Analysis of the Effects of Working Memory Capacity on L2 Vocabulary Recall and Speech Production
第二言語の語彙想起と発話における作業記憶能力の影響分析

Michael Kelland
ケランド・マイケル

***Abstract*:** The study addresses the question of how differences in the capacity of individuals' working memory (WM) effects ability to recall vocabulary items and aid in oral output over short and longer terms. Participants initially took an OSPAN test (an adapted version of Daneman's 1991 working memory speaking span test) using the PEBL battery with both OSCORE and TSCORE results used to create a high and low WM capacity experimental groups, followed by a vocabulary test, a treatment, and post-treatment testing. Tests were evaluated using a standardized university speaking test rubric with 2-way ANOVA run on the OSPAN outcomes of the top and bottom 30% of participants. Results show that although vocabulary test scores were equal at the pre-, post-, and delayed post-treatment testing stages between the two groups, high WM capacity participants consistently outscored lower capacity participants in speaking tests at all three testing stages.

***Keywords*:** working memory, assessment, fluency

***要旨*:** 本研究は、個人の作業記憶能力の差異が、語彙を想起して発話する能力に短・長期的に如何に影響するかを探究するものである。被験者には、まず PEBL ソフトを用いた OSPAN テスト（Daneman 1991 のスピーキング・スパン・テストを応用）を受けてもらい、そのスコアに応じて作業記憶能力の高いグループと低いグループに分けた上で、語彙とスピーキングのテスト、トリートメント学習、学習後テストを受けてもらった。テスト結果は大学の統一基準に基づいて評価し、OSPAN テスト結果の上位と下位 30%の被験者のスコアを比較する分散分析を行った。その結果、語彙テストに関しては、トリートメント学習前、直後、一週間後の両グループのスコアに差は見られなかった。しかし、スピーキング・テストに関しては、全ての段階において、作業記憶能力の高いグループが低いグループよりも高いスコアを得ていたことが分かった。

***キーワード*:** 作業記憶、評価、能力

1. Introduction

Working memory has been outlined in numerous ways, however can broadly be defined as a limited capacity cognitive system responsible for the temporary storage, processing, and manipulation of information necessary for a variety of complex cognitive tasks, including language comprehension, learning, and reasoning (Baddeley & Hitch, 1974; Miyake & Shah, 1999). It works as a processing source for the active maintenance of

task-relevant information while simultaneously processing the same or other information activated along with task operations (Swanson, Zheng, & Jerman, 2009).

Research on the topic of language acquisition with regards to WM capacity has indicated limitations may well be fundamental in processes connected to both the acquisition and use of first and second languages (Daneman, 1991; Fortkamp,1999; 2000; O'Brien, 2006) considering that the learning of a second language draws on a range of cognitive processes (Juffs and Harrington, 2011). These aforementioned studies have shown outperformance in favour of high WM capacity individuals in favour of those with lower capacity scores.

Daneman and Green (1986) and Daneman (1996), in research on language acquisition into both L1 and L2, acknowledge that learners' individual working memory capacities may be seen as independent constraints on the processes involved, therefore having direct implications on the ability to acquire language, and also the rate at which this happens. Several other research studies have clearly indicated correlations between working memory capacity and a variety of first language cognitive skills (Engle, Carullo & Collins, 1991; Daneman & Green, 1986), denoting parallels between high working memory capacity individuals and advantages in cognitive tasks, such as vocabulary acquisition.

2. The Present Study

The research question seeking to be answered through this research is whether students possessing higher working memory capacities are better able to both retain and recall vocabulary items, as well as aiding in oral output in both the short and long when compared to those with lower working memory capacities, therefore putting them at a potential advantage with regards to second language acquisition (SLA). The research focused on three main hypotheses: 1) It was expected that scores for high and low capacity participants on the pre-treatment test would be similar; 2) analysis would indicate that individuals who scored highest in the working memory test would do better in the post-treatment tests due to the increased ability to manage information within the working memory; 3) delayed post-treatment test scores would favor those with higher working memory capacities, however, it was unclear to what degree.

3. Method

The purpose of the research was to investigate the relationship between individual differences in working memory capacity (WMC) and second language acquisition (SLA) in terms of retention and recall ability of vocabulary items, and oral fluency by means of a testing against a standardized speaking test rubric. The research consisted of 4 parts, administered over three separate sessions. Session 1: Working memory capacity test, vocabulary test, speech test. Session 2: Treatment, post-treatment vocabulary test, post-treatment vocabulary speech test. Session 3: delayed post-treatment vocabulary test, delayed post-treatment speech test.

<table>
<tr><td></td><td colspan="2">TESTING</td><td>TREATMENT</td><td colspan="3">TESTING</td></tr>
<tr><td rowspan="2">CONTROL GROUP</td><td rowspan="2">WORKING MEMORY TEST</td><td rowspan="2">PRE-TEST VOCAB AND FLUENCY</td><td>NO TREATMENT</td><td rowspan="2">POST-TEST VOCAB AND FLUENCY</td><td rowspan="2">DELAYED POST-TEST VOCAB AND FLUENCY</td></tr>
<tr><td rowspan="2">EXPERIMENT GROUP</td><td>STANDARD TREATMENT</td></tr>
<tr><td colspan="2">ANALYSIS</td><td></td><td colspan="2">ANALYSIS</td></tr>
</table>

3.1 Participants

All participants were freshman students aged 18-19 years old, with a male/female ratio of 5:2, within two streamed Listening and Speaking classes at Tokai University, Japan. Research sessions were made up of sixty students in total, however because of the multi-sessional process nature of the study, due to a number of students missing one or more sessions resulting in their data becoming unusable, the total number of eligible participants reduced to thirty-four. At the analysis stage, the top and bottom 30% scorers in the OSPAN test using the OSCORE method were used as the high and low WM capacity experimental group. The mean score on the test was 48.44 with a standard deviation of 16.55.

3.2 Instruments

The experiment consisted of four separate sections: a test to compute participants' working memory capacity, a task to elicit vocabulary recall ability, a treatment, and a speech production exercise enabling analysis of scores against a rubric.

3.2.1 Working memory capacity test

Participants' working memory capacity was initially measured by means of the OSPAN speaking span test (Daneman, 1991), administered on PCs using the PEBL software battery. The test measures capacity by showing to-be-remembered items, interspersed with a mathematical distracting activity, with the tasks requiring serial recall of the to-be-remembered items, as well as a required correct mathematical answer threshold. After practice sessions, participants are presented with a single letter to memorise on a screen, followed by a maths equation to answer in numerical form, for example:

11 – 8 =

Participants are required to click the mouse when they have arrived at the answer, and on the following screen:

11 – 8 = 3

state whether the answer presented is correct or not by clicking on a "true" or "false" button. To ensure that participants focus not only on the letters, a maths score of under 85% would result in the participants' current session being cancelled, with data not recorded. This letter/maths process would be repeated with between two and seven sets, with participants being required to enter the letters in the correct order on an onscreen keyboard following each letter/maths question set. The results of the OSPAN are produced in the form of an "absolute OSPAN score" (OSCORE) and a "total number correct score" (TSCORE). The OSCORE being the total number of correct letters recalled in sets in which the entire data string is correct, with TSCORE being the total number of letters recalled in the correct position, not taking into account whether or not the entire set was recalled perfectly. For example, a 4-letter string of:

A G F D

would need all letters in the correct order to score 4 points, with one error, for example:

A G F C

scoring zero. Conversely, as 3 letters (A G F) were in the correct position within the string, scored via the TSCORE method, the participant would score 3, even though the entire string was not entered perfectly. The maximum score attainable on the OSPAN test is 80 for both OSCORE and TSCORE.

3.2.2 Vocabulary test

Participants were administered a vocabulary test containing fifteen vocabulary items connected to the main topic taken from a set textbook keyword list at CEFR A2 level.

They were asked to translate from Japanese to English. Spelling errors were not taken into account.

3.2.3 Treatment

The treatment phase of the research took place over 90 minutes, and gave exposure and practice of the keywords. Participants were first able to see answers to the translation test given previously, followed by a pronunciation/repetition phase, and a reading and gap-fill exercise based on the keywords. Participants were then divided into groups of three, with each provided with a different short reading on a specific sub-topic, containing the keywords, of between 100 and 130 words, and given 15 minutes alone to read. Participants were encouraged to memorise as much of the information as possible, in order to later be able to talk on their reading to other group members. After the allotted time, the three students within the group would come together, taking turns to either listen to other members' sub-topics, or talk about the information they had read. Students were each given five minutes to recall all information they had memorised, and were able to check their reading copy after to ensure all students were exposed to all information on all three sub-topics.

3.2.4 Speaking test

Following the treatment, three separate speaking tests were administered, pre-treatment (immediately before the treatment started), post-treatment (immediately after the treatment), and delayed post-treatment (one week after the treatment) where participants were asked to speak for 60 seconds on a given set question. In the case of the pre-treatment test, the topic was unknown to the participants, and no prior practice or keywords were studied. Voice recording applications on personal smartphones were used to record the tests, and the subsequently produced .mp4 files sent to the researchers by email.

4. Data Analysis Procedures

4.1 Vocabulary recall test

The vocabulary tests were marked by hand by the researcher, given a score out of 15 and recorded on a spreadsheet.

4.2 Speaking test

Speaking tests for all students at all three stages were graded by the same teacher, who was familiar with the rubric, yet was otherwise unconnected to the research and participants. The rubric consisting of pronunciation, vocabulary, structure and fluency sections.

4.3 Analysis

Data was analysed by means of two-way ANOVA (with Bonferroni comparisons), comparing the scores of the vocabulary test, pre-, post-, and delayed post-treatment speaking tests for both the high and low WM capacity participants.

5. Quantitative Findings

5.1 Vocabulary recall test

**: p < .10 *: p <. 05

Vocabulary Test Scores between the groups by two-way ANOVA (Bonferroni)									
			Low Group Average	High Group Average	Differences	SE	F	P	
Pre-test	Low	High	4.20	3.90	0.30	0.55	0.55	0.59	
Post-test	Low	High	12.40	12.80	0.40	0.55	0.73	0.47	
Delayed Post	Low	High	9.80	10.60	0.80	0.55	1.46	0.15	

Table 1. Vocabulary test analysis

The analysis shows that both high and low WM capacity students' pre-, post- and delayed post-treatment test scores on the vocabulary test remained on par throughout the three tests. This showed that participants within this experimental group all had comparable pre-knowledge of the vocabulary items, attaining similar scores (*D=0.3*), which, considering the standard test-passing focus placed on the curriculum in Japanese high schools, was not surprising. The results also indicated that having high or low WM capacity had little bearing on ability to recall items both in the short term following the treatment (*D=0.4*), and in the identical post-treatment delayed test administered one week later (*D=0.8*), showing that high and low WM capacity participants both gained and lost information at similar rates.

5.2 Speaking test

**: p < 0.1 *: p <. 05

Speaking Test Scores between the groups by two-way ANOVA (Bonferroni)

	TEST 1	TEST 2	TEST 1 AVG	TEST 2 AVG	DIFFE-RENCE	SE	F	P	
LOW WMC	PRE	POST	20	22.2	2.2	0.5871	3.7475	0.0019	**
	PRE	DELAYED	20	21.6	1.6	0.5871	2.7255	0.0296	*
	POST	DELAYED	22.2	21.6	(-)0.6	0.5871	1.0221	0.9407	
HIGH WMC	PRE	POST	23.5	25.4	1.9	0.5871	3.2365	0.0078	**
	PRE	DELAYED	23.5	26.5	3	0.5871	5.1103	0.0000	**
	POST	DELAYED	25.4	26.5	1.1	0.5871	1.8738	0.2073	

Table 2. Speaking tests analysis

Analysis shows that high WM capacity students significantly outscored the low capacity students at all three testing stages. However, the results of the first speaking test went against hypothesis in that, although placed in a level-streamed class and having scored equally in the vocabulary tests, high WMC students significantly outscored lower capacity students *(D=3.5)*, indicating that even before a treatment had been administered, these students were at an advantage to their lower capacity counterparts.

Comparing the differences before and after the treatment showed that, although high WMC students initially achieved superior scores, they both developed at similar rates, improving by approximately two points within each group.

In line with hypothesis made prior to the research, a principal difference between the two groups was noted between the post-treatment and the delayed post-treatment tests. Whereas the high WMC group continued to improve on past test scores *(D=1.1)*, the low WMC group saw a decline *(D=0.6)*, indicating that lower capacity students have a more drastic drop-off rate in the period directly following instruction, and are less able to recycle used and presented information to improve output levels in terms of testing against a speaking rubric.

The most significant difference that the analysis outlined when comparing the averages of the 2 groups was when scores were compared across all three tests. The low WMC group improved by a total of 1.6 points, from the first to the third test, however, the high WMC group improved by almost double, by 3 points. In terms of the data analysed for this experiment, we can see the clear natural advantages that students possessing high working memory capacity have over those within the same class who do not.

6. Discussion and Conclusion

Students' initial knowledge of vocabulary was very similar, and although no significant improvements for high over low WMC students was seen in the results of any of the vocabulary tests, the scores of the speaking tests yielded substantial differences. High WMC students scored higher at the post-treatment and delayed post-treatment speech tests stages, indicating not only an ability to hold onto and effectively utilize taught information in terms of language production, but also suggesting an initial advantage even before a treatment has even been delivered. The data alone suggests that higher working memory capacity students within a level-streamed class who have similar knowledge of vocabulary, regularly outscored their fellow students by significant margins in speech output tests.

7. Implications and Limitations

The pilot nature of the research in terms of student numbers limits the reliability of the research. However, parallel research on a larger scale could indicate similar results, potentially offering worthy contributions to the field. Even though the students were in a level-streamed class, a further limitation was that there would have been variations in the quality and detail of information relayed between students in the information-share section. An extension of this study could also include exploration of the recorded speaking tests through speech analysis software, indicating whether speech rates and articulation rates improved, giving a more rounded picture of actual spoken ability comparisons in terms of vocabulary output.

References

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.A. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 8, pp.47-90).

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. Psychonomic Bulletin and Review, 3, 422-433.

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

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in the role of working memory in comprehension and following directions in children. Journal of Educational Research, 84, 253-262.

Juffs, A. Harrington, M. (2011) Aspects of working memory in L2 learning. Cambridge University Press, 44.2, 137-166

Miyake, A., & Shah, P. (Eds.) (1999). Models of working memory: Mechanisms of active maintenance and executive control. New York: Cambridge University Press

Mueller, S. T. (2014). PEBL: The Psychology experiment building language (Version 0.14) [Computer experiment programming language]. Retrieved June 2014 from http://pebl.sourceforge.net.

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

Swanson, H.L., Zheng, X., and Jerman, O. (2009). Working Memory, Short-Term Memory, and Reading Disabilities: A Selective Meta-Analysis of the Literature.

Received on 24 November 2017