Applied Psycholinguistics **30** (2009), 79–99
Printed in the United States of America
doi:10.1017/S0142716408090048

When study-abroad experience fails to deliver: The internal resources threshold effect

GRETCHEN SUNDERMAN
Florida State University

JUDITH F. KROLL
Pennsylvania State University

Received: August 21, 2007 Accepted for publication: June 17, 2008

ADDRESS FOR CORRESPONDENCE
Gretchen Sunderman, Department of Modern Languages & Linguistics, Florida State University, Tallahassee, FL 32306-1540. E-mail: gsunderman@fsu.edu

ABSTRACT
Some second language (L2) learners return from study-abroad experiences (SAEs) with seemingly no change in their L2 ability. In this study we investigate whether a certain level of internal cognitive resources is necessary in order for individuals to take full advantage of the SAE. Specifically, we examine the role of working memory resources in lexical comprehension and production for learners who had or had not studied abroad. Participants included native English learners of Spanish. Participants completed a translation recognition task and a picture-naming task. The results suggest that individuals who lack a certain threshold of working memory resources are unable to benefit from the study-abroad context in terms of being able to produce accurately in the L2.

Adult second language (L2) acquisition in the United States often takes place primarily in college classrooms, with only brief opportunities to spend time immersed in the L2 in a study-abroad environment. It is commonly assumed that a study-abroad experience (SAE) will enhance the L2 learning process (Freed, 1995). Indeed, many of the individuals who return from such SAEs appear more fluent and confident in their L2. Yet, other individuals, even on the same study-abroad program as their peers, appear to have little change in their L2 abilities. Why might this be the case?

One possible explanation may be found in the internal cognitive resources an individual brings to a SAE. We know that adult L2 acquisition is highly variable and may be constrained by a wide range of individual differences such as age, aptitude, learning style, motivation, and personality (e.g., see reviews by Segalowitz, 1997, and Dörnyei & Skehan, 2003). We also know that processing in the L2 makes greater cognitive demands of an individual in terms of working memory (WM) resources compared to processing in first language (L1; Hasegawa, Carpenter, &

© 2008 Cambridge University Press 0142-7164/08 $15.00

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

Just 2002). Therefore, one factor that may partly explain the differential success of L2 learners could be differences in their ability to process and store information (i.e., WM). WM capacity has been posited as a possible factor affecting both L1 (e.g., Daneman & Carpenter 1980; Just & Carpenter, 1992) and L2 processing (e.g., Harrington & Sawyer, 1992; Kroll, Michael, Tokowicz, & Dufour, 2002; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Miyake & Friedman, 1998; Robinson, 2002; Tokowicz, Michael, & Kroll, 2004), although the results have not been as consistent regarding the relationship between WM and performance in an L2 (e.g., Juffs, 2004).

The absence of a simple benefit following study abroad may lie in the interactions between internal cognitive resources and information available in the external environment. Indeed, a recent set of studies (Collentine & Freed, 2004; Díaz-Campos, 2004; Dewey, 2004; Freed, Segalowitz, & Dewey, 2004; Segalowitz & Freed, 2004) provided rather consistent findings suggesting that no one context of learning (at home, study abroad, or immersion) is "uniformly superior to another for all students, at all levels of language learning, and for all language skills" (Collentine & Freed, 2004, p. 164) and called for the development of a "consistent research agenda that addresses the interaction of individual cognitive abilities and the differential aspects of learning contexts (Collentine & Freed, 2004, p. 165).

Accordingly, in this study we consider whether individual differences in internal resources and experiential factors jointly affect performance in L2 lexical comprehension and production. We focus on lexical processing given the research suggesting important relationships between vocabulary growth and SAE (e.g., DeKeyser, 1991). Specifically, we ask whether the benefit of learning context varies across learners who come to the task of L2 study with different levels of internal WM resources. Additionally, we investigate whether a critical level of internal memory resources is necessary for individuals to take full advantage of the SAE. We first review some of the recent empirical evidence exploring the relationship between cognitive abilities and study abroad learning contexts. We next consider the evidence for the role of WM in lexical processing. Finally, we present our study.

THE EFFECTS OF STUDY-ABROAD LEARNING CONTEXTS

Only a few previous studies have explored the interaction between cognitive abilities and different learning contexts. Segalowitz and Freed (2004) compared the oral performance of L2 learners with varying cognitive processing abilities who were either in a study-abroad context or classroom only. Their study considered how initial pretest performance on a variety of cognitive measures in the L2, including speed and automaticity of lexical access and speed and efficiency of attention control, could predict which learners would be better able to take advantage of different learning contexts. Segalowitz and Freed (2004) hypothesized that the speed and efficiency of L2-specific lexical access and attention control may serve as "readiness factors" for gains in oral ability. In other words, an individual's performance on these L2 cognitive measures may indicate whether they are able to process the input well. Below a certain threshold of readiness,

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

Segalowitz and Freed argue that individuals may not see gains in oral performance. In fact, they suggest that the demands of a study abroad environment for those individuals with cognitive abilities beneath a certain threshold may be overwhelming.

In their study they tested learners before and after their respective language-learning experiences (either in the L2 context or in the classroom in the L1 context.). In the first week, the participants were given a pretest version of an Oral Proficiency Interview (OPI) in Spanish as well as a set of online cognitive tests of lexical access and attention control. Thirteen weeks later after the participants' learning experiences, they repeated the posttest version of the same tasks. The results suggest that overall gains in oral proficiency as measured by the OPI were significantly related to L2 cognitive skills, although a specific threshold was not pinpointed. In terms of oral fluency, as measured by temporal and hesitation phenomena, the learners in the L2 context made greater gains than the classroom learners on some, but not all, oral fluency measures. In some cases, individual differences in cognitive processing abilities were negatively related to gains made in oral fluency, particularly among those who were in the study-abroad context. In other words, in some instances, individuals with higher cognitive processing abilities were less fluent than those individuals with lower cognitive resources, as measured in their L2.

Several studies have emerged as a result of the Segalowitz and Freed (2004) study discussed above. O'Brien, Segalowitz, Freed, and Collentine (2007) investigated the role of phonological memory in L2 fluency at home and abroad. Previous research in the laboratory (Papagno & Vallar, 1995; Papagno, Valentine, & Baddeley, 1991) and in classroom settings (Cheung, 1996; Service, 1992, Service & Kohonen, 1995) suggests that phonological WM correlates with L2 vocabulary acquisition. Those individuals with greater phonological memory capacity acquire words more easily that individuals with a more limited capacity. O'Brien et al. (2007) extend the conclusions from these previous studies to include an important role for phonological memory and L2 oral fluency development. However, they found that the role of phonological memory in language learning is independent of the learning context (i.e., at home or abroad). Their results suggest that those individuals with better phonological memory skills will outperform those with more limited skills in terms of L2 fluency development, regardless of the L2 environment. Along a similar vein, O'Brien, Segalowitz, Collentine, and Freed (2006) found that phonological memory plays a more important role in narrative development at early stages of L2 learning compared to later stages. Indeed, the studies discussed above highlight the complex interaction between what an individual brings to the learning context cognitively and experientially and the contribution of the learning context itself.

Further evidence for a context–learner interaction was reported by Tokowicz et al. (2004) in a study that examined the effects of SAE and WM capacity in the L1 on the types of errors made during single-word translation. They found that learners with high WM and SAE made more meaning-based errors (as opposed to nonresponse errors) when translating into their L2. Tokowicz et al. (2004) argued that learners with increased WM resources were better able to take advantage

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

of communicative strategies when immersed in the study-abroad context. Overall, there is limited evidence regarding the relationship between L1 internal and external variables in L2 processing, particularly in both comprehension and productions tasks.

INTERNAL RESOURCES: WM

WM is understood to be a system for storing and processing information. Several studies have looked at the role of WM in L2 lexical acquisition and development. For example, Daneman and Green (1986) examined the role of WM in vocabulary acquisition. In one of the studies, participants read a passage containing an unknown word and were asked to infer the meaning. The learners with higher memory span were better able to infer meaning from context. In a second study, they found that WM was correlated with the ability to produce sentences and synonyms from given words. The results of these two studies suggested that WM span is highly correlated with a learner's ability to acquire vocabulary and with verbal fluency (the ability to produce L2 words).

Kroll et al. (2002) compared performance on a word-naming task and translation task for a group of beginning language learners studying French or Spanish. All participants completed a memory-span test and were divided into groups according to their relative span score. They found that higher span participants were faster than the lower span participants to translate in both directions. However, this was only true for words that were noncognates. When translating cognates, the higher span learners were slower to translate cognates than the lower span learners. Kroll et al. (2002) argued that the higher span learners were able to use their additional memory resources to avoid being fooled by false cognates, but as a consequence, they suffered a processing cost in this task. In a sense this result parallels Segalowitz and Freed's (2004) findings with respect to oral fluency, such that having higher cognitive resources resulted in less fluent speech. In a word-naming task, Kroll et al. (2002) found a trend toward faster performance for the higher span participants in naming words in L2, but not in L1. Overall, the results of this study suggest that at the level of the lexicon, WM resources affect translation performance and, in particular, the strategies that L2 learners adopt to map new L2 words to their respective meanings.

The results of studies investigating the role of WM in L2 lexical processing suggest that those learners with more available WM resources may be better equipped to process different information simultaneously. At times, this may be helpful for higher memory-span individuals (in terms of increased speed of processing), and at other times it may be appear to be a hindrance (when aspects of form might otherwise facilitate processing). Overall, WM resources represent, at a minimum, an internal cognitive factor that affects language processing in both the L1 and L2.

Why might WM resources be a critical factor in the study abroad context? In an immersion setting, an individual is visually and auditorily surrounded by the L2. To be able to understand and produce in the L2, there must be some attenuation of the L1. One possibility is that the attenuation of L1 occurs naturally, as a consequence of the increased presence of L2. Another alternative is that the individual must

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

"shut off" to a certain extent the dominant L1 (for a review of inhibitory processes in bilinguals, see Meuter, 2005). WM resources then may play a role in allowing an individual to store and process information in the L2, while at the same time suppress the L1.

THE PRESENT STUDY

The goal of the present study was to investigate the role of WM resources in lexical comprehension and production for learners with and without SAE. Based on the limited research exploring the intersection between internal and external factor in lexical processing, we considered four alternative hypotheses. First, the *internal resource hypothesis* proposes a relationship between the internal L1 cognitive resources in terms of memory that individuals possess and their ability to comprehend and produce in the L2. We predict that individuals with greater WM resources will be faster and more accurate to both comprehend and produce in the L2. Second, the *external cue hypothesis* proposes a relationship between the external context of learning (e.g., the study abroad context) and an individual's ability to comprehend and produce in the L2. We predict that individuals with SAE will be able to comprehend and produce more quickly and accurately than those individuals without SAE. Third, the *interaction hypothesis* assumes that learner outcomes reflect a mix of external and internal factors, with individual differences in cognitive resources modulating the influence of external influences. Specifically, this hypothesis assumes that as internal resources increase, an individual will increasingly be able to benefit more from the SAE. Fourth and finally, the *threshold hypothesis* assumes that the benefits afforded by an immersion context during study abroad might not hold for all individuals. Specifically, the hypothesis assumes that learners above an internal resource threshold may be better able to exploit the information in the study abroad context than low span learners and will see an added benefit of SAE. Conversely, learners lacking a sufficient quantity of internal memory resources may not be able to benefit from SAE.

To test the above hypotheses, native English-speaking learners of Spanish with or without SAE performed a series of on-line psycholinguistic tasks measuring their speed of processing and accuracy. Support for the above hypotheses with respect to the benefit provided by internal and external resources would be reflected in faster reaction times (RTs) in milliseconds and higher accuracy scores. A complete description of the experimental method follows.

METHOD

General Procedure

The L2 learners were tested individually in a laboratory and paid for their participation. They completed a set of computer-based tasks measuring WM resources, language comprehension, and language production. Following the on-line tasks, each participant completed a language history questionnaire. Below we describe the participants as well as each measure, the materials used, and the specific procedure and scoring for each measure.

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

Table 1. *Language learner characteristics*

	Learners	
	Without SAE ($N = 34$)	With SAE ($N = 14$)
Age (years)	20.4 (0.8)	20.9 (0.9)
Time studied Spanish		
High school (years)	4.0 (0.7)	3.8 (0.1)
College (semesters)	4.7 (0.9)	4.7 (0.9)
Proficiency rating		
L2	5.9 (1.1)	6.6 (1.0)
L1	9.6 (0.6)	9.4 (0.5)
Time studied abroad (months)	0.0 (0.0)	3.8 (2.4)
Memory span (number of words recalled)	40.3 (11.5)	33.0 (13.5)

Note: Mean self-rating scales range from 1 (*low ability*) to 10 (*high ability*). Standard deviations are in parentheses. SAE, study-abroad experience; L2, second language; L1, first language.

Participants

There were 48 participants in the study. All were native English speakers recruited from Spanish language classes at a large university in the United States. The participants were all Spanish majors who had very similar amounts of classroom time studying Spanish, but some of them had participated in a study-abroad program. They completed a language history questionnaire reporting their L2 experience and rating their proficiency in both English and Spanish on a scale from 1 to 10, from least to most proficient. Table 1 provides a summary of the participant characteristics.

The L2 learners differed only with respect to their experiences studying abroad. Of the 48 participants, 34 had not studied abroad, but 14 had. The average time an individual studied abroad was approximately 3.8 months, with a minimum study abroad time of 2 months and a maximum of 10 months. There were no significant differences between the two groups based on their age, time spent studying Spanish (in either high school or college), or their self-ratings in Spanish or English. As such, and based on these measures, these two groups were similar, except with respect to their SAE. Most of the participants had spent the previous academic semester and/or summer abroad and then were tested in the early fall. Thus, the average time that had elapsed between the SAE and testing was approximately 12.8 weeks ($SD = 3.3$ weeks). All participants also performed a reading span task to assess their WM capacity.

Reading span task

The span task was adapted from a reading span measure designed by Waters and Caplan (1996). Participants viewed a series of 80 sentences on a computer screen and judged the semantic plausibility of the sentences. For example, they would read a sentence such as "The audience laughed at the joke that told the

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

puppet." If the sentence made sense, the participant pressed a button labeled *yes*. If a sentence did not make sense, the participant pressed a button labeled *no*. They were instructed to make the plausibility judgment as quickly and as accurately as possible. The 80 sentences were grouped into 20 sets. There were 4 sets of each of 2, 3, 4, 5, and 6 sentences. Half of the sentences were plausible and half were implausible. The number of words per sentence was matched so that it did not vary significantly across plausible versus implausible sentences or across the sentences in the 5 different set sizes. In addition, Waters and Caplan controlled for a number of characteristics of the words to be recalled; the words were matched on word length, frequency, familiarity, concreteness, and imageability.

After a set of two to six sentences, participants were instructed to recall as many of the last words of the sentences in that set as possible. They had a small booklet that contained a separate page for each set of sentences so that they were unable to see the words they had written in previous sets. There were six lines on each page of the booklet in order to eliminate any cue to how many sentences would be in each set. Each trial in each set proceeded as follows: a fixation point was presented on the screen for 300 ms, followed by a sentence that remained on the screen until the participant made a key press response. Each set ended with a screen with the word *RECALL*. The participants then wrote their responses in the booklet and when they were ready to proceed, they pressed a key to start the next trial. Following the scoring protocol of Kroll et al. (2002), the span score was based on the total number of words the participant could recall only on trials for which the plausibility judgment was correct. If the participant did not judge the sentence correctly but recalled the final word, this word was not included in the total number of items recalled. The span measure could range from 0 to 80.

Reading span results

The learners who had not studied abroad had an overall higher span (recalling on average 40 items) compared to those who had studied abroad (recalling on average 33 items); however, this difference was only marginally significant, $F\ (1, 46) = 3.65$, $p = .06$, and can be seen in Table 1. Overall, in terms of age, time spent studying Spanish, self-ratings in Spanish, and internal resources, the two groups were closely matched. In the discussion we later consider the consequences of other factors, such as motivation and proficiency, which were not controlled in the matching procedure.

Comprehension measure

The comprehension measure consisted of a translation recognition task.[1] In this task, participants were asked to decide if two words were translation equivalents of each other. The words were always presented so that the L2 (Spanish) word appeared first and was followed by the L1 (English) word. For example, the Spanish word "cara," meaning face, might be presented followed by the English word "card." The second word, in this example "card" remained on the screen until the participant pressed either the "yes" or the "no" button. In this exaxmple, participants would be expected to respond "no" because "cara" and "card" are not translation equivalents. They were instructed to make their responses as quickly

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

and accurately as possible. The RT was recorded to the nearest millisecond from the onset of the presentation of the second word. The presentation order of the word pairs was randomized for each participant by the PsyScope program (Cohen, MacWhinney, Flatt, & Provost, 1993). Ten practice trials were given prior to the start of the experiment. The materials for the translation recognition task consisted of 192 word pairs, 96 pairs that were accurate translation equivalents, and 96 pairs that were inaccurate.

Only correct responses were included in the RT analyses; the accuracy analyses included data from all trials. Accuracy was coded as either correct or error. RTs faster than 300 ms or slower than 3000 ms were removed from the analyses and treated as outliers. Means for the yes and the no trials were calculated separately for each participant. Based on these values, RTs that were 2.5 *SD* above or below the participant's mean RT for each of the overall conditions were excluded from the analyses and treated as outliers. Data trimming was done in this way because it is typically thought that extremely fast scores reflect anticipatory processes while extremely slow scores are because of lapses in attention or other processing strategies, and therefore do not reflect the processes of interest (Ratcliff, 1993). Less than 1% of the data were excluded as outliers.

Production measure

The production measure was a picture-naming task.[2] Participants were presented with 120 different pictures taken from the Snodgrass and Vanderwart (1980) picture norms. Only pictures that had at least 80% name agreement in both Spanish and English as determined by Sanfeliu and Fernandez (1996) were used in the experiment. The task was to name the picture as quickly as possible. All participants were tested individually and were given verbal instructions by the experimenter. A practice block of 10 pictures was given prior to the start of the experiment. Participants were presented with the randomized pictures one at a time at the center of the computer screen. A fixation point preceded each picture. The fixation remained on the screen until the participant pressed a button. It was replaced by a line drawing of an object. Participants were instructed to respond by speaking the picture's name aloud. The picture was replaced with a fixation point after registering a response from the participant.

Responses were tape recorded and later coded for accuracy. The speed and accuracy of naming were the dependent measures. Only correct responses were included in the RT analyses; the accuracy analyses included data from all trials. RTs faster than 300 ms or slower than 3000 ms were removed from the analyses and treated as outliers. Based on these values, RTs that were 2.5 *SD* above or below the participant's mean RT for each of the overall conditions were also excluded from the analyses and treated as outliers. Approximately 6% of the data were excluded as outliers.

RESULTS

Descriptive data

The goal of the present study was to examine the effects of external cues (SAE) and internal resources (WM) on comprehension and production performance.

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

Table 2. *Overall performance on comprehension and production tasks*

	Comprehension		Production	
Learners	RT (ms)	Accuracy (%)	RT (ms)	Accuracy (%)
Without SAE ($N = 34$)	980 (297)	87 (14)	1660 (601)	31 (24)
With SAE ($N = 14$)	920 (194)	93 (10)	1489 (436)	50 (26)

Note: Standard deviations are in parentheses. RT, reaction time; SAE, study-abroad experience.

However, as a point of departure, we first present in Table 2 a descriptive summary[3] of the performance of the learners with and without SAE on the comprehension and production task in terms of their RTs and percentage accuracy when responding in the L2 Spanish.

The learners with SAE were faster and more accurate than the learners without SAE in both comprehension and production. In comprehension, the learners with SAE were on average 60 ms faster and 5% points more accurate; in production the same individuals were 171 ms faster and 19% points more accurate. The impact of SAE is more striking in the production performance.

Data analysis

Pooling procedure. Within both the comprehension and the production tasks, there were a number of different experimental conditions. For example, the conditions for the comprehension experiment were type of response (yes/no) and, among the "no" responses, additional conditions included lexical relations (form type 1/form type 2/semantic), grammatical class (matched/unmatched), and relatedness (related/unrelated). The conditions for the production experiment were the cognate status of the picture's name (cognate/noncognate) and the type of block presentation (mixed/pure Spanish). For the purpose of the current study, we pooled participants' performance across all of the experimental conditions in each task (the comprehension and production task separately) and then statistically controlled for those conditions. In other words, we counted participants' performance on any given individual condition as an independent observation. For example, when we combined the data for our 48 participants across each of the combinations of conditions, our final unit of analysis was the participant-condition and resulted in 624 (48 subjects × 13 conditions) and 240 (48 subjects × 5 conditions) observations for comprehension and production, respectively. Pooling the data across each of the experimental condition allowed us to increase the number of observations that we had for each participant and thus increased the statistical power of the present analysis. However, combining data in this manner resulted in each participant's responses not being statistically independent across each of the combinations of conditions. To account for this statistical dependence, we estimated each of the following regression models with robust standard errors clustered within subject (Huber, 1967; White, 1980; Williams, 2000). In other words, we accounted for the fact that these multiple observations were coming

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

from the same individual. In addition, and as we further discuss below, we also included a series of control variables in our regression models to account for the various conditions described above that were in the experiments.

Regression analyses. We used ordinary least squares (OLS) regression to test our hypotheses. We used multivariate regression analyses for several reasons. Regression analysis allowed WM to be treated as a continuous variable. In the past, studies examining WM have split participants into groups of high and low span. Previous research (Waters & Caplan, 2004) has called this practice into question. Moreover, WM is conceptualized as a continuous distribution of cognitive resources that an individual possesses, not necessarily as a discrete resource that an individual either possesses or not. Therefore, we were able to estimate the effect of a one-unit change in span on RTs and accuracy. In other words, we were able to observe how performance for both comprehension and production are affected by increases in WM resources. Most importantly, with respect to the interaction hypothesis, using regression in conjunction with a continuous measure of memory span, we could estimate the precise level at which internal resources may or may not have consequences for performance.

Given that both of our dependent variables (RT and accuracy) were continuous ratio-level variables, we employed standard OLS multivariate regression. First, to test the external cue and internal resources hypotheses, we estimated additive models that included variables for SAE and WM resources. In each of the following tables we refer to these additive models of SAE and WM as Model 1. Second, to test the interaction hypothesis we also estimated multiplicative models that included variables for SAE and WM as well as an interaction variable. In each of the following tables we refer to these multiplicative models as Model 2. For ease of interpretation, we present these models side by side in each of the tables below. In addition, for each of the regression models that we estimated we also included a series of control variables to account for differences in both response RTs and accuracy across the unique conditions of the experiments described above. For the comprehension models, reported in Table 3, we included four control variables and for the production models reported in Table 4, we included two control variables. In each of the tables, the controls that had significant effects on RT and accuracy all were in the expected directions. Third, to test the threshold hypothesis of internal resources we employed the Brambor, Clark, and Golder (2006) technique in which we plot the estimated effect of SAE versus no SAE on RTs and accuracy for a given level of memory span with 95% confidence levels plotted around the relationship to determine whether there is a specific point at which internal resources become a critical factor.

Comprehension

RTs. In Table 3 we first turn our attention to the individual unstandardized coefficients on SAE in Model 1. This analysis shows that the benefit of SAE versus no SAE results in an average increased processing speed of 118 ms. This suggests that SAE has an independent effect (controlling for the presence of other variables) on the speed of comprehension. If we turn to the individual coefficients

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

Table 3. *Ordinary least squares regression models for comprehension*

	Reaction Time				Accuracy			
	Model 1		Model 2		Model 1		Model 2	
Variable	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
SAE	−118.63**	58.40	−272.20*	162.90	6.40***	1.73	9.42*	5.19
WM	−8.09***	2.11	−9.60***	2.76	0.16**	0.07	0.19*	0.10
SAE × WM interaction	—	—	4.32	4.26	—	—	−0.08	0.11
Controls								
Lexical relation	−9.48	13.15	−9.48	13.16	0.39	0.96	0.39	0.96
Grammatical	−11.67		−11.67		−4.69***		−4.69***	
Class		13.17		13.18		0.89		0.89
Relatedness	45.23***	10.26	45.23***	10.27	−5.56***	0.79	−5.56***	0.79
Response	117.89***	21.24	117.89***	21.26	8.85***	1.27	8.85***	1.27
Constant	1184.94***	98.50	1245.99***	124.74	77.76***	3.30	76.56***	4.64
R^2	0.160		0.168		0.163		0.164	
F value (*df*)	11.65 (6, 47)***		10.22 (7, 47)***		31.66 (6, 47)***		29.07 (7, 47)***	
Number of								
Observations	624		624		624		624	
Subjects	48		48		48		48	

Note: Observations are participant-conditions. Standard errors are clustered on participant. SAE, study-abroad experience; WM, working memory.
*$p < .10$. **$p < .05$. ***$p < .01$.

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

Table 4. *Ordinary least squares regression models for production*

	Reaction Time				Accuracy			
	Model 1		Model 2		Model 1		Model 2	
Variable	*B*	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
SAE	−213.96**	99.30	153.22	342.01	21.68***	4.02	0.75	14.54
WM	−5.93	4.38	−2.30	5.62	0.35*	0.20	0.14	0.23
SAE × WM interaction	—	—	−10.34	8.68	—	—	0.59*	0.34
Controls								
Cognate	−19.34	63.57	−19.34	63.70	19.02***	2.43	19.02***	2.43
Mixed	138.97**	56.10	138.97**	56.10	3.90	3.30	3.90	3.30
Constant	1823.03***	192.9	1677.06***	246.49	7.11	8.98	15.43	10.11
R^2	0.050		0.061		0.261		0.277	
F value (*df*)	2.81 (4, 47)**		2.61 (5, 47)**		27.14 (4, 47)***		25.79 (5, 47)***	
Number of								
Observations	240		240		240		240	
Subjects	48		48		48		48	

Note: Observations are participant-conditions. Standard errors are clustered on participant. SAE, study-abroad experience; WM, working memory.
*$p < .10$. **$p < .05$. ***$p < .01$.

on WM in Model 1, we see that an increase in memory resources resulted in faster times as well, with each additional 10-unit increment on the memory span measure leading to an 80 ms increase in processing speed. This suggests that, controlling for the presence of the other variable, internal cognitive resources affected the speed of comprehension. As seen in Table 3, the interaction for SAE and WM was not significant ($p = .38$). Additionally, there was no support for the threshold hypothesis when the Brambor et al. (2006) technique was utilized.

Accuracy. In examining Model 1 in comprehension accuracy, we see a similar pattern; both internal and external resources affect accuracy. With respect to the coefficients, individuals who had studied abroad had an average increased accuracy of approximately six percentage points. With respect to WM, every additional 10-unit increment increase in WM resources increased an individual's accuracy by an average of 1.6 percentage points. These results suggest that each of these internal and external variables have independent effects when controlling for the presence of the other. With respect to the interaction hypothesis, in Model 2 in comprehension accuracy we again see evidence that these two variables do not interact ($p = .44$). In addition, there was no support for the threshold hypothesis when the Brambor et al. (2006) technique was utilized.

In sum, the results of comprehension performance suggest that SAE and WM resources independently affected the speed and accuracy of lexical comprehension. Thus, both the internal resource and external cue hypothesis are supported. However, the interaction hypothesis and the threshold hypothesis are not supported.

Production

RTs. In Table 4 we examine the coefficients in Model 1 but now for performance in the picture naming production task.

With respect to production RTs, Model 1 suggests that only external resources are significant. Those individuals who had studied abroad were 214 ms faster compared to a learner who has not studied abroad. WM resources had no independent effect ($p = .18$). In Model 2 for production RTs, one can see that there is no evidence that the two variables interact ($p = .24$). Moreover, there was no support for the threshold hypothesis when the Brambor et al. (2006) technique was utilized.

Accuracy. In Model 1, the learners with SAE had on average a 21.6 percentage point increase in accuracy compared to those who did not study abroad. The effects of WM were marginally significant and suggest that with each 10-point increment in WM resources an individual gained an average of 3.5 percentage points in accuracy. In Model 2, there is an interaction effect, providing some evidence that one's internal resources may mediate the benefits of SA experience. The coefficient for the interaction term is positive and marginally significant at $p = .09$, suggesting that individuals with higher WM may benefit more from SA experience. However, given the marginal significance of the interaction, we cannot claim that increases in internal resources uniformly interact with SAE.

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

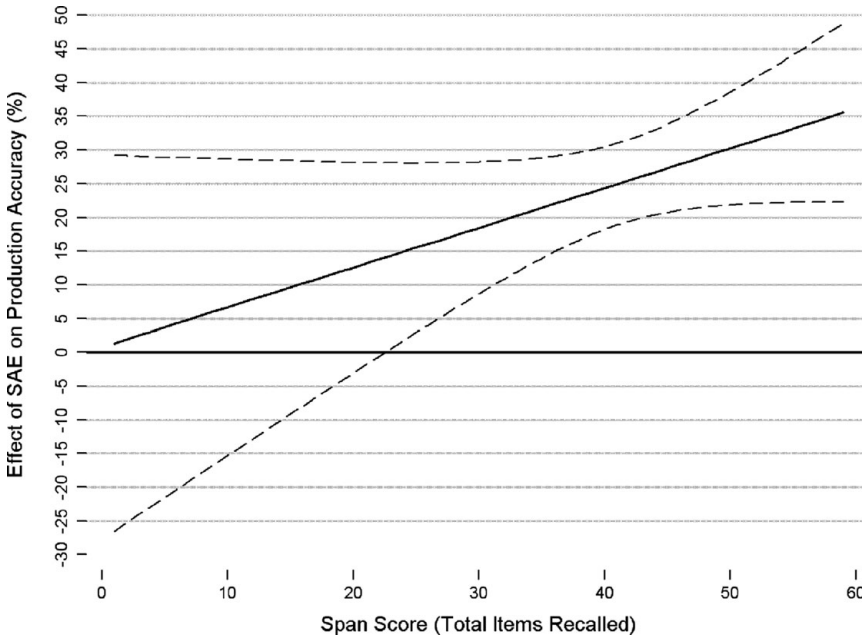


Figure 1. The effects of study-abroad experience (SAE) on working memory resources in second language production accuracy.

There was support, however, for the threshold hypothesis when the Brambor et al. (2006) technique was utilized. In using this technique, we plotted the estimated effect of SAE on production accuracy for a given level of memory span with 95% confidence levels plotted around this relationship.[4] In these figures, when the 95% confidence bands did not straddle zero (the *x* axis) the effect of SAE on production accuracy was significant for a given level of span; when the bands straddled the *x* axis, the effect was essentially zero for that level of span (Brambor et al., 2006). By conducting this test we were able to see at what point the interaction between WM and SAE was significant at the $p < .05$ level. Figure 1 illustrates the effects of SAE on WM resources in L2 production accuracy. Put another way, it displays, on average, the expected change in accuracy for an individual with SAE compared to an individual with no SAE across different levels of memory span. For a complete description of the mathematical equations used to create Figure 1, see Note 4.

Figure 1 shows that when WM scores were below approximately 24 words recalled (the point at which the bands straddle the *x* axis), there was no effect of SAE on production accuracy. This means that among those learners who had sufficiently low internal resources, SAE resulted in an estimated effect on production accuracy that was not different than zero. In other words, it was as if these individuals had not studied abroad. However, above this level of internal resources, the figure reveals a different story. Above a span score of approximately 24 words recalled (where

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

the bands did not straddle the *x* axis) the effect of SAE was statistically significant ($p < .05$) and positively influenced accuracy in production. This figure addresses the question as to how much benefit an individual would get by studying abroad. The answer is that it depends on the individual's memory span. An individual with a memory span of below 24 will not see any benefit in terms of production accuracy. For individuals with memory span scores of 40, for example, they would expect to have a 20 percentage point increase on their production accuracy.[5]

In sum, there was little support for the internal resource hypothesis in the production data. In terms of production, having high internal resources was simply not enough to affect the speed and to some extent the accuracy of lexical processing. The external cue hypothesis however was supported. Those individuals with SAE were faster and more accurate in producing in the L2. The interaction hypothesis was only marginally supported, perhaps suggesting a relationship between internal and external factors. Indeed, we see the exact relationship between those factors in the test for the threshold hypothesis that is supported by the data. Without an adequate level of internal resources, or a minimum threshold of WM, the benefits of SAE, in terms of production accuracy at the word level, are lost on these learners. It is as if these individuals had not spent several additional months receiving input in the study abroad context. It must be noted that the benefits of SAE that we are discussing here are in terms of accuracy in production, and that of course these individuals may have benefited from other aspects of the SAE. Interestingly, only by using regression, and specifically plotting the estimated means, were we able to examine the precise nature of the interaction between cognitive resources and external experience.

DISCUSSION

The goal of the present study was to investigate the relationship between internal cognitive resources and external experience in L2 processing. Specifically, we investigated the role of WM resources and SAE in both lexical comprehension and production. We considered four hypotheses: the internal resource hypothesis, the external cue hypothesis, the interaction hypothesis, and the threshold hypothesis. The *internal resource hypothesis* proposed that the ability to both comprehend and produce quickly and accurately in the L2 is dependent on WM resources, regardless of a learner's external environment. *The external cue hypothesis* proposed that information salient in the L2 environment would provide learners with increased opportunities to more directly process L2 in a way that is not possible in the classroom, thus resulting in faster and more accurate comprehension and production, regardless of a learner's cognitive abilities. The *interaction hypothesis* proposed that learner outcomes reflect a mix of external and internal factors, with individual differences in cognitive resources modulating the influence of external influences. Finally, the *threshold hypothesis* proposed that the benefits afforded by an immersion context during study abroad might not be expected to contribute an additional benefit to all individuals. It was hypothesized that higher internal resources may allow learners to better take advantage of the SAE, whereas a lack of internal resources may explain why some learners return from studying

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

abroad with little effect on their L2 ability. We discuss the results separately for comprehension and production.

In comprehension, the regression analyses suggest that both internal and external resources make independent contributions to processing speed and accuracy. Learners with high span were faster and more accurate than their low span counterparts regardless of learning context. Learners who had studied abroad were faster and more accurate than those individuals who had not studied abroad, regardless of internal resources. The interaction hypothesis in comprehension was not supported, indicating that the effects of WM resources and SAE had independent effects on language processing. The absence of an interaction stands in contrast to recent findings (Linck, 2005) that demonstrate that high span individuals without immersion experience are able to perform as well on comprehension-based measures as low span individuals who are currently immersed in the L2. However, it is important to note that the individuals in the current study were not immersed at the time of testing. Finally, there was no support in the comprehension data for any threshold effects of internal resources.

The pattern of results for the production data was quite different. There was no support for the internal resources hypothesis in terms of processing speed and only marginal support in terms of accuracy. WM resources alone were not sufficient to affect L2 production as they were in comprehension. This result differs from O'Brien et al. (2007) who found that phonological memory affected L2 oral fluency regardless of context. However, it is unclear whether phonological memory measures and WM measures are tapping into distinct or common processes. There was, however, support for the external cue hypothesis both in speed and accuracy of producing in the L2. The conclusions about production are qualified by a marginally significant interaction between internal and external factors. These data suggest that those individuals with greater internal resources were better able to take advantage of the benefits of SAE. These findings parallel those reported by Tokowicz et al. (2004). In both studies, individuals with both greater cognitive resources and SAE were better able to produce, or at least attempt to produce, more accurately in the L2.

Perhaps the most striking result emerged as a consequence of modeling WM as a continuous variable and plotting the effects of SAE on production accuracy for a range of WM. In doing so, we revealed that if an individual did not have sufficient internal resources, then the benefits of SAE were not realized for this individual. This "threshold effect" reflects the level of internal resources that a learner must possess in order to be able to exploit the effects of being immersed in the L2, at least in terms of L2 production accuracy. Although Segalowitz and Freed (2004) hypothesized that a threshold might exist, they did not in fact test this notion explicitly as we have done here. They found that gains in oral production were significantly related to initial L2 cognitive abilities, but did not indicate a threshold, but a more continuous relationship.

In considering the threshold effect, we must consider why someone with limited WM resources would not be able to take advantage of SAE. In a study-abroad context an individual is literally surrounded by the L2. As a consequence of the increased exposure to the L2, the L1 may be naturally, or passively inhibited. In other words, L1 activation may be reduced without requiring the use of internal

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

resources to direct attention selectively to suppress the L1. However, if there is a requirement to actively inhibit the L1, then WM resources could play a more central role. Having higher WM abilities may allow an individual to attend to a greater number of factors simultaneously (i.e., suppressing the L1 while attending to the L2). Thus, only individuals with adequate WM resources may be able to negotiate this required suppression. Perhaps, for those individuals with lower internal cognitive resources managing the cross-linguistic competition in the immersion environment becomes too challenging of a task and the individual simply reverts to speaking L1 or suffers other costs to language processing. In future research this alternative can be evaluated by examining the relative consequences of internal versus external factors for language processing in both the L1 and L2. An interesting hypothesis is that high span learners who are able to exploit immersion learning may suffer greater costs to their L1 than low span learners. Furthermore, it will be of interest to examine the performance of L2 learners enrolled in domestic immersion programs in which they are actively prohibited from using their L1. If the problem for low span learners is to effectively allocate cognitive resources to inhibit the L1, then the prohibition on L1 use in the domestic immersion context may compensate for the lack of internal resources.

As we consider the relationship between WM and SAE, we find that both factors are highly dependent on the nature of the task. In comparing the role of WM and SAE in comprehension and production, we see that WM plays a more critical role in comprehension than in production. In production, SAE was the salient factor affecting L2 performance, not WM. A possible interpretation that must be considered is that SAE is reflecting a gain in proficiency. However, given the threshold effect found in the production accuracy data, we know that proficiency alone cannot be the only explanation. An additional point to consider is why the threshold effect only emerged in the production accuracy data. We argue that being able to accurately select and produce a word in the L2 is by far the more difficult task compared to comprehension. Thus, the effects of SA and WM are most apparent in this condition.

CONCLUSION

The present study investigated whether individual differences in WM resources and SAE may explain the differential success in lexical processing among L2 learners. The results suggest that although both cognitive resources and L2 learning experience affect L2 processing, the interactive effects of both are strongest in L2 production. In particular, data from the production tasks suggest that individuals who lack a certain threshold of resources are unable to benefit from the study abroad context in terms of being able to produce accurately in the L2. Conversely, individuals with greater WM resources were better able to benefit from the SAE. Although the present study investigated the notion of a threshold of cognitive resources necessary to take advantage of an immersion experience, future research will want to model SAE continuously and, using the methods outlined here, determine whether there is a threshold of immersion experience needed to see gains in L2 comprehension and production. Overall, the results of this study provide support for Collentine and Freed's (2004) call that theories of bilingual

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

language processing must take into account the roles of the context of learning and individual differences in cognitive abilities during the L2 acquisition process. Future work examining immersion experiences will enable us to better understand the complex interaction between internal and external variables in second language acquisition.

ACKNOWLEDGMENTS

This research was supported in part by NSF Doctoral Enhancement Grant BCS-0111733 (to G.S. and J.F.K.), NSF Grants BCS-0111734 and BCS-0418071, and NIH Grant R56-HD053146 (to J.F.K.). We thank Rachel Varra and Asha Persaud for research assistance and our colleagues Joe Barcroft, Michael Leeser, Christopher Reenock, Natasha Tokowicz, and the anonymous reviewers for their helpful comments.

NOTES

1. Sunderman and Kroll (2006) include the data reported here together with a larger data set of learner performance on the comprehension measures. That paper does not address SAE or individual differences but rather developmental changes over a greater range of L2 experience than the present paper. It also examines issues of lexical form, meaning, and grammatical class relations in L2 processing that are not the focus of the present report.
2. Sunderman (2002) includes the data reported here together with a larger set of learner performance on the production measure. That paper does not address SAE but instead investigates the role of cognate status and language switching in L2 production.
3. The means reported in Table 2 are unconditional means that do not take into account additional control variables. The means in Tables 3 and 4 are slightly different because they are estimated conditional means (i.e., the estimated mean based upon a regression with controls being held constant.)
4. To test our proposition that a learner's memory span modulates the effect of SAE on production accuracy, we estimate multiplicative regression models in which production accuracy is a function of the interaction between study abroad and memory span. This hypothesis can be specified in an econometric regression model of accuracy as follows:

$$\begin{aligned}\text{production accuracy} = {} & \beta_0 + \beta_1 \text{ study abroad} + \beta_2 \text{ memory span} \\ & + \beta_3 \text{ (study abroad)(memory span)} + \sum \beta_j Z_j + \varepsilon,\end{aligned}$$

where Z_j is a set of exogenous control variables and ε is the disturbance term. This suggests that to evaluate our interaction and threshold hypotheses in a regression context we must estimate the marginal effect of SAE on accuracy and RT. (Note that the term "marginal effect" refers to the standard econometric definition of the change in the dependent variable that follows a unit change in the independent variable of interest, or, in mathematical terms, the first derivative. The word "marginal" in this term in no way connotes "unsubstantial" or "unimportant" but rather the expected change in Y given a 1-unit change in x.) For example, the marginal effect of *SAE* on production accuracy, or the expected change in accuracy for a unit shift from no SAE

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

to SAE is found by taking the first derivative of the above regression equation with respect to SAE. This yields a marginal effect of

$$\frac{d(\text{production accuracy})}{d(\text{SAE})} = \beta_1 + \beta_3 \text{ memory span},$$

where β_1 and β_3 are the estimated regression coefficients. This equation reveals the precise manner in which memory span conditions SAE's effect on production accuracy. For a memory span score of zero the effect of SAE on production accuracy is simply β_1 For increasingly higher memory span scores the effect of SAE on production accuracy is $\beta_1 + \beta_3$ (memory span), with the nature of the effect being determined by the sign and magnitude of the β_3 coefficient. The statistical significance of this marginal effect, however, is determined by its standard error. However, unlike additive models, in multiplicative models, the standard error of the marginal effect also varies across the level of conditioning variable, or in our case, memory span. In the above regression equation, if we denote SAE as SAE and memory span as MEM, then the conditional standard error of the marginal effect of SAE across values of MEM is given as

$$\text{conditional } SE \text{ of SAE} = \sqrt{\text{var}(B_1) + \text{var}(B_3)\text{MEM}^2 + 2\,\text{cov}(B_1 B_3)\text{MEM}},$$

where $\text{var}(B_x)$ represents the variance of the relevant regression coefficient and $\text{cov}(B_1 B_3)$ represents the covariance between the SAE and the interaction term. To determine the statistical significance of the marginal effect of SAE we then proceed as we would in a typical regression context. Except here we plot the marginal effect of SAE on production with its 95% confidence bands plotted around it. The confidence bands are calculated by taking estimated standard error of the marginal effect of SAE for a given level of MEM and multiplying it by 1.96. We then take this resulting width and add and subtract it to the estimated marginal effect. The result is a figure that displays the 95% confidence interval plotted around the marginal effect.

5. An anonymous reviewer pointed out that for span scores between 24 and 40, the difference from zero seems to get larger and larger and inquired whether there was an effect that continues to be in play after the "threshold" of the 24 span score is reached. To investigate this question, we split the file above 24 to see if the interaction between SAE and WM was significant. The logic being that if the interactive relationship continues above the 24 threshold, then we should observe a significant interaction between SAE and WM. If the interaction were not significant, then it would simply suggest a threshold effect. In conducting the additional analyses, we found that above the threshold, there was no significant interaction ($p = .835$). When we ran an additive model with just SAE and WM for those above 24, SAE was significant on its own ($p < .001$) and was positive with a coefficient of 26.2, which means that on average when individuals had a memory span score of higher than 24, they were 26% points higher on production accuracy. Thus, whether someone had a memory span score of 25 or 30 or 40, they all received a similar 26% accuracy boost. The apparent continuation of the SAE in Figure 1 is merely because of random error; SAE does not continue to have an increasingly higher benefit as memory span scores increases beyond the threshold of 24.

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

REFERENCES

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis, 14*, 63–82.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*, 867–873.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

Collentine, J., & Freed, B. F. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*, 227–248.

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

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

De Keyser, R. (1991). The semester overseas: What difference does it make? *ADFL Bulletin, 22*, 42–48.

Dewey, D. P. (2004). A comparison of reading development by learners of Japanese in intensive domestic immersion and study abroad contexts. *Studies in Second Language Acquisition, 26*, 303–327.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition, 26*, 249–273.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Malden, MA: Blackwell.

Freed, B. F. (Ed.). (1995). *Second language acquisition in a study abroad context*. Philadelphia, PA: John Benjamins.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition, 26*, 275–301.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25–38.

Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *NeuroImage, 15*, 647–660.

Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions*. Paper presented at the Fifth Berkeley Symposium on Mathematical Statistics and Probability.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*, 199–226.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18*, 137–171.

Linck, J. A. (2005). *Language and cognitive processing in a study abroad context*. Unpublished thesis, Pennsylvania State University.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181–209). Philadelphia, PA: John Benjamins.

Meuter, R. (2005). Language selection in bilinguals: Mechanisms and processes. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistics approaches* (pp. 349–370). New York: Oxford University Press.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

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

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

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language learning. *Journal of Memory and Language, 30*, 331–347.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology, 48A*, 98–107.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA. In P. Robinson. (Ed.), *Individual differences and instructed language learning* (pp. 211–266). Philadelphia, PA: John Benjamins.

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass–Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, and Computers, 28*, 537–555.

Segalowitz, N. (1997). Individual differences in second language acquisition. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 85–112). Mahwah, NJ: Erlbaum.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*, 173–199.

Service, E. (1992). Phonology, working memory and foreign-language learning. *Quarterly Journal of Experimental Psychology. A Human Experimental Psychology, 45A*, 21–50.

Service, E., & Kohonon, V. (1995). Is the relationship between phonological memory and foreing language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155–172.

Snodgrass, J. G., & Vanderwart, M. (1980). Standardized set of 260 picture-norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Sunderman, G. (2002). *A psycholinguistic investigation of second language lexical processing*. Unpublished doctoral dissertation, Pennsylvania State University.

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form meaning and grammatical class. *Studies in Second Language Acquisition, 28*, 387–422.

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition, 7*, 255–272.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology, 49*, 51–79.

Waters, G. S., & Caplan, D. (2004). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers, 35*, 550–564.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*, 817–838.

Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics, 56*, 645–646.

https://doi.org/10.1017/S0142716408090048 Published online by Cambridge University Press