Individual Differences in Working Memory Capacity and the Development of L2 Speech Production

Janaina Weissheimer
Federal University of Rio Grande do Norte

Mailce Borges Mota
Federal University of Santa Catarina

This study investigates the relationship between individual differences in working memory capacity and L2 speech development. Thirty-two undergraduate English as a Foreign Language students participated in this study, which involved two data collection phases, each consisting of a working memory test (the speaking span test) and a speech generation task, with a two-month interval between the two data collections. Participants' speaking samples were analyzed in terms of fluency, accuracy and complexity. The results show that only lower span individuals demonstrated a statistically significant improvement in working memory capacity and that such improvement was not a function of increased proficiency. In addition, although the speaking span test predicted fluency and complexity in participants' L2 speech, it was not a good indicator of the development of speech accuracy.

INTRODUCTION

Various researchers have claimed that L2[1] speaking should be investigated as a skill in its own right (e.g., Bygate, 2001; McCarthy, 1998), and that at least two factors have contributed to this claim. First, learners' proficiency in an L2 is usually directly associated with how fluently they speak the language (Lennon, 1990)[2]. As a result, most learners who engage in the process of acquiring a second language make the ability to speak it a top priority. Secondly, because speaking is generally the main goal, a considerable amount of the teaching and learning of a second language is devoted to it (Hieke, 1985). With this in mind, the need to understand L2 speech production is crucial to aid learners on thei path to proficiency.

One way to address L2 speech production is to investigate the impact of working memory capacity on speech performance. Working memory is considered a "brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning" (Baddeley, 1992, p. 556). It is also referred to as "those mechanisms or processes that are involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition" (Miyake & Shah, 1999, p. 450).

Research on speech production acknowledges that working memory capacity may be seen as a possible independent constraint on the process of speaking in both L1 (e.g., Daneman & Green, 1986; Daneman, 1991) and L2 (e.g., Fortkamp, 1999;

Issues in Applied Linguistics ISSN 1050-4273

Vol. 17 No. 2, 93-112

Harsuiker & Barkhuysen, 2006; Mizera, 2006).These studies have shown that in general, individuals with a higher working memory capacity tend to outperform those with a lower capacity in many aspects of speech production, including fluency, accuracy, complexity, lexical density and syntactic planning. Research has also devoted attention to the role of working memory in L2 speech *performance,* whereas, by comparison, little attention has been given to the role of working memory in adult L2 speech *development*[3](Fortkamp, 1999; 2003).

According to a number of researchers, the role of working memory capacity in the development of L2 speaking is a relevant question to be pursued, since it is believed that this capacity may vary in the course of L2 acquisition as a function of increased command of the target language and as a result of automatization within the language system (Berquist, 1998; Harrington, 1992; Harrington & Sawyer, 1992). This hypothesis is supported by Daneman and Carpenter (1980), who claim that working memory capacity (in their case, in L1) is task-specific and varies according to the individual's processing efficiency in a given task. Our aim is to contribute to research regarding the relationship of working memory capacity and speech performance and development, characterized by L2 fluency, accuracy and complexity in speech production.

THEORETICAL BACKGROUND

L1 and L2 Speech Production Models

One of the most pervasive models of speech production is Levelt's (1989) *blueprint for the speaker*, proposed to explain L1 speech production. Levelt's model has inspired various bilingual speech models (DeBot, 1992; Poulisse & Bongaerts, 1994) as well as theories of second language acquisition (Swain, 1995).

Levelt (1989) claims that the speaker performs two major roles, as an information processor and as an interlocutor. As an information processor, the speaker conceptualizes, formulates and articulates, so that overt speech can be generated. In the process of conceptualization, the speaker selects the appropriate information to convey an intention (macroplanning) and brings this information into perspective considering the status of the addressee and the focus of the message (microplanning). The formulator translates a conceptual structure, the preverbal message, into a linguistic structure by following two steps: grammatical encoding and phonological encoding. The grammatical encoder**,** a subcomponent of the formulator, accesses the appropriate lemma with syntactic and semantic information from the speaker's mental lexicon and adds it to the preverbal message which was generated by the conceptualizer. When all the relevant lemmas have been selected, a surface structure is generated as a product and is stored in a syntactic buffer. The phonological encoder, on the other hand, builds a phonetic or articulatory plan for each lemma and for the utterance as a whole. The articulator executes the phonetic plan by the musculature of the respiratory, the laryngeal and the supralaryngeal systems. Internal speech, the product of the formulator, is temporarily stored in the

articulatory buffer until successive chunks are retrieved and unfolded for execution. The product of articulation is, finally, overt speech.

The distinction between controlled and automatic processing is fundamental to a discussion of Levelt's speech production model, and despite their dichotomous nature, both controlled and automatic processing coexist within the act of speaking. According to Shiffrin and Schneider (1977), by now, classic proposal, automatic processes are executed without intention or conscious awareness, are usually quick and operate on their own resources. Controlled processes, on the other hand, demand attentional resources that are limited in working memory. They are usually serial in nature and therefore, take more time (Shiffrin & Schneider). In Levelt's blueprint, message generation (by the conceptualizer) and monitoring involve highly controlled processing in order to find the means to express intentions and to monitor internal and overt speech. All other components, including the formulator and articulator, are claimed to be largely automatic in L1 speech production. Working memory is also crucial in Levelt's speech production model because it stores intermediate representations of messages (preverbal message, surface structure and phonetic plan) and makes them available for further processing.

Arguing that Levelt's model (1989) is based on sound psycholinguistic research, de Bot (1992) adapted it to explain bilingual and multilingual speech production. In his proposal, which is considered the most comprehensive in the field to date (de Angelis, 2007), de Bot makes only those changes that he feels are necessary to explain bilingual speech production. Thus, in regard to the processes of the conceptualizer, de Bot suggests that only microplanning is language-specific. The L2 formulator, in turn, has different processing components for each language and produces two concurrent speech plans.[4] To describe the organization of the mental lexicon, de Bot elected to use Paradis' (1987) Subset Hypothesis and proposed that language-specific subsets are activated independently in one large lexicon which contains entries for both languages. Finally, according to de Bot, the formulator contains a common store for the syllables of the two languages, and syllabic patterns will be stored individually unless a matching pattern already exists. Recently, de Bot (2004) has updated the model to accommodate the claims of the Dynamic Model of Multilingualism (Herdina & Jessner, 2002), which emphasizes individual differences and variations and views languages as psycholinguistic systems.

Individual Differences in Working Memory Capacity

The view of working memory capacity as a source of individual differences in L1 development and use is accepted across the literature (Conway & Engle, 1996; Daneman & Carpenter 1980; Daneman & Green, 1986; Engle, Laughlin, Tuholski & Conway, 1999; Kane, Bleckley, Conway & Engle, 2000; Turner & Engle, 1989). A growing number of researchers have voiced claims in favor of investigating working memory capacity as a possible independent constraint on the process of second language acquisition. They claim that working memory capacity plays an even greater role in learning an L2 by the adult learner than it does in L1

acquisition (e.g., Berquist, 1998; Ellis &Sinclair, 1996; Fortkamp, 1999, 2003; Harrington, 1992; Harrington & Sawyer, 1992; Miyake & Friedman, 1998; Mizera, 2006; Sawyer & Ranta, 2001).

Research has found a relationship between measures of working memory capacity and measures of L2 performance. Harrington (1992) found a significant correlation between working memory capacity and performance in L2 vocabulary, grammar and reading. Similarly, Harrington and Sawyer (1992) found a significant correlation between participants' reading spans (a working memory capacity task) in L2 and their performance in the Grammar and Reading sections of the TOEFL. Berquist (1998) found significant correlations between L1 and L2 working memory measures and participants' performance in the Listening and Reading sections of the TOEIC. Fortkamp (1999) found a significant correlation between an L2 version of the Daneman's (1991) speaking span test and fluency and articulation in L2. In a follow-up study, Fortkamp (2003) found a significant correlation between L2 working memory capacity and various measures of fluency, accuracy, complexity and lexical density in L2 speech production. Mizera (2006) found a significant correlation between L2 working memory capacity and L2 oral fluency. Finally, Kormos and Sáfár (2008) found a significant correlation between phonological short-term memory capacity and the overall proficiency of pre-intermediate learners; and between the digit span test and the L2 acquisition of learners at a beginner's level of proficiency.

The present study aims at contributing to studies on the relationship between working memory capacity and L2 acquisition by investigating individual differences in working memory capacity and the development of L2 speech production. The specific questions addressed in this study were:

(1) Does participants' working memory capacity, as measured by the speaking span test, change over the course of L2 acquisition? If so, does this change have a greater impact on higher or lower span individuals in the course of speech development?

(2) Do higher and lower span individuals experience different gains in speech production, in terms of fluency, accuracy and complexity?

(3) Does working memory capacity, as measured by the speaking span test, predict L2 speech development?

METHOD

Participants

Thirty-two undergraduate English as a Foreign Language students at a major Federal University in the northeastern region of Brazil participated in this study. The cohort consisted of 11 male and 21 female participants, ranging in age from 18 to 35 years.

Proficiency Trial

Participants' L2 speaking proficiency was assessed using a speech generation task. Three experienced raters judged the speech samples on a speaking proficiency scale adapted from the Cambridge First Certificate in English speaking assessment scale and Iwashita, McNamara and Elder's scale (2002).

Pearson's Correlations, which were run for each of the three ratings in the proficiency trial, proved to be significant ($r(32)=.90$; .83; and .76, $p< .01$), showing consistency among the ratings provided by the three different raters. Cronbach's coefficient alpha for the proficiency test was .93, attesting internal consistency and reliability to the test.

In an attempt to guarantee sample homogeneity in terms of oral proficiency, participants whose scores were at least 1.5 standard deviation above or below the mean (M=2.86) were excluded from the sample.[5] As a result of the proficiency trial, twenty-eight out of the original thirty-two participants were included in the next phase of this study.

Instruments and Procedures

The Speaking Span Test (SSPAN)

Two versions of Daneman's (1991) speaking span test, each constructed with 120 unrelated words, were administered over an interval of eight weeks. The total number of words was organized in six sets, each consisting of subsets of 2, 3, 4, 5 and 6 words. Each word was presented individually in the middle of a computer screen for one second. Participants were instructed to read the words silently. After ten milliseconds, the next word in the set appeared in the same position on the screen as the previous word. This procedure was followed until the set ended and a black screen appeared with question marks on it. These marks signaled the number of words that had to be recalled and the number of sentences that had to be produced. Participants were instructed to use the words that appeared in the correct form and order to generate syntactically and semantically acceptable sentences aloud in English. No restrictions were placed on the length or complexity of the sentences. A training phase (60 words) preceded the testing phase (60 words), and the actual testing phase only took place when participants reported being comfortable and confident enough to perform the test.

Participants' speaking span was defined as the maximum number of words (out of 60) for which they could generate grammatically and semantically acceptable sentences in English. Following Daneman (1991) and Daneman and Green (1986), participants' responses were recorded, transcribed and analyzed. Two different speaking span scores were generated: a *strict speaking span*, whereby all the sentences the subject produced contained the target word in the exact form and order of presentation, and a *lenient speaking span*, whereby credit was given for sentences that contained the target word in a form that varied from how the word appeared in the presentation (e.g., target word being 'drug' and the word in the sentence produced being 'drugs'). No credit was given to ungrammatical

sentences in terms of syntax and semantics. Tests were counter-balanced to avoid order and task effects.

The mean of the two memory test scores (strict and lenient) was obtained and participants were then characterized as having a higher working memory span or a lower working memory span on the basis of scores that fell in the upper or lower tripartite of the total sample of subjects who took part in the experiment. Twelve participants were considered *lower span* subjects (*M*= 19.42, *SD*= 2.74) and eight participants were considered *higher span* subjects (*M*= 36.13, *SD*= 5.30). The data of eight subjects in the middle tripartite were omitted from the analysis. This is a common procedure when applying extreme-groups designs to prevent subjects from being misclassified as high when they should have been classified as low, and vice-versa.

The Speech Generation Task (SGT)

The speech generation task (in phases one and two of the study) consisted of participants describing a picture-cued narrative, administered right after participants had completed the memory tests in both phases. Participants were encouraged to narrate the story with as many details as possible, using their imagination, command of language and background knowledge to accomplish the task. No time limit was given to participants for performing the task, and they could continue to view the picture during the narration. All participants took at least one minute to perform the task. In order to control for task effects and to prevent differences in performance in the two phases due to task complexity, half of the participants were assigned speech generation task 1 and the other half was assigned speech generation task 2 in both phases of the study.

Measures of Speech Production in the Speech Generation Task

Fluency

Participants' fluency in the speech generation task was measured in terms of speech rate, *pruned* and *unpruned* (Lennon, 1990; Ortega, 1999). Speech rate *unpruned* was calculated by dividing the total number of semantic units (complete and partial words) produced, including repetitions, by the total time spoken (including pauses) in seconds. The result was then multiplied by 60 so as to express it by the number of words per minute. Speech rate *pruned* was calculated the same way, but all words that were repeated (excluding repetitions for rhetorical effect and repetitions for immediate correction) were excluded from the count. In both cases, contractions were counted as one word.

Accuracy

Two general measures of accuracy were used in this study:

Number of errors per 100 words (ACCURE) (Fortkamp, 2003; Mehnert, 1998) was obtained by dividing participants' total number of errors by the total number of words produced and multiplying the result by 100.

Percentage of error-free clauses (ACCURC) (Foster & Skehan, 1996; Mehnert, 1998) was obtained by dividing the number of error-free clauses by the total number of clauses in the speech sample.

The analysis to determine the number of errors and error-free clauses counted all errors in syntax, morphology and lexical choice, but did not consider errors in pronunciation and intonation. Errors that were repeated were counted twice. In both measures of accuracy, errors that were immediately corrected by the speaker were not counted.

Complexity

Because little agreement exists in speech production literature with regard to which measure of complexity captures the greatest amount of useful variance (Foster & Skehan, 1996; Mehnert, 1998), three different ways of looking at this aspect of speech production were operationalized in this study:

Number of dependent clauses per minute (COMPLT) was obtained by dividing the number of dependent clauses produced by the time in seconds and multiplying the resulting figure by 60. Following Mehnert (1998), the analysis of dependent clauses included finite and nonfinite subordinate clauses, coordinate clauses with subject deletion, and infinite constructions, excluding infinitives with modal verbs.

Number of subordinate clauses per T-unit (COMPLC) (Mehnert, 1998) was obtained by dividing the number of subordinate clauses by the total number of T-units produced by the speaker. A *T-unit* is described as "one main clause plus whatever subordinate clauses, phrases, and words happen to be attached to or embedded within it" (Mehnert, 1998, p. 90). The analysis of dependent clauses included finite and nonfinite subordinate clauses, coordinate clauses with subject deletion, and infinite constructions, excluding infinitives with modal verbs.

Number of clauses per c-unit (COMPLG) (Foster & Skehan, 1996) was obtained by dividing the total number of clauses in the speech sample (either a simple independent finite clause or a dependent finite clause or nonfinite clause) by the number of *c-units* (independent utterances providing referential or pragmatic meaning). According to Foster and Skehan (1996), the c-unit (Brock, 1986) is a more sensitive measure of complexity for spoken language.

For all measures of speech production, the researchers and an additional rater performed the analysis. In cases where the coders had doubts or were unsure, another rater intervened so that an agreement could be reached for the data analysis.

RESULTS

Working Memory Capacity and Measures of Speech Production

Working memory scores had a significant positive correlation with measures of accuracy in L2 speech production in both phases of the experiment, as can be seen in Tables 1 and 2. As for the two measures of accuracy utilized – errors per

100 words (ACCURE) and percentage of error-free clauses (ACCURC) – the first proved to be sensitive to the data in both phases of the experiment (first phase $r(28)=-.45$, $p<.05$ strict score, $r(28)=-.40$, $p<.05$ lenient score; second phase $r(28)=-.38$, $p<.05$ strict, $r(28)=-.44$, $p<.05$ lenient), whereas the second measure showed higher statistical significance but only in the second phase of the experiment ($r(28)=.46$, $p<.01$ strict score, $r(28)=.53$, $p<.01$, lenient score). Since these results don't indicate that one general measure of accuracy was clearly superior, we maintained both measures of accuracy for the next steps in the analysis.

Table 1. Correlations Between Measures of Working Memory Capacity and Speech Production in Phase One

	SRUN1	SRP1	ACCUR1E	ACCUR1C	COMPLC1	COMPLT1	COMPLG1
WM1S	0.14	0.14	-0.45*	0.34	-0.06	0.44*	0.03
WM1L	0.15	0.15	-0.40*	0.29	0.04	0.37*	0.07

Note. N=28. WM1S = working memory strict in first phase; WM1L = working memory lenient in first phase; SRUN1 = speech rate unpruned in first phase; SRP1= speech rate pruned in first phase; SRP2 = speech rate pruned in second phase; ACCUR1E = accuracy (number of errors in hundred words) in first phase; ACCUR1C = accuracy (percentage of error-free clauses) in first phase; COMPLT1 = complexity (subordinate clauses per minute) in first phase; COMPLG1 = complexity (clauses per c-unit) in first phase.
* p<0.05.

Table 2. Correlations Between Measures of Working Memory Capacity and Speech Production in Phase Two

	SRUN2	SRP2	ACCUR2E	ACCUR2C	COMPLC2	COMPLT2	COMPLG2
WM2S	0.25	0.27	-0.38*	0.46**	0.24	0.41*	0.08
WM2L	0.28	0.30	-0.44*	0.53**	0.23	0.40*	0.10

Note. N=28. WM2S = working memory strict in second phase; WM2L = working memory lenient in second phase; SRUN2 = speech rate unpruned in second phase; SRP2 = speech rate pruned in second phase; ACCUR2E= accuracy (number of errors in hundred words) in second phase; ACCUR2C = accuracy (percentage of error-free clauses) in second phase; COMPLT2 = complexity (subordinate clauses per minute) in second phase; COMPLG2 = complexity (clauses per c unit) in second phase.
* p<0.05.
** p<0.01.

By contrast, the correlations between measures of working memory capacity and fluency scores (SRUN, SRP), as displayed in Tables 1 and 2, failed to show significance. When correlations were run separately for each memory group (Table 3), they reached significance for the lower span group ($r(12)=.60$, $p<.05$ for speech rate unpruned, $r(12)=.58$, $p<.05$, for speech rate pruned), but not for the higher span group ($r(8)=.29$, $p<.05$ speech rate unpruned, $r(8)=.28$, $p<.05$ for speech rate pruned). Therefore, the pattern of correlation between working memory capacity

and speech rate is clearly broken among higher span individuals. Moreover, the standard deviation of the higher span group is higher than that of the lower span group (*SD*=5.30 and *SD*=2.74, respectively), indicating that the former is probably a more heterogeneous group. Thus, the lack of a significant correlation might be due to the smaller size higher span sample (only 8 participants).

Table 3. Correlations between Working Memory Capacity and Speech Rate for Higher and Lower Spans

Lower Spans[a]			Higher Spans[b]		
	SRUN1	SRP1		SRUN1	SRP1
WM1S	0.60*	0.58*	WM1S	0.29	0.28
WM1L	0.61*	0.59*	WM1L	0.30	0.29

Note. SRUN1 = speech rate unpruned in first phase; SRP1 = speech rate pruned in first phase; WM1S =working memory strict in first phase; WM1L = working memory lenient in first phase.
an=12. bn=8.
*p< 0.05.

As for the correlations among the three measures of complexity, only the number of dependent clauses per minute showed a significant correlation with measures of working memory capacity in both phases of the experiment ($r(28)=.44$, $p<.05$ for the strict score, $r(28)=.37$, $p<.05$ for the lenient score). Therefore, the number of dependent clauses per minute was the only measure of complexity that was taken into account in the analysis, as it proved to be the most sensitive to the data in question.

Taken together, these results can be interpreted as initial evidence that participants' working memory capacity scores significantly correlated with the thee measures of speech production: fluency (only in the case of lower span individuals), accuracy, and complexity, in both phases of the experiment.

Working Memory Scores in the Two Phases

In order to provide an overview of the performance of learners in both phases of the experiment, Table 4 displays the mean and standard deviation of working memory capacity for the entire group of participants.

Table 4. Means and Standard Deviations fo Working Memory Scores in the Two Phases

Paired Samples					
		Mean	N	SD	SEM
Pair 1	WM1S	26.04	28	8.011	1.514
	WM2S	27.93	28	7.64	1.444
Pair 3	WM1L	26.96	28	8.011	1.514
	WM2L	28.96	28	7.748	1.464

Note. WM1S = working memory strict in first phase; WM2S = working memory strict in second phase; WM1L= working memory lenient in first phase; WM2L = working memory lenient in second phase.

Although the results reported in Table 4 show that there were indeed gains in working memory scores from phase one to phase two, we ran paired t-tests to check whether these were significant or not. Paired t-tests showed that there were no significant gains when the entire group of participants was considered. When gains were analyzed in terms of higher and lower span groups, however, a different and more interesting pattern emerged. Table 5 shows that lower span individuals showed statistically significant improvement in the two working memory scores applied: the strict score and the lenient score ($t(12)=4.33, p<.05$, for the strict score, $t(12)=4.75, p<.05$, for the lenient score). The same pattern did not hold true for the higher span individuals who did not show any significant change in working memory performance from phase one to phase two of the experiment.

Table 5. Paired-samples Test for Working Memory Scores

Paired Samples							
			Paired Differences		T	df	Sig. (2-tailed)
WM_G			Mean	*SD*			
Lower	Pair 1	WM1S – WM2S	-4.33	6.53	-2.30	11	0.04*
	Pair 3	WM1L – WM2L	-4.75	6.59	-2.50	11	0.03*
Higher	Pair 1	WM1S – WM2S	4.50	7.69	1.66	7	0.14
	Pair 3	WM1L – WM2L	3.50	7.15	1.38	7	0.21

Note. WM1S = working memory strict in first phase; WM2S = working memory strict in second phase; WM1L = working memory lenient in first phase; WM2L = working memory lenient in second phase

*p< 0.05.

These results can be taken as evidence that while the lower span group experienced some sort of increase in working memory scores, the higher span group did not show significant improvement in working memory scores between the two phases of the experiment.

Speech Production Scores in the Two Phases

In Table 6, mean and standard deviations scores for all measures of speech production in the study show that lower span participants experienced gains in the following measures of speech production within the two phases: speech rate

unpruned (*M*= 80.20, *SD*=23.88; *M*=80.93, *SD*=28.95), speech rate pruned (*M*= 75.78, *SD*=24.96; *M*=76.27, *SD*=30.90) and accuracy - percentage of error-free clauses (*M*= 0.57, *SD*=0.26; *M*=0.60, *SD*=0.29). Conversely, the scores presented reduced measures from the first to the second phase in terms of accuracy (number of errors per hundred words) and complexity. Higher spans, on the other hand, showed gains in performance regarding speech rate pruned (*M*=82.55, *SD*=22.86; *M*=93.94, *SD*=25.18), speech rate unpruned (*M*=86.46, *SD*=21.67; *M*=99.37, *SD*=24.30) and complexity (*M*= 3.99, *SD*=1.87; *M*=5.23, *SD*=2.35), as measured by the number of dependent clauses per minute. Both measures of accuracy indicated reduced performance.

Table 6. Descriptive Statistics of Speech Production

Paired Samples Statistics			Mean	N	SD	SEM
Lower	Pair 1	SRUN1	80.20	12	23.88	6.89
		SRUN2	80.91	12	28.95	8.35
	Pair 2	SRP1	75.78	12	24.96	7.20
		SRP2	76.27	12	30.90	8.92
	Pair 3	ACCUR1E	11.17	12	7.64	2.20
		ACCUR2E	9.31	12	7.53	2.17
	Pair 4	ACCUR1C	0.57	12	0.26	0.07
		ACCUR2C	0.6	12	0.29	0.08
	Pair 5	COMPLT1	3.36	12	2.28	0.66
		COMPLT2	2.82	12	2.84	0.82
Higher	Pair 1	SRUN1	86.46	8	21.67	7.66
		SRUN2	99.37	8	24.30	8.59
	Pair 2	SRP1	82.55	8	22.86	8.08
		SRP2	93.94	8	25.48	9.01
	Pair 3	ACCUR1E	5.41	8	4.45	1.57
		ACCUR2E	4.93	8	6.38	2.25
	Pair 4	ACCUR1C	0.78	8	0.18	0.06
		ACCUR2C	0.76	8	0.26	0.09
	Pair 5	COMPLT1	3.99	8	1.87	0.66
		COMPLT2	5.23	8	2.35	0.83

Note. SRUN1 = speech rate unpruned in first phase; SRUN2 = speech rate unpruned in second phase; SRP1= speech rate pruned in first phase; SRP2 = speech rate pruned in second phase; ACCUR1E = accuracy (number of errors in hundred words) in first phase; ACCUR2E= accuracy (number of errors in hundred words) in second phase; ACCUR1C = accuracy (percentage of error-free clauses) in first phase; ACCUR2C = accuracy (percentage of error-free clauses) in second phase; COMPLT1 = complexity (subordinate clauses per minute) in first phase; COMPLT2 = complexity (subordinate clauses per minute) in second phase.

Similar to the gains in working memory capacity scores reported above, differences in measures of speech production in the two phases did not show significance if looked at as a group. However, as can be seen in Table 7, if the division into higher and lower spans is considered in the analysis, one may notice that, although not yet significant, the lowest *p* levels are among higher span participants in the measures of speech rate unpruned and complexity (number of dependent clauses per minute), indicating that the highest level of variance was among these participants.

Table 7. Paired-samples Tests for Measures of Speech Production

Paired Samples Test

WM G			Paired Differences				
			Mean	SD	T	df	Sig (2-tailed)
Lower	Pair 1	SRUN1 – SRUN2	-0.71	20.22	-0.12	11	0.91
	Pair 2	SRP1 – SRP2	-0.49	19.91	-0.09	11	0.93
	Pair 3	ACCUR1E – ACCUR2E	1.86	4.96	1.30	11	0.22
	Pair 4	ACCUR1C – ACCUR2C	-0.02	0.21	-0.40	11	0.70
	Pair 5	COMPLT1 – COMPLT2	0.54	3.59	0.52	11	0.61
Higher	Pair 1	SRUN1 – SRUN2	-12.91	24.73	-1.48	7	0.18
	Pair 2	SRP1 – SRP2	-11.40	23.97	-1.35	7	0.22
	Pair 3	ACCUR1E – ACCUR2E	0.48	4.34	0.32	7	0.76
	Pair 4	ACCUR1C – ACCUR2C	0.02	0.14	0.37	7	0.72
	Pair 5	COMPLT1 – COMPLT2	-1.24	2.13	-1.65	7	0.14

Note. SRUN1 = speech rate unpruned in first phase; SRUN2 –speech rate unpruned in second phase; SRP1 = speech rate pruned in first phase; SRP2 = speech rate pruned in second phase; ACCUR1E = accuracy (number of errors in hundred words) in first phase; ACCUR2E = accuracy (number of errors in hundred words) in second phase; ACCUR1C = accuracy (percentage of error-free clauses) in first phase; ACCUR2C = accuracy (percentage of error-free clauses) in second phase; COMPLT1 = complexity (subordinate clauses per minute) in first phase; COMPLT2 = complexity (subordinate clauses per minute) in second phase.

More specifically, the mean speech rate unpruned for the higher span participants in the second phase (M= 99.37, SD=24.30) was higher (t(8)= 12.91, p= 0.18) than that of the first phase (M=86.36, SD=21.67) and this increase in performance almost reached statistical significance. The same was also true for the measure of complexity (number of dependent clauses per minute). The mean number of dependent clauses per minute in the second phase (M=5.23, SD=2.35) was higher (t(8)= 1.24, p=0.14) than that of the first phase (M=3.99, SD=1.87). Again, these results are expected to gain significance once the sample is increased. At any rate, none of the differences in performance among lower span participants reached significance or were close to reaching statistical significance (as in the case of higher spans).

DISCUSSION

To reiterate, the first research question addressed the developmental nature of the working memory system. The conflated results reported in the previous section indicate that only lower span participants showed a significant increase in working memory capacity scores and that working memory development did not follow speech development in this study.

Although we acknowledge that working memory capacity may develop in the course of L2 speech acquisition, based on the results of lower spans in this study, this development cannot be attributed to the improvement of speech production in these subjects, especially since they did not show any significant increase in speech scores within the different phases. Higher span subjects, conversely, revealed the exact opposite pattern. Even though they did experience (almost significant) gains in speech production scores within the phases, in terms of fluency and complexity, their working memory scores remained stable across tests. Together, these results cannot support the claim that working memory capacity in L2 is a function of an increased command of language, as proposed by Harrington (1992), Berquist (1998), and Harrington & Sawyer (1992). Contrary to that hypothesis, results in this study indicate that, even though working memory development in L2 seems to take place, it is apparently not brought about by language proficiency alone.

The fact that lower spans experienced gains in working memory scores seems to lend empirical credence to a number of studies claiming that working memory capacity – more specifically the integration of controlled and automatic processes – plays a greater role in earlier stages of acquisition (Anderson, 1983; McLaughlin & Heredia, 1996; Perlow, Jattuso, & Moore, 1997; Woltz, 1988).

Due to the correlation found between working memory scores and speaking proficiency in this study, lower span subjects are assumed also to be less proficient as compared to higher span subjects. Consequently, it then seems possible to claim that lower span subjects in this study are also in a relatively lower, more controlled stage in the acquisition of L2 speech, which relies on a greater deal of declarative knowledge and lacks automaticity of the processes involved. Higher spans, to the contrary, display more efficient processing during the execution of the skill, resulting in faster and less error-prone performance.

Our argument here, although speculative, is that the greater degree of variance in working memory test scores of lower span subjects between the two phases may reflect improvement at coordinating the controlled and automatic processing of language. Inasmuch as the speaking span test was sensitive to that developmental aspect of the working memory system, it is possible that the lower span subjects may have shown increased efficiency in the task of integrating processes simply because they had more room for improvement in that area. Higher spans, in contrast, may have been initially more efficient in this area, and therefore there was no variation in scores within the phases.

An alternative explanation for the increase in the lower spans' working memory scores acknowledges the possibility of a methodological limitation. Accordingly, the efficiency with which lower and higher spans approached the two memory tests may, in turn, explain their performance improvement or lack thereof. That is, higher spans were originally more efficient at storing and processing information, which allowed them to learn the memory task faster and achieve better results in the very first moment, therefore no improvement was shown from the first to the second trials. Lower span subjects, on the other hand, started the first test

with significantly less processing efficiency and might have benefited more from performing the working memory test twice. The strategies they developed for dealing with the test could very well result in improvements within the two phases.

Developmental psychologists have suggested that improved performance in working memory tests may result from more opportunities to practice and develop expertise than from growth of working memory capacity per se (Lépine, Barrouillet, & Camos, 2005; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). In our opinion, individual differences in working memory improvement, as in the case of the lower spans in this study, would more likely result from the ability to employ efficient strategies for dealing with the test than an improvement in working memory capacity itself. In this sense, practice effects and strategy use would cause working memory to artificially inflate, masking the true relationship between working memory capacity and complex cognition (Turley-Ames & Whitfield).

As the main focus of this study is on working memory *capacity*, the alternative proposed above would represent a methodological limitation. Further research would have to be conducted to find ways of minimizing the effects of practice and strategy use, because once the possibility of dealing strategically with the task is reduced, its predictive value is increased (Lépine, Barrouillet, & Camos, 2005).

In sum, the discussion presented above raises a compelling paradox. If gains in working memory capacity among lower spans indeed represent their increased efficiency at coordinating controlled and automatic processing in the production of speech, why didn't such processing improvement result in more efficient speech production in the speech generation task, since, after all, the two tasks arguably share the same underlying processes? For this reason, the second explanation, which acknowledges the effects of practice and strategy use on lower spans' working memory scores seems, at least at this point, more suitable to account for the results in this study.

The second research question addressed the relationship between working memory capacity and specific measures of speech production (fluency, accuracy, and complexity) in the two phases of the study. In the present study, working memory capacity significantly correlated with measures of accuracy and complexity for all participants. These findings corroborate those of Fortkamp (2003) who, similarly, found that higher span individuals are more capable of producing complex speech and are less prone to making mistakes as compared to lower spans.

By contrast, the correlations between measures of working memory capacity and fluency scores of higher span subjects failed to show significance and are, therefore, at odds with previous results reported in the literature (Fortkamp, 2003; Mizera, 2006). Thus, an intriguing question can be raised when considering the results of higher span subjects: Why didn't higher span subjects perform more fluently than their counterparts in this study? Higher span individuals, it seemed, could only attain more accurate and complex speech at the expense of less fluent speech. In other words, the attention allocated by higher spans to achieve two goals (speaking more accurately and using more complex language) appeared to reduce

the capacity remaining for achieving a third goal (speaking fluently). As a consequence, accuracy and complexity became preeminent goals, but fluency suffered to some degree. This explanation is in line with other empirical results from studies in the speech production area (Bygate, 2001; Foster & Skehan, 1996; Skehan & Foster, 1995, among others) that have, similarly, reported trade-off effects among features of speech production.

The reason for higher spans prioritizing accuracy instead of fluency may have to do with the nature of the experimental conditions. For example, despite the fact that participants were not in a testing situation, the participants performed in a lab, had no interlocutor and knew that their performance would be evaluated. This situation might have resembled a testing situation. In testing situations, regardless of the conditions under which participants are performing, a focus on a more conservative orientation – not making mistakes – may be expected rather than a focus on being fluent (Iwashita, McNamara, & Elder, 2002). This might have been the case of the participants of this study.

Research question 3 addressed the predictive power of the speaking span test in L2 speech development. Although gains in speech performance failed to reach significance, a pattern was revealed when considering fluency and complexity measures of higher spans.

More specifically in the case of fluency development, increases in fluency may be attributable mainly to increases in the degree of proceduralization of knowledge in the L2 formulator (Towell, Hawkins, & Bazergui, 1996). Towell, Hawkins, and Bazergui showed that the gains in the L2 speech production by their participants were noticed in terms of faster and more complex speech and they claimed that this possibly resulted from the proceduralization of processes in the formulator. With this in mind, we will argue that the gains in fluency and complexity, reported in this study, are seen more obviously among higher span individuals because, being more proficient than lower spans, they must rely more on procedural knowledge and automatic L2 processes than lower spans in generating speech. Consequently, more resources could potentially be allocated to the development of specific aspects of speech production, in this case fluency and complexity.

Although working memory scores predicted fluency and complexity development, when the issue of accuracy was analyzed, what was found is not exactly what had been expected. Neither higher spans nor lower spans showed gains in accuracy scores in terms of the number of errors per hundred words from phase one to phase two.

The first explanation is related to the nature of the accuracy measures applied. Global measures of accuracy, such as number of errors per hundred words and percentage of error-free clauses, may not have been the most appropriate to use in order to predict gains in participant performance within the two phases in this study. Although general measures of accuracy have been considered efficient ways of assessing speech production when *performance* is the focus of the analysis (Foster & Skehan, 1996), it could be that accuracy should be looked at in a more

specific way when the *development* of speech production is at stake. If that is the case, for future research, a task that provokes the generation of specific forms would perhaps be a better choice for this experiment. For instance, Ellis (1987) used particular measures of accuracy, such as measuring error rates with three past-tense verb forms, and applied a task in which the past tense verb form played a key role. This was done to generate data about past tense verb forms and errors that occur when considering these verbs. A similar methodological implementation might be considered in the replication of the present study.

The second explanation aims at justifying the lack of accuracy improvement in the higher spans and refers to interlanguage development. The fact that higher spans showed an increase in complexity measures from phase one to phase two may indicate a greater willingness, on the part of this group of speakers, to experiment and take risks when restructuring their linguistic system (Foster & Skehan, 1996; Skehan, 1996). As a consequence of their attempt to use forms closer to the highest potential levels of their interlanguage, mistakes may occur naturally in this process, and these, in turn, could explain the lower accuracy scores in phase two in relation to phase one. In other words, participants' ability to exploit new linguistic forms may have pushed them to levels of syntax beyond their control.

CONCLUSION

The present study focused on investigating the role of working memory capacity in L2 speech development. We set out to investigate whether working memory scores would increase over trials as a result of a greater command of cognitive processes related to language production. We also investigated whether the speaking span test would predict participants' fluency, accuracy, and complexity development in L2 speech. The results, in general terms, show that only lower spans had a statistically significant improvement in working memory scores over trials and that such improvement was not a function of increased proficiency. In addition, inasmuch as the speaking span test predicted the development of fluency and complexity in speakers' L2 speech, it was not proven to be a good indicator of accuracy.

In considering the results of this study, the possibility that the increase in working memory capacity could be attributed to L2 proficiency was disregarded. In attempting to explain lower spans' working memory scores' improvement across testing phases, it was argued that this improvement reflects speakers' improvement at coordinating controlled and automatic processing during speech production and that the speaking span test was more sensitive to that aspect of the working memory system. Further, lower spans might have benefited more from practice and strategy use when taking the speaking span test and therefore, showed increased performance in the second test.

As for the predictive power of the speaking span test, because higher spans execute the tasks of processing and storing information more effectively when

speaking, they may direct a greater amount of their total cognitive resources toward the task of acquiring new speaking strategies and production rules. In addition, given that higher spans were also the most proficient participants in this study, we concluded that since the processes involved in speech production are more proceduralized and automatized among these participants, they were able to speak more fluently and to develop more complex structures.

This study has several limitations. First of all, the relatively small sample size, in particular, of higher span subjects, may have been responsible for the weakening of some correlations and the lack of significant gains in performance. Secondly, the fact that the working memory measure used in this study was not sensitive to a predicted variance in working memory scores among the higher span subjects, a more in-depth discussion of the findings was not possible. Thirdly, there was no control for practice effects or strategy use during the memory test, which may limit attributions of performance improvement to gains in working memory capacity. Finally, complementary measures of speech production, such as pauses, specific measures of accuracy and weighted lexical density must be implemented in order to draw more useful conclusions on participants' L2 speech development.

Despite its limitations, the result of the present study is relevant due to the fact that it goes beyond the general assumption that working memory is an efficient predictor of L2 speech performance. This study represents a step forward by showing how working memory capacity is involved in L2 speech development. The results reported in this study have implications for understanding the nature of capacity limits in working memory as well. The fact that the speaking span test was only sensitive to variance in the working memory performance of lower spans suggests that distinct mechanisms responsible for variance among the performance of higher spans might exist. To confirm this possibility, this area must be studied further.

Our goal in this study was not to resolve these issues but to demonstrate the importance of working memory capacity in the development of L2 speech production. Additional research in this field needs to be conducted. Future investigation can benefit from the results of this initial study and from its methodological implementations and hopefully, it will ultimately lead to a better understanding of the working memory system and its influence over the mechanisms underlying the development of L2 speech.

NOTES

[1] In the present study, the terms second and foreign language are used interchangeably and are referred to as L2.

[2] Fluency is defined in this study as "continued performance in real time" (Skehan, 1996).

[3] The terms speech development and speech acquisition will be used interchangeably in this paper and mean 'gains in speech performance within testing phases'.

[4] More recent findings in Neurocognition have shown that although the same areas in the brain seem to be used for both languages, L1 processing seems to be more efficient than

L2 processing, even in the case of very proficient learners (Stowe & Sabourin, 2005).
[5] The advised criteria of 1.25 points for a participant to be excluded from the sample could not be applied, because the number of participants to be divided into high and low spans in the next phase of the analysis was not sufficient.

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556-559.

Berquist, B. (1998). *Individual differences in working memory span and L2 proficiency: Capacity or processing efficiency?* Paper presented at the American Association for Applied Linguistics 1998 Annual Conference, Seattle, WA.

Brock, C. (1986).The effects of referential questions on ESL classroom discourse. *TESOL Quarterly, 20*, 47-59.

Bygate, M. (2001). Speaking. In R. Carter and D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 14-20). Cambridge: Cambridge University Press.

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory, 4*, 577-590.

Daneman, M., & Carpenter, P.A. (1980).Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25*, 1-18.

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

De Angelis, J. (2007). *Third or additional language acquisition*. Clevedon: Multilingual Matters.

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

De Bot, K. (2004). The multilingual lexicon: Modeling selection and control. *The International Journal of Multilingualism*, *1*, 17-32.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition*, *9*, 12-20.

Ellis, N. C., & Sinclair, S.G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49*(1), 234-250.

Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, *128*, 309-331.

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

Fortkamp, M. B. M. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos*, *24*, 69-104.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299-323.

Harrington, M. (1992). Working memory capacity as a constraint on L2 development. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp.123-135). Amsterdam:

Elsevier.
Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25-38.
Hartsuiker, R., & Barkhuysen, P. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, *21*, 181-204.
Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: Perspectives of change in psycholinguistics*. Clevedon: Multilingual Matters.
Hieke, A. E. (1985). A componential approach to oral fluency evaluation. *The Modern Language Journal*, *66*, 135-141.
Iwashita, N., McNamara, T., & Elder, C. (2002). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. *Language Learning*, *51*, 401-436.
Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2000). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General, 130*, 169-183.
Kormos, J., &Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*, 261-271.
Lépine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review, 12*, 165-170.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, Massachusetts: The MIT Press.
Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*, 387-417.
McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
McLaughlin, B., & Heredia, R. (1996). Information-processing approaches to research on second language acquisition and use. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 213-225). San Diego, CA: Academic Press.
McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, *29*, 10-17.
Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*, 83-108.
Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, N.J.: Lawrence Erlbaum.
Mizera, G. J. (2006). *Working memory and L2 oral fluency*. Unpublished doctoral dissertation, University of Pittsburgh.
Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, *21*, 109-148.
Paradis, M. (1987). *The assessment of bilingual aphasia*. Mahwah, N.J.: Lawrence Erlbaum.
Perlow, R., Jattuso, M., & Moore, D. (1997). Role of verbal working memory in complex skill acquisition. *Human Performance, 10*, 283-302.
Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15*, 36-57.
Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences and instructional de-

sign. In P. Robinson (Ed.), *Cognition and second language instruction* (pp.319-353). Cambridge: Cambridge University Press.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.
Skehan, P., & Foster, P. (1995). Task type and task processing as influences on foreign language performance. *Working Papers in English Language Teaching*, *3*, 139-188.
Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*, 38-62.
Stowe, L. A. & Sabourin, L. (2005). Imaging the processing of a second language: Effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching*, *43*(4), 329-353.
Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B.Seidlhofer (Eds.), *Principle & practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.
Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of English. *Applied Linguistics, 17*, 84-119.
Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, *49*, 446-468.
Turner, M., & Engle, R. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127-54
Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology, 117*, 319-331.

Janaina Weissheimer is Associate Professor of English and Applied Linguistics in the Department of Languages and Arts, Federal University of Rio Grande do Norte, Brazil. Her main areas of research in SLA are speech production, memory and neurocognition.

Mailce Borges Mota is Associate Professor of English and Applied Linguistics in the Department of Foreign Languages and Literature, Federal University of Santa Catarina, Brazil. Her main areas of research are the neurocognitive aspects of SLA and individual differences in working memory capacity.