# Effects of second language proficiency and working memory span on novel language learning

Kaitlyn L. Zavaleta and Janet L. Nicol
University of Arizona

In this paper we examine the effects of prior language learning experience and working memory capacity on learning a novel language. Participants with a range of proficiency in a second language were trained on a third language. They were presented with 20 Turkish words (to test word-learning) and their plural counterparts (to test rule-learning). After training, participants were asked to judge whether a given label correctly matched the pictured object (comprehension) and to verbally produce the correct label for a given object (production). The comprehension test took place immediately after training and again a week later. The production test was administered after the second comprehension test. We found that performance on the initial comprehension test was predicted by working memory span but not L2 proficiency; L2 proficiency predicted learning performance only for the delayed tests. This suggests that the two variables – L2 Proficiency and Working Memory – play different roles in L3 learning.

**Keywords:** language learning, bilingualism, second language acquisition (SLA), proficiency, language production, assessment, working memory

## Introduction

The research we present in this paper was motivated by two widely accepted views that have emerged from recent research on cognitive aspects of second language acquisition: (1) bilinguals are better than monolinguals at learning a new language (e.g., Nation & McLaughlin, 1986; Cenoz & Valencia, 1994; Van Hell & Candia Mahn, 1997; Keshavarz & Astaneh, 2004; Sanz, 2000; Kaushanskaya & Marian, 2009; Abu-Rabia & Sanitsky, 2010; Cenoz, 2013; Lado et al., 2017); and (2) better working memory span facilitates the learning of a new language (e.g., Baddeley, Papagno, & Vallar, 1988; Papagno, Valentine, & Baddeley, 1991; Service, 1992; Abu-

https://doi.org/10.1075/jsls.17025.zav
*Journal of Second Language Studies* 1:1 (2018), pp. 79–105. issn 2542-3835 | e-issn 2542-3843

Rabia, 2001; Speciale, Ellis & Bywater, 2004; Linck &Weiss, 2011; Williams, 2011; Linck et al., 2014). In the present research, we consider working memory span and extent of bilingualism as learner attributes that may each contribute to greater language learning success. In order to explore whether these attributes contribute in different ways to language learning, we make a number of comparisons. We contrast the learning of vocabulary with the learning of a grammatical rule, we measure learning in both immediate tests and delayed tests, and we assess learning via comprehension and production tests.

## Bilingualism and L3 learning

A large number of past studies have shown a bilingual advantage in learning a new language. Many of these studies have focused on how well students operate in a new language in an ongoing language learning context, often in school. For instance, Keshavarz and Astaneh (2004) tested Persian monolinguals and bilinguals who spoke both Persian and another language on a test of English vocabulary within a cloze context. The groups were similar with respect to age, sex, nationality, and level of proficiency. Keshavarz and Astaneh found that bilinguals did better than monolinguals, and that among the bilinguals, performance was better for those who had academic experience with the second language. Thomas (1988) tested both vocabulary and grammar. She compared English-Spanish bilinguals and English monolinguals on their performance on French vocabulary and grammar tests after one semester of French instruction. Bilinguals scored higher than the monolinguals, and bilinguals whose education had included classroom instruction in Spanish performed better on the grammar test than those whose did not. She suggests that bilingualism and classroom instruction lead to greater metalinguistic awareness, which in turn facilitates the learning of new languages. Sanz (2000) tested Catalan-Spanish bilinguals and Spanish monolinguals on their acquisition of English in a school setting. She examined performance on vocabulary and grammar sections of an English proficiency test, both of which revealed better performance by the bilinguals, due to a variety of factors, including enhanced working memory and metalinguistic awareness. Cenoz and Valencia (1994) considered even more global performance. They compared language learning in high school students who were early bilinguals (Basque-Spanish speakers) or Spanish monolinguals. Both groups were tested on their proficiency in a common foreign language, English. Cenoz and Valencia examined students' performance on tests which measured speaking ability, listening and reading comprehension, writing, vocabulary, and grammar skills. They found that bilingualism, in addition to other factors (motivation, intelligence,

exposure to English, and a participant's age) predicted performance on the English tests. They discuss a number of possible reasons for this, including greater metalinguistic awareness, cognitive control, and communicative competence. A study by Abu-Rabia and Sanitsky (2010) focused on orthographic knowledge. They tested two groups of Israeli sixth-graders on their knowledge of English. One group consisted of Hebrew-speaking monolinguals; the other consisted of Russian-Hebrew bilinguals. They found that knowledge of multiple orthographies facilitated learning a third, possibly due to enhanced phonological awareness. Degree of bilingualism may also matter. In a review paper, Cenoz (2013), reports that level of bilingualism has been shown to correlate with third language proficiency (in some domains but not others).

More relevant to the current research are the findings from studies that examine the initial stages of learning a new language. These are largely laboratory studies that control precisely what aspects of a new language are to be learned. For instance, in one study, Van Hell and Candia Mahn (1997) compared "experienced" learners (those with extensive classroom experience of three foreign languages for multiple hours each week) and "inexperienced" learners (monolingual speakers living in a predominantly monolingual region with minimal foreign language experience). The experienced learners were students from a Dutch university, whereas the inexperienced learners were students at an American university. Participants learned word pairs in which an L1 word was paired with a word from the novel language (Spanish for the native Dutch speakers, and Dutch for the native Spanish speakers). Results showed that experienced learners recalled more words and retrieved them more quickly than inexperienced learners. The authors suggest that experienced learners may have better phonological short-term memory and develop more detailed phonological representations.

Nonce words and artificial grammars have also been tested. Kaushanskaya and Marian (2009) taught novel words to two groups of early bilinguals (Spanish-English and Chinese-English bilinguals) and a group of English monolinguals. Forty-eight words were created from an inventory of phonemes consisting of four non-English, non-Spanish phonemes, and four English phonemes. These words were taught as paired-associates with their English translations. Participants were tested on their comprehension of these items via recall and recognition tests in which they were asked to provide the English translation to an item, either without a cue (recall) or with four additional alternatives to choose from (recognition). They were tested immediately after training and one week later. Both groups of bilinguals performed significantly better than the monolinguals. A number of potential explanations for this are considered, including the possibility that bilinguals have enhanced cognitive control, and a phonological system that is more "tolerant" than a monolingual system. In similar work, Antoniou et al. (2015)

trained English monolinguals and Mandarin-English bilinguals (bilingual from birth) on a small set of novel words by presenting pictures paired with auditory labels. The new words contained phonemes that did not occur in either English or Mandarin. Testing followed immediately and involved selecting the picture that matched a spoken label. The bilinguals significantly outperformed the monolinguals. Here, too, the explanation revolves around the possibility of specific cognitive advantages in bilinguals.

In the domain of grammar learning, there are relatively few studies that examine the influence of bilingualism on learning. A study by Nation and McLaughlin (1986) examined grammar acquisition in monolinguals, bilinguals, and multilinguals, using letter-strings generated by two Markov grammars. Participants were administered an implicit learning task and an explicit learning task, and following each one they were asked to make grammaticality judgments on grammatical and ungrammatical strings. Results showed a multilingual advantage, but only for the implicit test. (Performance by the bilinguals was no better than the performance of the monolinguals.) The authors suggest that multilingualism could lead to greater language sensitivity or to a greater engagement with linguistic stimuli. In similar research, Nayak, Hansen, Krueger, and McLaughlin (1990) trained monolinguals and multilinguals on a miniature artificial language. Half of their participants were asked to memorize the items, and half were asked to figure out the rules determining word order. During training, participants were asked to verbalize how they were going about the learning task. Results of a grammaticality judgment task showed that in the rule extraction condition, multilinguals outperformed monolinguals, possibly because they were more flexible in using different learning strategies (as shown by their verbalizations).

By far the majority of the studies described so far tested distinct groups of participants: monolinguals, bilinguals, and sometimes multilinguals. Relatively few have looked within the bilingual/multilingual groups and considered "degree" of bilingualism with respect to proficiency in a second language. One exception is a study by Lado et al. (2017), who grouped their English-speaking participants according to the level of their coursework in university Spanish classes (Beginning, Intermediate, Advanced, and Very Advanced). They then taught participants Latin nouns and then Latin sentences that contained the nouns along with case endings signaling subjecthood and objecthood. In one experiment, training was accompanied by "metalinguistic" feedback; in another experiment, the feedback only indicated whether responses were "correct" or "incorrect". Tests of sentence interpretation (using a sentence-picture matching task) revealed that with metalinguistic feedback, the three higher-level groups all outperformed the beginners on initial testing, but only the Very Advanced group showed an advantage after a two-week delay. With the non-explanatory feedback, the highest profi-

ciency group outperformed the other groups, which all behaved similarly. This research suggests that, at least for the acquisition of case-marking in the L3, less than high proficiency in L2 does not help.

Overall, these studies have shown that bilingualism (and multilingualism) leads to better language learning. Many reasons for this have been advanced, including the possibility that regularly using two languages leads to the enhancement of particular cognitive abilities, better working memory (see below), and other benefits.

But among the so-called "sequential" bilinguals, those who have experience learning a second language explicitly, greater L2 experience may lead to the development of more effective learning strategies (and in turn to better learning of L3 words), and to greater metalinguistic awareness (and hence, possibly to better acquisition of grammatical rules) (see Cenoz, 2013 for further review and discussion). It seems to us that these two benefits ought to be incremental: greater proficiency in L2 should lead to better word learning, and potentially, to better rule learning. The research we present here tests this hypothesis.

## Working memory and learning

The term "working memory" describes a memory buffer in which information can be stored and manipulated. It comes into play during the rehearsal of information (such as a telephone number) and during activities such as producing and comprehending language (Baddeley & Hitch, 1974). Working memory is measured in different ways. Some measures, such as the repetition of increasingly long strings of nonsense words tap into "phonological short-term memory", the ability to maintain sound strings in memory. Others, such as "alphabet span" and "backward digit span", also have a phonological component, but additionally test the ability to both maintain and manipulate information, in this case, holding onto the initial string while mentally rearranging the order of elements. Still others, such as reading or listening span tasks, require the comprehension of sentences and the simultaneous maintenance of information.

One would expect that individual differences in working memory capacity would have an impact on learning, and in the realm of second language learning, this has certainly been attested (Baddeley, Papagno, & Vallar, 1988; Papagno, Valentine, & Baddeley, 1991; Service, 1992; Abu-Rabia, 2001; Speciale, Ellis & Bywater, 2004; Linck & Weiss, 2011; Williams, 2011; and Linck et al., 2014, among many others; though cf. Akumatsu, 2008).

In a study that is similar in some respects to the current research, Martin and Ellis (2012) conducted a word- and grammar-learning study in which participants

had to learn nouns, verbs, and adjectives, along with prefixes that signaled either a dual or plural entity. Participants were also administered three tests of working memory, two of which targeted phonological working memory (nonword repetition and recognition tasks) and one of which probed working memory capacity via a listening span task. There were multiple measures of learning, including spoken translations to English of the novel sentences, and production of novel sentences to describe scenes. Overall, working memory performance correlated significantly with performance on the learning measures. Interesting, they also considered foreign language experience (in terms of number of years studying a foreign language), but found that it did not affect language learning.

Given such findings, it is useful to include a working memory measure in any study of language learning, and we do so here. Another reason to include a working memory measure is that there have been reports of a correlation between working memory and bilingualism, which we discuss just below.

## Bilingualism and working memory

During the last decade, there have been many reports of bilingualism's effects on executive functioning. The term "executive function" refers to a cluster of skills having to do with regulation of behavior, including the allocation of attention, the ability to inhibit task-irrelevant stimuli and responses, and the ability to switch easily from one task to another. Working memory has been implicated in executive functioning, though the relationship between working memory and other components of executive function is not clear.

Many studies have reported that bilingualism and good executive functioning correlate (e.g., Bialystok, Craik, Klein & Viswanathan, 2004; Bialystok, Craik & Ryan, 2006; Costa, Hernández, & Sebastián-Gallés, 2008; Prior & MacWhinney, 2010; Soveri et al., 2011, among many others). One explanation for this is that being bilingual strengthens executive functioning due to the fact that bilinguals need to inhibit one of their languages while speaking to monolinguals, and be able to switch between languages while speaking to bilinguals. (They also need to constantly monitor the language status of their interlocutors.)

But does bilingualism lead to better working memory? The results are mixed. A positive finding was reported by Papagno and Vallar (1995), who examined whether acquisition of translation-equivalents in a new language was affected not only by past language experience, but also by other cognitive skills and abilities. They tested "polyglots" and "non-polyglots" on a series of cognitive and linguistic tasks, including an intelligence test, vocabulary test in the participants' L1, visuospatial tests, and measures of phonological working memory. All participants

had exposure to foreign languages in school. "Polyglots" were defined as individuals self-reporting as fluent in three or four languages; "non-polyglots" were defined as individuals who had acquired only one foreign language in school, and had no additional extensive language experience. During training, high frequency, concrete nouns were paired with their (transliterated) translation equivalents from a novel language (Russian). At test, participants were presented with the first member of the word pair and asked to provide the second word. They found that polyglots outperformed non-polyglots. Polyglots also outperformed non-polyglots in measures of phonological working memory (in both a digit span and nonword repetition test), leading Papagno and Vallar to suggest that polyglots have larger phonological working memory spans, which allow them to be more successful in acquiring a novel language.

Other studies have found differences between monolinguals and bilinguals only for complex working memory tasks that require attentional control or other resources (e.g., Yang, Yang, Ceci, & Wang, 2005, Soveri, Laine, Hamalainen, & Hugdahl, 2011; Stafford, 2011; and Weinbach & Henik, 2012).

Some studies have found no effects at all. In a recent study, Ratiu and Azuma (2015) tested 105 participants – half monolingual English speakers and half English-Spanish bilinguals – on a series of verbal and nonverbal working memory tests, including complex, resource-demanding tasks. In no instance did they find a "bilingual advantage". It important to note that this failure to replicate the finding of a bilingual advantage on executive function tests is not an isolated occurrence. A number of recent papers have suggested that the bilingual advantage in executive function tasks is not as broadly reliable as it once seemed (e.g., Hilchey & Klein, 2011; Paap & Greenberg, 2013; and Gathercole et al., 2014).

## The present study

Our primary research questions are as follows:

1. Does the extent of L2 proficiency predict L3 learning performance? (Specifically, do learners need high proficiency in L2 to show a "bilingual advantage" in L3 acquisition, or is there a "graded" effect?)
2. Does working memory span correlate with L2 proficiency? If not, do working memory span and L2 proficiency have different effects on learning?

We also address two secondary questions:

3. Does good word-learning correlate with good rule-learning?
4. Do different methods of assessment reveal the same pattern of performance?

The first research question is whether the extent of past experience with two languages has an effect on language learning. In the studies cited above, the focus was on early bilinguals (Cenoz & Valencia, 1994; Nation & McLaughlin, 1986; Sanz, 2000), or participants with high levels of proficiency in their L2's (Papagno & Vallar, 1995; Nation & McLaughlin, 1986). The participants in our study are sequential bilinguals with varying degrees of L2 proficiency. They do not use two languages regularly except for a few who are highly proficient in L2. In this sense, they are unlike the early bilinguals in previous studies, who have vast experience managing two languages in everyday life.

The second research question is about the role of working memory in learning a novel language. Previous research has shown that working memory capacity (possibly especially phonological short-term memory) plays a role in learning new vocabulary, especially in the early stages of second language acquisition, when new representations are being created (e.g., Baddeley et al., 1998; Gathercole & Masoura, 2003). In this study, we test working memory span to determine whether it correlates with L2 Proficiency. If the two learner attributes are independent, we can examine whether they play a role in different aspects of learning.

Our third research question has to do with the scope of the "bilingual advantage". Language learning is a complex process: it requires both memorizing many words and grammatical morphemes and the conscious and/or unconscious acquisition of rules and the ability to deploy them during testing. However, most studies of novel language learning focus on either word learning or grammar learning. Some studies, such as that by Cenoz & Valencia (1994), have tested both, but not during the initial stage of learning. The study by Martin and Ellis (2012), described above, examines both, but does not directly test L2 Proficiency.

The fourth research question addresses method of assessment. Most laboratory studies of L3 language learning assess language learning skill using comprehension measures (e.g., Sanz, 2000; Kaushanskaya & Marian, 2009). This is unsurprising, given that comprehension is easy to test and comprehension data can provide a quick and clean measure of language knowledge. However, as anyone with personal experience in a second language knows, it is one thing to comprehend the novel language and another entirely to produce the novel words on demand. This study tests participants' ability to produce the novel words they have acquired, as well as their comprehension. As an additional comparison, comprehension tests were conducted immediately after training and one week later, allowing us to examine ability to remember new items in the short term and also over time.

## Method

### Participants

105 students participated in this study for course credit in an introductory psychology course at a large southwestern U.S. university. All participants were dominant English speakers who had had at least some exposure to either Spanish (93/105) or French (12/105).

### Materials and procedures

Before beginning the experiment, participants completed a language background questionnaire and a proficiency test to measure their experience in a second or third language. The language background questionnaire included a variety of questions, including some that probed whether and in what setting they acquired a second or third language, judgments about their proficiency levels in these languages, and whether they code-switched.

#### *Measures of learner characteristics*

There were two measures.

*Proficiency tests*

For most students, their second language is either Spanish or French. Therefore, two proficiency tests were created that probed proficiency in these languages. Each test consisted of 30 fill-in-the-blank sentences with multiple options for each blank. Sentences were designed to measure knowledge of L2 grammar and language use in an offline task, similar to those used by Marian and colleagues in their research (e.g., Marian, Blumenfeld, & Kaushanskaya, 2007). Knowledge required to fill in the blanks ranged from that a beginning speaker ought to know (i.e., basic conjugation of verbs, as in "Claire et Brigitte ___ à Chicago" (*Claire and Brigitte ____ in Chicago),* with the options "habite, habites, habitant, and habitent" *(live/lives)*), to more specific information that only a more advanced speaker would be likely to know, such as the subjunctive (i.e., "Si ____ riche, je voyagerais beaucoup" *(If ____ rich, I would travel a lot)* with the alternatives consisting of "je" (*I)* and some form of "être" (*to be)).* The sentences included in the proficiency tests covered grammar and content found in the teaching materials used for teaching French and Spanish at the university in which the research was conducted. Questions that required a great deal of experience with L2 would require approximately 3–4 years of college-level language courses, those demanding intermediate knowledge would require approximately 2–3 years of experi-

ence, and those probing only beginning levels of proficiency would require fewer than two years' experience of a language. To ensure comparability across tests, the French test was created first and then translated into Spanish. The sentences were cross-checked with advanced speakers of both languages for (a) accuracy and (b) representativeness of a range of proficiency levels.

*Working memory test*

We used an alphabet span test (Craik, 1986), which requires both the storage of information and the manipulation of remembered items. This test consisted of a series of auditorily-presented words, first in a set of two words (e.g., *test, desk*), and increasing after five trials to a set of three (e.g., *cat, ball, desk*), and so on, up to sets of eight items. Participants were told to remember the words and repeat them back in alphabetical order. They were told that if they forgot a word, they should say, "blank" in the location of the missing word. The test was presented via DMDX software (Forster & Forster, 2003) and responses were recorded via cassette tape for later scoring. Responses were scored according to how many words were produced in the correct order in a given set. For example, in a trial of three items, a participant might be presented with *chair, book, grass*. Each word had to be repeated in the appropriate order: *book, chair, grass*. Thus, if all three items were presented in the correct order, then a score of 3 was given for that item. Participants we instructed to say "blank" to indicate a held space, such as *book, blank, grass.* In this situation, two of the three items are in the correct order, so a score of 2 would be given. All trial scores were summed for each participant. The number of total items across all trials was 175, although no participants earned a perfect score. The test took approximately 10 minutes to complete.

*Language learning task*

The language learning task was divided into three parts: training, immediate testing (comprehension only), and delayed testing (comprehension and production tests). All tasks were presented using DMDX software (Forster & Forster, 2003), which recorded response accuracy.

We trained participants on 20 (picturable) nouns from a new unnamed language (Turkish) with each word given in its singular and plural forms (see Table 2 for a complete list). Turkish was selected due to its similar orthography to English, its linguistic dissimilarity to English, Spanish, and French (the languages with which participants were familiar) and the fact that it has a regular pluralization rule based on vowel harmony. We only selected words that had no cognates in English, French, or Spanish. Pictures were taken from a set normed by Rossion, B. & Pourtois, G. (2004).

In Turkish, a plural is formed by suffixing *-lar* or *-ler* to the singular form of the noun. Generally speaking, if a noun contains a back vowel as its final vowel, the correct ending is *-lar*, otherwise the correct ending is *-ler.* This allowed us to examine whether participants could acquire an aspect of the grammar; i.e., the morphophonological "rule" that dictates the form of the plural inflection. Half the words took the *-lar* suffix and half took the *-ler* suffix.

**Table 1.** Turkish nouns included in language training and testing

| **+ ler Items** | | | **+ lar Items** | | |
|---|---|---|---|---|---|
| English | *Turkish (sg.)* | *Turkish (pl.)* | English | *Turkish (sg.)* | *Turkish (pl.)* |
| butterfly | *kelebek* | *kelebekler* | banana | *muz* | *muzlar* |
| bread | *ekmek* | *ekmekler* | book | *kitap* | *kitaplar* |
| church | *kilise* | *kiliseler* | finger | *parmak* | *parmaklar* |
| cow | *inek* | *inekler* | candle | *mum* | *mumlar* |
| house | *ev* | *evler* | ear | *kulak* | *kulaklar* |
| camel | *deve* | *develer* | girl | *kız* | *kızlar* |
| boat | *tekne* | *tekneler* | woman | *kadın* | *kadınlar* |
| man | *erkek* | *erkekler* | star | *yıldız* | *yıldızlar* |
| window | *pencere* | *pencereler* | chicken | *tavuk* | *tavuklar* |
| dress | *elbise* | *elbiseler* | flag | *bayrak* | *bayraklar* |

The materials contained a mix of monosyllabic and multisyllabic words. The *-lar* words contained a final (or only) vowel that was a back vowel, such as /u/ (written as 'u'), /a/ (written as 'a'), or /ɪ/ (written as 'ı', contrasting with /i/). The *-ler* words contained a final (or only) front vowel, such as /e/ (written as 'e'). In order to keep the novel words as familiar as possible to English speakers, the training stimuli were limited to words containing only these vowels (and in some instances, /i/ as a non-final vowel); vowels such as 'ü' (in IPA as /y/) were not used. For the same reason, all words were pronounced by a native English speaker with a slightly Americanized pronunciation, in order to facilitate acquisition.

*Training*

During training, which lasted approximately 20 minutes, participants saw a picture in the center of the computer screen, saw the Turkish word written below it, and heard it spoken (see Figure 1). After hearing the word, they were encouraged to repeat it in order to remember it. Note that while participants were learning the practice items at the beginning of the experiment, the experimenters strongly encouraged them to repeat the noun out loud to facilitate learning. The entire session was tape-recorded. During data analysis, the experimenters confirmed that

the majority of participants did, in fact, repeat the words as they practiced them. All nouns were presented in the singular form twice and in the plural form twice; no plural form was presented before its singular counterpart.

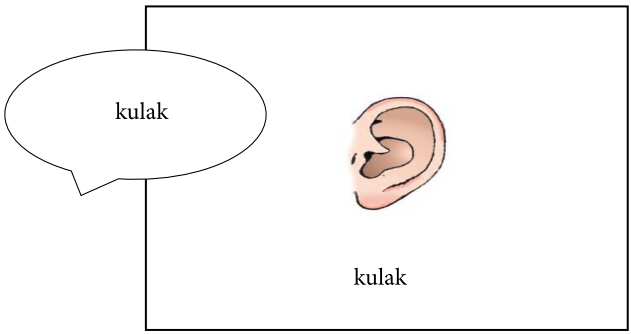


**Figure 1.** Training example for 'ear' (kulak)

*Language learning measures*

There were 4 tests: (1) an immediate comprehension test; (2) an additional comprehension test for most subjects; (3) a delayed comprehension test; and (4) a (delayed) production test. All items presented at test were trained during the training portion of the experiment.

*Immediate comprehension test*

The immediate comprehension test directly followed training and served both as an immediate measure of learning and as a criterion test. During this test, participants were required to judge whether the word written below a picture (the same pictures and words from training) was the correct or incorrect word for that picture (there were no audio presentations). For example, if the word *tekne* 'star' appeared for the picture of an ear (*kulak*), then the participant would press the button indicating "incorrect" (see Figure 2). Participants were given feedback on their accuracy.

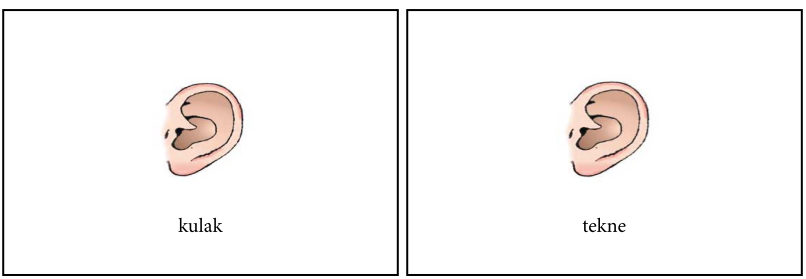


**Figure 2.** Comprehension test example showing a correct and incorrect picture-word pair

The first half of the comprehension test included only singular items (to test word learning) and the second half of the test focused on plural items (to test whether participants had learned the correct ending for each word). For example, if during the plural portion of the test, a picture of three books were paired with the word *kitapler*, then the participant would indicate that this was incorrect, because the correct form is *kitaplar* (see Figure 3). Each picture was presented four times in the singular form (twice paired with the correct word, twice paired with an incorrect word) and was presented four times in the plural form (twice paired with the correct ending, twice paired with an incorrect ending). Accuracy was recorded.

Note that because the test of plural forms included only correct stems, this portion of the test provides additional training of the stem forms. The entire comprehension test lasted approximately 10 minutes.

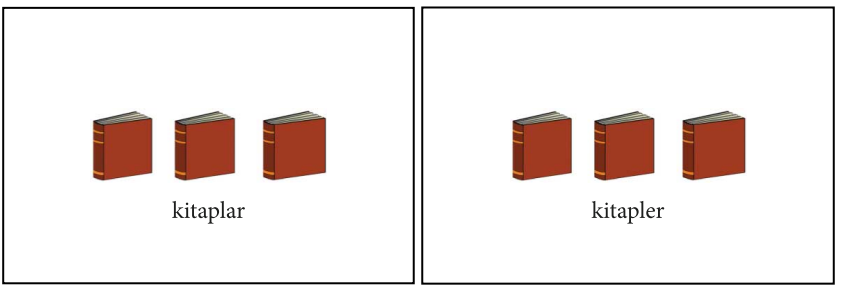


**Figure 3.** Correct (left) and incorrect (right) forms for plural of kitap ('book')

*Additional comprehension test*

Participants who did not reach at least 80% accuracy on the immediate comprehension test completed a second round of training (identical to the first). The purpose of this second round of training was to get as many participants as possible above 80% so that we could then examine retention over time. As such, we do not consider this a measure of learning. However, we do consider the mean number of training experiences to be a measure of language learning. Below, we report participants' ***Final Comprehension*** scores, that is, the scores for the last comprehension test of Session 1: this includes scores from the Immediate Comprehension Test for participants who scored higher than 80% and scores from the Additional Comprehension Test for participants who scored below 80% in the Immediate Comprehension Test.

*Delayed comprehension test*

When participants returned approximately one week later, they were once again administered the comprehension test. This test was identical to the immediate comprehension measure, although no feedback was provided this time, as it was

considered to be an absolute test and not an opportunity for learning. They were told to do their best, as they would not be re-trained on the stimuli from the previous week.

*Production test*

This was completed immediately following the Delayed Comprehension test and provided a different (and more stringent) measure of language learning. Participants had been told at the beginning of the experiment that they would be required to say the words aloud later in the experiment and were encouraged to learn them with that goal in mind. During the Production task, participants were presented with the same pictures from training and asked to say the names out loud. Whereas the Comprehension test was blocked by number (i.e., singular items were presented before all plural items), the Production test included singular and plural items in a pseudo-randomized order. Each picture appeared four times, twice in the singular form and twice in the plural form. Importantly, participants were encouraged to "do their best" when producing the names. In order to reduce potential inhibitions about making errors, they were told that errors were perfectly acceptable and they were allowed to correct themselves if they realized they had made an error. Spoken utterances (for the production task and working memory task) were recorded via cassette tape and later transcribed and scored for accuracy.

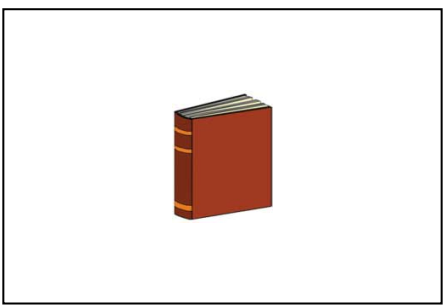

**Figure 4.** Example of stimuli presented during production test for 'kitap'

Utterances were scored according to the accuracy of the responses as labels for the pictured objects. There was some allowance for slight deviations from the target pronunciation; however, the produced word could differ no more than one phoneme from the target (e.g., *yaldız* for *yildız*) to be scored as correct. For plural items, however, the ending did have to be correct in order to be scored as correct. For the image for a single 'book', *kitap* would be correct. The participant would receive a correct score for that trial if they (a) produced the correct label, and (b) did not add a plural ending. For the image of multiple 'books', *kitaplar*

would be correct. A correct score for the trial would be given if the participant (a) produced the correct stem for the noun, (b) correctly added the appropriate ending *-lar* instead of *-ler.* Each trial was given one point for a correct or nearly-correct response and no points for an incorrect response (either due to an incorrect stem, incorrect use of singular or plural form, incorrect pluralization choice, no response at all, or a completely invented response). Scores for the two instances of each test item were summed.

The order of task completion was as follows. All participants completed the language background questionnaire as the first task during the first experimental session. Participants were then trained on the novel items and completed the immediate comprehension test once. If they did not pass criterion (80%) during the Immediate Comprehension test, they repeated both training and the Immediate Comprehension test. Following this, participants completed the working memory test and, if time allowed, the proficiency test. During the second experimental session, participants began with the Delayed Comprehension test, followed by the Production test and, if not already completed, the proficiency test.

## Results

### Measures of learner characteristics

The mean score on the L2 Proficiency tests was 15.37/30 (scores ranged from 6 to 29, and the S.D. was 5.12). The Spanish and French tests showed similar means (15.5 and 14.6), ranges (6–28 and 9–29), and S.D.s (5.2 and 5.4).[1] For the Working Memory test, the mean score was 85 (the range was 20–121; the S.D. was 17.81).

A series of correlational analyses were conducted to examine the relationship between learning scores and L2 Proficiency on one hand, and Working Memory on the other. First, however, we considered whether L2 Proficiency and Working Memory span were correlated. They were not: $r = .1341$, $p = .17$. Below we report the results of correlational analyses between each learner variable and the language learning measures.

1. In order to assess equivalency between the Spanish and French proficiency tests, we also compared the proficiency test results with self-ratings of proficiency. (Not all of our participants provided this information, but 88/93 L2-Spanish speakers and 11/12 L2-French speakers did.) The average self-ratings were 3.6 (on a 5 point scale) for L2-Spanish speakers and 3.4 for L2-French speakers. Average proficiency scores were also similar: the mean was 15.5 for L2 Spanish speakers and 14.5 for French speakers.

## Correlations between learner characteristics and language learning tests

### *Number of trainings*

As mentioned above, participants completed either one or two training sessions based on their score on the immediate comprehension test. If participants completed the first training and received less than 80% accuracy during the immediate comprehension test, then they went through training a second time. Of the 105 participants, 36 had an accuracy of 80% or higher after the first training session, whereas the remaining 69 participants completed a second training session.

Correlational analyses indicate that Number of Trainings correlates negatively with both WM Span and L2 Proficiency; however only the correlation with WM Span is significant ($r = -.254$, $p = .009$).

### *Comprehension tests and production tests*

Below we consider first word learning (the data for the singular items) then rule learning (the data for the plural items).

### *Vocabulary learning*

Table 2 displays descriptive statistics for the three tests of word learning, and the data for the Final Comprehension test. The Final Comprehension test scores are the learning scores that participants obtained either on the Immediate Comprehension test, or on the Comprehension Test that followed a second round of training. Recall that the purpose of the additional training was to maximize learning so that we could more effectively examine retention of vocabulary in memory. Thus, the Final Comprehension scores are not a direct measure of learning *per se*, and are considered only in comparison to the Delayed Comprehension scores; this analysis is discussed after we examine the results of the learning tests.

**Table 2.** Descriptive statistics for the three word learning tests and the Final Comprehension test

| **Mean proportion correct** | | | |
|---|---|---|---|
| **Session 1** | | **Session 2** | |
| **Immediate comprehension** | **Final comprehension** | **Delayed comprehension** | **Production** |
| .84 (*SD* = .08, range .55–.99) | .93 (*SD* = .06, range .61–1.00) | .86 (*SD* = .09, range .48–1.00) | .61 (*SD* = .18, range .15–1.00) |

### *Immediate comprehension test*

A correlation between accuracy on the Immediate Comprehension test and L2 Proficiency scores was not significant ($r = .075$, $p = .446$). Thus, immediately fol-

lowing training, we see no effect of L2 proficiency on learning performance. But Working Memory Span (WM) does correlate with learning ($r = .342$, $p < .001$).

*Delayed comprehension test*
A week later, accuracy scores correlated significantly with L2 Proficiency ($r = .2189$, $p = .0255$), but not WM ($r = .154$, $p = .116$).

*Production test*
There was a significant correlation between L2 Proficiency scores and Production scores ($r = .292$, $p = .003$) and between WM and Production scores ($r = .206$, $p = .035$).

*Rule learning*

**Table 3.** Descriptive statistics for the three rule learning tests (and the Final Comprehension Test)

| **Mean proportion correct** | | | |
|---|---|---|---|
| **Session 1** | | **Session 2** | |
| **Immediate comprehension** | **Final comprehension** | **Delayed comprehension** | **Production** |
| .64 (*SD* = .02, range .36–.97) | .84 (*SD* = .13, range .45–1.00) | .84 (*SD* = .12, range .46–1.00) | .47 (*SD* = .22, range 0–.95) |

*Immediate comprehension test*
Just as for word learning (i.e., learning of singular items), there was no correlation between L2 Proficiency and accuracy scores for plural items ($r = -.146$, $p = .138$) but a significant correlation between WM and accuracy on plural items ($r = .284$, $p < .003$).

*Delayed comprehension test*
Unlike the findings for singular items, accuracy scores on the Delayed Comprehension test show no correlation with L2 Proficiency ($r = .066$, $p = .501$). However, there is a marginal correlation with WM ($r = .167$, $p = .088$).

*Production test*
Here, like the findings for singular items, both learner variables correlate with performance. For L2 Proficiency, $r = .341$, $p < .001$, and for WM, $r = .241$, $p = .013$.

*Session 1 vs. session 2*

One of our goals was to determine whether retention of learned words varied as a function of L2 Proficiency or Working Memory. Because we repeated the comprehension test, we can simply subtract the scores on the Delayed test from the scores

on the Final test to determine each participant's decline in memory over time. We consider only the scores for the singular items because only they differed significantly from one session to the next ($t(104) = 9.58$; $p < .001$).

Correlational analyses revealed a significant correlation between extent of decline and L2 Proficiency ($r = -.201$, $p = .04$), but not WM ($r = -.073$, $p = .461$).

*Singular and plural forms*

Here we consider both the differences in performance on singulars and plurals on the comprehension tests and production test, and correlations between scores for singulars and scores for plurals.

Note that the mean accuracy scores displayed in Tables 2 and 3 show that plural forms are initially harder to learn. Paired t-tests reveal a significant difference between singular and plural items for both the Immediate Comprehension test, $t(104) = 13.2$, $p < .001$, and Final Comprehension test, $t(104) = 6.73$, $p < .001$.

These results suggest that early in learning, at least some participants learn plural forms as whole units. If all the participants had been able to extract the rule for plural formation, the comprehension test for plurals (which presented a word with either a correct or incorrect ending) would have been quite easy; they could do well on such a task even if they had not yet learned the stem form. The fact that the performance on plurals was worse than for singulars suggests that not only had at least some participants not deduced the plural rule, they had not learned the plural variants of the set of (singular) words that they had learned.

Interestingly, by the Delayed Comprehension test, the difference between performance on singulars and plurals was no longer significant ($t(104) = .92$, $p = .357$.) In fact, the performance on plurals showed no decline at all.

For both the Immediate test and the Delayed test, performance on singulars correlated significantly with performance on plurals ($r = .392$, $p < .001$ for the Immediate Comprehension test and $r = .248$, $p = .011$ for the Delayed test). This suggests that, overall, good word learners are good at learning plural forms. But are they better *rule learners*? In order to explore this, for each participant, we subtracted scores for plural items from scores for singular items. A negative score meant that a participant's performance on plural items was better than on singular items, suggesting that the plural rule had been deduced, at least to some extent. A significant correlation between these "rule learning" scores and scores for singular items could mean that good word learners are also good rule learners. On the Immediate test, these did not correlate ($r = -.107$, $p = .276$). However, the Final test did show a significant correlation ($r = -.227$, $p = .02$), and so did the Delayed test ($r = -.428$, $p < .001$), suggesting, perhaps, that participants needed additional exposure to the plural forms in order to extract the rule.

## Discussion

Our findings are summarized in Table 4.

**Table 4.** Result summary table for all language learning measures

| | Language learning measures | | | | | | |
|---|---|---|---|---|---|---|---|
| **Learner variables** | **No. of trainings** | **Immediate comprehension** | | **Delayed comprehension** | | **Production** | |
| | | **Singular** | **Plural** | **Singular** | **Plural** | **Singular** | **Plural** |
| **L2 Proficiency** | | | | ✓ | | ✓ | ✓ |
| **Working Memory** | ✓ | ✓ | ✓ | | | ✓ | ✓ |

Note: Significant correlations between the Learner Variables and Learning Measures are indicated with a checkmark.

We will discuss first the results for vocabulary learning (the singular items), then rule learning (the plural items).

### Word learning

During Session 1, there were two measures of learning (number of trainings, and performance on the Immediate Comprehension Test) and both correlated significantly with WM scores. This is not surprising: working memory underpins the learning process and is crucial in encoding new lexical items.

Interestingly, L2 proficiency did not correlate significantly with either measure of language learning during Session 1. But L2 proficiency did affect retention. This shows up in the analysis of the Delayed Comprehension test, and in the analysis of the decline in comprehension scores (the difference between the Final Comprehension scores and the Delayed Comprehension scores). As raised in the Introduction, this is possibly because participants with greater proficiency in L2 have developed strategies for remembering new words over a longer time interval.

Overall, then, learners with better working memory are better able to maintain new words in a working buffer during initial learning, but then past experience with language learning helps learners consolidate these new words in long term memory so that at the later test date, participants with higher L2 Proficiency perform better on both the comprehension and production tasks.

The idea that working memory effects disappear after the initial learning phase is broadly consistent with a study by Atkins and Baddeley (1998), who had their English-speaking participants learn English-Finnish translation pairs. The

first session included training and tests; the focus was on how much training was required to learn the words. Performance correlated significantly with tests of working memory (digit span and letter span). One week later, participants were asked to recall the words that they had learned. Performance on this task failed to correlate with working memory span.

We found that accuracy scores on the production task correlated significantly with both L2 Proficiency and Working Memory. Note that proficiency in an L2 could be additionally helpful for language production. This is because the act of speaking an L2 requires the development of inhibitory control so that when a speaker intends to speak in L2, L1 is not inadvertently produced. Therefore, the correlation between L2 Proficiency and the production measure could be due both to better consolidation of the L3 and to less interference during production.

But why does production performance also turn out to correlate with WM? There are two plausible explanations for this. One is that, because the production task immediately followed the comprehension test, some participants benefited from being able to maintain some of the word-forms in working memory. But another possibility is that language production in general requires working memory, and speaking in a brand new L2 is demanding enough that a relatively larger working memory capacity is beneficial, and its effects are observable (Green, 1998; Hernandez & Meschyan, 2006; Linck et al., 2014).

## Rule learning

In the comprehension tests of plural items, a picture appeared with the correct stem and either the correct or incorrect suffix. There were two ways to do well on the comprehension tests of plural items. One was to learn the plural forms as whole words; the other was to infer the rule. In the former case, one would predict that the same learner variables that affect learning words in their singular forms would play a role in learning plurals. In the latter case, the prediction is much less straight-forward. Acquiring a rule about the co-occurrence a class of vowels in a word stem and a particular vowel in a suffix could occur implicitly or explicitly, the number of trials needed for rule learning could vary from few to many, and the rule that is acquired could be incomplete.

A relatively larger working memory plausibly helps with this because a generalization needs to be made over specific (remembered) instances, and also because during testing, participants may be invoking their memories of what the pluralized items sounded like during training. Prior language learning experience, on the other hand, is not so clearly helpful: it is not as though a native English-speaker's experience with Spanish or French is likely to result in a strategy for

learning about vowel harmony. That may be why L2 Proficiency does not correlate with comprehension performance on the Immediate Test.

Performance on the Delayed Comprehension test did not correlate significantly with either L2 Proficiency or Working Memory (though there was a trend for the latter). But if some of the participants had learned plurals as whole forms, wouldn't we expect to see some effect of L2 Proficiency, just as we did for the singular forms? To explore this question, we subtracted participants' scores on plural items from their scores on singular items (as discussed in the Results section) and divided the group into two. In one group ($N=47$) were those whose difference scores showed better performance on singulars than plurals (the "plural memorizers"), and in the other group ($N=54$) were those whose difference scores showed better performance on plurals than singulars (the "rule learners") (five participants showed difference scores of zero). For the first group, we found that L2 Proficiency correlated significantly with their scores on the plural items. (Working Memory again showed a marginal effect; this does not have an obvious explanation). For the second group, neither variable correlated significantly with comprehension scores.

For the Production task, there is only one way to do well on the plural items: the stem needs to be produced with a good degree of accuracy (just as in the case of singular words), and the right suffix needs to appended. One would expect performance to correlate with the same learner variables as performance on singular items, and that is what we found.

## Comprehension vs. production

As can be seen in Tables 2 and 3, the pattern of performance for the Delayed Comprehension test and Production test differ: accuracy is higher in the comprehension tests and performance on singulars and plurals is similar, whereas in the Production test, performance on singulars is significantly better than performance on plurals.

At the outset, we expected the performance on the comprehension task to exceed performance on the production task, due to the fact that recognition memory is superior to recall memory (e.g., Vallar and Papagno, 2002). However, it is difficult to assess how much better the comprehension results are, given two facts: (1) that chance performance is 50% in the comprehension tasks, and roughly zero in the production task, and (2) that the comprehension test in this study may *overestimate* learning. Here is our reasoning: in order to correctly reject, for instance, "tekne" as the label for the pictured "ear", a participant only needs to know that the correct label ("kulak") begins with a "k", or contains an "l" or a "u" or an "a". Par-

ticipants' phonological (and orthographic) representations of the Turkish words may be – and quite likely *are*, given the production data – incomplete or otherwise faulty. But due to the way in which the comprehension tests were designed, participants could do quite well even with fragmentary phonological knowledge of the words. (Even our most phonologically similar pairing – 'butterfly' (*kelebek*), labeled with *kadin* ('woman') – would be easy to reject based on partial phonological knowledge). In contrast, participants can only do well in the production task if they have acquired complete phonological representations, or near-complete representations (our scoring protocol allowed for a one-phoneme difference a naming response and the target word).

It is worth noting that the language learning tests in much of the previous research have the same character as our comprehension test. In many of these studies, a newly learned word is presented and participants either provide the L1 translation or select the L1 translation from among a set of choices. Thus, the participants in such studies would not have had to develop full, rich, phonological representations of the new words in order to perform well.[2]

With respect to the finding that there is better performance on singulars than plurals in the Production task but equivalent performance in the Delayed Comprehension task (see Tables 2 and 3), we believe there are a number of explanations. First, in the Comprehension task, participants who have acquired the plural rule can do very well even if they have not learned the stem forms. But in the Production task, that is not possible. Second, the Production task is generally more demanding than the Comprehension task, and this difference may be even greater for the plural forms: the Production task requires the addition of a syllable, so plural forms will always be more demanding to produce than singulars, but the Comprehension task only requires participants to pay attention to the final syllable of the plural form, which may be no more onerous than attending to singular forms.

## Conclusion

The main focus of this study was on the relationship between specific learner characteristics and the ability to learn vocabulary and a grammatical rule in a

**2.** In the typical language learning experiments, one does not see a research scenario in which the set of to-be-learned words contains many phonologically similar words; possibly in part because natural language learning is unlikely to involve clusters of similar-sounding words, and also possibly because global phonological similarity would likely create massive interference, making the learning task too difficult.

novel language. Specifically, we asked (1) whether learners need high proficiency in L2 to show a "bilingual advantage" in L3 acquisition, or whether there might be a "graded" effect, and (2) whether any beneficial effect of L2 proficiency was due to better working memory.

The answer to the first question is that there appears to be a graded effect, particularly with respect to vocabulary acquisition, with L3 language learning gains increasing with increased L2 proficiency. The answer to the second question is a clear "no". We found that working memory span does not correlate with L2 proficiency, and the two variables appear to have different effects. A larger working memory facilitates initial learning, as shown by performance in our Immediate Comprehension Test, but L2 Proficiency is linked to better retention, as shown by performance in the Delayed Comprehension Test. Working Memory span also affects performance in the much more cognitively demanding production task.

In addition, we designed the study to be able to examine two secondary questions. One was whether good word-learning correlates with good rule-learning. Data from the Immediate Comprehension test did not show a correlation, but data from the Final Comprehension test did, suggesting that additional exposure to the stimuli was required for rule-learning. The other question was whether different methods of assessment reveal the same pattern of performance. There are differences in performance, as discussed above. We believe that the Production task is a truer measure of learning new words because it requires that the learner acquire complete phonological representations. But testing the production of plural forms cannot, in itself, distinguish between the learning of whole forms vs. the learning of stem forms plus the application of a rule. Therefore, rule-learning is better tested via a comprehension test, or even better, a generalization test, in which participants would be asked to judge the proper pluralization of untrained words.

## References

Abu-Rabia, S. (2001). Testing the interdependence hypothesis among native adult bilingual Russian-English students. *Journal of Psycholinguistic Research*, 30(4), 437–455. https://doi.org/10.1023/A:1010425825251

Abu-Rabia, S., & Sanitsky, E. (2010). Advantages of bilinguals over monolinguals in learning a third language. *Bilingual Research Journal: The Journal of the National Association for Bilingual Education*, 33(2), 173–199. https://doi.org/10.1080/15235882.2010.502797

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics*, 29, 175–193. https://doi.org/10.1017/S0142716408080089

Antoniou, M., Liang, E., Ettlinger, M. & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition*, 18, 683–695. https://doi.org/10.1017/S1366728914000777

Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, 19, 537–552. https://doi.org/10.1017/S0142716400010353

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

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27(5), 586–595. https://doi.org/10.1016/0749-596X(88)90028-9

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

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19(2), 290–303. https://doi.org/10.1037/0882-7974.19.2.290

Bialystok, E., Craik, F. I. M., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32, 1341–1354.

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *The International Journal of Bilingualism*, 7(1), 71–87. https://doi.org/10.1177/13670069030070010501

Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching, 46*, 71–86. https://doi.org/10.1017/S0261444811000218

Cenoz, J., & Valencia, J. F. (1994). Additive trilingualism: Evidence from the Basque country. *Applied Psycholinguistics*, 15, 195–207. https://doi.org/10.1017/S0142716400005324

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, 106(1), 59–86. https://doi.org/10.1016/j.cognition.2006.12.013

Craik, F. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive abilities* (pp. 409–422). Amsterdam: Elsevier.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, 35, 116–124. https://doi.org/10.3758/BF03195503

Gathercole, S., & Masoura, E. (2003). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory*, 12(3–4), 422–429.

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas-Guasch, N., Roberts, E. J., Hughes, E. K., & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, 5(11), 1–14.

Green, D. W. (1998). Mental control of the bilingual lexicosemantic system. *Bilingualism: Language and Cognition*, 1, 67–81. https://doi.org/10.1017/S1366728998000133

Hernandez, A. E., & Meschyan, G. (2006). Executive function is necessary to enhance lexical processing in a less proficient L2: Evidence from fMRI during picture naming. *Bilingualism: Language and cognition*, 9(2), 177–188. https://doi.org/10.1017/S1366728906002525

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, 18(4), 625–658. https://doi.org/10.3758/s13423-011-0116-7

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, 16(4), 705–710. https://doi.org/10.3758/PBR.16.4.705

Keshavarz, M. H. & Astaneh, H. (2004). The impact of bilinguality on the learning of English vocabulary as a foreign language (L3). *International Journal of Bilingual Education and Bilingualism*, 7(4), 295–302. https://doi.org/10.1080/13670050408667814

Lado, B., Bowden, H. W., Stafford, C., & Sanz, C. (2017). Two birds, one stone, or how learning a foreign language makes you a better language learner. *Hispania*, 100(3), 361–378. https://doi.org/10.1353/hpn.2017.0064

Linck, J. A., & Weiss, D. J. (2011). Working memory predicts the acquisition of explicit L2 knowledge. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 101–114). Washington, DC: Georgetown University Press.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

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

Nation, R., & McLaughlin, B. (1986). Novices and experts: An information processing approach to the "good language learner" problem. *Applied Psycholinguistics*, 7, 41–56. https://doi.org/10.1017/S0142716400007177

Nayak, N., Hansen, N., Krueger, N., & McLaughlin, B. (1990). Language-learning strategies in monolingual and multilingual adults. *Language learning*, 40(2), 221–244. https://doi.org/10.1111/j.1467-1770.1990.tb01334.x

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive psychology*, 66(2), 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 48(1), 98–107. https://doi.org/10.1080/14640749508401378

Papagno, G., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, 30, 331–347. https://doi.org/10.1016/0749-596X(91)90040-Q

Prior, M., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13(2), 253–262. https://doi.org/10.1017/S1366728909990526

Ratiu, I. & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology*, 27(1), 1–11. https://doi.org/10.1080/20445911.2014.976226

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. https://doi.org/10.1068/p5117

Sanz, C. (2000). Bilingual education enhances third language acquisition: Evidence from Catalonia. *Applied Psycholinguistics*, 21, 23–44. https://doi.org/10.1017/S0142716400001028

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 45(1), 21–50. https://doi.org/10.1080/14640749208401314

Soveri, A., Laine, M., Hamalainen, H., & Hugdahl, K. (2011). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism: Language and Cognition*, 14(3), 371–378. https://doi.org/10.1017/S1366728910000118

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

Stafford, C. (2011). Bilingualism and enhanced attention in early adulthood. *International Journal of Bilingual Education and Bilingualism*, 14(1), 1–22. https://doi.org/10.1080/13670050903568209

Thomas, J. (1988). The role played by metalinguistic awareness in second and third language learning. *Journal of Multilingual and Multicultural Development*, 9(3), 235–246. https://doi.org/10.1080/01434632.1988.9994334

Vallar, G. & Papagno, C. (2002). Neuropsychological impairments of verbal short-term memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), *The handbook of memory disorders* (2nd ed.; pp. 249–270). Chichester, UK: Wiley & Sons.

Van Hell, J. G., & Candia Mahn, A. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experience and inexperienced learners. *Language learning*, 47(3), 507–546. https://doi.org/10.1111/0023-8333.00018

Weinbach, N., & Henik, A. (2012). Temporal orienting and alerting – same or different? *Frontiers in Psychology*, 3, 1–3. https://doi.org/10.3389/fpsyg.2012.00236

Williams, J. N. (2011). Working memory and SLA. In S. M. Gass & A. Mackey (Eds.), *The handbook of second language acquisition* (pp. 427–441). New York, NY: Routledge.

Yang, H., Yang, S., Ceci, S., & Wang, Q. (2005). Effects of bilinguals' controlled-attention on working memory and recognition. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (pp. 2401–2404). Somerville, MA: Cascadilla Press.

## Address for correspondence

Janet L. Nicol
Departments of Linguistics and Psychology
Program in Cognitive Science
University of Arizona
1503 E. University Blvd.
Tuscon AZ 85721
USA
nicol@email.arizona.edu

## Co-author information

Kaitlyn L. Zavaleta
Department of Psychology
University of Arizona
klzavaleta@gmail.com