The Effect of Working Memory Capacity on Instructed Vocabulary Learning

Caroline J. Handley
Michael Kelland
Tokai University

Previous research has shown a small to moderate but robust effect of working memory (WM) on both first language (L1) and second language (L2) learning and comprehension. However, few studies have addressed the relationship between WM and vocabulary learning in a naturalistic (classroom) setting. In this study, we report on a multi-site experiment that assessed the effect of WM capacity on L2 vocabulary learning. The target items were embedded in a highly supportive learning context which included both input and output activities. Immediate and delayed posttests showed that even under such conditions, WM was positively associated with vocabulary learning outcomes. These results extend findings that WM influences initial word learning (e.g., Martin & Ellis, 2012). In contrast with the findings of Yang et al. (2017), the association between WM and vocabulary learning remained at delayed posttest, suggesting that WM may affect retention as well as learning rate.

先行研究によると、ワーキングメモリ（WM）は第一言語（L1）および第二言語（L2）の学習と理解の両方に対して、小から中程度ながら確たる効果を持っていることが示されている。しかし、教室などの自然な授業環境でWMと語彙学習との関係を扱った研究はほとんどない。本研究では、L2の語彙学習に対するWM容量の影響を評価するため、多地点で行われた実験の1つを報告する。対象となる項目は、インプットとアウトプットの活動を含む高度に支援的な学習コンテクスト内に組み込まれており、このような条件下でも、WMは語彙学習の成果と正の相関があることが、直後テストと遅延事後テストによって示された。これらの結果は、WMが初期の単語学習に影響を与えるという研究結果（Martin & Ellis, 2012）をさらに発展させるものである。また、Yang et al.（2017）の調査結果とは対照的に、WMと語彙学習との関連性は遅延事後のテストでも維持された。このことは、WMが学習速度のみならず記憶保持にも影響を与える可能性があることを示唆している。

https://doi.org/10.37546/JALTTLT48.3-1

Working Memory and Language Learning

Working memory (WM), as first proposed by Baddeley (Baddeley & Hitch, 1974; Baddeley, 1986, 2000), is “our mental sketchpad, where we hold information ‘in mind’ and process it” (Miller, 2013, p.411). It is also assumed to be critical for enabling the consolidation of information in long-term memory (e.g., Daneman & Carpenter, 1980). Importantly, WM has been shown to have limited capacity; people can actively maintain only a few items in WM before the system becomes overburdened (e.g., Baddeley & Hitch, 1974; Cowan, 2014). Numerous studies have demonstrated the effect of individual differences in WM capacity on learning outcomes, including studies on first language (L1) and second language (L2) acquisition (for reviews, see Juffs & Harrington, 2011; Szmalec et al., 2012).

In a meta-analysis of L1 research, Daneman and Merikle (1996) found WM was a significant predictor of language comprehension. They also found that simple WM tasks, which assess short-term storage ability or phonological short-term memory (PSTM), are more weakly associated with comprehension than complex tasks, which involve an intervening task, thus also assessing processing ability or executive WM.

In a meta-analysis based on 79 studies of WM and L2 processing and proficiency, Linck et al. (2014) estimated the population effect size of WM to be $r = .255$, which is generally considered to be a small to medium effect (see Cohen, 1992). Most studies reported a positive correlation between WM and learning outcomes, although, as in L1 studies, complex span tasks were better predictors than simple ones. However, the studies in the meta-analysis looked at general and specific learning outcomes and assessed different language knowledge and skills. In general, few studies have examined WM in relation to explicit L2 vocabulary learning.

Working Memory and Explicit Vocabulary Learning

Some studies suggest that WM is related to vocabulary acquisition at early stages of language learning. For example, Kempe et al. (2010) found that WM was associated with word learning in a task in which 47 participants with no prior exposure to the language learned to match Russian words to pictures. In a similar design in which 40 participants learned German words, Speciale et al. (2004) found that PSTM was associated with performance in an immediate posttest when translating the words into German, but not when translating from German into the L1 (English). In a study by Martin and Ellis (2012), in which 40 participants learned words (and grammar) in an artificial language, both

PSTM and WM were significant predictors of word learning at immediate posttest. In another study, Cheung (1996) assessed whether performance on two simple-span WM tasks predicted the number of trials needed to learn three English words through direct translation. After dividing the 84 participants into low and high English vocabulary subgroups, he found that the nonword span task was a significant predictor for the low vocabulary group. Finally, in a study with children who had been learning L2 English for more than three years, Masoura and Gathercole (2005) found that vocabulary knowledge, not PSTM, predicted the number of repetitions necessary for children to learn eight new words. However, children were divided into four groups (low and high vocabulary and PSTM), each with fewer than 20 children, which means that, given the estimated effect size (Linck et al., 2014), there was a low probability of detecting an effect of PSTM, even if one existed.

In addition, two studies that used complex WM tasks found conflicting results. Kormos and Sáfár (2008) measured the association between WM and English language gains following an intensive English course. They found that PSTM was related to language use scores (grammar and vocabulary) for a subset of 21 students who had some prior knowledge of English, but not in their larger group of 100 beginner learners. In addition, 45 students from this larger group also completed a complex WM task, which showed an association with language use scores. Elgort et al. (2018) assessed word learning in 47 Chinese and 50 Dutch advanced learners of English, who learned low frequency words and pseudowords by reading each item in three sentences, either inferring the meaning or writing the words. In immediate posttests, WM was associated with knowledge of form but not meaning, in both learning conditions for the Dutch participants and in the writing condition for the Chinese participants.

To our knowledge, Yang et al. (2017) is the only study of WM and vocabulary learning that has used naturalistic instruction and included a delayed posttest. Their study involved three groups of advanced-level English language learners. However, they primarily investigated the effect of different post-reading activities. All groups read the same text, including eight target items, followed by a sentence writing, gap-fill, or non-vocabulary-related (control) activity. WM was associated with vocabulary learning at immediate posttest only, except for the sentence writing group. The researchers concluded that WM modulates vocabulary learning rate but not decay rate. However, as each group had only 18-26 participants, significant effects of WM could only be detected at larger effect sizes than the meta-analytic effect size calculated by Linck et al. (2014) and non-significant results were reported for associations of the predicted effect size of WM at delayed posttest in the sentence writing and control group. Thus, the generalisability of these results is uncertain.

The Current Study

As reviewed above, the few studies that have examined WM and explicit L2 vocabulary learning have mostly measured WM using simple span tasks, focused on beginner-level language learners, involved fewer than 50 participants, and assessed only immediate learning gains. Although in their meta-analysis of WM and L2 learning, Linck et al. (2014) found no evidence of publication bias, suggesting the effect of WM to be robust, discrepancies in studies of WM and L2 vocabulary learning may be due to underpowered designs. The minimum power recommended by Cohen (1992) is 80% power at $\alpha = .05$ (i.e., an 80% chance of detecting an effect if there is a true effect, accepting a 5% risk of making a Type I error and rejecting a true null hypothesis). Based on the meta-analytic effect size of WM of $r = .255$ calculated by Linck et al. (2014), power analysis in G*Power 3 (Faul et al., 2007), shows that, assuming a positive or no effect of WM, 93 participants are needed to reach this power (to consider a potential negative effect of WM, 118 participants would be needed). Therefore, we designed a relatively large, multi-site study to explore whether WM, as measured by a complex span task, is related to explicit L2 (English) vocabulary learning in intermediate-level language learners, and whether WM is associated only with immediate learning rate or also with longer-term retention. To ensure ecological validity, we embedded the target words in a 90-minute communicative lesson plan, which was taught by the second researcher.

Method

Participants

A total of 111 first-year students in five intermediate-level classes at three Japanese universities took part in the study, after providing written informed consent in Japanese. However, the data were discarded for nine students who failed to attend all three classes. The data were also excluded for two additional students, who may have been disadvantaged on the translation-based vocabulary test because their L1 was not Japanese. The data for another three students were excluded as they scored zero points on the immediate posttest, suggesting that they had misunderstood the test, given that no other student scored fewer than three points. In the end, the final number of participants was 97.

Working Memory Test

An operation span (OSpan test; Daneman, 1991) from the PEBL battery of psychological tests (Mueller, 2012) was administered to assess each student's WM capacity. This complex span task was chosen as it does not involve language skills beyond decoding single letters, thus raising the likeliness of reflecting L2 speakers' genuine WM (Sanchez et al., 2010). In complex span tasks, target stimuli are interleaved with a distractor task to assess the ability to maintain information without rehearsal (Conway et al., 2005).

In this test, participants tried to remember strings of randomised single letters, presented individually and separated by a mathematical equation which they had to solve, such as 6 + 2 − 3 = ? The participants had 2 seconds to click the mouse to indicate they knew the answer, and then a number was displayed on the next screen. The participants indicated whether the number was the correct answer by clicking on the True or False button next to it. They were then prompted to type the letter string in the presented order. This letter/maths process was repeated with strings of two to seven letters. The test was restarted if the maths score was under 85%. A final WM score was calculated based on the recall accuracy of the letter strings.

Final scores included "absolute OSPAN score" (OSCORE) and "total number correct score" (TSCORE). The OSCORE is the total number of letters correctly recalled in sets in which the entire data string is entered in the correct order, whereas the TSCORE is the total number of letters recalled in the correct position, regardless of whether the entire set was recalled perfectly. The maximum attainable score is 80 for both OSCORE and TSCORE. The TSCORE was used here, as it is considered more reliable (Conway et al., 2005).

The test was conducted under controlled conditions in a university computer room. All necessary software from the PEBL battery was pre-installed, and the students familiarised themselves with the program through practice sessions before the test started.

Target Vocabulary Selection

To select 12 target vocabulary items, we first chose 20 words that could fit naturally into the teaching materials and did not fall within the first 2000 highest frequency English words on Nation's (2017) BNC/COCA headword lists. The initial 20 words were translated into Japanese by a bilingual colleague and given to a separate class of students with a similar English level as the study participants, who were asked to translate them into Japanese. These vocabulary tests were marked by the second author and a Japanese colleague; all items that were correctly translated by three or more students were excluded. The final set of 12 target vocabulary items—amphibian, carnivore, conspicuous, disguise, entice, evade, habitat, mate, mimic, offspring, venom, vibrant—were all medium-frequency words (from the 4000-8000 headword bands), except for the word "mate," which was unfamiliar in its scientific meaning.

Teaching Intervention

At the beginning of the lesson, students were given a pretest consisting of 12 short English sentences containing the target items, which provided minimal contextual clues to their meaning, and then asked to translate each word into Japanese (see Appendix B). The same instrument was used as a posttest and delayed posttest vocabulary test. The target vocabulary was taught in one 90-minute lesson, which included presentation and practice of the target words, and was adapted from a lesson plan used in a previous study by Kelland (2018). The students were first shown the correct answers to the pretest, followed by a pronunciation/repetition phase. They then read an introductory text on how animals use colour, in which the target vocabulary was boldfaced and underlined, and completed a gap-fill exercise, using pictures to assist them. Next, they were divided into groups of three, and each student read one of three different short texts (130 words) about how a particular animal uses colour (see Appendix A for all teaching materials). Each text included four target words, such that all 12 items appeared once across the texts. Texts were run through the Lextutor vocabulary profiler software (Cobb, n.d.) to ensure that all the words, apart from the target items and animal names (species), were within the first 2000 highest frequency words and so should be familiar and not cause additional learning load. The students read their text and underlined the four new words. They then prepared to tell their group about their animal, with instructions to use all four new words. As the students shared their information, they were instructed to also listen for the new words. Although the students were asked to recall the information from memory, afterwards they could check their text and share any forgotten information. This design meant that all students were exposed to all the information on all three animals and to all 12 target words during the speaking and listening activity.

Vocabulary Tests

A vocabulary test, identical to the pretest, was given as an immediate posttest at the end of the teaching session and as a delayed posttest two weeks later (see Appendix B). A meaning-recall translation test was used following Webb (2008), who argued that multiple-choice (recognition) tests may inflate scores (see also Stoeckel et al., 2021).

Results

Analyses were conducted in RStudio (RStudio Team, 2020), run on R version 4.2.2 (R Core Team, 2022) using a variety of packages: car (Fox & Weisberg, 2019), ggplot2 (Wickham, 2016), lme4 (Bates et al., 2015), moments (Komsta & Novomestky, 2022), psych (Revelle, 2022), QuantPsyc (Fletcher, 2022), tidyverse (Wickham et al., 2019).

Working Memory Test

The mean score on the WM test was 57.61 ($SD = 11.34$), from a maximum TSCORE of 80. Scores showed a significantly non-normal negative skew (Shapiro-Wilk test, $W = 0.96$, $p = .007$) and a boxplot revealed two extreme outliers (scores of 23 and 25 points). The data from these participants were retained, as removing them from the analyses revealed similar correlations.

Vocabulary Test Scores

On the vocabulary pretest, 76 students gave no correct answers, 17 gave one, five gave two, and one student gave three correct answers ($M = 0.31$, $SD = 0.62$), indicating that the 12 target words were largely unknown before the teaching intervention. Thus, all the words were assumed to be equally difficult, and the differences between raw scores on the tests were treated as linear.

The vocabulary posttest scores were high, $M = 8.82$, $SD = 2.67$, showing that students learned the new words during the lesson. The delayed posttest scores, although lower, showed that this new knowledge was maintained over the following two weeks, $M = 6.00$, $SD = 3.15$. As expected, vocabulary test scores were oppositely skewed, with predominantly higher scores on the posttest and lower scores on the delayed posttest (posttest $W = 0.91$, $p < .001$; delayed posttest $W = 0.94$, $p < .001$).

Working Memory and Vocabulary Learning

As the data are interval data and not normally distributed, a Kendall correlation was calculated (a non-parametric correlation which is more accurate with tied ranks), see Table 1. TSCORE values, our measure of WM, were correlated with vocabulary gains at both immediate posttest ($\tau = .180$, $p = .014$) and delayed posttest ($\tau = .189$, $p = .009$). These results are similar to the meta-analytic effect size of WM on L2 acquisition, $r = .255$ (Linck et al., 2014), being approximately equivalent to $r = .279$ and $r = .294$, respectively (see Gilpin, 1993). Due to the non-normal distribution, robust 95% confidence intervals (CIs) were calculated for each correlation via the bias-corrected and accelerated (BCa) bootstrap (Efron, 1987), with 10,000 samples. This analysis confirmed the positive correlations between TSCORE and vocabulary test scores.

Table 1.

Kendall Correlations for WM and Vocabulary Test Scores

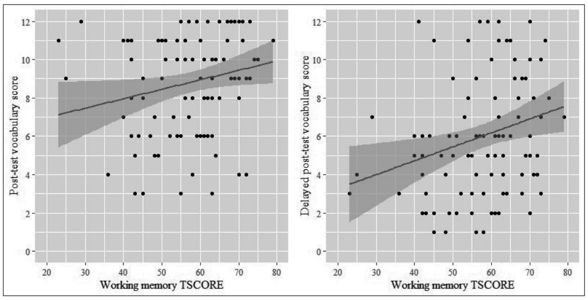
	Kendall's tau (τ)	p	95% CI BCa bootstrap
TSCORE – Posttest	.180	.014	.027, .317
TSCORE – Delayed	.189	.009	.056, .305
Posttest – Delayed	.400	< .001	.260, .516

The association between WM, as measured by TSCORE, and posttest and delayed posttest vocabulary scores is visualized in Figure 1. The regression line (surrounded by 95% confidence intervals) shows high uncertainty in the association between WM and vocabulary learning for those students who scored very low on the WM test. This may indicate that these students' performance on the WM test was adversely affected by uncontrolled factors, such as low task motivation, resulting in measurement error (Conway et al., 2005).

In addition, we examined whether fitting a linear mixed-effects model would improve this simple model. Adding random intercepts for class did not improve model fit for the association between WM and vocabulary scores, indicating that correlations were similar across the different classes. However, pretest vocabulary scores were significantly correlated with posttest scores, as shown in Table 2. Importantly, there was no correlation between vocabulary pretest score and TSCORE, indicating independent correlations between these variables and posttest scores. The calculation of standardized β coefficients showed that as TSCORE increased by 1 SD, posttest and delayed posttest vocabulary scores increased by 0.2 and 0.25 SD, respectively. This means that scoring an additional 11 points on the WM test was associated with correctly recalling an extra 0.5 and 0.8 words.

Figure 1.

Association Between WM and Vocabulary Learning



Scatterplots showing the correlations between WM and posttest and delayed posttest vocabulary scores. The shaded areas around the blue linear regression lines represent 95% confidence intervals.

Discussion

We examined the association between WM and instructed L2 vocabulary learning within a classroom context. The results showed a small but statistically significant effect on both the posttest and delayed posttest scores, confirming previous research suggesting that WM may be a predictor of L2 vocabulary learning (Yang et al., 2017), but with a larger sample size and multi-site design. In contrast with previous studies (e.g., Martin & Ellis, 2012), we found that WM was associated with learning outcomes in intermediate-level learners. Notably, this effect was observed in a high-support learning context, which included several explicit strategies for teaching the new words (visual images, definitions, and example sentences), an input task (reading), and a controlled but interactive output task in which students either produced or heard each target item. This design was informed by research showing that explicit pre-teaching of vocabulary is more effective than post-teaching, especially when accompanied by visual images (Alamri & Rogers, 2018). To our knowledge, this is the largest range of pedagogical tasks used in such a study, yet the association between WM and vocabulary learning outcomes was comparable to the meta-analytic effect size (Linck et al., 2014).

The association between WM and vocabulary learning points to the importance of this element of individual aptitude in L2 learning, especially given that learning new words may exert relatively low demand or load on WM resources. Sweller (2010) suggested that WM load in the presentation of new knowledge is dependent not on the number of different elements to be learned, but on the interactions between them. He described learning individual words as a "low element interactivity task" rather than a complex task, suggesting it may incur a relatively low WM load. Although the reading activities would have imposed a higher load, due to the syntactic interactions between words in sentences and the meanings created as the text unfolds, the texts were carefully controlled to contain only high-frequency (known) words, thereby minimizing extraneous load. In sum, the teaching materials enabled students to focus their cognitive resources on learning the target vocabulary while gradually promoting deeper processing through progressively more complex tasks, which should have mitigated WM effects. These findings contradict those of Li et al. (2019), suggesting that WM affects learning outcomes even when cognitive load is low, reinforcing their observation that researchers and, in our opin-

Table 2.

Association of WM and Pretest Scores with Vocabulary Test Scores

		Posttest vocabulary scores			
term	Coefficients	SE	t value	*p*	95% CI BCa bootstrap
(Intercept)	5.70	1.32	4.31	<.001	2.691, 8.291
Pretest	1.34	0.41	3.25	.002	0.762, 1.992
TSCORE	0.05	0.02	2.09	.040	0.003, 0.096
	Delayed posttest vocabulary scores				
(Intercept)	1.43	1.50	0.95	.34	-0.775, 3.834
Pretest	1.87	0.47	3.98	<.001	0.816, 2.647
TSCORE	0.07	0.03	2.71	.008	0.029, 0.109

ion, also teachers should give more consideration to the effects of individual cognitive differences on learning outcomes.

Finally, a few limitations of this study should be noted. First, the students were treated as homogenous groups regarding English language knowledge and ability, based on university assessment. Although the students' knowledge of the target items was pretested, overall prior vocabulary knowledge may have influenced learning outcomes. In a study with advanced English language learners, phonological memory was only associated with L2 vocabulary knowledge in lower proficiency learners (Hummel, 2009). Second, environmental factors that affect learning and WM were not considered. For example, sleep deprivation has a significant negative influence on WM (for a review, see Blasiman & Was, 2018, pp. 203-204).

Finally, although we recruited participants from three different universities, achieving a larger group sample size than in previous studies, the wide confidence intervals around the associations reported here indicate the uncertainty about the true effect size of WM on the students' word learning. Thus, our results highlight the need for future large-scale multi-site studies in research into the influence of WM capacity on L2 learning to achieve reliable estimations.

Conclusion

We examined whether WM was associated with vocabulary learning in a valid ecological setting, namely within a communicative classroom context. Given the noise of uncontrolled variables and the simplicity of the learning task, any association between WM and vocabulary learning was predicted to be small. Therefore, we used a within-subjects multi-site design to try to maximise the reliability of the findings. Our results suggest that vocabulary learning is associated with WM capacity, even when various instructional tasks are used, underscoring the relevance of individual differences to foreign language learning outcomes.

References

Alamri, K., & Rogers, V. (2018). The effectiveness of different explicit vocabulary-teaching strategies on learners' retention of technical and academic words. *The Language Learning Journal, 46*(5), 622–633. https://doi.org/10.1080/09571736.2018.1503139

Baddeley, A. (1986). *Working memory*. Oxford University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). Academic Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Blasiman, R. N., & Was, C. A. (2018). Why is working memory performance unstable? A review of 21 factors. *Europe's Journal of Psychology, 14*(1). 188–231. https://doi.org/10.5964/ejop.v14i1.1472

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary language. *Developmental Psychology*, *32*(5), 867–873. https://doi.org/10.1037/0012-1649.32.5.867

Cobb, T. (n.d.). *Vocab profiler-classic* [computer program]. Accessed 12 February 2024 at https://www.lextutor.ca/vp/eng/

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://dx.doi.org/10.1037/0033-2909.112.1.155

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/bf03196772

Cowan N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197–223. https://doi.org/10.1007/s10648-013-9246-y

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research, 20*(6), 445–464. https://doi.org/10.1007/BF01067637

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433. https://doi.org/10.3758/BF03214546

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82*(397), 171–185. https://doi.org/10.2307/2289144

Elgort, I., Candry, S., Boutorwick, T. J., Eyckmans, J., & Brysbaert, M. (2018). Contextual word learning with form-focused and meaning-focused elaboration. *Applied Linguistics, 39*(5), 646–667. https://doi.org/10.1093/applin/amw029

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fletcher, T. D. (2022). QuantPsyc: Quantitative psychology tools_. R package version 1.6. https://CRAN.R-project.org/package=QuantPsyc

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd Ed.). Sage publications. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics, 30*(2), 225–249. https://doi.org/10.1017/S0142716409090109

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137–166. https://doi.org/10.1017/S0261444810000509

Kelland, M. (2018). The effects of working memory capacity on L2 vocabulary recall and production. *The MindBrainEd Journal, 1*, 47–57.

Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning, 60*(1), 127–153. https://doi.org/10.1111/j.1467-9922.2009.00553.x

Komsta, L., & Novomestky, F. (2022). moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.14.1. https://cran.r-project.org/package=moments

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*(2), 261–271. https://doi.org/10.1017/S1366728908003416

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics, 40*(3), 693–722. https://doi.org/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*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

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*(3), 379–413. https://doi.org/10.1017/S0272263112000125

Masoura, V. M., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*(3/4), 422–429. https://doi.org/10.1080/09658210344000323

Miller, E. K. (2013). The "working" of working memory. *Dialogues in Clinical Neuroscience, 15*(4), 411–418. https://doi.org/10.31887/DCNS.2013.15.4/emiller

Mueller, S. T. (2012). The Psychology Experiment Building Language, Version 0.13. Retrieved from http://pebl.sourceforge.net

Nation, I. S. P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. Available from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Revelle, W. (2022). *psych: Procedures for personality and psychological research*. Northwestern University. https://CRAN.R-project.org/package=psych

RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio. http://www.rstudio.com/

Sanchez, C. A., Wiley, J., Miura, T. K., Colflesh, G. J. H., Ricks, T. R., Jensen, M. S., & Conway, A. R. A. (2010). Assessing working memory capacity in a non-native language. *Learning and Individual Differences, 20*(5), 488–493. https://doi.org/10.1016/j.lindif.2010.04.001

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics, 25*(2), 293–321. https://doi.org/10.1017/S0142716404001146

Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, *43*(1), 181–203. https://doi.org/10.1017/S027226312000025X

Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Szmalec, A., Brysbaert, M., & Duyck, W. (2012). Working memory and (second) language processing. In J. Altarriba & L. Isurin (Eds.), *Memory, language, and bilingualism: Theoretical and applied approaches* (pp. 74–94). Cambridge University Press. https://doi.org/10.1017/CBO9781139035279

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies on Second Language Acquisition, 30*(1), 79–95. https://doi.org/10.1017/S0272263108080042

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System, 70*, 38–49. https://doi.org/10.1016/j.system.2017.09.012

Caroline J. Handley is a research assistant at an institute in Spain. She taught English at universities in Japan for seven years, as well as in China and Spain. She has a PhD in Applied Linguistics from Swansea University, UK, in which she investigated lexical knowledge in first and second language speakers.

Michael Kelland is a lecturer at Tokai University in Japan. He has taught EFL for over 20 years in both Europe and Asia. His current research involves the role of working memory capacity in second language learning. He also investigates government-funded study abroad research programs.

Appendix A

Texts used in the teaching intervention (reformatted for publication).

(a) Introductory text

How Do Animals Use Colour?

The fur, skin, or feathers of animals can be many different colours. Some animals have dark or plain coloured bodies, but others have bright colours and patterns. This is because some animals use colour to hide, but others use colour to stand out. Colour helps animals in 3 important ways**:** (1) to hide, (2) to signal, or (3) to attract a partner.

(1) *To hide*

Many animals use colour to deceive hungry **carnivores** that want to eat them. There are 4 main ways animals use colour to hide. Sometimes, an animal hides itself against a background of the same colour, other animals have spots, stripes, or patterns that help them **evade** animals that like to eat them. Animals may also try to look the same as the area they live in. They use colour as a **disguise** to make them look like another object. The last way animals use colour is to **mimic** other animals, so they protect themselves by looking like other dangerous or bad-tasting animals.

(EXAMPLE 1:)
(EXAMPLE 2:)

(2) *To Signal*

Instead of hiding, some animals use colour to make themselves very **conspicuous** to other animals. Their colour sends out clear warning messages that they are dangerous, and should not be eaten. Many insects use colour to signal, as well as some **amphibians** and fish. Such animals may taste bad or they may produce **venom** which could kill other animals. Animals use bright colours to signal, such as red and yellow, which are very easy to see in their natural **habitat**.

(EXAMPLE 1:)
(EXAMPLE 2:)

(3) *To Attract a partner*

In many animals, males have developed different behaviours, such as singing, to attract females. Some male animals use colour to attract a **mate**, even though such colours may also make it difficult to hide from animals that want to eat them. The males have **vibrant** colours that can **entice** females, as they show that the male animal is strong and healthy, which means they will produce strong and healthy **offspring**.

(EXAMPLE 1:)
(EXAMPLE 2:)

(b) Gap-fill exercise

Keyword Practice

1. Actors ___________ the voices of different people.
2. The male lion was searching for a ___________.
3. Animals that eat other animals are called ___________.
4. Many super heroes ___________ themselves so that people don't know who they are.
5. A monkey's ___________ is the jungle.
6. The woman was ___________ due to her bright pink hair.
7. ___________ , such as frogs and toads, live on land and in water.
8. The ___________ of cats are called kittens.
9. The ___________ from some snakes can kill people.
10. The killer ___________ the police for many years before he was caught.
11. The artist loved to use bright, ___________ colours in her paintings.
12. Some people ___________ wild animals to their garden with food.

(c) Jigsaw reading texts

Bowerbird

The male Bowerbird, which lives in Australia, is an example of an animal that uses colour to find a mate. They are called Bowerbirds because they collect vibrant coloured objects to decorate a structure called a bower. They do this very strange behaviour to entice a female partner. The male spends many hours putting brightly coloured objects, such as shells, leaves, flowers, pieces of plastic, stones, or glass, in and around the bower. They will then dance next to the bower, and the female will choose the bird with the most beautiful bower and the best dance as her mate. Some bowerbirds can attract up to 30 mates in one season, and so have many offspring. However, males with badly decorated bowers might not be chosen by even one female.

Poison Dart Frog

The poison dart frogs' habitat is the jungles of South America. They are an example of a very conspicuous animal. These amphibians use very bright colours to help them to survive. There are over one hundred different types, or species, of poison dart frogs, and they have many different colours and patterns. The most common colours are yellow, red, green, blue, and black. Although poison dart frogs are only 5cm long, they are dangerous and can hurt other animals very badly. They all have poison, contained in their skin, that will make an animal sick or even kill it. One species has such strong venom that it could easily kill a large animal or person. The bright colours of these frogs warn other animals that they should not eat them.

Stick Insect

The stick insect, with a body length of between 1.5cm and 30cm, is an example of an insect found around the world that uses colour to help it survive. They are usually green or brown, which allows them to disguise themselves with the natural colours of the forest. In this way they can evade animals that want to eat them. Some types, or species, of stick insect can even change colour to match their environment. One species can keep their bodies completely straight, so they look like a stick, and others can even move their bodies from side to side to mimic the movement of the trees or leaves. They also only ever feed at night, when they can hide in the darkness from carnivores who want to eat them.

Appendix B

Vocabulary test used as pretest, posttest and delayed posttest.

Keyword Check

Look at the words below. Write the Japanese translation for the **KEYWORD** only.

Example:
TREE: The **TREE** is tall.
ANSWER= 木

1. **CARNIVORE**: It is a carnivore
   ANSWER = ______________________________
2. **EVADE**: He evaded everyone
   ANSWER = ______________________________
3. **CONSPICUOUS**: I like to be conspicuous
   ANSWER = ______________________________
4. **AMPHIBIANS**: She likes all amphibians
   ANSWER = ______________________________
5. **VENOM**: Some venom is very strong
   ANSWER = ______________________________
6. **HABITAT**: This is its habitat
   ANSWER = ______________________________
7. **VIBRANT**: It is a vibrant place
   ANSWER = ______________________________
8. **OFFSPRING**: Its offspring are over there
   ANSWER = ______________________________
9. **MIMIC**: She mimicked her friend
   ANSWER = ______________________________
10. **MATE**: It searched for a mate
    ANSWER = ______________________________
11. **DISGUISE**: That's a good disguise
    ANSWER = ______________________________
12. **ENTICE**: It enticed her
    ANSWER = ______________________________

2024 Board of Directors Election Candidate Statements can be found on page 51 of this *TLT*