Bilingualism: Language and Cognition 21 (3), 2018, 585–597 © Cambridge University Press 2017 doi:10.1017/S1366728917000141

Learning written word vocabulary in a second language: Theoretical and practical implications*

BRENDAN STUART WEEKES
Laboratory for Communication Science,
University of Hong Kong

School of Psychological Sciences, Faculty of Medicine,
Dentistry and Health Sciences, University of Melbourne,
Melbourne, VIC, Australia

(Received: March 27, 2016; final revision received: February 18, 2017; accepted: February 20, 2017; first published online 27 June 2017)

Short-term memory (STM) is required for second language learning. However, it is not clear what components of STM are necessary for the acquisition and lexicalisation of new written words. Studies suggest that memory for serial order is a critical cognitive process in spoken word acquisition although correlated mechanisms such as executive control also play a role. In this study, bilingual Cantonese–English speakers who are learning written expert words in a non-native language were tested over a one year period in their first year of instruction. Written word lexicalisation was measured using lexical decision and spelling to dictation tasks. Results showed measures of executive control (Stroop performance) and serial order memory capacity predict recognition and recall of written expert words at different stages. Whereas serial order memory predicts improvements to lexical decision accuracy, executive control predicts spelling to dictation performance after one year. The conclusion is that STM processes do constrain written word lexicalisation in a second language. However, executive control and serial order memory capacity have differential effects during word lexicalisation.

Keywords: serial order short term memory, executive control, fluid intelligence, expert word learning, Cantonese–English bilinguals

Vocabulary learning is a necessary component of second language acquisition. However, the cognitive processes used to learn new words are myriad (Majerus, 2013). Cognitive processes including executive control and memory for serial order are required to recall spoken words in a native and a non-native language (Leclercq & Majerus, 2010; Majerus, Poncelet, Van der Linden & Weekes, 2008; Majerus, Heiligenstein, Gautherot, Poncelet & Van der Linden, 2009; Majerus, 2013; Mikan, 2013). Studies have focussed on the immediate recall of spoken foreign vocabulary, low frequency words and nonwords. By contrast, the processes used to learn written words are relatively understudied (Andrews, 2015; Weekes, 1991; 1993; 1994).

We know there is a significant correlation between verbal short term memory (STM) tasks and spoken word learning in first (L1) and second (L2) languages (Gathercole, Service, Hitch & Martins, 1999; Massoura & Gathercole, 2005; Mikan, 2013). STM processes such as serial order memory capacity and measures of executive control predict vocabulary learning (Majerus et al., 2008; 2009). Studies of spoken vocabulary learning in a non-native language are mostly performed in laboratory settings. However, vocabulary learning typically involves study of written texts, often printed in a non-native language. This is a growing phenomenon particularly when English is the medium of instruction (Dearden, 2015). Although we might expect the processes used to learn spoken words in a second language are also necessary for the acquisition of written words, no study has tested this hypothesis directly. The goal of this study is to investigate vocabulary learning in a natural learning environment (tertiary studies) when the medium of instruction is English as in Hong Kong). To identify the significance of correlated processes that contribute to new vocabulary learning, multiple regression methods are useful. However, if multi-colinearity between variables is large, alternative designs can be used. The approach adopted here is to test hypotheses using a longitudinal design over a one year period in a naturalistic rather than a laboratory setting. The advantage of this approach is that effects of correlated variables might be revealed at different time points during learning.

Many theories of new vocabulary learning assume a causal role for STM and moreover some posit that executive control and serial order memory capacity have independent effects on spoken word learning (Gupta, 2003). Although the latter claim is not contested, no theory explicitly assumes differential effects of correlated cognitive processes on learning of written words. STM models that assume related mechanisms for executive

---

* This study was supported by Research Grant Number 754412 from the General Research Fund, Research Grants Council of Hong Kong. Thanks to Nathalie Ip, Josephine Tze, Kevin Lee and Loan Vuong for assistance with data collection and analysis.

Address for correspondence:
Brendan S. Weekes, Laboratory for Communication Science, Faculty of Education, University of Hong Kong, Hong Kong,
weekes@hku.hk

control and serial order memory capacity (e.g., Botvinick & Plaut, 2006; Burgess & Hitch, 2006), generate predictions about how serial order memory capacity constrains vocabulary learning. There is mounting evidence that serial order memory capacity – estimated by spoken recall of digits, words and nonwords – predicts vocabulary learning in native and non-native languages (Majerus, 2013) including Cantonese speakers learning in English (Cheung, Kemper & Leung, 2000). Serial order memory capacity predicts acquisition of spoken French words in native French speakers and English and native German speakers (Majerus et al., 2006: Majerus et al., 2008a, 2008b; Mikan, 2013). However, we do not yet know if tests of serial order memory capacity and theoretically related cognitive processes predict written word acquisition in a second language. We do know that phonological skills predict the acquisition of literacy (reading and writing) in languages that use an alphabet. However, this is not true in languages that use a nonalphabetic script such as Chinese (Barrett, 2011; Tong & McBride-Chang, 2010). The consolidation of written words into long term memory, i.e., lexicalisation, takes time for adult English speakers and undergraduates often do not attain spelling proficiency for words taught in the curriculum (Andrews, 2015; Holmes, Malone & Redenbach, 2008; Weekes, 1991, 1994, 1996). The question posed here is do the STM processes that we know contribute to spoken word learning in a second language also predict written word acquisition in a second language? The answer to this question has implications for models of second language learning and pedagogy when the medium of instruction is not the native language. This is a growing trend with the use of English globally (Dearden, 2015).

Stadthagen-Gonzalez, Bowers and Damian (2004) investigated the lexicalisation of written words by adult monolingual English speaking undergraduates learning new expert words. Expert words were extracted from the written texts used in academic disciplines and presented according to whether they were known before study and how often they were exposed to students during tertiary studies. Stadthagen-Gonzalez et al. (2004) assessed written expert word recognition using a lexical decision task and found that expert word recognition became faster and more accurate after one year of exposure. Although they assumed lexicalisation of expert words was achieved, it is not certain that expert words were fully lexicalised. A more reliable and valid test of expert word lexicalisation can be achieved using a writing to dictation task. One goal of the present study, therefore, is to test expert word lexicalisation using a spelling task.

Stadthagen-Gonzalez et al. (2004) did not offer a theoretical account of the cognitive processes used to lexicalise written expert words and therefore no predictions can be derived about how these processes contribute to learning of expert written expert words in a second language. However, given written words are typically presented simultaneously with spoken words in tertiary curricula, it might be assumed that the mechanisms used to lexicalise expert written words rely in part on STM. Therefore, one testable hypothesis is that performance on verbal STM tasks will be associated with expert word lexicalisation. To test this hypothesis, the present study recruited Cantonese speakers who are learning written expert words in their second language – English (L2). The study had three phases. In Phase One, participants were given a battery of tasks to measure cognitive abilities and a lexical decision task with expert written words and nonwords derived as minimal pairs e.g., presbycusis and bresbycusis to assess expert word knowledge. In Phase Two, expert word knowledge was reassessed to distil a metric of written expert word learning (Phase 2 – Phase 1 lexical decision performance). In Phase Three, expert word lexicalisation was assessed after one year via writing to dictation. The expectation was that STM measures would predict lexicalisation of written expert words.

The specific hypothesis was that measures of executive control and serial order memory capacity would be correlated with written word lexicalisation. To assess serial order memory capacity, verbal and nonverbal STM tasks were used including the serial order reconstruction paradigm developed by Majerus et al. (2008). In this paradigm, participants listen to familiar digits ranging from 3–7 items and then reconstruct the same order of presentation by pointing to index cards, i.e., nonverbally. Two main predictions were derived: (1) performance on tests of executive control and serial order memory capacity will predict the lexicalisation of expert words, and (2) these effects would be equivalent across tests of lexicalisation, i.e., recognition and recall of written expert words. An alternative outcome is that cognitive processes impact on lexicalisation at different stages in the learning trajectory with written expert words. One reason to expect this outcome is that STM could be less significant for acquisition of words if extant lexical networks, i.e., comprising words established in long term memory, already contain similar words (Lanfranci & Swanson, 2005). Cognates and synonyms allow lexical and semantic bootstrapping thus reducing the demands on STM processes. To minimise extant vocabulary, Cantonese–English speakers in Hong Kong were specifically recruited because Chinese uses a non-alphabetic writing system (Leck, Weekes & Chen, 1995) making direct translations of written expert words unlikely. Although extant vocabulary cannot be controlled completely in this sample – as undergraduates in Hong Kong must attain a level of English literacy to enter tertiary studies – this sample does allow a relatively strong test of the predictions. Hong Kong undergraduates are highly literate according to standardised testing (Program for International Student Assessment, 2012). Therefore, sampling Cantonese–English bilingual speakers who are learning written expert words when the medium of instruction is a non-native language (English), provides reasonable control over possible effects of extant vocabulary on written expert word lexicalisation in a second language. However, to control for the potential effects of extant vocabulary on new word learning, measures of English proficiency were also taken.

Method

Participants

Twenty-eight participants (7 males, mean age = 19.11, SD = 0.57) were recruited from the first-year cohort of a four-year tertiary programme. All participants were native Cantonese (L1) speakers with age of English acquisition ranging from 3 to 9 years with normal auditory and visual acuity and language development. The method of teaching was Problem Based Learning (PBL) for all students (Albanese & Mitchell, 1993). 1st-year students were recruited to minimise pre-instruction knowledge of written expert words. However, formal testing did not begin until the course had commenced. Testing of expert word recognition and recall was performed 12 months after Phase One with 25 participants (89%) of the sample recruited.

Preparatory study

Pilot testing established the validity of targets in the following way: twenty-one participants (five male, sixteen female) were recruited from an earlier first year undergraduate cohort. All were native Cantonese-speaking students aged between 18 and 19 years. Inclusion criteria were: (1) Cantonese first language; (2) no history of neurological disorders or developmental delays; (3) normal auditory and visual acuity; and (4) normal language development with no learning difficulties. All participants had learned English from six years of age and all had at least 12 years experience using English as a second language including sufficient mastery in spoken and written English to enter tertiary studies, i.e., all were in the top 5% of their cohort.

A total of 14 tasks were administered. In order to reduce demands on oral production, a majority of tasks chosen were nonverbal, i.e., required no spoken word response. Nonverbal is not equivalent to non-linguistic since all tasks reported could be performed with linguistic operations. To assess knowledge of English target words, a written lexical decision task with expert words and nonword minimal pairs was devised. Prior testing with a comparable group of participants from a previous year cohort had identified 500 expert words (e.g., echolalia) in the readings and course notes in the same curriculum. A sample of 72 English expert words were selected according to the criterion that were identified as words no better than chance (50%) in the first month of the course and were defined as targets (shown in Appendix One).

Executive control was assessed using the Flanker (Fan, McCandliss, Sommer, Raz & Posner, 2002) and Stroop tasks (Stroop, 1935). Serial order memory capacity was measured via the serial reconstruction task devised by Majerus et al. (2008a) consisting of digits presented auditorily to a participant individually and in a random order by a pre-recorded female voice. Each block had 3 to 7 digit sequences presented of increasing length and six trials were presented for each sequence length until criterion was attained (defined as the sequence in which a participant could not produce the correct order in more than half the trials). Each participant was presented with stimulus cards (5cm x 5cm) with a digit printed on a card and asked to arrange them linearly in the same sequence after a brief delay according to the horizontally arranged configuration. The longest sequence recalled was a measure of serial order capacity. Performance was close to ceiling. This may be due to larger digit span reported for Chinese–English undergraduates in Hong Kong (Chincotta & Hoosain, 1995). Verbal measures of STM including digit span, nonword span and word span in Cantonese and in English were tested. Measures of extant English vocabulary knowledge were also collected from each participant (see details below).

Items were chosen as targets if 90% of participants did not recognise a word in lexical decision performance (at least 10% error rate). To test the sensitivity of experimental tasks, target recognition was correlated with cognitive and language test performance. Regression analyses identified four variables predicting expert word acquisition: Digit-Symbol matching ($\beta = 0.824$), English vocabulary production ($\beta = -0.668$), nonword recall ($\beta = 0.353$) and target recognition from words translated into Cantonese ($\beta = -0.320$). Additional analysis found that serial order reconstruction (in English) was significantly correlated with performance on the same task conducted in Cantonese $r = 0.45$, $p = 0.04$ and a measure of Cantonese vocabulary production $r = -0.51, p = 0.02$ confirming the sensitivity and validity of experimental tasks both in Cantonese and English. Serial order reconstruction was significantly correlated with (1) delayed memory for Cantonese words $r = -0.51$, $p = 0.02$ and (2) English words $r = 0.53$, $p = 0.01$ and performance on Corsi Blocks $r = 0.56$, $p < .05$. A test of writing to dictation found fewer than 20% of targets were correctly spelled suggesting targets had not been lexicalised. Serial order reconstruction did not predict target recognition or recall. This is due to a ceiling effect.

Experimental procedure

Twenty-eight participants were tested on an individual basis in a quiet location. Each session lasted for 1.5 hours maximum. In Phase 1, participants performed tasks assessing expert word knowledge, serial order reconstruction, verbal and nonverbal executive control, verbal STM, intellectual ability, self-rating of Cantonese and English exposure and number of hours per week studying in English. In Phase 2, expert word knowledge was reassessed with a lexical decision task. In Phase 3, expert word knowledge was reassessed with a lexical decision task and writing to dictation task. Phase 3 was conducted 12 months after the end of Phase 2.

Materials

Expert word recognition

A lexical decision task was devised using E-prime software. A total of 72 word and 72 nonword trials were presented visually on a computer screen one at a time in random order. Participants were instructed to press a mouse button (counterbalanced) as quickly and accurately as possible to indicate whether the stimulus was a word or a nonword. Decision accuracy (percentage correct) and decision latencies (in msec) were recorded for subsequent analyses. Mean accuracy of 50% correct represents chance level of performance. Therefore, acquisition was defined as better than chance performance. Successful acquisition of expert words was further defined according to the correct rejection of nonwords above a criterion of 80%. Subsequent analyses of expert word lexicalisation refer to word trials only.

Serial order reconstruction task

As described in the pilot study – however the range of digits was increased from 3–7 to 5–9 due to ceiling effects with 3–7 digits found in previous testing.

Flanker task

The arrow version of the Flanker task was administered via the Presentation Software (Ridderinkhof, van der Molen, Band & Bashore, 1997). Target stimuli were arrows presented in the centre of the computer screen. The target was surrounded by distractors on both sides (right and left) either arrows or straight lines. In the neutral condition (32 trials) the target was flanked by straight lines; in the congruent condition (32 trials), flankers were arrows pointing the same direction as targets; in the incongruent condition (32 trials), flankers were arrows pointing the opposite direction. Participants decided if the arrow direction was congruent by pressing a right or left mouse button (counterbalanced). Accuracy and reaction time (RTs) were recorded.

Stroop Colour-Word task (English)

A computerised version of the Stroop Colour-Word task (Stroop, 1935) was administered. Stimuli were presented on a computer screen one at a time. Participants pressed a key to indicate ink colour of the stimulus target. Response accuracy (% correct) and reaction times were recorded and computed collapsing across the neutral and incongruent trials. The task consisted of colour word trials presented in a different colour (i.e., "BLUE" in red colour) and trials in the same colour. A word was presented one at a time on the computer screen in a random order. Participants identified word colour by pressing the corresponding button on a keyboard (e.g., 'red' by pressing 'R') as quickly and accurately as possible (counterbalanced). Response accuracy and reaction times (RTs) were recorded. Note that no oral response was required.

Digit span (English)

Software program 'D-Span' with pre-recorded voice audio of digits was used (native male English speaker). Lists with 3 to 11 digits were presented to participants with increasing length and inter-stimuli interval of 1000 msec. Participants were instructed to type digits recalled on a keyboard. No oral response was required. Familiar sequences were excluded (e.g., 911). Each set of digits consisted of five trials. The point at which a participant made 3 errors on 5 trials was treated as their maximum digit (nonverbal) recall capacity. The total number of digits correctly recalled was recorded as a participant's maximum digit span.

Word span (English)

Word lists contained 4 to 15 syllables with 3 trials for each syllable span. Stimuli were presented auditorily using a pre-recorded female voice. Each participant recalled (orally) the auditorily presented stimuli with increasing length starting from four syllables. The condition in which a participant made two errors out of three trials was treated as the maximum score and the largest number of syllables recalled was the participant's span.

Nonword span

Nonword stimuli were pseudo-syllables derived by eliminating gaps in the phonemes of English words according to phonotactic rules. Nonword sets consisted of 5 to 12 syllables. Stimuli were presented auditorily in a pre-recorded female (English speaker) voice. Participants recalled auditorily presented stimuli (orally) with increasing length beginning at 5 syllables. The condition in which a participant made 2 errors out of 3 trials was defined as the maximum score and greatest number of syllables recollected the total nonword span.

Corsi-block tapping

Nine plastic blocks (5 x 5 x 5 cm) were used. The participants were asked to observe a sequence of blocks tapped by a confederate and repeat the sequence tapping each block in the correct order. The task began with 2 blocks and increased to a maximum number of 9 blocks. Each sequence length contained 3 trials. The condition in which a participant made errors on 2 out of 3 trials was treated as a maximum level non-verbal span. The longest sequence recalled was computed and served as the participant's visuo-spatial nonverbal span.

Raven's Progressive Matrices

The test included Set A and Set B of the Standard Progressive Matrices (Raven, 1984). Participants saw 24 sequentially presented visual patterns with one pattern piece missing. They were asked to choose a target from among 6 choices to complete each pattern. A total score out of 24 trials was calculated and the raw scores were computed.

Controlled Oral Word Association Test (COWAT)

Participants were asked to name words from a semantic category (animals) in one minute (Benton, Varney & Hamsher, 1978). Total number of correct English animal names retrieved was defined as a measure of English word fluency.

British Picture Vocabulary Scale (BPVS)

A Short Form of the BPVS (Dunn, Dunn, Whetton & Pintilie, 1982) was administered. A computerised version was used consisting of 32 trials presented in increasing difficulty (according to word frequency). On each trial, participants heard a spoken word and were asked to match the word with one of four pictures. Number of correct responses was recorded and used as a measure of vocabulary. No oral response was required.

Cantonese Naming

Stimuli were taken from the set of colorized Snodgrass and Vanderwart (Roisson & Pourtois, 2004). A stimulus was presented on a computer screen. There were 30 trials. Each participant was asked to name pictures in Cantonese within 2 seconds. Naming accuracy was calculated. Total number named correctly was used to estimate word fluency.

English Naming

The procedure was identical to the Cantonese Naming task but alternative items were presented. Items were matched to the Cantonese task for relevant variables. There were 30 trials. Each participant was asked to name pictures in English within 2 seconds. Naming accuracy was calculated. Total number named correctly was the estimate of fluency.

Cantonese translation

Words with variable low, medium and high frequency in Cantonese were presented auditorily to each participant who was asked to translate a target into English orally. A total of 90 trials were given. Translation accuracy was a measure of L2 vocabulary.

English translation

Identical to the Cantonese translation task except an English word was translated into Cantonese. There were 90 trials in total. Translation accuracy was calculated.

Self-rated language questionnaire

Participants were asked to rate L1 and L2 proficiency using a 7-point scale (1 represents least fluent and 7 represents most fluent); number of hours exposed to Cantonese and English, respectively per day; age of onset and duration of English vocabulary learning; and exposure (in number of hours) to academic materials in English.

Results

Learning words in a second language probably draws on knowledge of phonological forms in the language network (Majerus et al., 2006; Masoura & Gathercole, 2005; Thorn & Gathercole, 2001) as well as cognitive skills that vary across individuals (Bartolotti, Marian, Schroeder & Shook, 2011; Cowan, 1995; Kaushanskaya & Marian, 2009). Therefore, the contribution of correlated variables was partialled out using regression analyses to determine the significance of STM measures on written expert word learning. Given a small sample size ($n = 28$), a maximum of three variables were introduced into analyses. It is important to note, however, that predictor variables were highly correlated with written expert word learning.

Mean expert word recognition accuracy in Phase One was 80.9% ($SD = 6.5$), which was significantly above chance level of 50% $p < 0.001$. This shows participants had lexicalised some target words before testing commenced. Nevertheless, when re-tested six months later (Phase Two) mean accuracy was 87.3% ($SD = 5.1$) a significant increase $t(27) = 8.88$, $p < 0.001$. For analysis of latencies, only correct trials were included. Participants averaged 976 msec ($SD = 246$) in Phase One and 859 msec ($SD = 166$) in Phase Two, a significant decrease in reaction time $p < 0.001$. On tests of executive function, accuracy was at ceiling on the Stroop task ($mean = 97.7$ and 98.4% on incongruent and neutral trials respectively) and the Flanker task ($mean = 95.8$ % and 99.7% on incongruent and neutral trials, respectively). Executive control was determined by subtracting RTs in the congruent condition from the incongruent conditions in each task. On the Flanker Task, average RT was 471 msec ($SD = 57$) for the congruent trials and 562 msec ($SD = 66$) for the incongruent trials. A paired *t*-test confirmed a Flanker

Table 1. *Descriptive Statistics for the main measures (N = 28)*

Measures	Tasks	Mean (*SD*)	Range
Expert word	LDD accuracy (phase 2- phase1)	7.49 *(5.48)*	0.00 to 26.39
	LDD RT (phase 1-phase 2)	117 *(123)*	-94to 364
Attention control	Flanker RT effect	91 msec *(28)*	37 to 168
	Stroop RT effect	227 msec *(131)*	86 to 650
Verbal STM	Serial order	7.93 *(0.98)*	6 to 9
	Digit Span	6.89 *(1.13)*	5 to 9
	Non-word Span	6.75 *(1.04)*	5 to 9
	Word Span	10.29 *(1.86)*	7 to 13
Visuo-spatial memory	Corsi-block tapping	5.75 *(1.29)*	4 to 9
General cognitive ability	Raven's Progressive Matrices	23.21 *(0.99)*	21 to 24
Semantic fluency	COWAT	20.11 *(3.91)*	14 to 28
Other background measures	BPVS	23.61 *(2.27)*	16 to 28
	Cantonese Naming	92.9% *(3.71)*	86.7 to 100.0
	English Naming	72.3% *(7.81)*	53.3 to 86.7
	Cantonese Translation	91.0% *(4.68)*	78.9 to 96.7
	English Translation	90.9% *(5.22)*	75.6 to 107.8
	Exposure to Cantonese (hr/day)	6.74 hrs *(2.23)*	1.5 to 10.0
	Exposure to English (hr/day)	7.62 hrs *(2.56)*	2.5 to 13.0
	Exposure to English for Academic purpose (hr/day)	5.99 hrs *(1.39)*	2.5 to 9
	Self-rated fluency in Cantonese 1–7)	6.64 *(0.49)*	6 to 7
	Self-rated fluency in English (1-7)	4.79 *(0.69)*	4 to 6

effect i.e., a significant RT difference $t(27) = -17.03$, $p < 0.001$. On the Stroop Task, average RT was 1039 msec (SD = 192) for the congruent trials and 1266 msec (SD = 242) for the incongruent trials. Paired *t*-test confirmed a Stroop effect i.e., a significant RT difference $t(27) = 9.15$, $p < 0.001$. Performance on other experimental tasks refers to the number of trials correct or else as stated.

A summary of the descriptive statistics from all experimental tasks is shown in Table 1 and a summary of correlation coefficients between experimental tasks is shown in Table 2. The difference in lexical decision time between Phase Two and Phase One was used to form the index of expert word learning for analyses (accuracy was significantly correlated with the index $r = -0.42$, $p = 0.03$). Ceiling effects were noted for performance on Raven's Progressive Matrices and self-rated fluency of Cantonese (oral and written vocabulary) but no other task.

Significant positive correlations were observed between expert word learning scores and STM measures, including serial order reconstruction $r = 0.52$, $p < 0.01$, digit span $r = 0.50$, $p < 0.01$, word span $r = 0.44$, $p < 0.02$, and number of hours of exposure to academic materials in English $r = 0.52$, $p < 0.01$. Significant positive correlations were also observed between serial order reconstruction and digit span $r = 0.69$, $p < 0.01$ and with nonword span $r = 0.64$, $p < 0.01$ and nonword span with word span $r = 0.82$, $p < 0.01$ showing that recall performance on verbal and nonverbal tasks was correlated. Regression analyses tested significant predictors. Results are summarized in Table 3. Only performance on serial order reconstruction and exposure to academic materials independently predicted expert word learning performance in Phase Two.

Re-testing of expert word knowledge was performed twelve months after completion of Phase One with 25 participants (89%) of the original sample recruited. Three tasks were presented: a nonword delayed repetition task requiring verbal production, expert word lexical decision, and writing to dictation. Non-word delayed repetition was adapted from Majerus et al. (2006) to test verbal STM. Materials were 30 monosyllabic nonwords taken from Gupta, Lipinski, Abbs, Lin, Aktunc, Ludden, Martin and Newman (2004). Participants heard spoken nonwords presented one at a time and were asked to repeated each item after a 6 second delay during which they continuously repeated blah, blah (articulatory suppression). Responses were recorded with a digital recorder and were scored via offline transcription. The number of correct responses (nonwords recalled) was calculated for each participant. Mean expert word lexical decision accuracy was 89%

Table 2. *Correlation matrix for all variables*

FULL	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A																				
B	0.72**																			
C	0.57**	0																		
D	0	0.05	0.24																	
E	–0.25	–0.12	–0.33	–0.19																
F	0.87**	0.50**	0.66**	–0.02	–0.29															
G	0.43*	0.45*	0.22	0.19	0.08	0.19														
H	0.74**	0.36	0.55**	0.08	0.12	0.82**	0.57**													
I	0.66**	0.52**	0.39*	–0.06	–0.29	0.69**	0.32	0.64**												
J	–0.26	–0.30	0.10	0.15	–0.07	–0.20	–0.05	–0.41*	–0.22											
K	–0.07	–0.15	0.21	0.09	0.06	0.12	0.15	0.23	0.13	0.07										
L	0.24	0.51**	–0.12	0.23	–0.05	0.12	0.34	0.11	0.05	0.11	–0.04									
M	0.12	0.38	0.11	0.29	0.04	0.46*	0	–0.06	–0.11	0.28	0.26	0.18								
N	0.10	0.07	0.05	–0.10	0.02	–0.11	–0.15	0.42*	0.14	0.03	0.37	0.89**	0.20							
O	–0.03	–0.13	0.19	0.26	0.13	–0.15	–0.14	–0.02	–0.24	0.12	0.16	0.10	0.19	0.56**						
P	0.32	–0.29	0.14	0.05	0	0.18	0.03	0.27	0.02	–0.18	0.19	0.39*	0.12	0.58**	0.68**					
Q	–0.01	0.04	–0.03	0.18	0.12	–0.11	0.08	0.09	0.02	0.02	0.04	0.22	0.03	0.38*	0.32	0.23				
R	0.21	0.25	0.11	0.17	0.06	0.02	0.04	0.11	0	–0.04	0.38*	0.05	0.39*	0.58**	0.26	0.31	0.17			
S	0.52**	0.55**	0.17	0.01	–0.08	0.25	0.26	0.22	0.20	–0.15	0.02	0.24	0.19	0.38*	0.09	0.24	0.29	0.69**		
T	–0.30	–0.10	–0.20	0.03	0.02	–0.21	–0.09	–0.33	–0.29	–0.15	0.16	0.25	0.39*	0.12	0.13	0.26	–0.14	0.10	–0.17	
U	0.10	0.02	0.14	0.03	0.30	0.07	0.24	0.23	0.08	0.06	0.02	0.06	0.04	0.26	0.49**	0.43*	0.26	0.25	0.06	0.21

A Overall accuracy in lexical decision task
B Accuracy in word trials in lexical decision task
C Accuracy in nonword trials in lexical decision task
D Flanker (Reaction time measures)
E Stroop (Reaction time measures)
F Digit Span
G Word Span
H Nonword Span
I Serial order memory
J Corsi-blocking tapping task
K Raven's Progressive Matrices
L British Vocabulary Picture Scale
M Cantonese Naming
N English Naming
O Cantonese Translation
P English Translation
Q Cantonese Exposure
R English Exposure
S Exposure to academic materials
T Self-rated proficiency in Cantonese
U Self-rated proficiency in English
$^{*}p < .05$, $^{**}p < .01$

(SD = 7.1), which was not significantly improved from performance in Phase Two. The writing to dictation task comprised of seventy expert words (Appendix One), recorded by a female English speaker. Participants heard expert words presented one at a time, and wrote the word using a computer keyboard. The writing to dictation task was presented after completion of the lexical decision task for all participants. Mean writing-to-dictation accuracy was 61% correct ($SD = 15$, range 45–79%), which is a significant increase from performance in Phase One $t(22) = 4.48$, $p < 0.01$.

In regression analyses, performance on lexical decision and writing to dictation tasks were dependent measures and predictors were serial order reconstruction, Flanker, Stroop and Ravens scores (measured in Phase One) and delayed nonword repetition (measured in Phase Three). Due to the significant correlations between delayed nonword repetition (verbal) and serial order

Table 3. *Multiple regression results predicting L2 written expert word learning (Phase 2-Phase 1)*

Predictor variables	β	t	p
Serial reconstruction	.60	4.08	.001*
Academic materials	.39	2.87	.01*
Flanker	.01	0.06	.95
Stroop	–.01	–0.06	.96
BPVS	.11	0.81	.43

*$p < 0.05$

Table 4. *Multiple regression results predicting L2 written expert word lexicalisation (Phase 3)*

	Lexical Tasks					
	Lexical decision			Writing-to-dictation		
Predictor variables	β	t	p	β	t	p
Serial reconstruction	–.01	.06	.96	.12	.62	.55
Nonword repetition	–.02	–.10	.93	–.18	–.93	.37
Flanker effect	.06	.24	.81	–.13	–.71	.49
Stroop effect	–.37	–1.4	.17	.42	2.2	.045*
Raven's matrices	.19	.74	.47	.43	2.3	.035*

*$p < 0.05$

reconstruction (nonverbal) $r = 0.64$, $p < 0.01$, variables were analysed separately. Results showed significant effects of Ravens score $\beta = 0.48$, $p < 0.05$ and Stroop score $\beta = 0.42$, $p < 0.05$ only on writing to dictation. No other variable had a significant effect (see Table 4).

Discussion

Results show that executive control and serial order memory capacity both predict the lexicalisation of L2 written words. However, these effects varied across task: lexical decision (written expert word recognition) and writing to dictation (written expert word recall). Serial order memory capacity (nonverbal digit reconstruction) significantly predicted written word learning but was not significant after one year – whereas nonverbal executive control (Stroop performance) did not predict written word recognition but did predict written word recall. Exposure to academic materials (assessed in Phase One) predicted written word recognition but did not predict written word recall and nonverbal reasoning predicted written word recall. The results support the hypothesis that cognitive processes used to acquire spoken words in a second language predict the learning of written words. It is of interest to note that almost all tasks predicting expert written word learning were non-verbal, i.e., required no oral production in English – except for the correlation between word learning and word span (see Table 2). If it is assumed that writing to dictation draws on phonological representations, the results are surprising since nonverbal tasks predict written word recall after one year of exposure. Verbal STM could predict spoken word learning using an auditory expert word lexical decision task.

The results can be understood with reference to Gupta's (2003) model of vocabulary acquisition, with the caveat that the model was not proposed to explain written word learning. This model assumes that serial order STM predicts spoken language acquisition. STM allows for the activation and rehearsal of phoneme sequences and strengthens formation of lexical and phonological representations. Burgess and Hitch (2006) proposed a similar account based on a Hebbian adjustment of long term connection weights between serial order STM and extant language representations, including lexical and sublexical language networks. When a word is presented for learning, phonemes are activated in the sublexical system and a node is formed in the lexical system. Serial order memory encodes and temporarily stores phonemes allowing transfer of sound patterns to the language network subsequently building up long-term phonological representations. The lexical system stores the phonological representations whereas the sublexical system stores subword phonological information. Both systems may be relevant to expert word learning in a second language, but this is not considered in models.

Majerus et al. (2008a) report data from English speaking undergraduates and Majerus et al. (2008b) and data from German speakers showing serial order memory capacity predicts learning of spoken French words (L2) in bilingual speakers. The present results show that serial order memory capacity predicts learning of written words learning of written English words in bilingual speakers. Surprisingly, nonverbal processes appear to be sufficient to learn written words in a second language, i.e., verbal mediation may not be necessary. The effects of serial order memory capacity and nonverbal executive control on written word lexicalisation in a second language is therefore relevant to theoretical models of word learning, i.e., oral production may not be necessary for written word learning. This suggests that strategies such as rote repetition might have a minimal impact on learning written words. Of course it is not possible to exclude verbal processes as significant predictors of written word lexicalisation in a second language based on the present results, e.g., verbal word span was also correlated with written word recognition. However, the results suggest that verbal STM may not be necessary or at least not may not be significant for teaching new written words in second language. This assertion may be qualified by the unique type of written words learned here, i.e., expert words.

The results suggest that serial order memory capacity has greatest impact in the early stages of written word lexicalisation whereas executive control impacts on lexicalisation over the longer term. This is also of interest theoretically since written word recognition and recall are universally assumed to depend on the quality of phonological knowledge (both lexical and sublexical) in languages that use an alphabetic script such as English. Another way to interpret the differences across task however is to assume that word recall is a more difficult task than word recognition. Both tasks require lexical retrieval (Weekes, 2010), which in turn requires the inhibition of competitor responses within the language network. Inhibition of competitors may be more demanding in written word recall, particularly when a word cannot be spelled correctly via sub-lexical (phoneme to grapheme) procedures alone. As the correct spelling of expert words in English requires a degree of lexical knowledge (e.g., presbycusis), inhibition of competing representations for written word output must be necessary. Deep dysgraphia is an extreme consequence of damage to such a mechanism, resulting in tendency to produce semantic errors in writing to dictation (Weekes, 2006). The finding that Stroop performance predicts writing to dictation supports the claim that inhibition is necessary. On the other hand, lexical decision could be performed based on multiple criteria (orthographic, phonological, semantic) that reflect recollection rather than recall, perhaps making it an easier task. Lexical decision performance was therefore (not surprisingly) close to ceiling here. Prior studies have used lexical decision as a measure of expert word lexicalisation (Stathagen-Gonzalez et al., 2004) rather than written word LEARNING. The results suggest that written word recall may be a more demanding and thus more reliable test of lexicalisation that written word recognition. It is important to note that the present findings might be different if spoken word learning were the measure of lexicalisation, since verbal STM processing may also be a significant constraint.

The results suggest that the process of lexicalisation for written words in English as a second language is incremental and furthermore that lexicalisation is not complete after one year of study (spelling is little better than 60% correct). Written word recall is a demanding test (Andrews, 2015; Burt, 2006; Holmes et al., 2008; Weekes, 1994; Weekes, Castles & Davies, 2006) and is vulnerable to brain damage (Weekes, Davies, Parris & Robinson, 2003; Weekes, 2005; 2012). Less is known about the written word recall in a second language although some studies have looked at Cantonese speaking children learning both alphabetic and non-alphabetic scripts (Tong & McBride-Chang, 2010). The evidence shows that phonological knowledge is one significant predictor of literacy in English for native Cantonese speaking children but does not generalise to literacy in Chinese readily (Barrett, 2011; McBride-Chang, Tong, Shu, Wong, Leung & Tardif, 2008; Tong & McBride-Chang, 2010). The present results suggest that verbal STM processes tested in English and in Cantonese do not predict written recall of English words in Cantonese–English adults. This may be specific to native Cantonese speakers. However, one possibility is that acquisition of literacy in Cantonese–English speakers depends in part on nonverbal STM processes. These are not often assessed in studies of reading and writing in this population. There is evidence that such processes predict spoken word learning in young Cantonese speakers (Ooi, 2016). Given the relationship between spoken and written language, this is an interesting hypothesis.

As noted by Majerus and others (Lanfranci & Swanson, 2005; Majerus, 2013), verbal STM may be most critical for the acquisition of new words when the lexical network contains few similar sounding words. Cantonese and English share very few cognates, etymology and NO direct translations of written words. The results therefore allow inferences about relative contributions of correlated cognitive processes on lexicalisation of new words without extant knowledge. The present results show that proficiency in English (naming, rated exposure, translation) did not significantly predict lexicalisation of written words. The sample tested is therefore a suitable one for testing theories of vocabulary learning in a second language with maximum control over the influence of known words on lexicalisation. Hong Kong samples are therefore well suited for laboratory based studies of cognitive processes in word learning.

The results suggest the mechanisms for inhibitory control and memory are dissociable (Hamilton & Martin, 2005; Majerus, 2013). To learn a foreign vocabulary, semantic features (meanings) must be linked to novel orthographic and phonological representation. Concepts are assumed learned via L1 and then linked to L2 (Tian & Macaro, 2012). Expert words are unique because extant concepts are rarely available before study (in L1 or L2). Expert word learning in an undergraduate sample therefore likely builds on new vocabulary incrementally. In the immediate stage of learning, conceptual mediation is probably necessary. This would explain why Stroop performance (conceptually mediated) predicts expert word lexicalisation whereas Flanker performance (not conceptually mediated) does not. An additional conjecture in the case of second language learning is that greater executive control inhibits the activation of L1 when learning new words. For example, when learning the word "drooling", students might inhibit competing Cantonese words such as /lau$_4$/ /hau$_2$/ /seoi$_2$/ relying instead on their conceptual knowledge, e.g., saliva running from the mouth – that is also language independent. Learning via conceptual mediation alone might be an exceptional case, however, when rarely used expert words are studied for lexicalisation and English (L2) is a medium of instruction.

Other cognitive processes may be critical to expert word learning in another language. Mental shifting, which includes the ability to switch between languages (Yim & Bialystok, 2012) is necessary to code-switch between two different languages. Tian and Macaro (2012) reported that vocabulary learning is better when instructors code-switch during lessons. To the extent that Stroop performance here reflects switching, relationships between Stroop and writing to dictation performance is consistent with such reports. Code switching might also predict expert word lexicalisation. Results from pilot testing showed that lexical decision was significantly correlated with performance on Digit Symbol Matching, which is a nonverbal test of code-switching ability. Further studies could test relationships between code switching and written word learning by manipulating code switching explicitly in the learning context.

Regression analysis found that measures of nonverbal problem solving (Ravens score) predicted lexicalisation after one year. This is notable given a restricted range of scores on the task (observed range = 20–24 possible range = 0–24). Pishghadam and Khajavy (2013) reported that Raven scores accounted for 12% of variance in foreign language proficiency. Nonverbal problem solving can be defined as mental operations including inferences, concept formation and hypothesizing to identify novel relations to solve problems (McGrew, 2009). Participants with better nonverbal problem solving abilities may form relations within expert vocabulary by relating new words with extant concepts more readily reflecting a wider use of strategies. The relationship between exposure to academic materials and word learning here suggests that students who memorise expert words may lexicalise new written words more efficiently. An alternative is that students who use a wide variety of nonverbal problem solving strategies learn better than students who do not use such learning strategies. These strategies might be more efficient than memorising and reciting new words when learning L2 expert vocabulary. If this assertion is correct, the results suggest that second language teachers could distinguish between conceptually mediated and domain general inhibitory control when planning lessons.

There are limitations to the study. Despite extensive pilot testing, it was quite difficult to titrate expert word learning using a recognition task. Participants were learning new expert words between Phase One and Two but had not learned more words months later at Phase 3. Although writing to dictation improved over a year, this task should have been administered in Phase Two. Unsupervised learning also presented problems with experimental control. It is better that experiments ensure that the targets are not known before training. Using rare, low frequency words and nonwords is one possibility but the learning context then becomes more constrained and less ecologically valid. Presenting written words from a different language family as done here offers some opportunities, e.g., English speakers learning Chinese (Ehrich & Meuter, 2009; Shen, 2013; Shen & Xu, 2015; Wang, Ying &Perfetti, 2004; Xu, Chang & Perfetti, 2014; Yao, 2005). The paradigm developed here also requires further refinement in terms of the time course of lexicalisation with repeated tests of writing to dictation over time.

Executive control had no impact on written word recognition (lexical decision). This may be due to lack of sensitivity in the lexical decision task given that performance was close to ceiling in all phases. It is more surprising however that serial order memory capacity had no impact on writing to dictation given sublexical processes are engaged in English (Weekes, 2006). This may reflect under-reliance on phonological processes to spell words in a second language. That is not to say that serial order memory capacity has no impact on writing to dictation in a second language. Spelling nonwords, for example, will draw on such processes.

Extant vocabulary in L1 and L2 had no impact on written word lexicalisation as was expected for expert words presented in English. Research on spoken vocabulary distinguishes vocabulary size and depth, i.e., how well people know word meanings before learning words (Cobb, 1999) and only vocabulary size was measured here. Nassaji (2006) reports that depth of vocabulary allows metalinguistic inferences that could enhance second language learning. Depth of vocabulary in Cantonese and English could be assessed more fully in future studies.

Hong Kong students are among the most literate in the world (PISA, 2012). Concepts are regularly taught using English technical vocabulary in Hong Kong classrooms. However, all variables tested here explain less than 30% variance in expert written word lexicalisation. There is therefore scope for more research. One issue is gender differences. Girls outperform boys on tests of literacy and this difference is persistent according to PISA (2012) yet also report more foreign language anxiety (FLA) than boys (Liu, Liu & Su, 2010). Future studies could test theoretically motivated hypotheses about interactions between STM processes and a wider range of outcome measures (Kim et al., 2016; Xu & Padilla, 2013; Jubera, 2016).

In sum, STM measures predict learning of written expert words in a second language. However, component processes have differential effects on recognition and recall of words, suggesting written word lexicalisation is an incremental process. Hong Kong undergraduates are sophisticated and regular users of second (and often more languages), reading and writing English on a daily basis. The present results have implications for second language learning and pedagogy in environments where English is the medium of instruction but implications may be limited to Chinese speaking students. However, when multiple languages are spoken and written, a variety of cognitive processes might be used to lexicalise written expert words.

Appendix

The list of words used in the writing to dictation task

Item number	Words	Item number	Words
1.	hoarseness	37.	frication
2.	babble	38.	choking
3.	motherese	39.	cleft
4.	utterance	40.	affricate
5.	repertoire	41.	consonant
6.	articulation	42.	peekaboo
7.	prematurity	43.	echolalia
8.	paraphasia	44.	approximant
9.	generalization	45.	congenital
10.	resonance	46.	discourse
11.	videofluoroscopy	47.	jargon
12.	hypoglossal	48.	receptive
13.	adverbial	49.	polyp
14.	hyperactivity	50.	retrieval
15.	syntax	51.	unilateral
16.	acquisition	52.	vowel
17.	anomia	53.	wernicke
18.	vagus	54.	pragmatic
19.	otoscopy	55.	cue
20.	parkinson	56.	amplification
21.	broca	57.	presbycusis
22.	maneuver	58.	stuttering
23.	drooling	59.	acoustics
24.	narration	60.	semantic
25.	innate	61.	modelling
26.	dysarthria	62.	dysphagia
27.	expressive	63.	bilateral
28.	breathiness	64.	audiogram
29.	aspiration	65.	gliding
30.	harshness	66.	nodule
31.	prosody	67.	milestone
32.	cognition	68.	tympanometry
33.	psychogenic	69.	intelligibility
34.	formant	70.	morpheme
35.	aphasia	71.	intonation
36.	frication	72.	transcranial

References

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation. *Academic Medicine, 68*(*1*), 52–81.

Andrews, S. (2015). Individual differences among skilled readers: The role of lexical quality. In Alexander Pollatsek, Rebecca Treiman (Eds.), *The Oxford Handbook of Reading*, pp. 129–148. Oxford: Oxford University Press.

Barrett, E. A. (2011). *The Effects of Second Language Proficiency and Linguistic Distance on Cognitive Abilities in Bilingual Children*. (PhD Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR. Retrieved from https://doi.org/10.5353/th_b4722941

Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Bilingualism: Cognitive and Language, 522*, 146–155.

Benton, A. L., Varney, N. R., & Hamsher, K. D. (1978). Visuospatial judgment. A clinical test. *Archives of Neurology, 35*(*6*), 364–7.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113*(*2*), 201–233.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*, 627–652.

Burt, J. S. (2006). Spelling in adults: The combined influences of language skills and reading experience. *Journal of Psycholinguistic Research, 35*(*5*), 447–470.

Cheung, H., Kemper, S., & Leung, E. (2000). A phonological account for the cross- language variation in working memory processing. *The Psychological Record, 50*, 373–386.

Chincotta, D., & Hoosain, R. (1995). Reading rate, articulatory suppression and bilingual digit span. *European Journal of Cognitive Psychology, 7*(*2*), 201–211.

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning, 12*(*4*), 345–360

Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. Oxford University Press, New York.

Dearden, J. (2015). *English as a Medium of Instruction – A Growing Global Phenomenon*. The British Council, Oxford.

Dunn, L. M., Dunn, L. M., Whetton, C., & Pintilie, D. (1982). *British Picture Vocabulary Scale*. NFER-Nelson, Windsor.

Ehrich, J. F., & Meuter, R. F. I. (2009). Acquiring an artificial logographic orthography: The beneficial effects of a logographic L1 background and bilinguality. *Journal of Cross-Cultural Psychology, 40*(*5*), 711–745.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(*3*), 340–347.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*, 966–979.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13*, 65–77.

Gathercole, S. E., & Masoura, E. V. (2003). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*(*3-4*), 422–429.

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology, 56*, 1213–1236.

Gupta, P., Lipinski, J., Abbs, B., Lin, P. H., Aktunc, E., Ludden, D., Martin, N., & Newman, R. (2004). Space aliens and nonwords: Stimuli for investigating the learning of novel word-meaning pairs. *Behavior Research Methods, Instruments and Computers, 36*(*4*), 599–603.

Hamilton, A. C., & Martin, R. C. (2005). Dissociation among tasks involving inhibition: A single-case study. *Cognitive, Affective and Behavioral Neuroscience, 5*(*1*), 1–13.

Holmes, V. M., Malone, A. M., & Redenbach, H. (2008), Orthographic processing and visual sequential memory in unexpectedly poor spellers. *Journal of Research in Reading, 31*, 136–156.

Jubera, J. F. (2016). *Working Memory and Phonological Short-term Memory in Second Language Learning*. Universidad de La Rioja, Servicio de Publicacione.

Kaushanskaya, M., & Marian, V. (2009). Bilingualism reduces native-language interference during novel-word learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*(*3*), 829–839.

Lanfranchi, S., & Swanson, H. L. (2005). Short-term memory and working memory in children as a function of language-specific knowledge in English and Spanish. *Learning and Individual Differences, 15*, 299–319.

Leclercq, A. L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development: Evidence from a longitudinal study. *Developmental Psychology, 46*(*2*), 417–427.

Leck, K. J., Weekes, B. S., & Chen, M. J. (1995). Visual and phonological pathways to the lexicon: Evidence from Chinese readers. *Memory and Cognition, 23*(*4*), 468–476.

Liu, L., Liu, M. H., & Su, D. L. (2010). An investigation of Chinese undergraduates' foreign language anxiety: Changes and differences. *Asian Journal of English Language Teaching, 20*, 1–12.

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience, 7*(*357*), 1–16.

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology, 93*, 95–119.

Majerus, S., Poncelet, M., Van der Linden, M., & Weekes, B. (2008a). Lexical learning in bilingual adults: the relative importance of short-term memory for serial order and phonological knowledge. *Cognition, 107*(*2*), 395–419.

Majerus, S., Belayachi, S., De Smedt, B., Leclercq, A. L., Martinez, T., Schmidt, C., Weekes, B., & Maquet, P. (2008b). Neural networks for short-term memory for order differentiate high and low proficiency bilinguals. *NeuroImage, 42*, 1698–1713.

Majerus, S., Heiligenstein, L., Gautherot, N., Poncelet, M., & Van der Linden, M. (2009). The impact of auditory selective attention on verbal short-term memory and vocabulary development. *Journal of Experimental Child Psychology, 103*, 66–86.

Masoura, E. V., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*, 422–429.

McBride-Chang, C., Tong, X.-L., Shu, H., Wong, A.M.-Y., Leung, K.-W., & Tardif, T. (2008). Syllable, phoneme, and tone: Psycholinguistic units in early Chinese and English word recognition. *Scientific Studies of Reading, 12*(*2*), 171–194.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(*1*), 1–10.

Mikan, K. A. M. (2013). *Verbal Short-term Memory and Vocabulary Learning*. Doctoral thesis (PhD), University of Sussex.

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners lexical inferencing strategy use and success. *Modern Language Journal, 90*, 387–401.

Ooi, C.-W. (2016). Phonological Short term Memory and Relative Clause Comprehension in Cantonese Speaking Children with SLI. Doctoral thesis (PhD), University of Sussex.

PISA (2012). http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm

Pishghadam, R., Khajavy, G. H. (2013). Intelligence and metacognition as predictors of foreign language achievement: A structural equation modeling approach. *Learning and Individual Differences, 24*, 176–181.

Raven, J. C. (1984). *Raven's Coloured Progressive Matrices*. London: H. K. Lewis & Co. Ltd.

Ridderinkhof, K. R., van der Molen, W., Band, G., & Bashore, T. R. (1997). Sources of interference from irrelevant information: A developmental study. *Journal of Experimental Child Psychology, 65*, 315–341.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217–236.

Shen, H. H. (2013), Chinese L2 literacy development: Cognitive characteristics, learning strategies, and pedagogical interventions. *Language and Linguistics Compass, 7*, 371–387.

Shen, H. H., & Xu, W. (2015). Active learning: Qualitative inquiries in vocabulary instruction in Chinese L2 classrooms. *Foreign Language Annals, 48*, 82–99.

Stadthagen- Gonzalez, H., Bowers, J. S., & Damian, M. F. (2004). Age of acquisition effects in visual word recognition: Evidence from expert vocabularies. *Cognition, 93*, 11–26.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Thorn, A. S. C., & Gathercole, S. E. (2001). Language differences in short-term memory do not exclusively originate in the process of subvocal rehearsal. *Psychonomic Bulletin and Review, 8*, 357–364.

Tian, L., & Macaro, E. (2012). Comparing the effect of teacher code switching with English-only explanations on the vocabulary acquisition of Chinese university students: A Lexical Focus-on-Form study. *Language Teaching Research, 16*(*3*), 367–391.

Tong, X., & McBride-Chang, C. (2010). Chinese-English bi-scriptal reading: Cognitive component skills across orthographies. *Reading and Writing, 3–4*, 293–310.

Wang, M., Ying, L., & Perfetti, C. A. (2004). The implicit and explicit learning of orthographic structure and function of a new writing system. *Scientific Studies of Reading, 8*(*4*), 357–379.

Weekes, B. S. (1991). *A Cognitive Neuropsychological Investigation of the Orthographic Processing System.* Macquarie University, PhD thesis.

Weekes, B. S. (1993). A test of episodic visual memory. *Perceptual and Motor Skills, 77*(*3*), 1091–1096.

Weekes, B. S. (1994). Spelling skills of lexical readers. *British Journal of Psychology, 85*, 245–257.

Weekes, B. S. (1996). Surface dyslexia and surface dysgraphia: Treatment studies and their theoretical implications. *Cognitive Neuropsychology, 13*(*2*), 277–315.

Weekes, B. S. (2005). *Acquired Dyslexia and Dysgraphia Across Scripts*. IOS press, Amsterdam, The Netherlands

Weekes, B. S. (2006). Deep dysgraphia: Evidence for a summation account of written word production. *Brain and Language, 99*, 30–31.

Weekes, B. S. (2010). Lexical retrieval in alphabetic and non-alphabetic scripts: Evidence from brain imaging. In N. Brunswick (Ed). *The Role of Orthographies in Reading and Spelling*. Hove, UK: Psychology Press.

Weekes, B. S. (2012). Reading and writing across scripts. *Behavioural Neurology, 25*(*3*), 159–163.

Weekes, B. S., Davies, R. A., Parris, B. A., & Robinson, G. A. (2003). Age of acquisition effects on spelling in surface dysgraphia. *Aphasiology, 17*(*6-7*), 563–584.

Weekes, B. S., Castles, A. E., & Davies, R. A. (2006). Effects of consistency and age of acquisition on reading and spelling in developing readers. *Reading and Writing, 19*(*2*), 133–169.

Xu, Y., Chang, L., & Perfetti, C. A. (2014). The effect of radical-based grouping in character learning in Chinese as a foreign language. *The Modern Language Journal, 98*, 773–793.

Xu, X., & Padilla, A. M. (2013). Using meaningful interpretation and chunking to enhance memory: The case of Chinese character learning. *Foreign Language Annals, 46*, 402–422.

Yao, M. [姚君霓]. (2005). *How English Speakers Learn Chinese Characters*. (PhD Thesis). University of Hong Kong. Retrieved from https://doi.org/10.5353/th_b3160179

Yim, O., & Bialystok, E. (2012). Degree of conversational code-switching enhances verbal task switching in Cantonese–English bilinguals. *Bilingualism: Language and Cognition, 15*(*04*), 873–883.