of which were lexical complexity measures. The multivariate partial $\eta 2$ values indicate that approximately 9% of multivariate variance of WM level is associated with MSTTR and 7.5% of variance is associated with GI (see Tables 6-10).

6. Discussion

Advanced L2 learners in EAP and EMI contexts, such as the participants of this study, are asked to produce fluent and accurate speech with enriched lexical and structural complexity on daily basis. In order to achieve this, they should possess cognitive skills that help them to store and retrieve linguistic information efficiently. Therefore, when learners with the same level of linguistic knowledge possess varying levels of working memory, differences between their performances are expected. In accordance with this, the current investigation aimed to examine the role of L2 working memory capacity in L2 oral performance. Temporal qualities of EAP students' oral performance were compared according to participating students' level of WM capacity in L2 which was assessed in a test that directly dealt with speaking skill (Daneman, 1991). Results showed that there was a significant difference in L2 lexical complexity scores of low and high WM groups. However, no significant differences were observed for accuracy, fluency and syntactic complexity. In other words, participants' capacity to hold long stretches of words in short term memory did not help them to produce utterances faster or more accurately with enhanced level of structural (syntactic complexity); nevertheless, participants who possessed the ability to store and process long stretches of speech were able to produce lexically enhanced utterances.

These findings are partly in contrast to previous research that found a significant association between WM and L2 oral fluency (Fortkamp 2000; Gilabert & Munoz 2010); but they also confirm previous research that found partial relationship between WM and oral fluency measures (Georgiadou & Roehr-Brackin, 2017; Kormos & Safar, 2008; Mizera, 2006). These studies used different measures to assess WM capacity in L2. For instance, Georgiadou and Roehr-Brackin (2017) used a backward digit span test and a listening span test, while Mizera (2006) used Daneman's (1991) speaking span test. Daneman hypothesized that fluent speakers have a larger WM capacity that allowed them to store linguistic information and efficiently allocate cognitive resources to processes involved in production. She also claimed that WM is a capacity that could distinguish skilled and unskilled speakers. All things considered, the current findings are not surprising in that different measures of WM have so far revealed different associations with oral fluency measures which also varied vastly in previous studies. In the current study, nonconfounding temporal measures were used to assess fluency. Results showed that for EAP students, having a low or high speaking span in L2 did not create a difference in pauses, self-corrections or speed of speech. This finding can be explained by advanced learners' temptation to use formulaic language to increase their fluency (Kormos, 2006; Segalowitz, 2010). Indeed, previous research found that learning formulaic language can increase oral fluency in EAP and EMI contexts (Nergis, 2018). This finding is also confirmed by the current findings of the study that showed having a longer speaking span in L2 helped EAP students to produce lexically complex utterances, which will be further discussed below.

Previous research has also produced contradicting results about the relationship between oral accuracy and complexity and WM capacity. For example, Zalbidea (2017) found that high WM capacity helped accuracy in writing and syntactic complexity in speaking. On the other hand, Kormos and Safar (2008) found low correlation between WM (measured with backward digit span) and L2 speaking (measured through raters' assessment). While discussing the results regarding speaking, they emphasized the role of WM in chunking, meaning that advanced learners have better ability to store and retrieve longer chains of speech. They asserted that the relationship between advanced learners' WM capacity and lexical complexity of their speech is linked to each other. A similar idea has been underlined by Kormos (2006), who asserted that the capacity to store bits of speech before producing an utterance helps L2 speakers plan and successfully encode their message, increasing the quality of their speech. Similarly, Segalowitz (2010) explained that WM plays a significant role between the planning and articulation stage of language production. L2 speakers usually find speaking in an L2 difficult because while they are planning the overall message, they have to deal with mental lexicon racing in their minds. This cognitive load that occurs during micro-planning is one of the reasons that make L2 speech less fluent or faltered. This was not observed in the current study; although high WM group lagged behind in terms of fluency, accuracy, and syntactic complexity, they outperformed the low WM group with respect to lexical complexity.

The association with high level of WM and lexical complexity of the participants' speech might be related to the nature of the WM test and the lexical measures used. Daneman's speaking test used in this study deals with storing of lexical units in mind and producing stretches of speech. This level of storage and processing capacity can be associated with the two lexical complexity scores of the study (Guiraud's index and Mean segmental type-token ratio), both of which deal with number of different word forms used in speech. This result is not surprising in that WM test also deals with this capacity. Indeed, enhanced vocabulary knowledge has been nominated to be a better indicator of lexical access and retrieval capacity than WM (Gilabert & Munoz, 2010; Mizera, 2006). Furthermore, higher WM could help L2 speakers to retrieve words for the incoming message more efficiently, by allowing them to reach vocabulary store in their minds with ease and also, possibly, increasing their chance to consider competing lexical alternatives (Gilabert & Munoz, 2010).

6. Conclusion

L2 speaking is not a monolithic construct; a myriad of subskills and processes are involved, indicating the need for examining components of speech in relation to each other. According to Skehan and Foster (1997), all components of speech compete with each other, probably due to limited human capacity on attention. For instance, when attention is on speaking faster, grammatical accuracy might be forfeited even for advanced L2 speakers. This effect can be observed in instructional settings very often: when a learner is pushed to speak more accurately, controlled processes, rather than automatic processes, are encouraged and students might speak more accurately, but less fluently (Ahmadian & Tavakoli, 2011). The present study showed that the role of WM in oral production of advanced L2 learners, as in the case of the participating EAP students in this study, is also something to think about when lexical richness is taken into consideration. Students who are at a similar level of proficiency, while completing a typical EAP task such as the argumentation task used in this study, can differ with respect to lexical complexity based on their speaking span capacity.

References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity and fluency in EFL learners' oral production. *Language Teaching Research, 15*(1), 35-59.

Anderson, J. R. (1995). *Language and memory: An integrated approach*. New York: Wiley.

Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics 19*, 537-552.

Bosker, H.R., Pinget, A.-F., Quené, H., Sanders, T., & de Jong, N.H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*(2), 159-175.

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy.* Cambridge: Cambridge University Press.

Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. *System, 72*, 85-98.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research, 20*(6), 445-464.

de Jong, N. H., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61*(2), 533-568.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*(2), 385-390.

Fortkamp, M. B. M. (1999). Working memory capacity and aspects of L2 speech production. *Communication and Cognition, 32*, 259-296.

Fortkamp, M. B. M. (2000). *Working memory capacity and L2 speech production: An exploratory study* (Unpublished doctoral dissertation). Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition, 18*, 299-324.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375. http://doi.org/10.1093/applin/21.3.354

Fulcher, G. (2003). *Testing second language speaking*. London: Longman/Pearson Education.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review, 61*, 325-353.

Georgiadou, E., & Roehr-Brackin, K. (2017). Investigating executive working memory and phonological short-term memory in relation to fluency and self-repair behavior in L2 speech. *Journal of Psycholinguistic Research, 46*, 877-895.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching, 45*, 215-240. http://doi.org/10.1515/iral.2007.010

Gilabert, R., & Munoz, C. (2010). Differences in attainment and performance in a foreign language: The role of working memory capacity. *International Journal of English Studies, 10*(1), 19-42.

Hulstijn, J. H., Schoonen, R., de Jong, N. H., Steinel, M. P., & Florijn, A. F. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the common European framework of reference for languages (CEFR). *Language Teaching, 29*(2), 202-220.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching 44*, 137–166.

Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly, 9*(3), 249-269.

Kormos, J. (2006). *Speech production and second language acquisition*. New Jersey: Lawrence Erlbaum Associates.

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

Kormos, J., & Safar, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*(2), 261-271.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387-417.

McLaughlin, B. (1995). Aptitude from an information processing perspective. *Language Testing, 11*, 364–381.

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

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healey & L. J. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum.

Mizera, G. J. (2006). *Working memory and L2 oral fluency* (Unpublished doctoral dissertation). University of Pittsburg.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal Instruction and study abroad learning contexts. *TESOL Quarterly, 46*(4), 610-641.

Nergis, A. (2018). *The effect of explicit instruction of formulaic sequences on L2 oral fluency* (Unpublished doctoral dissertation). Yeditepe University, Istanbul, Turkey.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578. http://doi.org/10.1093/applin/amp044

O'Brien, I. N., Segalowitz, J., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27*, 377-402.

O'Brien, I. N., Segalowitz, J., Collentine, J., & Freed, B. (2007). Phonological memory predicts L2 oral fluency gains in adults. *Studies in Second Language Acquisition, 29*, 557-582.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*, 38-62.

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., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 1-27.

Suzuki, S., & Kormos, J. (2019). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency. *Studies in Second Language Acquisition, 42*(1), 1-25. http://doi.org/10.1017/S0272263119000421

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*(2), 439-473

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: John Benjamins.

Tavares, M. G. G. (2008). *Pre-task planning, working memory capacity and L2 speech performance* (Unpublished doctoral dissertation). Universidad de Federal De Santa Catarina, Florianópolis, Brazil.

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly, 50*(2), 369-393.

Ur, P. (1981). *Discussions that work: Task-centered fluency practice*. Cambridge: Cambridge University Press.

VanPatten, B. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12*, 287-301.

Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics, 38*(1), 90-111. http://doi.org/10.1093/applin/amv002

Weissheimer, J. (2007). *Working memory capacity and the development of L2 speech production* (Unpublished doctoral dissertation). Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil.

Wen, Z. (2014). Theorizing and measuring working memory in first and second language research. *Language Teaching, 47*(2), 174-190.

Wen, Z. (2016). Phonological and executive working memory in L2 task-based speech planning and performance. *The Language Learning Journal, 44*(4), 418-435. http://doi.org/10.1080/09571736.2016.1227220

Wood, D. (2012). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. London: Continuum International Publishing.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 24*(1), 1-27.

Zalbidea, J. (2017). 'One task fits all'? The roles of task complexity, modality, and working memory capacity in L2 performance. *Modern Language Journal*, *101*(2), 335-352.

APPENDIX 1

Oral Argumentation Task

In this interview, you will talk about YOUR IDEAS about some problems in the world. Note that your speech will be recorded.
Here is a list of problems that need attention as determined by world leaders in the year 2000 in a meeting organized by the United Nations:

- Children's health
- Mothers' health
- HIV/AIDS and other diseases
- Gender equality issues
- Extreme poverty and hunger
- Education
- Environmental sustainability
- Developing partnership between the world governments

Examine the list. Is there anything that you do not understand?

Which one of these important problems do you think needs the most immediate attention and which one needs the least immediate attention and why?

You have thirty seconds to plan your speech and three minutes to complete the task.

APPENDIX 2

Words Used in the Speaking Span Test (Daneman, 1991)

Trial sets

a) pumpkin, balance
b) fingers, noticed
c) machine, results, compass

2-word sets

a) kitchen, farmers
b) signals, thirsty
c) perfume, giraffe
d) healthy, rewards
e) biscuit, shampoo

3-word sets

a) pencils, observe, journey
b) nervous, quickly, younger
c) trumpet, windows, believe
d) earning, dentist, tallest
e) parking, succeed, whisper

4-word sets

a) butcher, wrinkle, ceiling, glasses
b) certain, warning, mittens, husband
c) diapers, special, instant, plastic
d) explain, stylish, garbage, request
e) trouble, bending, advance, roasted

5-word sets

a) teacher, stomach, foreign, cousins, quarter
b) jealous, monthly, arrange, sweater, treated
c) growing, surfing, ashamed, lettuce, cushion
d) damaged, respect, private, clearly, witness
e) useless, helping, passive, buttons, tonight

6-word sets

a) student, careful, reduced, vandals, orchard, ignored
b) morning, village, traffic, islands, handles, patient
c) chimney, achieve, cookies, explode, feather, address
d) knuckle, chicken, working, storage, injured, playful
e) lawyers, mailbox, freezer, release, lightly, fragile