Longitudinal Effects of Phonological Short-Term Memory and Working Memory Capacity on L2 Grammar Knowledge

Adriana Biedroń[1], Mauricio Véliz-Campos[2], and Katarzyna Zychowicz[1]

[1] Pomeranian University in Słupsk, Poland

[2] Faculty of Education Sciences, Universidad de Talca, Chile

ABSTRACT

Working memory (WM) has been found to play a major role in learning L2 grammar (Li et al., 2019). However, there is little research into the longitudinal effects of phonological short-term memory and WM capacity on L2 grammar knowledge development (Sagarra, 2017). The current longitudinal study investigated the relationship between phonological short-term memory, WM capacity, and the development of L2 grammar knowledge over the period of two years. This report is part of an ongoing larger-scale study including the components of reading, writing, and speaking. Participants were 107 Year 1, 2, and 3 Polish university students majoring in English as an L2. The measurements included two phonological short-term memory capacity tests, two WM capacity tests, and four tests of grammar knowledge. The results indicated that grammar tests correlated with nonword, listening, and reading spans. However, latent growth models showed that only WM capacity positively predicted changes in L2 grammar knowledge over time.

KEYWORDS

phonological short-term memory
working memory
working memory capacity
L2 grammar knowledge
longitudinal

INTRODUCTION

Models of working memory (WM) differ in their theoretical conceptualization of this term. However, most of them agree that WM is the limited memory capacity to temporarily store, process, and maintain a restricted amount of information while performing mentally demanding tasks (Cowan, 2014; Wen & Li, 2019). The best-known model of WM is the multicomponential model proposed by Baddeley and Hitch (1974), which suggests the existence of two storage systems: a phonological loop and a visuospatial sketchpad, regulated by a supervisory attention-limited control system. In 2000, Baddeley added a new component, the episodic buffer, which temporarily stores multimodal information and provides a link to long-term memory (Baddeley, 2003). The components which are most relevant to L2 learning are the storage system and the executive control system, that is, the phonological loop and the central executive. The former, also named phonological short-term memory (PSTM), is responsible for temporarily holding sound-based information through the process of articulatory rehearsal, which aids the learning of the phonological forms of new words (Baddeley et al., 1998). Typical measures of PSTM are simple span tasks, such as the digit span or the nonword-span. The latter, also referred to as working memory capacity (WMC), performs executive functions involved in controlling, allocating, and inhibiting attentional resources during higher-level cognitive tasks. It is usually measured with complex span tasks such as reading, listening, speaking, or operation span, which involve simultaneous processing of information. According to Kane et al. (2004), WMC tasks largely reflect a domain-general factor which results in the broad predictive utility of WM span measures, whereas PSTM tasks, based on the same stimuli as the WMC tasks, are much more domain-specific.

Working memory is mostly characterized by its limited capacity. Yet, no consensus has been reached as to how limited this capacity is, that is, its capacity to hold information readily available to our consciousness (Conway et al., 2008) without losing it immediately.

There is a substantial body of evidence (Li et al., 2019; Linck et al., 2014; Martin & Ellis, 2012; Pawlak & Biedroń, 2021; Wen, 2019; Wen & Li, 2019) suggesting that WM, in the form of PSTM and WMC, affects the acquisition of L2 grammar knowledge while being modified by cognitive demands of the processing component of WM tests, L2 proficiency, and task demands.

Much of the research done thus far into the effects of PSTM and WMC on L2 grammar knowledge has been done from a cross-sectional perspective. Consequently, there is scant data on the role of the aforementioned cognitive factors in second language acquisition (SLA)

Corresponding author: Mauricio Véliz-Campos, Faculty of Education Sciences, Universidad de Talca; 380, María Auxiliadora St., Linares, Chile.
E-mail: mauricio.veliz@utalca.cl

DOI • 10.5709/acp-0353-1

from a longitudinal perspective, despite its relevance for L2 theory and research. To date, only a few studies have examined the longitudinal effects of both PSTM and WMC on grammatical development and the results are varied and unconvincing (see Sagarra, 2017; Serafini & Sanz, 2016, for a review). These studies have employed different research designs, memory tests, and grammar tasks, as well as methods of analysis, which makes it difficult to arrive at definitive conclusions regarding their results. Moreover, most of these few studies have focused on low versus high proficiency learners, with intermediate learners being excluded from the studies. The current study seeks to answer the question whether PSTM and WMC are related to grammatical knowledge development over two years of instruction among lower-intermediate-advanced learners of English. Our independent variables included two complex tests of WMC, two simple tests of PSTM, as well as years of study. The grammar knowledge development was assessed with four tests, completed after the first, second, third, and fourth semesters, six months apart from one another.

The Relationship Between Working Memory and Grammar

Working memory, one of the so-called individual differences in L2 learning, is generally thought of as the human cognitive ability dealing with storing and processing a small amount of "task-relevant information in our heads during some cognitive activities such as mental calculation, logical reasoning, planning, and language comprehension" (Lu & Wen, 2022, p. 1). As such, it plays a pivotal role in most, if not all, everyday activities ranging from remembering directions to finding one's way when visiting a new city or remembering a phone number if no pen and paper is on hand. By the same token, WM seems to play a significant role in both L1 and L2 learning. As far as the latter is concerned, WM can be said to comprise two components, namely, PSTM, which deals with the individual's ability to retain verbal information, and WMC, which deals with the individual's ability to process information in real time. Working memory has been found to predict L2 vocabulary and grammar learning, as well as reading and listening comprehension (Baddeley, 2000; Daneman & Hannon, 2007; Wen, 2016). However, its role in SLA is not free from controversies. For example, in a recent study by Gagné et al., (2022), WMC and PSTM played only a minor role in learners' L2 fluency outcomes.

A number of cross-sectional studies provide evidence for a relationship between PSTM and grammar learning (Ellis, 1996; Ellis & Sinclair, 1996; French, 2006; Martin & Ellis, 2012; Williams & Lovatt, 2003). However, the number of studies indicating the impact of WMC on grammar learning is far greater. They have provided evidence for a correlation between performance on complex span tests, usually reading, listening, speaking or operation span, and grammar scores (Fortkamp, 1999, 2003; Harrington & Sawyer, 1992; Li et al., 2019; Martin & Ellis, 2012; Pawlak & Biedroń, 2021; Santamaria & Sunderman, 2015; Sanz et al., 2014; Suzuki & DeKeyser, 2017a; Tagarelli et al., 2015). Moreover, stronger influence of WMC, as opposed to PSTM, on grammar learning has been repeatedly voiced in the literature (Martin & Ellis, 2012; Pawlak & Biedroń, 2021). Taking all the evidence into account, it seems that both WMC and PSTM are involved in learning grammar. Nonetheless, this relationship is mediated by a number of factors, such as learning conditions, type of task, learner age, proficiency level, and other individual factors, together with the ways in which grammar knowledge is tested. Conversely, there are correlational studies in which no relationship between WM and L2 grammar has been found (see Antoniou et al., 2016; Ettlinger et al., 2014). For example, Foryś-Nogala (2021) and Suzuki and DeKeyser (2017b), among others, have found no correlation between PSTM and L2 grammar, thus questioning the role of this type of memory in grammar learning and processing.

Longitudinal Effects of Working Memory on Grammar Processing and Production

A number of studies have suggested that both the PSTM and the WMC have unique effects on various aspects of L2 grammar, which are modified by the cognitive demands of the processing component of WM tests, L2 proficiency, and task demands (Li, et al., 2019; Martin & Ellis, 2012; Pawlak & Biedroń, 2021; Wen & Li, 2019). Nevertheless, most of such studies have used a cross-sectional perspective. There have been only a few attempts to study the longitudinal – understood as the examination of variables over a period of time – effects of whichever component of WM on learning grammar (Faretta-Stutenberg & Morgan-Short, 2018; French & O'Brien, 2008; Grey et al., 2015; Linck & Weiss, 2015; O'Brien et al., 2006; O'Brien et al., 2007; Sagarra, 2017; Serafini & Sanz, 2016). Evidence that individual differences in WMC and PSTM predict L2 grammar longitudinally is controversial and research results are conflicting. O'Brien et al. (2006, see also O'Brien et al., 2007) found a significant impact of PSTM on measures of adult L2 grammar learning, operationalized as the use of free grammatical morphemes and subordinate clauses, both at the beginning and at the end of a semester-long course in Spanish. Linck and Weiss (2015) found WMC effects on a metalinguistic grammar test using an operation span in a sample of beginners. Serafini and Sanz (2016) compared learners from a range of L2 proficiency levels, from beginner to advanced, and found a correlation between WMC and PSTM, and performance on both a grammaticality judgment test and an elicited oral imitation in lower, but not higher, proficiency learners over one semester. Their results showed that learner skill level, demands related to task performance, testing time, and item grammaticality were key factors mediating the impact of WM components. Sagarra (2017) examined the longitudinal effects of WMC on grammar and reading abilities and found that only WM tests with a processing component (performed under time pressure) yielded longitudinal WM effects in beginning learners over the course of one semester, as assessed by both recall (storage) and response time (processing) scores. Sagarra suggested that the presence of a taxing WM processing component, L2 proficiency level of the participants, and task demands explain the conflicting findings of longitudinal effects of WM in L2 grammar learning. Faretta-Stutenberg and Morgan-Short (2018) examined the role of WM in processing of L2 Spanish syntax before and after a semester of study in a traditional classroom context and a study abroad context. The impact of WM was observed in study-abroad learners, but not in at-home learners.

In sum, most longitudinal studies applying a WMC test conducted among low proficiency learners reveal WMC effects on L2 grammar development (Linck & Weiss, 2015; Sagarra, 2017; Serafini & Sanz, 2016). On the contrary, longitudinal studies on high proficiency learners reveal no WMC effects on L2 grammar development (Grey et al., 2015; Serafini & Sanz, 2016). As far as PSTM is concerned, some longitudinal studies reveal its effects on grammar learning in advanced learners (O'Brien et al., 2006, O'Brien et al., 2007) and beginners and intermediate learners (Serafini & Sanz, 2016), whereas others do not (Grey et al., 2015). This shows that there is a need for further research the relationship between WMC, PSTM, and longitudinal L2 grammar learning conducted in a finer-grained manner. The current study was an attempt to contribute to this line of inquiry.

THE CURRENT STUDY

The current study was motivated by a gap in the literature regarding the longitudinal relationship between grammar knowledge and two components of WM, that is WMC and PSTM, in intermediate L2 learners. In other words, we sought to explore whether these components of WM contributed to improving grammar knowledge over time. Unlike most previous studies, which focused on either PSTM or WMC, we tested both components of WM, as both seem to affect grammar learning (O'Brien et al., 2006; Pawlak & Biedroń, 2021; Serafini & Sanz, 2016). To make our study more fine-grained, we applied two tests of PSTM (digit span and nonwords) and two tests of WMC (reading span and listening span). Moreover, we decided to test intermediate level adult learners progressing from the B2 level to the C1 level. The available evidence suggests that longitudinal effects can be observed in lower-level but not advanced L2 learners (Sagarra, 2017). Longitudinal studies incorporating intermediate L2 learners (Serafini & Sanz, 2016) are much needed. This is theoretically relevant, as more and richer data can be gathered over a two-year period, which would allow for more complex L2 grammar development trajectories to emerge as well as for the evaluation of the possible endurance and fluctuations of correlations between WM and L2 grammar development. To heed Sagarra's (2017) call, we used WMC tests with a demanding processing component, that is, time pressure. Additionally, we applied a variety of grammar tests with an increasing level of difficulty, namely, testing gradually more difficult and complex structures introduced and practiced each semester to observe the potential improvement of the learners as modified by WM. Thus, the research question we set out to answer in this study was as follows: Will higher levels of WM contribute to improvement in grammar knowledge over the period of two years?

To answer this research question, we used a nonexperimental longitudinal design that measured in a first wave (T1) all participants' grammar knowledge and WM. Then, in three subsequent waves, six months apart from each other (T2-T4), we measured the same participants in terms of grammar knowledge. Using latent growth models (Wang & Wang, 2012), we explored whether there were changes in grammar knowledge over time and whether these changes could be explained by the initial levels of WM.

Participants

The study took place at the Pomeranian University in Słupsk, Poland. The sample consisted of 107 students (70 females). At the beginning of the study, their ages ranged from 19 to 23 years old (M = 19.50, SD = .87), and they had been studying English as a foreign language both in and out school for 4 to 11 years (M = 7.64, SD = 1.95). The participants' English language level was lower-intermediate (B1/B1+ following the Common European Framework of Reference) and achieved the C1 level by the end of the study, which is considered advanced. At university, they attended content classes in English, including subjects in linguistics and literary studies, and also took practical English classes, including grammar, pronunciation, and the four language skills. Over the two-year period the study covered they completed approximately 1200 class hours.

Measures

GRAMMATICAL KNOWLEDGE

As the main outcome variable, we administered four tests (six months apart) to have a longitudinal measure of grammatical knowledge, operationalized as the use of different grammatical forms on tests tapping into grammar knowledge. Each test was of increasing difficulty, which means that it included more complex structures, and incorporated the contents covered in the preceding semester. This means that the first test required grammar knowledge at B1+, the second at B2, the third at B2+, and the fourth at C1 level of proficiency. Each test contained receptive and productive tasks, including multiple choice, verb form filling, key word paraphrasing, and open cloze tasks. Thus, the similarity of grammatical task types safeguarded test-related reliability. The scoring ranged from 0, 0.5 to 1 point, depending on the seriousness of the error. The tests were coded and marked by two grammar lecturers and inter-rater reliability was at .89. The internal consistency of these tests, determined by calculating Cronbach's α, ranged between 0.70-0.91, depending on the test. The evaluation of each one of these tests was based on the accuracy of the answers; the scores ranged from 2 to 5, with 2 indicating fail, 3 – satisfactory, 4 – good, and 5 very good, which is the universal grading system at Polish universities. The composite score on each of the tests was 15 points. The threshold value for passing each test was 53%.

Task examples from the test taken after the fourth semester:

1. For each of the following sentences, write a new sentence as similar as possible to the original sentence. Use the words given.

   a. It's quite simple for a locksmith to copy your keys. Copied

2. Fill each of the blanks with a suitable word.

   a. I left with the distinct feeling of...........been..........for granted.

3. Choose the correct answer.

As the tree was too high to climb, the boys………their ball down only by throwing sticks at it.

   a. would have knocked

b. will have to knock
c. could knock
d. were able to knock

4. Put the words in brackets into the correct form.

The Mystery of the Loch Ness Monster continues (1)…………………..(fascinate) scores of tourists every year, who come to Scotland in the hope of (2)………………….(capture) on film some proof that the monster exists.

PHONOLOGICAL SHORT-TERM MEMORY (PSTM)

The first independent variable was PSTM, which was measured through a digit span (an adaptation of the Wechsler Intelligence Scale by Wechsler, 1997, for use with the Polish population by Brzezinski et al., 1996) and the Polish nonword span (PNWSPAN, Zychowicz et al., 2018). The digit span includes sets of digits to repeat initially forwards and then backwards. Split-half reliabilities for the WAIS-R (Pl) were .88-.93 for the full scale. In the nonword span, all the nonwords were two-syllable, phonologically possible sequences of five Polish sounds. They were prerecorded and presented to participants in sets of two, three, four, five, and six, three trials per stage, in the order from two to six items, yielding a total of 60 nonwords. Participants were asked to repeat the nonwords in the correct order. Each item was assigned from 0 to 3 points, depending on the quality of its recall. The Cronbach's αvalue of the PNWSPAN was .68.

WORKING MEMORY CAPACITY (WMC)

The second independent variable was WMC. It was measured through a Polish Reading Span (PRSPAN) test (Biedroń & Szczepaniak, 2012) and a Polish Listening Span test (PLSPAN, Zychowicz et al., 2017). The PRSPAN includes eight sets of sentences, which contain from three to ten sentences in Polish. The length of each sentence is approximately 10 words. The maximum score is 52. The PLSPAN has the same structure but contains nine sets of sentences, with a total score of 54. The test involves determining whether or not each sentence in a set is sensible and, at the same time, in the case of the listening span, remembering for later recollection the last word of each sentence, which was a two-syllable noun. In the case of the reading span, it was an additional word at the end of a sentence not related to it. Partial scoring was used in both tests to reflect the number of remembered words. Only learners who reached 80% accuracy in the processing task were included in subsequent analyses. The Cronbach's αfor the internal consistency for the PRSPAN was .69. The Kuder-Richardson αfor PLSPAN was .76. Both tests were computerized.

Procedure

Three sources of data were collected for the purpose of this study: (a) tests of PSTM and WMC taken at a single session at the beginning of the first semester; (b) four measures of grammar knowledge taken over two years, six months apart, after the first, second, third, and fourth semester; and (c) self-reported length of studying English (in years).

RESULTS

Data were analyzed using latent growth models (Wang & Wang, 2012). These models are appropriate when analyzing longitudinal data for the change over time and its determinants. This analytical approach allowed us to compute an intercept, which is understood as the mean of the construct being analyzed, and a slope, which is understood as the rate of change. The determinants can predict both elements in these models. Given the small sample size in this study, which was below the values suggested in the literature for structural equation models (Hoyle, 2012), we used a maximum likelihood estimation with robust standard errors (Jaccard, 2017). All the analyses were conducted using the library lavaan (Rosseel, 2012) for *R* (R Core Team, 2013). We decided to include the self-reported length of studying English, as L2 exposure appears to be a significant factor in longitudinal studies (cf. Saito et al., 2019).

In order to analyze whether control variables were necessary in the models, we analyzed the bivariate correlations between the four grammar knowledge measures and years studying English. Except for digit-span, all the tests correlated with grammar measures, with the correlations being weak to moderate. As shown in Table 1, only years of studying English significantly correlated with the four measures of grammar knowledge. For this reason, we ran robustness check analyses

TABLE 1.
Descriptive Statistics and Correlations Matrix

	M	*SD*	Min	Max	1	2	3	4	5	6	7	8	9
1. Grammar knowledge (T1)	3.83	0.77	2	5	1								
2. Grammar knowledge (T2)	3.99	0.82	2	5	.81***	1							
3. Grammar knowledge (T3)	3.86	0.81	2	5	.76***	.84***	1						
4. Grammar knowledge (T4)	3.87	0.87	2	5	.64***	.78***	.84***	1					
5. PSTM Digit Span	12.7	3.09	7	20	-.01	.02	.09	-.01	1				
6. PSTM Non-words	76.07	14.37	41	103	.19	.29**	.36***	.27**	.37***	1			
7. WMC Reading span	23.66	7.28	6	43	.19	.15	.25**	.22*	.32***	.44***	1		
8. WMC Listening span	28.09	5.77	17	47	.16	.28**	.35***	.30**	.38***	.39***	.60***	1	
9. Length Studying English	7.64	1.95	4	11	.58***	.49***	.44***	.39***	-.14	.09	.07	.09	1

Note. T1,T2,T3 and T4 indicate subsequent grammar tests.

*** $p < .001$; ** $p < .01$; * $p < .05$

TABLE 2.
Latent Growth Models with Grammar Knowledge as the Dependent Variable

Models without covariates																	
		Model 1: PSTM digit span				Model 2: PSTM non words				Model 3: Reading span				Model 4: Listening span			
		b	*SE*	*z*	*p*	*b*	*SE*	*z*	*p*	*b*	*SE*	*z*	*p*	*b*	*SE*	*z*	*p*
Intercept	Intercept	3.84	0.32	12.09	< .001	3.00	0.40	7.40	< .001	3.43	0.24	14.33	< .001	3.17	0.37	8.60	< .001
	Slope	−0.03	0.09	−0.36	0.721	−0.20	0.11	-1.82	0.069	−0.07	0.08	−0.94	0.350	−0.25	0.09	−2.74	0.006
Covariance	Intercept-Slope	−0.02	0.02	−0.98	0.330	−0.02	0.02	-1.24	0.217	−0.02	0.02	−1.13	0.260	−0.03	0.02	−1.38	0.168
Regressions	Intercept on independent variable	0.00	0.02	0.11	0.911	0.01	0.01	2.30	0.022	0.02	0.01	1.95	0.051	0.03	0.01	1.99	0.047
	Slope on independent variable	0.00	0.01	0.40	0.693	0.00	0.00	1.88	0.060	0.00	0.00	1.01	0.311	0.01	0.00	2.86	0.004
X2		18.52 (7)			0.01	18.40 (7)			0.01	17.28 (7)			0.016	17.90 (7)			0.012
CFI		0.964				0.966				0.970				0.968			
TLI		0.949				0.951				0.957				0.955			
RMSEA		0.124				0.134				0.125				0.129			
Models with covariates																	
		Model 1: PSTM digit span				Model 2: PSTM non words				Model 3: Reading span				Model 4: Listening span			
		b	*SE*	*z*	*p*	*b*	*SE*	*z*	*p*	*b*	*SE*	*z*	*p*	*b*	*SE*	*z*	*p*
Intercept	Intercept	1.83	0.40	4.61	< .001	1.50	0.40	3.72	< .001	1.85	0.31	6.02	< .001	1.68	0.38	4.39	< .001
	Slope	0.12	0.13	0.91	0.361	−0.07	0.13	−0.58	0.562	0.07	0.11	0.64	0.522	−0.11	0.12	−0.96	0.339
Covariance	Intercept - Slope	−0.01	0.02	−0.46	0.647	−0.01	0.02	−0.71	0.480	−0.01	0.02	−0.63	0.529	−0.01	0.02	−0.80	4.240
Regressions	Intercept on independent variable	0.02	0.02	1.18	0.236	0.01	0.00	2.29	0.022	0.01	0.01	1.90	0.058	0.02	0.01	1.85	0.064
	Slope on independent variable	0.00	0.01	0.20	0.845	0.00	0.00	2.12	0.034	0.00	0.00	1.15	0.252	0.01	0.00	3.19	0.001
	Intercept on years studying English	0.23	0.03	8.11	< .001	0.22	0.03	8.07	< .001	0.22	0.03	8.23	< .001	0.22	0.03	8.22	< .001
	Slope on years studying English	−0.02	0.01	−1.62	0.105	−0.02	0.01	−1.85	0.065	−0.02	0.01	−1.78	0.076	−0.02	0.01	−2.07	0.039
X2		19.81 (9)			0.019	20.42 (9)			0.037	19.27 (9)			0.023	19.82 (9)			0.019
CFI		0.970				0.970				0.973				0.972			
TLI		0.954				0.953				0.959				0.957			
RMSEA		0.106				0.116				0.108				0.112			

Note. In this table, the variables in the title (e.g., "Model 1:…") refer to the main independent variable used in the analysis.

including this variable as a control variable. Table 2 illustrates longitudinal changes in grammar knowledge.

The model with PSTM digit span as the independent variable showed that the intercept was significantly different from zero, but neither the slope nor the association between the independent variable and the slope and intercept were significant. In other words, there were no longitudinal changes in grammar knowledge over time predicted by PSTM. When adjusting for years studying English, the only difference was that this variable predicted the intercept. This means that students with more years studying English showed higher initial levels of grammar competence and that the difference between students with higher versus low levels of grammar competence did not change over time.

The model with PSTM nonwords as the independent variable showed similar results, yet this variable predicted the intercept. In other words, students scoring high on PSTM nonwords showed higher grammar knowledge. Nevertheless, the slope was not statistically significant, which means that there were no significant differences over time in grammar knowledge related to the PSTM nonword score. When adjusting for years studying English, the results showed that PSTM nonwords score predicted both the intercept and the slope. The fact that both coefficients were positive means that participants scoring high on PSTM nonwords had initial higher values in grammar knowledge and showed increments over time, although this effect size was rather small. In addition, years studying English predicted the intercept, reflecting that students high on this variable had higher initial values in grammar knowledge.

The model with reading span as the independent variable showed only the intercept as statistically significant. The inclusion of years studying English as a covariate only modified the results by showing that this variable predicted the intercept. In other words, students with more years studying English had higher initial values in grammar knowledge.

Finally, when using listening span as the independent variable, the results showed that this variable predicted both the intercept and the slope, although the effect sizes were small. This means that students with higher values in the listening span had higher initial values in grammar knowledge and reported an increment over time. When adjusting for years studying English, the results were similar, but the listening span only predicted the slope (i.e., changes over time). In addition, students with more years studying English had higher initial values in grammar knowledge, but they showed a slight drop over time.

DISCUSSION

The results of the current study examining the dynamics of grammar knowledge development as modified by WMC and PSTM revealed that only WMC, particularly the listening span test, predicted change over time. With reference to our research question, we found that both PSTM and WMC correlate with grammar tests scores, with correlations going from weak to moderate (cf. Serafini & Sanz, 2016; Suzuki & DeKeyser, 2017a). The only test that was not correlated was the digit span. This might be linked to the fact that although the digit span is considered a reliable PSTM test, its verbality has been questioned (Linck et al., 2014) as the verbal input is easily transferable to visual input. This might indicate that the digit span measures the capacity of not only the phonological loop but also the visuo-spatial sketchpad.

Years of studying English were more important than WM, with significant correlations with all grammar tests. However, the role of years of study seems to decrease over time, which suggests that due to intensive language training, this factor is gradually minimized and compensated by practice and instruction. This can be considered an important pedagogical implication. It is worth noting that none of the tests, save for the PLSPAN, predicted the slope, that is, the change over time. The participants with higher scores on the listening span had higher initial values in grammar knowledge and reported an increase over time. No longitudinal effect of the PRSPAN and small size effects of the PLSPAN indicate that WMC has little influence on grammar development in intermediate L2 learners, which contrasts with the results obtained by Serafini and Sanz (2016). It seems that although both tests are reliable measures of WMC, they are not equivalent. These results suggest that the divergence in the results of different studies can be attributed to the test properties and the measurement procedures. We also observed the lack of impact of PSTM as measured by nonwords on grammar knowledge changes over time, a result which resonates with Grey et al.'s (2015), results, but is different from those obtained by O'Brien et al. (2007). This discrepancy might result from methodological differences between grammar tasks. Tasks requiring spontaneous production (e.g., O'Brien et al., 2006) or elicited imitation (Serafini & Sanz, 2016) are more likely to overload WM in learners at advanced proficiency levels than written production or reception tasks (cf. Grey et al., 2015).

Another problem concerns the precise operationalizations of proficiency levels. Beginner, intermediate, and advanced levels are not clearly defined in many studies. Therefore, learners' unobserved variability in proficiency levels might lead to misinterpretation of results. In our study, the proficiency level of the participants was somewhat uneven and grammar knowledge at the outset was not fully controlled for, which is regarded as a study limitation. Lastly, the available research indicates that learner proficiency levels may mediate the extent to which WMC constrains adult learners' ability to process grammar knowledge. It seems that the factor of proficiency, especially at higher levels should be more thoroughly controlled for.

A limitation that could have affected the results was the small sample size in our study. A larger sample size might have allowed us to include in the same models measures of both PSTM and WMC to control for their shared variance, in addition to the covariates found to be statistically significantly associated to grammar knowledge. In fact, different criteria have been proposed, such as number of participants per parameter estimated or variables included in the model, with five independent variables (i.e., two PSTM dimensions, two WMC dimensions, and the control variable) and four dependent variables (i.e., grammar knowledge at T1-T4), which would require a larger sample size (for a review see Hamilton et al., 2003). In addition, a larger sample size, together with a complete longitudinal design (i.e., measuring PSTM and WMC in all waves), might have allowed for more sophisticated longitudinal analyses in a structural equation model framework, isolating individual stability in the constructs measured in our study (Hamaker et al., 2015).

There are some methodological implications for further research into the relationship between WM and grammar. A large body of evidence indicates that WMC affects SLA in general and grammar development in particular, as opposed to PSTM, which seems to be of minor importance, especially at more advanced levels of proficiency. This is probably related to WM overload, which is significantly higher in complex span tasks and which affects language processing (cf. Sagarra, 2017). Therefore, it seems that further research should use complex span tests to investigate grammar development longitudinally rather than simple spans, particularly at more advanced levels of proficiency. Moreover, including other covariates into the design, such as motivation and other measures of aptitude, rather than WM alone could better account for changes in grammar proficiency development.

Conclusions

We investigated the longitudinal effects of WM on the development of L2 grammar over the period of two years in young adult intermediate foreign language learners. The results confirmed the correlation between both WMC and PSTM and grammar knowledge scores. However, only WMC predicted changes in grammar scores over time. Moreover, years of studying English turned out to be a better predictor of grammar scores than WM, with its role decreasing over time. These results demonstrate that WMC is a weak predictor of grammar knowledge development at an intermediate level of proficiency. An important pedagogical implication for students majoring in English is that WM plays a minor role in learning outcomes and that the role of years of study decreases over time, which suggests that due to intensive language training these factors can be compensated by practice and instruction. Also, this study demonstrated that there is a need for further research into the relationship between WMC and PSTM and longitudinal L2 grammar learning, conducted in a finer-grained manner. We hope that our ongoing large-scale study will contribute to this line of enquiry.

ACKNOWLEDGEMENTS

This work was supported by the Chilean National Research and Development Agency [FONDECYT grant number 1220209].

DATA AVAILABILITY

Data and materials for the experiments reported here are available from the corresponding author on reasonable request

REFERENCES

Antoniou, M., Ettlinger, M., & Wong, P. C. M. (2016). Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PLoS ONE 11*, e0158812. doi: 10.1371/journal.pone.0158812

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423. doi: 10.1016/s1364-6613(00)01538-2

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189–208. doi: 10.1016/s0021-9924(03)00019-4

Baddeley, A., & Hitch, G. J. (1974). Working memory. In: G. A. Bower (Eds.), *Recent advances in learning and motivation* (pp. 47–90). Academic Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173. doi: 10.1037/0033-295X.105.1.158

Biedroń, A.,& Szczepaniak, A. (2012). Polish reading span test–an instrument for measuring verbal working memory capacity. In: J. Badio, & J. Kosecki (Eds), *Cognitive processes in language. Lodz studies in language* (pp. 29–37).

Brzezinski, J., Gaul, M., Hornowska, E., Machowski, A.,& Zakrzewska, M. (1996). *Skala inteligencji Wechslera dla dorosłych. Wersja zrewidowana. WAIS-R (Pl). Podręcznik* [Wechsler Intelligence Scale for Adults. Revised version. Manual]. Pracownia Testów Psychologicznych (PTP).

Conway, A. R. A., Jarrold, Ch., Kane, M. J., Miyake, A., & Towse, J. N. (2008). Variation in working memory. An introduction. In A. R. A. Conway, Ch. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *variation in working memory* (pp. 3–17). Oxford University Press.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review, 26*, 197–223. doi: 10.1007/s10648-013-9246-y

Daneman, M., & Hannon, B. (2007). What do working memory span tasks like reading span really measure? In: N. Osaka, H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 21–42). Oxford.

Ellis, N. C. (1996). Sequencing and SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition, 18*, 91–129. doi: 10.1017/S0272263100014698

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, 49A*, 234–250. doi: 10.1080/713755604

Ettlinger, M., Bradlow, A. R., & Wong, P. C. M. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics, 35*, 807–831. doi: 10.1017/S0142716412000586

Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research, 34*, 67–101. doi: 10.1177/0267658316684903

French, L. M. (2006). *Phonological working memory and second language acquisition: A developmental study of francophone children learning English in Quebec.* Edwin Mellen Press.

French, L., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29*, 463–487. doi: 10.1017/S0142716408080211

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.

Foryś-Nogala, M. (2021). Cognitive aptitudes and processing of L2 grammar: Exploring the role of rule inferencing skills and working memory. *Advances in Cognitive Psychology, 17*, 310–319. doi: 10.5709/acp-0340-1

Gagné, N., French, L. M., & Hummel, K. M. (2022). Investigating the contribution of L1 fluency, L2 initial fluency, working memory and phonological memory to L2 fluency development. *Language Teaching Research*. – doi: 10.1177/13621688221076418

Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *The Modern Language Journal, 99*, 137–157. doi: 10.1111/modl.12190

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*, 102–116. doi: 10.1037/a0038889

Hamilton, J., Gagne, P. E., & Hancock, G. R. (2003). The effect of sample size on latent growth models (ED476862). ERIC. Retrieved from https://files.eric.ed.gov/fulltext/ED476862.pdf

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25–38. http://dx.doi.org/10.1017/S0272263100010457

Hoyle, R. H. (2012). *Handbook of structural equation modeling.* The Guilford Press.

Jaccard, J. (2017). *Structural equation modelling made accessible: Sampling error, standard error, and estimation in SEM. Applied scientific analysis*.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217. doi: 10.1037/0096-3445.133.2.189

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics, 40*, 693–722. doi: 10.1017/S0142716418000796

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*, 861–883. doi: 10.3758/s13423-013-0565-2

Linck, J. A., & Weiss, D. J. (2015). Can working memory and inhibitory control predict second language learning in the classroom? *SAGE Open, 5*, 1–11. doi: 10.1177/2158244015607352

Lu, B., & Wen, Z. (2022). Short-term and working memory capacity and the language device: Chunking and parsing complexity. In: J. Schwieter, & Z. Wen (Eds.), *The Cambridge handbook of working memory and language* (pp. 1–22). Cambridge.

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological STM and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*, 379–413. doi: 10.1017/S0272263112000125

O'Brien, I., Segalowitz, N., 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. doi: 10.1017/S0142716406060322

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29*, 557–581. doi: 10.1017/S027226310707043X

Pawlak, M., & Biedroń, A. (2021). Working memory as a factor mediating explicit and implicit knowledge of English grammar. *Annual Review of Applied Linguistics, 41*, 118–125. doi: 10.1017/S0267190521000052

R Core Team. (2013). *R: A Language and environment for statistical computing. R foundation for statistical computing*. Retrieved from http://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36. doi: 10.18637/jss.v048.i02

Sagarra, N. (2017). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research, 33*, 341–363. doi: 10.1177/0267658317690577

Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience, and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition, 41*, 201–225. doi: 10.1017/S0272263117000432

Santamaria, K., & Sunderman, G. (2015). Working memory in processing instruction: The acquisition of L2 French clitics. In: Z. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 205–223). Multilingual Matters.

Sanz, C., Lin, H.-J., Lado, B., Stafford, C. A., & Bowden, H.W. (2014). One size fits all?Learning conditions and working memory capacity in ab initio language development. *Applied Linguistics, 37*, 669–692. doi: 10.1093/applin/amu058

Serafini, E., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition, 38*, 607–646. doi: 10.1017/S0272263115000327

Suzuki, Y., & DeKeyser, R. (2017a). Exploratory research on second language practice distribution: An Aptitude × Treatment interaction. *Applied Psycholinguistics, 38*, 27–56. doi: 10.1017/S0142716416000084

Suzuki, Y., DeKeyser, R. (2017b). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning 67*, 747–790. doi: 10.1111/lang.12241

Tagarelli, K. M., Borges Mota, M., & Rebuschat, P. (2015). Working memory, learning conditions, and the acquisition of L2 syntax. In: Z. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing: theory, research and commentary* (pp. 224–247). Multilingual Matters

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.

Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale – Third edition (WAIS III)*. The Psychological Corporation.

Wen, E. Z. (2016). *Working memory and second language learning:*

Towards an integrated approach. Multilingual Matters.

Wen, E. Z. (2019). Working memory as language aptitude: The phonological/executive model. In: Z. E. Wen, P. Skehan, A. Biedroń, S. Li, & R. Sparks (Eds), *Language aptitude: Advancing theory, testing, research and practice* (pp. 187–215). Routledge.

Wen, Z., & Li, S. (2019). *Working memory in L2 learning and processing.* In: J. Schwieter, & A. Benati (Eds). *The Cambridge handbook of language learning* (pp. 365–389). Cambridge University Press.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53*, 67–121. doi: 10.1111/1467-9922.00211

Zychowicz, K., Biedroń, A., & Pawlak, M. (2017). Polish Listening SPAN: A new tool for measuring verbal working memory. *Studies in Second Language Learning and Teaching, 7*, 601–618. doi: 10.14746/ssllt.2017.7.4.3

Zychowicz, K., Biedroń, A., & Pawlak, M. (2018). Polish Nonword Span (PNWSPAN): A new tool for measuring phonological loop capacity. *Glottodidactica. An International Journal of Applied Linguistics, 45*, 309–327. doi: 10.14746/gl.2018.45.2.18

RECEIVED 09.01.2022 | ACCEPTED 16.04.2022