The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL

Simón Ruiz[1]
Patrick Rebuschat[1,2]
Detmar Meurers[1,3]

[1] LEAD Graduate School & Research Network, University of Tübingen, Tübingen, Germany
[2] Department of Linguistics and English Language, Lancaster University, Lancaster, United Kingdom
[3] Department of Linguistics, University of Tübingen, Tübingen, Germany

Correspondence concerning this article should be addressed to Simón Ruiz, University of Tübingen, English Department, Wilhelmstraße 50, 72074 Tübingen, Germany.
E-mail: simon.ruiz-hernandez@sfs.uni-tuebingen.de

The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL

Abstract

The extent to which learners benefit from instruction may be largely dependent on their individual abilities. However, there is relatively little work on the interaction between instructional effectiveness in second language learning and learner individual factors. In this study, we investigated the relationship between instruction, individual differences in cognitive abilities (working memory and declarative memory), and second language vocabulary acquisition in the context of web-based intelligent computer assisted language learning (ICALL). To this end, 127 adult learners of English, predominantly advanced-level, German-speaking learners, read news texts on the web for about two weeks using an ICALL system under two instructional conditions: form-focused and meaning-focused instruction. Learners in the form-focused condition read and completed automatically-generated multiple-choice gaps where phrasal verbs appeared in the text, while learners in the meaning-focused condition simply read and did not complete any gaps. Mixed-effects regression analyses showed that working memory was associated with vocabulary acquisition and that this association depended on the instructional context, with working memory being predictive of learning only in the form-focused condition, suggesting an aptitude-treatment interaction. Furthermore, declarative memory abilities were related to learning only as measured by the Continuous Visual Memory Task, and the relationship was not moderated by instructional condition. Overall, the study contributes to accounting for variability in second language learning in general, as well as in instructional contexts supported by intelligent CALL.

I Introduction

Substantial empirical evidence confirms the facilitative effects of instruction on second language (L2) acquisition (Goo et al., 2015; Long, 1983; Norris and Ortega, 2000; Spada and Tomita, 2010; Sok et al., 2018), for example, by increasing learning rate and pushing ultimate attainment (Kang, et al., 2018, for a recent meta-analysis). Instruction has been broadly categorized into meaning-focused and form-focused instruction (Ellis, 2012; 2016; Norris and Ortega, 2000). Meaning-focused instruction refers to any type of intervention that requires learners to attend to the content of communication. Learning in this case is primarily the result of both exposure to rich and comprehensible input and to interaction with teachers or peers (Ellis, 2012). In contrast, form-focused instruction is a type of intervention that seeks to draw learners' attention to linguistic form (Ellis, 2016). The rationale is that mere exposure to input may be insufficient for acquisition to occur (Long, 1991), and form-focused instruction might be necessary for the acquisition of at least some aspects of the L2 (Robinson et al., 2012). The benefits of instruction may, however, depend on learners' individual differences, such as their cognitive ability to maintain and manipulate information simultaneously (Baddeley, 2017), or their motivation for learning (Csizér, 2017).

Individual differences are regarded as critical factors in acquiring and processing language (Kidd et al., 2018), and they moderate and/or mediate[1] the effects of instruction (Li, 2017a). In second or foreign language (L2) learning, they have been shown to have great explanatory power in accounting for learning outcomes (Pawlak, 2017). However, despite extensive research on the potential effects of individual differences on L2 learning (see Li, 2017a; Pawlak, 2017, for recent reviews), there are still relatively few studies that have explored the interaction between individual differences and instructional conditions or treatments (e.g., Ando et

---

[1] One reviewer suggested we clarify the distinction between *moderation* and *mediation*. A *moderator* refers to a variable that modifies the magnitude of the relationship between variables, whereas a *mediator* refers to a variable that explains the relationship between variables (Baron and Kenny, 1986). Or, to put it differently, moderation is seen as *an interaction effect* between two variables (e.g., "For which group of learners a given instructional treatment was effective?"), while mediation is considered as an *indirect effect* on the outcome variable through a mediating or intervening variable (e.g., "Does anxiety mediate the relationship between instruction and learning?") In this study, we addressed moderation, or interaction effects.

al., 1992; Tagarelli et al., 2015; 2016; Ruiz et al., 2018; Benson and DeKeyser, 2018; Faretta-Stutenberg and Morgan-Short, 2018; Indrarathne and Kormos, 2018; Malone, 2018; Suzuki and DeKeyser, 2017; see Granena and Yilmaz, 2018; Vatz et al., 2013, for reviews). This kind of interactional research (Li, 2017a) examines the extent to which pedagogical treatments are more or less effective depending on learners' personal characteristics or aptitudes, such as personality (Jackson, 2019), learning styles (Grey et al., 2015), language aptitude (Li et al., 2019), working memory (Indrarathne and Kormos, 2018), and declarative memory or procedural memory (Faretta-Stutenberg and Morgan-Short, 2018), to name but a few. Examining these aptitude-treatment interactions (Cronbach and Snow, 1977) will potentially inform the design of pedagogical interventions that can suit learners' characteristics, thereby optimizing L2 teaching and learning. Crucially, aptitude-treatment interaction research may also provide insights into the nature of L2 learning processes by ascertaining the possible facilitative or hampering role of certain cognitive abilities in the success or failure of specific instructional situations (DeKeyser, 2012).

In the current study, we aimed to expand this important line of inquiry by investigating the moderating role of two cognitive individual differences, namely, working memory and declarative memory, on L2 vocabulary acquisition, and its potential interaction with instructional conditions (meaning-focused versus form-focused) in a web-based intelligent computer-assisted language learning (ICALL) context. Working memory and declarative memory are deemed crucial sources of individual variation in L2 learning (Hamrick et al., 2018; Li, 2017a). Intelligent CALL, or ICALL, can be understood as the use of natural language processing (NLP) technology in computer-assisted language learning (CALL). NLP, a branch of artificial intelligence, is considered to be the applied side of computational linguistics that supports a potentially rich interface with language teaching and learning (Meurers and Dickinson, 2017; Lu, 2018). Given the relatively lack of L2 input in classroom contexts (European Commission Eurydice Report, 2017; Suzuki et al., 2019), web-based ICALL systems can provide learners with much needed exposure as well as access to authentic materials outside of the classroom. Here, a web-based ICALL platform, an adapted version of the WERTi system (Meurers et al., 2010), was employed to carry out an entire L2 experiment on the web.

I Working memory and L2 learning

Working memory is an attentional limited-capacity system that supports complex cognitive processing (Cowan, 2017). Although several models have been proposed to account for the functioning of working memory (e.g., Baddeley & Hitch, 1974; Cowan, 2010; Engle, 2018; see Adams et al. 2018, for a review), in this study we follow Baddeley and Hitch's model (1974, see also Baddeley, 2000), as it has served as the most influential theoretical framework used to examine the relationship between L2 learning and working memory. According to Baddeley and colleagues (e.g., 2017), working memory is a system that comprises storage subsystems responsible for temporarily holding and processing both verbal and visual-spatial information (i.e., the phonological loop and the visuospatial sketchpad, respectively); a domain-general component responsible for controlling and regulating attention (i.e., the central executive); and an episodic buffer that serves as a link between the storage subsystems and long-term memory. Depending on whether the emphasis is on examining the storage capacity (the phonological loop, in particular), or both storage and processing aspects of working memory (the phonological loop plus the central executive), a distinction is generally made between *phonological short-term memory* and *complex working memory*, respectively (Li, 2017a, p. 398). In this study, we focus on complex (executive) working memory (Wen, 2015), and thus we operationalize it as the ability to store and manipulate immediate information even in the face of possible interference from other cognitive activities (Williams, 2012). This operationalization is in line with both more recent contemporary views that regard executive functions (e.g., attention regulation) as the primary source of individual differences in working memory (e.g., Engle, 2018; Linck & Weiss, 2015), as well as the more general-domain (i.e., the central executive) view of working memory adopted by more recent second language acquisition (SLA) research on the topic (e.g., Doughty, 2019; Faretta-Stutenberg & Morgan-Short, 2018; Li et al., 2019; Suzuki & DeKeyser, 2017). For this reason, in the current work we used a complex span task (an adaptation of the operation span task [OSpan]; Unsworth et al., 2005) (see Methodology) to measure processing (attention) and storage (memory) capacity (i.e., *working memory capacity*) in learners.

Working memory is involved in many cognitive processes key to L2 learning, such as attention control, analogical reasoning, explicit deduction, information retrieval and decision making (Tagarelli et al., 2015; 2016), as well as the retention of metalinguistic information as learners comprehend and produce L2 language (Roehr, 2008). It is not surprising, therefore, that there have been calls to incorporate it as one of the components of language aptitude (e.g., Doughty, 2019; Robinson, 2005; Wen, 2019). The position assumed here is that working memory is an important factor in understanding individual differences in L2 learning, and as such, can be considered a crucial underlying component of language aptitude. In this sense, the limited nature of working memory capacity acts as a bottleneck that controls what information becomes part of subsequent long-term storage (i.e., learning) (Juff and Harrington, 2011), and thus individual differences in working memory

capacity can be predictive of success or failure in L2 acquisition (Linck et al., 2014). For instance, it has been reported that having more working memory capacity helps learners deal more successfully with learning tasks that are attention-demanding, as they require simultaneously processing of meaning, form, and use of language forms, as it is the case in form-focused instruction (Doughty, 2001; Li, 2017a). Likewise, complex working memory has also been found to be facilitative of learning that is conscious, intentional, and explicit (Tagarelli et al., 2015; Indrarathne & Kormos, 2018), as explicit processes are strongly associated with attention (Robinson et al., 2012).

On the other hand, working memory can also hinder L2 learning if, for example, instruction overloads learners' working memory resources, especially for learners with smaller working memory capacity (Biedroń & Pawlak, 2016). For instance, communicative language instruction, where the emphasis is on developing aural and oral skills, can be unfavorable for low working memory learners (Juff and Harrington, 2011). Similarly, the effectiveness of different instructional treatments (e.g., input enhancement, input flood), including corrective feedback, can also be dependent on learners' working memory abilities (Biedroń & Pawlak, 2016). To illustrate, implicit corrective feedback via recasts (i.e., reformulation of the learner's utterances) may be less suitable for learners with low working memory capacity, as they need to *notice* the recast and compare it to the erroneous language form they have just produced while also carrying out a language task at the same time; for learners with more working memory capacity, this type of instruction can be more favorable, as they are more cognitively adept to allocate and shift attention to form and meaning simultaneously when performing learning tasks (Li, 2017b). In this regard, however, despite the vast amount of research on the role of working memory in L2 acquisition (see Linck et al. 2014, for a meta-analysis), there is still a relative paucity of research on how working memory capacity interacts with different types of L2 instruction (i.e., aptitude-treatment interactions), especially with regard to form-focused instruction (e.g., Ando et al., 1992; Tagarelli et al., 2015; Erlam 2005; Faretta-Stutenberg and Morgan-Short, 2018; Goo, 2012; Indrarathne and Kormos, 2018, Lado, 2017; Li et al., 2019; Li and Roshan, 2019; Malone, 2018; Suzuki and DeKeyser, 2017; Yang et al., 2017).

In an often-cited study on how individual differences in working memory interact with different types of instruction, Ando et al. (1992) detected an aptitude-treatment interaction when investigating L2 grammar (basic English) learning under different instructional conditions, namely, meaning-focused and form-focused instruction (as in the current study). In a classroom-based context with Japanese children, learners in the meaning-focused condition engaged in implicit communicatively-oriented instruction (i.e., communicative tasks), whereas learners in the form-focused instruction received explicit grammar teaching (i.e., learning and applying abstract rules). To measure learners' working memory capacity, the researchers used complex span tasks (both listening and reading span tests). Ando et al. found that high working memory learners benefited more from form-focused instruction, whereas low working memory learners benefited more from meaning-focused instruction, suggesting the interaction between individual differences in working memory and the effectiveness of L2 instruction in more explicit instructional conditions.[2]

In a more recent study, Indrarathne and Kormos (2018) also examined the interplay between working memory and instructional treatments. More specifically, intermediate learners of English received instruction on L2 grammar (causative passive) under implicit and explicit instructional conditions. In the implicit conditions, learners read a story with either the target feature textually enhanced in bold font (Implicit 1) or not (Implicit 2); neither group received information about the targeted form. In the explicit conditions, learners read the textually enhanced story; one group additionally received an explicit metalinguistic explanation (Explicit 1), whereas the other group was further instructed to pay attention to the textually enhanced parts (Explicit 2). Both the storage capacity and the attention regulation functions of working memory were assessed using four different tasks (i.e., the forward Digit-Span test, the Keep Track task, the PlusMinus task, and the Stroop task), and then a composite score was calculated in a regression analysis to explore how working memory interacted differently with learning under the four different instructional conditions (i.e., aptitude-treatment interactions). Results showed that working memory was associated with learning in all four conditions, with a significant stronger association in the more explicit conditions. Furthermore, learners with higher working memory abilities performed better overall than did learners with less working memory capacity.

Malone (2018) also investigated the relationship between working memory and instructed language learning. In particular, the study focused on L2 vocabulary learning, a very under-researched area within aptitude-treatment studies in SLA (see Granena and Yilmaz, 2018, for a recent review), and more importantly, the object of study in the present work. Intermediate learners of English read stories containing low frequency words and received aural support or not while reading the stories (i.e., listening-while-reading). Tests of both

---

[2] In explicit instructional conditions, learners are encouraged to pay attention to certain linguistic forms in the input, and/or they receive metalinguistic information (i.e., rules or terminology) either deductively or inductively, and thus promoting intentional learning. By contrast, in implicit instructional conditions, learners receive no instructions to attend to particular language forms, nor do they receive any metalinguistic information (Graneda and Yilmaz, 2018).

phonological short-short term memory (i.e., the nonword span task) and complex working memory (the OSpan and the Shapebuilder memory task) were used. To examine aptitude-treatment interaction effects, however, only combined nonword span and Ospan scores were employed to compute a composite score. Learners were assessed on their knowledge of word form and word meaning. The results showed that working memory was predictive of learning outcomes, specifically, word form recognition, only for learners who had received aural input enhancement (Sharwood-Smith, 1991; 1993), indicating the presence of an aptitude-treatment interaction effect (Cronbach and Snow, 1977).

In sum, while considerable work has documented the role of individual differences in working memory on L2 learning, there is still a relative lack of research on how learners' individual differences in (complex) working memory interact with the kind of instruction they receive during instructed language learning, particularly, studies examining the acquisition of L2 lexical knowledge (Li, 2017a). To address this research gap in the literature, the experiment reported here extends previous research by assessing the influence of working memory on L2 lexical development, specifically, English phrasal verbs, and the possible interaction with two different instructional conditions, namely, meaning-focused versus form-focused, in a web-based ICALL setting.

2 Declarative memory and L2 learning

Declarative memory is increasingly being investigated as another individual difference factor that may affect L2 acquisition (Ullman, 2015; 2016). Declarative memory refers to the capacity to consciously remember and learn information. As a general-purpose learning system, the declarative memory system is involved in the learning and use of knowledge about facts and events, i.e., semantic and episodic knowledge (Squire, 2004). Learning occurring in this long-term memory system is attention-demanding, intentional, quick, and mostly explicit (Ullman and Lovelett, 2018).

Declarative memory is implicated in both first and second language acquisition (Ullman; 2015, 2016). In L2, declarative memory underpins the learning, storage and processing of both grammar and vocabulary. Regarding L2 grammar, declarative memory has been found to have a facilitative role, at least in the earliest phases of acquisition (Hamrick, 2015; Morgan-Short et al., 2014), as rule learning is initially learned in the declarative memory system (Ullman and Lovelett, 2018). As for L2 vocabulary learning, declarative memory is even more crucial and beneficial, as it is the memory system that specializes in rapid learning of arbitrary associations (Ullman; 2015, 2016). In fact, learning lexical/semantic information (e.g., word meanings) is *always* dependent on the declarative memory system, implying that L2 learners rely on this system to learn foreign words, regardless of their proficiency level (Ullman; 2015, 2016). Of relevance here is also the fact that the declarative memory facilitates the acquisition of both complex language forms, which are learned as chunks, and multi-word language units, such as the meaning of idioms (Ullman and Lovelett, 2018). In this sense, English phrasal verbs (e.g., *go on*, *put off*, *break up*), the target form in the current study, are both complex and multi-word language constructions (Schmitt & Redwood, 2011). Therefore, one can predict that individual differences in declarative memory can account for the differential learning of phrasal verbs.

Moreover, the relationship between declarative memory and L2 learning is influenced by the context of learning, where declarative memory might be particularly important for learners in more explicit instructional contexts (Faretta-Stutenberg and Morgan-Short, 2018). Given the prominence of explicit (classroom-like) instruction in L2 learning, declarative memory is of particular importance, as learners are thus expected to heavily rely on this memory system to learn the target language (Ullman and Lovelett, 2018). However, declarative memory may also have a detrimental effect on L2 learning, at least for grammar and for advanced learners, as the learning in the declarative memory system may impede the learning that occurs in the procedural memory system, the brain system responsible for the acquisition of more automatic, implicit (not conscious) knowledge (Hamrick et al., 2018)[3].

Finally, most previous research examining the role of declarative on L2 learning has been focused on grammar (see Buffington and Morgan-Short, 2019; Hamrick et al., 2018, for reviews), but little has been done so far regarding word acquisition. To the authors' knowledge, there is only one (lab-based) study that has examined the possible association between declarative memory abilities and the learning of foreign vocabulary

---

[3] One reviewer rightly points out that, according to DeKeyser's work (e.g., DeKeyser, 2015), automatization is possible, once declarative knowledge becomes procedural knowledge through practice. However, DeKeyser does not equate "knowledge" with "memory", as knowledge is the result of learning (as cited in Buffington and Morgan-Short, 2019). The notion that the declarative memory system can inhibit learning in the procedural memory system stems from the competition hypothesis postulated by Ullman (Ullman, 2015; 2016), whereby the declarative and procedural memory systems competitively interact (see Buffington and Morgan-Short, 2019, for a thorough discussion on the three main theoretical perspectives on the role of declarative and procedural memory in L2 acquisition).

(Ruiz et al., 2018). In this lab-based study, participants were asked to intentionally learn pseudowords, that is, participants learned the foreign words in explicit instructional conditions (DeKeyser, 1995). Declarative memory was assessed using a nonverbal measure (the Continuous Visual Memory Task). However, despite theoretical predictions that suggest that declarative memory is associated with L2 vocabulary learning (Ullman, 2015; 2016), the study failed to find evidence for such an association, clearly indicating the need for more research in this area. In order to fulfil this research gap and to contribute to this emerging line of L2 research, the present study examined the relationship between declarative memory and contextualized lexical acquisition, namely, English phrasal verbs, under different instructional conditions in a more real-life setting, i.e., web-based learning.

3 The current study

As indicated above, there is an extant literature on the predictive and explanatory role of individual differences to L2 acquisition (Li, 2017a), but relatively less work has been undertaken on how the effectiveness of instruction is affected by learner individual characteristics. Therefore, more research is needed to examine not only which individual differences are predictive of L2 learning, but also how instruction and individual differences *interact* to account for L2 acquisition (DeKeyser, 2012; Spada, 2011). From an instructional perspective, this type of aptitude-treatment interaction research can potentially inform how L2 instruction can be enhanced. Theoretically, it allows for the possibility to draw inferences about the underlying processes in L2 processing and learning (e.g., which aptitudes can facilitate or hinder learning across different instructional contexts) (DeKeyser, 2012). Thus, the purpose of the present study was to examine whether individual differences in working memory and declarative memory would predict L2 lexical acquisition in the context of web-based ICALL, and whether the potential influence of learners' individual differences in working memory and declarative memory on L2 vocabulary learning was associated with different instructional conditions, namely, meaning-focused (i.e., reading authentic news texts online) and form-focused (i.e., reading authentic news texts plus completing automatically-generated multiple-choice gaps in the online texts) instruction.

The study focused on the lexical learning of English phrasal verbs. The reason for choosing this subgroup of English vocabulary was that, despite their high frequency and importance, they are particularly difficult for non-native speakers to learn, even at advanced levels of proficiency (Mitchell, 2018). With this in mind, our research questions were as follows: Do working memory and declarative memory impact the acquisition of English phrasal verbs (Research question 1)? Does the potential role of these individual differences vary depending on the instruction type (meaning-focused versus form-focused) (Research question 2)? Can an ICALL system be used to conduct large-scale, experimental research addressing the previous two research questions (Research question 3)?

Regarding research question 1, it was predicted that working memory would be positively associated with the acquisition of L2 lexical knowledge, since working memory has been found to be an important predictor of L2 learning (e.g., Linck et al., 2014, for a meta-analysis). Considering the theoretical predictions that indicate that declarative memory underlies the acquisition of L2 vocabulary, including multi-word lexis (Ullman, 2015; 2016), it was expected that declarative memory would favorably affect the lexical learning of phrasal verbs. For research question 2, it was hypothesized that the role of working memory on lexical learning would be different under different instructional conditions (meaning-focused versus form-focused), and that the moderating effect would be more likely to occur in the form-focused condition, as learners with more working memory capacity are expected to better handle the cognitive, attentional demands of form-focused instruction (e.g., Ando et al., 1992; Indrarathne and Kormos, 2018). As for declarative memory, it was predicted that the effect would be observed in the form-focused condition, as learning in this brain system is associated with more explicit, attention-driven learning processes (Poldrack and Packard, 2003). As for research question 3, based on a previous SLA study that employed a similar ICALL system (Ziegler et al., 2017), it was expected that the all-encompassing web-based ICALL system used here in the present study (see Methodology) would also support the collection of experimental data to answer research question 1 and research question 2.

II Methodology

I Participants

A power analysis based on a preliminary study with a similar design (see Ziegler et al., 2017) indicated that at least 40 participants per condition (80 in total) would be needed to detect a hypothesized effect size of .32 (5% significance level, 80% power). In the actual experiment, a total of 127 participants ($M_{age} = 24.3$; 91 women) took part in the study. Because of technical difficulties, the number of participants varied for each cognitive test (see

description below): working memory (OSpan) = 102; visual declarative memory (CVMT) = 89; and verbal declarative memory (MLAT5) = 92. The participants were randomly assigned to two experimental conditions, meaning-focused and form-focused. The groups did not differ across the variables gender or age, all *p* > .05. Most participants were native speakers (L1) of German (78%), followed by Chinese, Italian, Russian (3% each), Lithuanian, Nepali, Polish, Spanish (2% each), and Bulgarian, Indonesian, Macedonian, Portuguese, Serbian, Swedish, Vietnamese (1% each). Results of a C-test (see description below) indicated that most participants were mostly advanced learners of English (61%), followed by intermediate (29%) and basic (9%). According to the DIALANG reading test (Alderson, 2005), the majority of learners were advanced readers in English (64%), followed by intermediate (35%).[4] All subjects gave informed consent and received €60 Amazon e-vouchers for participating.

2 Linguistic target

As noted earlier, the linguistic focus was on English phrasal verbs. These language forms consist of a verb followed by an adverbial particle that behaves as a syntactic and lexical unit (e.g., The small start-up was *snapped up* [bought] by a bigger company) (Larsen-Freeman and Celce-Murcia, 2015). Phrasal verbs are very typical and important features of the English language (Thim, 2012), and thereby crucial for L2 instructors to teach and for L2 learners to master (Garnier and Schmitt, 2016; Larsen-Freeman and Celce-Murcia, 2015). In fact, knowledge of phrasal verbs is taken as indicative of L2 proficiency (Liu, 2012).

However, phrasal verbs are notoriously difficult for L2 learners to acquire, even for advanced learners (Mitchell, 2018), and, as such, have been recently investigated in instructed L2 studies (e.g., Lee and Levine, 2018). Their difficulty, which stems from syntactic, semantic, and pragmatic considerations (Larsen-Freeman and Celce-Murcia, 2015), is such that learners avoid using phrasal verbs (Liao and Fukuya, 2004), or underuse them (Sung, 2017). From a syntactic point of view, for instance, the placement of the particle in phrasal verbs is problematic for learners, as phrasal verbs can be transitive and intransitive at the same time, and particles can be separated from their verbs by different elements, such as pronouns and adverbs, as in (1) and (2), respectively (Schmitt and Redwood, 2011).

(1) They *picked* **them** *up* (collected) from the concert.
(2) The man *came* **straight** *over* (visited) to see his children.

From a semantic perspective, phrasal verbs are also difficult for L2 learners because the same combination of verb and particle can have several possible senses or meanings, that is, they are *polysemous* (Gardner and Davies, 2007) as in (3) and (4):

(3) The family *put up* (stayed) for a night.
(4) The manager *put up* (displayed) a notice.

In fact, the semantic aspect of phrasal verbs is considered to be the most challenging for L2 learners (Neagu, 2007), as nonnative speakers, unlike native speakers who process both the verb and the particle together as a single lexical chunk (Cappelle et al., 2010), often perceive the association between a verb plus particle to be purely arbitrary and at random (Side, 1990), with the particle particularly regarded as illogical (Thibeau, 1999). Therefore, the focus of instruction in this study was on the development of both receptive and productive semantic knowledge of targeted phrasal verbs, that is, knowledge of meaning of both verb and particle as one syntactic and lexical *unit* (i.e., vocabulary acquisition, or lexical learning) (Larsen-Freeman and Celce-Murcia, 2015).

To collect experimental texts containing target vocabulary, a corpus-based study supported by NLP technology was conducted. NLP allowed for the automatic retrieval of texts from the web, specifically news articles from Reuters. The procedure to gather the corpus was as follows. First, a list of potential reading topics was collected via an online survey (http://purl.org/icall/werti-topics). Then, the list was used to automatically search for articles, resulting in a corpus of 11,747 texts. From this initial automatically gathered corpus, using both frequency of occurrence and news relevance (e.g., becoming outdated quickly) as criteria for inclusion, 40 texts were selected to serve as the experimental reading corpus in the study. The experimental texts included a variety of topics including technology, health, and sports news, and their length was approximately two pages per text. The experimental 40-text reading corpus contained 165 phrasal verbs in total, 30 of which were selected as instructional targets. The selection of these targeted phrasal verbs was done according to frequency of appearance

---

[4] Given that most of the participants were L1 German and advanced-level learners of English, there was therefore not enough variability in the data to statistically tease out the effects of neither L1 nor proficiency on the learning of the targeted phrasal verbs.

in the 40-text reading corpus, with the number of occurrences of each targeted phrasal verb ranging from two to nine across the experimental texts. Most of the 30 targeted phrasal verbs were considered to have a figurative, non-transparent meaning (63%), as suggested by consulted native speakers of English (see Appendix 1). As the reading corpus was composed of authentic, natural occurring news texts from the web, the sequence in which the targeted phrasal verbs were presented to the learners was therefore determined by the natural order in which they originally occurred in the texts; there was no intervention by the researchers to alter this original sequence other than organizing the 40 texts in the reading corpus according to number of phrasal verbs in each text, resulting in a fixed-order of reading materials with the texts containing the highest numbers of phrasal verbs being presented first.

3 Instruments:

a WERTi interface

The entire online experiment, from the pretest stage to the posttest stage, was housed in a customized version of the WERTi (Working with English Real Texts interactively; http://sifnos.sfs.uni-tuebingen.de/WERTi/) system, a web-based tool that aims to support language learning and teaching by providing automatic enhancement of L2 texts (Meurers et al., 2010). More specifically, the system uses NLP techniques (e.g., part-of-speech annotation, syntactic parsing) to identify and visually enhance targeted language items in webpages.

The tailored adaptation of the WERTi interface employed in this study supported the deployment of both testing procedures and instructional treatments in one web location (see Appendices 2 through 6). The interface also allowed for explicit logging of all learner-system interactions during treatment, which was based on a login that automatically and randomly assigned every participant to either the meaning-focused or the form-focused groups (see Appendix 2). In this sense, the WERTi system maintained a learner log file for each participant, recording information about the news article; the information filled by the learner in the reading questionnaire; and the time when each article was completed. In the case of the form-focused group (see Procedure), the system additionally registered the total number of correct and incorrect answers in the multiple-choice gaps (this scoring was performed automatically), the total number of attempted items (i.e., items left incorrect after receiving color coded feedback), and the number of hints clicked or cheat instances (i.e., clicking on the smiley face) (see Appendix 4).

b Pre-tests and post-tests

To measure semantic knowledge of English phrasal verbs before and after the treatments, two types of tests were designed. More specifically, the tests measured the *meaning* or *sense* of the targeted phrasal verbs, that is, the meaning or sense of the combination of a base verb and a particle (e.g., *carry out* = complete), and, because some phrasal verbs can have several meanings, only the most frequent meaning found across the 40-text reading corpus was assessed (see Appendix 1).

Given the importance of assessing both productive and receptive mastery of L2 vocabulary (González-Fernández and Schmitt, 2017), a cloze-like task to measure productive knowledge and a multiple-choice test to measure receptive knowledge were developed, respectively. Both instruments included similar items in the form of a single sentence testing the same target phrasal verb productively and receptively (for examples, see Appendix 3). The same testing items were used in both pretest and posttest, with the order randomized by the computer for each participant. The productive test required participants to provide the targeted linguistic form themselves, which is more cognitively demanding than recognizing target items in a receptive test (Schmitt and Redwood, 2011). In this test, the first letter of both the verb and the particle was given. As previously noted, in the receptive test, the same target phrasal verb from the productive test was assessed using the same sentence. However, unlike the productive items, the first letter prompts in the receptive items were omitted and five multiple-choice alternatives were added, including a Don't know ('?') option in order to reduce guessing. Following Schmitt and Redwood (2011), the productive test was administered first because it is feasible to assume that if the participants knew a phrasal verb productively, they would also know it receptively, as comprehension precedes production (Hopman and MacDonald, 2018).

Each test contained 40 items, 30 assessed the target phrasal verbs, and 10 verb-noun collocations (e.g., break loose) served as distractors. Scoring of the receptive test was done automatically by the computer, as one point was awarded if the participant had chosen the correct option. The productive test, on the other hand, was

manually scored for meaning by the first author.[5] If part of the answer was incorrect, that is, either the verb or the particle was wrong, a zero was given. Spelling errors were not counted as incorrect as long as they did not change the meaning of the phrasal verbs; the tense of the verb was not considered either.

c Individual difference measures

a Working memory. Participants' working memory capacity was gauged by means of a web-based adaptation of the Automated Operation Span Task (OSpan; Unsworth et al., 2005) (see Appendix 4). Based on the Klingon Span task devised by Hicks et al (2016), the adaptation consisted of substituting the letter stimuli in the original OSpan task with Klingon symbols. Participants initially practiced both recalling the symbols in the order they appeared on the screen, and solving a series of simple math operations (e.g., 9 * 2 + 6 =?), first separately and then together as they would perform it later in the actual test (i.e., symbol recall while solving the equations). In the real trials, participants were shown 75 symbols and 75 math problems in total, 15 sets, 3-7 randomized symbols that appeared mixed together with the equations. At the end of each trial, participants were tasked with recalling the Klingon symbols in the order they had been shown. The OSpan score represented the total number of correct symbols that were recalled in the correct order.

b Declarative memory. Web-based versions of the Modern Language Aptitude Test, subtest 5, word pair associates (MLAT5; Carroll and Sapon, 1959), and the Continuous Visual Memory Task (CVMT; Trahan and Larrabee, 1988) served as verbal and nonverbal measures of declarative memory, respectively (Buffington and Morgan-Short, 2019; Hamrick et al., 2018) (see Appendix 4). In the MLAT5, participants were asked to memorize pseudo-Kurdish words and their corresponding meanings in English. After studying and practicing 24-word association pairs, participants completed a timed multiple-choice test (four minutes). For each correct response, one point was given, resulting in a total score of 24 points. In the CVMT, participants were required to commit to memory complex figures in a sequence and to indicate whether they had seen the figures in the sequence before or not ("OLD or NEW?"). All figures were presented in a random but fixed order, each one appearing for two seconds. For each participant, a d' (d-prime) score was computed. The scores of both tests were analysed separately (see below).

c General English proficiency. A C-test, a former placement test at a university language center, was employed to measure learners' general English proficiency (see Appendix 5). The test consisted of a set of five short texts with different degrees of difficulty that were converted into C-test (gap filling) format. Each test followed the construction principles of C-tests (Klein-Braley, 1985): The second half (i.e., about 50%) of every second word in a sentence was deleted and had to be completed. The first and last sentences of the text were, however, left intact. Thus, as such, the answers had already been pre-established for the computer to automatically score the C-test. For each gap filled correctly, the computer awarded one point, resulting in a maximum of 124 points. Participants' scores were then mapped to one of proficiency levels of the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001): A (basic; 0-64), B (intermediate; 65-84), and C (advanced; 85 onward). Reliability (Cronbach's alpha) for the C-test was 0.93.

d English reading proficiency. The reading subtest of the DIALANG test (Alderson, 2005) was administered as a measure of participants' reading ability in English. The test consisted of 30 multiple-choice items and assessed three reading subskills, namely, understanding the main idea, making inferences, and understanding specific details. The test reported a number between 1 and 6, which was then fit into one of the proficiency levels of the CEFR using the following mapping: A (basic) = 1-2; B (intermediate) = 3-4; and C (advanced) = 5-6.

4 Procedure

a Treatment

After completing the pretest block (DIALANG test, C-test, and pretests), and the cognitive testing block (OSpan, CVMT, and MLAT5) in two separate days, learners in both the *meaning-focused* (*n* = 63) and the *form-focused*

---

[5] Due to logistical reasons, it was not possible to have a second rater, and thus we were not able to measure inter-rater reliability for the productive test.

($n = 64$) groups read the target 40 news texts on the web for about two weeks (M =16.34; SD = 2.8 days)[6]. In the meaning-focused group, participants read for meaning. In the form-focused group, as a form of automatic input/output enhancement (Ziegler et al., 2017; Sharwood Smith, 1991, 1993; Swain, 1995), learners additionally interacted with the texts via selecting automatically generated multiple-choice options (see Appendix 6). Specifically, they completed the phrasal verbs in the text by providing the corresponding particle (e.g., *out*, *off*) in the gaps. For every gap, the system provided four options: one option was the correct particle that completed the meaning of the phrasal verb in the context of the sentence; two options were semantically similar distractors that were chosen using distributional semantic methods (i.e., automatic word extraction from texts based on frequency of co-occurrence information) (see Skeppstedt et al., 2013); and a fourth option which was selected from a list of manually collected lexical items that similarly function as particles (e.g., prepositions). If the answer was correct, it turned green (*correct*). If the answer was incorrect, it was highlighted in red (*incorrect*). If the learner clicked on the smiley face next to the blank space, the answer was given by the system, and was shown in green (*cheats*). The smiley face was to be used as a last resort. All these interactions with the texts (i.e., *correct*, *incorrect*, and *cheats*) were recorded in logs by the WERTi system (see Ziegler et al., 2017, for a similar procedure). Upon finishing reading the 40 articles, learners were immediately directed to final part of the experiment, where they took both productive and receptive posttests (about 20 min), and completed a debriefing questionnaire (about 5 min).

b Reading questionnaire

Both groups completed a reading questionnaire after reading each news article (see Appendix 7). The reading questionnaire had two parts, one assessing global understanding of the text (Cronbach's alpha 0.66), and the other collecting participants' feedback. In the first part, learners were mainly required to select the main idea of the article in a four-option multiple-choice question. In the second, participants reported perceived text difficulty, enjoyment of the text, and familiarity with text topic using a Likert-type scale from 1 to 5 (Pino-Silva, 2009). This part allowed to ensure that the texts were somewhat similar across groups.

5 Statistical analyses

All data were analysed using the statistical software package R version 3.3.2 (R Core Team, 2016). In order to examine the association between instruction, individual differences in working memory and declarative memory, and the lexical learning of the targeted phrasal verbs, logistic mixed-effects regression models (Jaeger, 2008) with first-order interaction terms (i.e., three-way interactions) (DeKeyser, 2012) were built. To that end, the *glmer* (generalized linear mixed model) function in the lme4 package (Bates et al., 2015) was used. In order to avoid nonconvergence during model estimation, the BOBYQA algorithm was employed as the optimizer (see Linck and Cunnings, 2015). In the mixed-effects regression models, the dependent variable was defined as a binary variable representing accuracy (correct = 1, or incorrect = 0) in both receptive and productive tests (i.e., item-based learning). Scores from the cognitive tests (OSpan, MLAT5, and CVMT); instructional condition (meaning-focused vs. form-focused); type of knowledge (receptive vs. productive); and type of language test (pretest vs. posttest) were included as fixed effects; learners and phrasal verbs were included as random crossed (i.e., independent) effects (Baayen et al., 2008). To ensure a parsimonious analysis (Bates et al., 2015), only one cognitive predictor at a time was included in each regression model. Furthermore, in order to reduce multicollinearity (i.e., high correlations among predictors) when calculating interactions, as well as to facilitate the interpretation of results (Aiken and West, 1991), the raw scores of the continuous variables working memory (OSpan) and verbal declarative memory (MLAT5) were mean-centered; for the nondeclarative memory test (CVMT), the d' scores (see above) were already mean-centered, as they were calculated from z-scores (i.e., M = 0, SD = 1). Additionally, all factorial fixed effects (i.e., instructional condition, type of knowledge, and type of language test) were contrast-coded (see Table 2) (Cohen et al., 2003). Missing data in the cognitive tests needed no special treatment because mixed-effects regression models are robust against unbalanced data (Baayen et at., 2008; Jaeger, 2008).

It should be noted that, in the regression analysis, the variable type of language test represented the difference in performance between pretest and posttest (i.e., learning effects; see below); therefore, the three-way interaction term of this variable with instructional condition and the learner cognitive factors under investigation (working memory and declarative memory) captured the effect of these variables on L2 vocabulary learning. Consequently, if this three-way interaction term were found to be significant, it would

---

[6] One reviewer asked about whether the amount of time on task was equal for all participants. However, due to the online nature of the study, it was not possible to assess effective time on task.

indicate that the cognitive factor in question was associated with the lexical learning of phrasal verbs in the study, and thus it would provide evidence to positively answer research question 1. Further resolving the significant three-way interaction by means of follow-up simple slope analyses (Aiken and West, 1991) would indicate whether the effect of instruction on the lexical learning of phrasal verbs was moderated by the instructional condition or not, and hence these follow-up analyses would supply evidence to address research question 2.

III Results

Table 1 presents the descriptive statistics summarizing participants' performance on both receptive and productive tests.

Table 1. Descriptive statistics for pretest and posttest comparison of percentage scores.

Test		Meaning-focused		Form-focused	
		Pre	Post	Pre	Post
Receptive	*M*	83.67	87.87	84.86	89.70
	SD	12.11	11.05	12.57	12.99
Productive	*M*	44.09	54.73	47.07	59.12
	SD	15.85	15.66	13.44	16.18

In order to determine learning effects, paired t-tests were used to compare performance on pretests and posttests. Pretest-posttest comparisons showed that, for the receptive test, there was a statistical difference between pre and post testing for both the meaning-focused group, $t(62) = -3.60$, $p < .001$, $d = -.45$, and the form-focused group, $t(63)= -4.29$, $p < .001$, $d = -.54$. For the productive test, there was also a statistical difference between pre and post testing for both the meaning-focused group, $t(62) = -7.49$, $p < .001$, $d = -.94$, and the form-focused group, $t(63)= -8.31$, $p < .001$, $d = -1.04$. Overall, results suggest that both groups demonstrated a learning effect from pretest to posttest in both tests.

The first research question concerned whether there was an association between instruction, individual cognitive abilities, namely working memory and declarative memory, and the lexical learning of targeted phrasal verbs. Regarding working memory, the mixed-effect regression analyses showed that there was significant three-way interaction between instructional condition, type of language test, and learners' working memory capacity, indicating that working memory was related to lexical learning (estimate = 0.023, $SE = 0.008$, $p = .004$) (see Table 2). In regard to declarative memory, the effect on lexical learning depended on the type of measure. More specifically, for the nonverbal declarative memory measure (CVMT), there was a significant three-way interaction with instructional condition, type of language test, and declarative memory (estimate = 0.456, $SE = 0.232$, $p = .050$), suggesting an association of nonverbal declarative memory and L2 vocabulary acquisition. For the verbal declarative measure (MLAT5), on the other hand, there was no evidence for such an association (estimate = 0.018, $SE = 0.028$, $p = .522$), nor was there significant evidence for an effect of verbal declarative memory on lexical learning (estimate = -0.005, $SE = 0.014$, $p = .705$). In sum, the results indicate that the answer to the first research question is affirmative for both working memory and (nonverbal) declarative memory.

Table 2. Linear mixed-effects regression for working memory.

Fixed effect	Estimate	SE	p
Intercept	1.496	0.311	< .001
Type of knowledge	2.757	0.066	< .001
Type of language test	0.659	0.058	< .001
Instructional condition	0.246	0.202	.224
Working memory	0.010	0.007	.164
Type of knowledge:Working memory	-0.001	0.004	.844
Type of knowledge:Type of language test	-0.231	0.115	.045
Type of language test:Working memory	0.002	0.004	.687
Type of knowledge:Instructional condition	-0.093	0.118	.428
Instructional condition:Working memory	0.033	0.014	.019
Type of language test:Instructional condition	0.116	0.115	.314
Type of knowledge:Type of language test: Working memory	0.003	0.008	.727
Type of knowledge:Instructional condition:Working memory	0.025	0.008	.002
Type of knowledge:Type of language test:Instructional condition	0.166	0.230	.470
Instructional condition:Type of language test:Working memory	0.023	0.008	.004
Type of knowledge:Instructional condition:Type of language test:Working memory	0.005	0.016	.759

Random effects	Variance	SD	
Phrasal verb	0.905	0.951	
Participant	2.576	1.605	

Note. All factors were contrast-coded, as follows: Type of knowledge (–0.5 = productive, 0.5 = receptive), Type of language test (–0.5 = posttest, 0.5 = pretest), Instructional condition (–0.5 = meaning-focused, 0.5 = form-focused).

The second research question asked whether the potential effect of individual differences in working memory and declarative memory on lexical learning could be dependent on instructional condition (meaning-based versus form-focused instruction). In regard to working memory, the results of the follow-up (i.e., simple slope) analyses (Aiken and West, 1991) showed that, while the effect of working memory was negative and not significant for participants in the meaning-focused group (estimate = -0.010, *SE* = 0.005, *p* = .063), the working memory effect was positive and significant in the form-focused group (estimate = 0.013, *SE* = 0.006, *p* = .024). With regard to declarative memory, recall that there was a significant three-way interaction only for nonverbal declarative memory as measured by the CVMT, and, therefore, the results reported here solely concern this test. In this regard, while the effect of nonverbal declarative memory was negative and not significant for participants in the meaning-focused group (estimate = - 0.197, *SE* = 0.182, *p* = .279), the nonverbal declarative memory effect was positive and approached significance in the form-focused group (estimate = 0.259, *SE* = 0.145, *p* = .073). In summary, the effect of working memory on the lexical learning of phrasal verbs was moderated by instructional condition, such that working memory was predictive of L2 vocabulary acquisition only for participants in the form-focused condition, but not for the meaning-focused group, suggesting the presence of an aptitude by treatment interaction (Cronbach and Snow, 1977). As for declarative memory, there was no such a moderating effect, although there was a tendency for a positive effect in the form-focused condition. Hence, the answer to the second research is positive for working memory, but negative for declarative memory.

The third question focused on the affordances that the web-based ICALL system used in the present work would offer to address both research question 1 and research question 2. The analysis provided above clearly indicates that the ICALL system allowed for the collection of experimental data to examine these two research questions, and thus the answer to research question 3 is affirmative. Moreover, the system also afforded the opportunity to gather data on the process of form learning (i.e., vocabulary learning) during instructional treatment. In this sense, the following sections present the analysis of these data.

Computer log analysis

As described earlier, the WERTi system kept log files containing information about interactions with the text via multiple-choice gaps (i.e., *correct*, *incorrect*, and *cheats*, see Methodology) in the form-focused group, and responses to the reading questionnaire. In order to provide a more detailed, incremental measurement of learner development during instructional treatments, the learner log data were analysed (Ziegler et al., 2017). The results of this analysis are given below.

a Development of accuracy in the form-focused group

The scatterplot in Figure 1 gives an overall impression of the development of the accuracy with which individual learners completed the multiple-choice gaps in the form-focused group ($n = 64$). Each mark corresponds to a text read by a learner, with the x-axis showing the cumulative number of items a learner had interacted with up to that point. Results showed that the accuracy of learners often reached ceiling, and that learners interacted up to 400 items in the 40 texts they read.

[insert Figure 1.]

Figure 1. Scatterplot of accuracy by interacted items.

In order to determine whether time as a variable predicted performance in the multiple-choice gaps, a mixed-effects analysis with time as both a linear and a quadratic predictor was conducted. The results showed that there was a significant negative nonlinear (curvilinear) relationship between time and accuracy (estimate = -0.3846, $SE = 0.187$, $p = .037$), indicating that relationship was concave, i.e., accuracy increased up to approximately half the treatment time (texts 20-21) and then declined.

b Reading questionnaire

Descriptive statistics for global reading comprehension (i.e., understanding the meaning of the L2 text) are reported in Table 3. Pretest-posttest comparisons indicated that there was no significant difference between groups, $t(119.05) = 1.06$, $p = .293$, $d = .19$.

Table 3. Descriptive statistics for global reading comprehension.

Group	M	SD
Meaning-focused	62%	11%
Form-focused	60%	14%

To examine whether individual differences in cognitive abilities, reading proficiency in English, and instructional conditions had a combined effect on global reading comprehension, mixed-effects models with three-way interaction terms were developed. Results showed that none of the three-way interaction terms were significant. However, the interaction between group and reading proficiency was significant (estimate = -0.854, $SE = 0.272$, $p = .002$), signifying that the effect of the instructional condition was different depending on learners' reading proficiency level. A simple slope analysis (Aiken and West, 1991) revealed that while the group effect was negative and significant for intermediate learners (estimate = -0.617, $SE = 0.221$, $p = .005$), the effect of group was positive and not significant for advanced learners (estimate = 0.238, $SE = 0.158$, $p = .133$). Further inspection of means found that advanced learners in the meaning-focused group had a mean (intercept = 0.679, $SE = .252$) that did not differ from the one in the form-focused group for these participants (intercept = 0.916, $SE = .239$). However, for intermediate learners, scores on reading comprehension were higher in the meaning-focused group (intercept = 0.548, $SE = .278$) than in the form-focused group (intercept = -0.069, $SE = .259$). Overall, the results suggest that reading comprehension was affected for intermediate, not advanced learners, particularly in the form-focused condition.

Finally, the results of participants' feedback showed that, on a scale from 1 to 5, the mean difficulty of the texts rating was 2.93 ($SD = 1.23$), while the mean enjoyment rating was 2.93 ($SD = 1.23$), and the mean familiarity rating was 2.35 ($SD = 1.25$). The groups did not differ across the variables difficulty, enjoyment, and familiarity ($p > .05$).

IV Discussion

The current study examined the relationship between instruction, cognitive individual differences, and L2 vocabulary acquisition in a web-based ICALL environment. Understanding this interplay is important because the extent to which learners profit from different instructional treatments may significantly depend on their

individual abilities. As such, the study had three research questions whose answers, on the basis of the results described above, are addressed in more detail in what follows.

Research question 1: *Do working memory and declarative memory impact the acquisition of English phrasal verbs?*

The results suggest that both working memory and (nonverbal) declarative memory and were related to the lexical learning of target phrasal verbs. In regard to the former, the present positive findings are in line with previous research that has attested the role of working memory as an important predictor of L2 learning (e.g., Ando et al. 1992; Indrarathne and Kormos, 2018; Juff and Harrington, 2011; Linck et al., 2014; Malone, 2018; Yang et al., 2017). Concerning the latter, the results support the theoretical predictions (Ullman, 2015, 2016) that suggest a role for declarative memory in the acquisition of L2 word knowledge (Buffington and Morgan-Short, 2019; Hamrick et al., 2018; Ullman, 2015, 2016; Ullman and Lovelett, 2018), but contradict earlier lab-based research that found no evidence for such a role (Ruiz et al., 2018). However, the association between declarative memory and lexical learning depended on the kind of test used to measure learners' declarative memory abilities. In particular, the association was significant for nonverbal declarative memory, as measured by the CVMT, but not for verbal declarative memory, as indexed by the MLAT5. In their recent review on the topic, Buffington and Morgan-Short (2019) concluded that the role of declarative memory and procedural memory on L2 grammar acquisition may be somewhat dependent on type of cognitive tests used, at least for procedural memory. It stands to reason, then, that the same could be the case for L2 vocabulary acquisition and for declarative memory. However, as indicated in the literature review, there is currently limited prior research on the association between declarative memory and L2 word learning, and thus more empirical work is required to further explore this relationship.

Research question 2: *Does the potential role of these individual differences vary depending on the instruction type (meaning-focused versus form-focused)?*

As regards working memory, the present study provides evidence to suggest a differential role of working memory in the development of phrasal verb knowledge under different instructional conditions in an ICALL-supported environment. In particular, the moderating effect of working memory was only evident in the form-focused group, that is, learners who read and filled multiple-choice gaps where the phrasal verbs appeared in the texts, suggesting the presence of an aptitude-treatment interaction (Cronbach and Snow, 1977). These findings lend support to the argument that factors such as instructional context can moderate the effect of working memory on L2 acquisition (e.g., Tagarelli et al., 2015; Ando et al., 1992; Faretta-Stutenberg and Morgan-Short, 2018; Indrarathne and Kormos, 2018, Malone, 2018; Suzuki and DeKeyser, 2017; Yang et al., 2017). Likewise, the results are consistent with i) the theoretical perspective that the role of working memory is larger when instruction is more cognitively demanding (e.g., Linck et al., 2014), as well as ii) previous research that has found that working memory interacts with instructional treatments (e.g., Tagarelli et al., 2015; Ando et al., 1992; Faretta-Stutenberg and Morgan-Short, 2018; Indrarathne and Kormos, 2018, Malone, 2018; Suzuki and DeKeyser, 2017; Yang et al., 2017). Moreover, the interaction between working memory and more explicit instructional conditions reported here is similar to what has been observed in artificial language research conducted in the lab (e.g., Tagarelli et al., 2015), but in a more realistic setting. Therefore, this study provides further evidence for the predictive validity of working memory for L2 learning in more ecologically relevant contexts (Ando et al. 1992; Linck and Weiss, 2015).

In regard to declarative memory, the results showed that the effect of this individual factor on L2 vocabulary learning was not moderated by the instructional condition. However, there was a trend for a beneficial effect of (nonverbal) declarative memory on lexical learning in the form-focused condition. It is possible that with a larger sample size, this trend would have proved to be significant. Therefore, this study needs to be replicated using a larger group of learners. On the other hand, it may be possible that the form-focused condition might not have not been explicit enough for declarative memory to emerge as a predictor of word learning, despite that learning in the declarative memory system has been associated with explicit, attention-demanding learning processes (Poldrack and Packard, 2003; Hamrick et al. 2018). Recall that, in this instructional condition, learners were asked to pay attention to form, as they were instructed to notice the blanks next to verbs in the texts and to click on the dropdown menus to complete the gaps. Another plausible explanation for nonsignificant results may be that, while declarative memory affects grammar learning differently as a function of proficiency (e.g., Hamrick, 2015; Morgan-Short et al., 2014), acquiring L2 lexical knowledge is predicted to be always dependent on the declarative memory system, regardless of learners' L2 proficiency level (Ullman; 2015, 2016). Consequently, having greater declarative memory capacity may not

provide a significant advantage when learning more arbitrary features of language (i.e., lexical learning) (Ruiz et al., 2018). It follows that strict interpretations of theoretical predictions may not be as clear-cut. Different variables (e.g., instructional context, gender, proficiency) may strengthen or weaken the potential effect on declarative memory on lexical development (Hamrick, 2015; Morgan-Short et al., 2014; Ullman, 2015, 2016). More research is warranted to continue investigating the potential influence of declarative memory on L2 word learning.

Research question 3: *Can an ICALL system be used to conduct large-scale, experimental research addressing the previous two research questions?*

The final research question and set of findings concern whether an ICALL environment could provide the affordances necessary to carry out large-scale, experimental research addressing the study's first two research questions. The above discussion clearly shows that it was possible to collect experimental L2 data to address these two research questions, as expected. In this regard, one of the methodological advantages offered by the use of ICALL systems for SLA research when compared to lab-based or classroom-based research is the possibility of collecting data in order to explore the incremental process of learning during treatment (Meurers and Dickinson, 2017). In this study, such an advantage allowed for the analysis of language development in the form-focused group, and the assessment of a potential three-way interaction between cognitive individual differences, L2 reading proficiency, and instructional conditions. It will be recalled that learners in the form-focused condition read news articles and completed multiple-choice gaps as a medium of automatic input/output enhancement (Ziegler et al., 2017; Sharwood Smith, 1991, 1993; Swain, 1995), and that learners' global reading comprehension (i.e., understanding of main ideas) in both instructional conditions was assessed.

Regarding the development of accuracy during treatment in the form-focused condition, the results indicated that development was not linear, but parabolic, implying that accuracy first progressed and then regressed as a function of time. An explanation for these results is the notion that the process of learning an L2 is nonlinear and incremental, in which small changes occur over time (Murakami, 2016), thereby coinciding with previous research which has found evidence for nonlinearity in L2 development (e.g., Bulté and Housen, 2018; Murakami, 2016). Another conceivable explanation for the observed nonlinearity may be that, as the experiment progressed, learners might have experienced fatigue, boredom or any other non-developmental individual factors (Bulté and Housen, 2018). It is important to note, however, that the results of the reading questionnaire showed that, on average, learners mostly enjoyed reading the experimental texts. Finally, it could be the case that learners were presented with less complex texts in the first half of the experiment (texts 20-21) than in the second half, which could explain why accuracy first increased and then decreased. However, again, learners reported that the texts were generally not very difficult to read. The fact that the texts were shown in fixed order during treatment might have introduced a difficulty bias. Future research should consider randomizing the order of the experimental texts so as to reduce the possible influence of differences in text difficulty.

In regard to the potential three-way relationship between individual differences, L2 reading proficiency, and instructional conditions, the results indicated that these factors did not interact to predict overall reading comprehension. However, the results showed that reading comprehension was differently affected by instructional condition depending on learners' proficiency level. In particular, reading comprehension was negatively impacted for intermediate learners, especially for those in the form-focused condition. For these learners, performing a dual task (i.e., reading while paying attention to linguistic form) seemed to have had a detrimental effect on their reading comprehension. It is worth pointing out that, in the present study, the term *form learning* was used to refer to what students learn about the target language, in this case, the meaning of the phrasal verbs as a lexical item, and the term *comprehension* to refer to what they learn about the content or the meaning of the L2 text. The results here are in line with previous research which has found a trade-off for simultaneous processing of form and meaning (in this instance, the meaning of the text) (e.g., Boers, et al., 2017; VanPatten, 1990). Moreover, these findings also align with the notion that the (unfavorable) effects of paying attention to form and meaning simultaneously is related to learners' proficiency level. Particularly, learners at lower levels of proficiency are more likely to have difficulty simultaneously attending to form and meaning, as they need to invest more attentional resources to making form-meanings connections than more advanced learners (VanPatten, 1990).

Limitations of the study and future research directions

Despite positive results, there are some limitations to this study. One potential limitation is that there was no control group, which could have been used to account for potential testing effects, as learners might have

learned from having taken the same tests twice. Nonetheless, if this had been the case, the testing effect would have been observed in both treatment groups, but this was not found in the statistical analyses. Notwithstanding, the inclusion of a control group could have prevented such a possibility completely. Therefore, future research should replicate this study using a control group. A reviewer pointed out that the learning tasks used in the present study, especially in the form-focused condition (i.e., filling multiple choice gaps to complete phrasal verbs in the text), were focused on the development explicit rather than implicit knowledge of the targeted phrasal verbs. However, vocabulary learning is mostly regarded as item-based learning (Schmitt, 2010, p. 185), signifying that learners either know the lexical item or they do not know it. As such, then, word learning largely amounts to the development of declarative, explicit knowledge of lexical items (Ullman, 2015, 2016, Hulstjin, 2007). Further research could consider using other tasks that measure the development of implicit lexical knowledge as well (Sonbul and Schmit, 2013). Another study limitation is that, due to logistical reasons, it was not possible to assess inter-reliability for the productive test, as the scoring of this test involved only one rater. Future research needs to involve more than one rater. Lastly, a further limitation concerns the lack of a delayed posttest. The test would have indicated whether the findings here would hold across time (e.g., Ando et al., 1992). Therefore, future work would benefit from the inclusion of a delayed posttest.

V Conclusions

The findings of the study suggest that working memory and declarative memory may be particularly important in ICALL-mediated interventions. In particular, working memory appears to have a facilitative role in instructional conditions whereby learners' attention is drawn to linguistic form, that is, form-focused instruction, in comparison to instructional conditions whereby the emphasis is only on communication, that is, meaning-focused instruction (Ellis, 2012). More generally, the results are in line with the overall findings of individual difference research, in that learners' cognitive individual differences are crucial predictors of L2 acquisition (e.g., Li, 2017a). Hence, they should be taken into account when explaining differential attainment of language learners in general, as well as when implementing instructional treatments in web-based learning contexts.

From a pedagogical perspective, using a web-based intelligent CALL system such as the one used here can certainly serve as a complement to what is done in the classroom (see Meurers et al., 2019, for an instance of an ICALL system already being used in a school setting). One of the problems with learners in the classroom is the fact that there is insufficient L2 input, since time in class is limited (European Commission Eurydice Report, 2017; Suzuki et al., 2019). Furthermore, there is also the problem that learners tend to have little or no access to authentic materials in the target language. And even if the materials are authentic, they are not individually catered to the learner. In general, for instance, learners are unable to select the texts they wish to read, which in turn may affect their motivation to learn the L2 language. In this sense, NLP-powered technology in the form ICALL systems like the one shown here is necessary to supplement the language classroom by providing authentic materials, and plenty and enriched input that is adapted to individual learners (see Ziegler et al., 2017; Chen and Meurers, 2019).

The present study is an important first step toward describing what type of learner is more likely to benefit from language instruction delivered via ICALL systems (Ziegler et al., 2017). On the other hand, it provides evidence that it is possible to run web-based intervention studies on adult language learning in a scalable and reliable manner. The collection of experimental data via the Internet can contribute significantly to the study of second language acquisition by facilitating the collection of substantially greater amounts of data in shorter time periods than is usually possible in lab-based experiments (MacWhinney, 2017; Meurers and Dickinson, 2017). Overall, the study contributes to explaining why some people are more successful than others when learning a foreign language in general, as well as in instructional contexts aided by intelligent CALL.

References

Adams, E.J., Nguyen, A.T. and Cowan, N., (2018) Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose. *Language, Speech, and Hearing Services in Schools*, *49*(3), pp. 340-355.

Aiken LS and West SG (1991) *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: SAGE.

Alderson JC (2005) *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.

Ando J, Fukunaga N, Kurahachi J, Suto T, Nakano T and Kage M (1992) A comparative study on the two EFL teaching methods: The communicative and the grammatical approach. *Japanese Journal of Educational Psychology* 40(3): 247–256.

Baayen RH (2008) *Analyzing linguistic data: A Practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baddeley AD (2017) Modularity, working memory and language acquisition. *Second Language Research* 33(3): 299–311. DOI: 10.1177/0267658317709852.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

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

Baron, R.M., and Kenny, D.A. (1986) The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

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

Benson S and DeKeyser RM (2018) Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research* DOI: 10.1177/1362168818770921.

Biedroń, A. and Pawlak, M., 2016. The interface between research on individual difference variables and teaching practice: The case of cognitive factors and personality. *Studies in Second Language Learning and Teaching*, *6*(3) pp. 395-422.

Boers F, Warren P, He L and Deconinck J (2017) Does adding pictures to glosses enhance vocabulary uptake from reading? *System* 66: 113–129. DOI: 10.1016/j.system.2017.03.017.

Buffington J and Morgan-Short K (2019) Declarative and procedural memory as individual differences in second language aptitude. In: Wen Z., Skehan, P., Biedron, A., Li, S., & Sparks, R. (eds.), *Language aptitude: Advancing theory, testing, research and practice*. New York: Routledge, pp. 215–237.

Bulté B and Housen A (2018) Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics* DOI: 10.1111/ijal.12196.

Cappelle, B., Shtyrov, Y., and Pulvermüller, F. (2010) Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain and Language*, *115*(3), 189–201.

Carroll JB and Sapon SM (1959) *Modern Language Aptitude Test*. New York: The Psychological Corporation/Harcourt Brace Jovanovich.

Chen, X. and Meurers, D., (2019). Linking text readability and learner proficiency using linguistic complexity feature vector distance. *Computer Assisted Language Learning*, *32*(4), 418-447. DOI: 10.1080/09588221.2018.1527358.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: Erlbaum.

Council of Europe (2001) *A Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Cowan N (2017) The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review* 24(4): 1158–1170. DOI: 10.3758/s13423-016-1191-6.

Cowan, N., (2010) The magical mystery four: How is working memory capacity limited, and why? Current directions in psychological science, *19*(1), pp.51-57.

Cronbach LJ and Snow RE (1977) *Aptitudes and instructional methods: A handbook for research on interactions*. England: Irvington.

Csizér, K (2017) Motivation in the L2 classroom. In: Loewen S and Sato M (eds) *The Routledge Handbook of instructed second language acquisition*. New York: Routledge, pp. 418–432.

DeKeyser, R. M. (2015). Skill acquisition theory. In: B. VanPatten, & J. Williams (eds) *Theories in second language acquisition: An introduction* (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates, pp. 94-112.

DeKeyser RM (2012) Interactions between individual differences, treatments, and structures in SLA. *Language Learning* 62: 189–200. DOI: 10.1111/j.1467-9922.2012.00712.x.

DeKeyser, R.M., (1995) Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in second language acquisition*, *17*(3), pp.379-410.

Doughty, C.J., (2019) Cognitive language aptitude. *Language Learning*, *69*, pp.101-126.

Doughty, C. J. (2001) Cognitive underpinnings of focus on form. In: P. Robinson (ed) *Cognition and second language instruction*. Cambridge: Cambridge University Press, pp. 206–257.

Ellis R (2012) *Form-focused instruction and second language learning*. Oxford: Wiley-Blackwell.

Ellis R (2016) Anniversary article Focus on form: A critical review. *Language Teaching Research* 3(20): 405–428. DOI:10.1177/1362168816628627

Engle, R.W., (2018) Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*(2), pp.190-193.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147–171.

European Commission, Eurydice (2017). Key Data on Eurydice Report Teaching Languages at School in Europe 2017 Edition.

http://viaa.gov.lv/library/files/original/Key_Data_on_Teaching_Languages_2017_Highlights.pdf

Faretta-Stutenberg M and Morgan-Short K (2018) The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research* 34(1): 67–101. DOI: 10.1177/0267658316684903.

Gardner, D., and Davies, M. (2007).Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, *41*(2), 339–359.

Garnier, M., & Schmitt, N. (2016) Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, *59*, 29–44.

Grey, S., Williams, J.N. and Rebuschat, P., 2015. Individual differences in incidental language learning: Phonological working memory, learning styles, and personality. *Learning and Individual Differences*, *38*, pp. 44-53. DOI: 10.1016/j.lindif.2015.01.019.

González Fernández B and Schmitt N (2017) Vocabulary acquisition. In: Loewen S and Sato M (eds) *The Routledge Handbook of instructed second language acquisition*. New York: Routledge, pp. 280–298.

Goo, J. (2012) Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, *34*, 445–474.

Goo, J., Granena, G., Yilmaz, Y., and Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In: P. Rebuschat (ed), *Implicit and explicit learning of languages*. Amsterdam: John Benjamins Publishing, pp. 443–482.

Granena, G., & Yilmaz, Y. (2018) Aptitude-treatment interaction in L2 learning: A research synthesis. *Studies in English Education*, *4*, 803–830.

Hamrick P (2015) Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences* 44: 9–15. DOI: 10.1016/j.lindif.2015.10.003.

Hamrick P, Lum JAG and Ullman MT (2018) Child first language and adult second language are both tied to general purpose learning systems. *Proceedings of the National Academy of Sciences* 115(7): 1487–1492. DOI: 10.1073/pnas.1713975115.

Hicks KL, Foster JL and Engle RW (2016) Measuring working memory capacity on the web with the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* 5: 478–489. DOI: 10.1016/j.jarmac.2016.07.010.

Hopman EWM and MacDonald MC (2018) Production practice during language learning improves comprehension. *Psychological Science* DOI: 10.1177/0956797618754486.

Hulstijn, J.H., 2007. Fundamental issues in the study of second language acquisition. *EuroSLA Yearbook*, *7*(1), pp.191-203.

Indrarathne B and Kormos J (2018) The role of working memory in processing L2 input: Insights from eye-tracking. *Bilingualism: Language and Cognition* 21(2): 355–374. DOI: 10.1017/S1366728917000098.

Jackson, D.O. (2019) The potential relationship between openness and explicit versus implicit L2 knowledge. *Journal of Psycholinguistic Research*, *48*(2), pp.289–306.

Jaeger TF (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4): 434–446. DOI: 10.1016/j.jml.2007.11.007.

Juffs, A., & Harrington, M. (2011) Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137–166.

Kang EY, Sok S and Han Z (2018) Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research* DOI: 10.1177/1362168818776671.

Kidd E, Donnelly S and Christiansen MH (2018) Individual differences in language acquisition and processing. *Trends in Cognitive Sciences* 22(2): 154–169. DOI: 10.1016/j.tics.2017.11.006.

Klein-Braley C (1985) A cloze-up on the C-Test: A study in the construct validation of authentic tests. *Language Testing* 2(1): 76–104. DOI: 10.1177/026553228500200108.

Larsen-Freeman D and Celce-Murcia M (2015) *The grammar book: Form, meaning, and use for English language teachers*. Boston: National Geographic Learning.

Lado, B. (2017). Aptitude and pedagogical conditions in the early development of a nonprimary language. *Applied Psycholinguistics*, 38(3), 679–701.

Lee, J.H. and Levine, G.S., (2018) The effects of instructor language choice on second language vocabulary learning and listening comprehension. *Language Teaching Research*.

Li S (2017a) Cognitive differences and ISLA. In: Loewen S and Sato M (eds) *The Routledge handbook of instructed second language acquisition*. New York: Routledge, pp. 396–417.

Li, S. (2017b). The effects of cognitive aptitudes on the process and product of L2 interaction: A synthetic review. In L. Gurzynski-Weiss (Ed.), Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors. Amsterdam: John Benjamins, pp. 41–70.

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics 40*(3), 693–722.

Li, S. and Roshan, S., (2019) The associations between working memory and the effects of four different types

of written corrective feedback. *Journal of Second Language Writing, 45*, pp.1-15.

Liao Y and Fukuya YJ (2004) Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning* 54(2): 193–226. DOI:10.1111/j.1467-9922.2004.00254.x.

Linck, JA, and Cunnings, I (2015). The utility and application of mixed-effects models in second language research. *Language Learning, 65*(S1), 185–207.

Linck JA, Osthus P, Koeth JT and Bunting MF (2014) Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review* 21(4):861–883. DOI: 10.3758/s13423-013-0565-2.

Linck JA and Weiss DJ (2015) Can working memory and inhibitory control predict second language learning in the classroom? *SAGE Open* 5(4): 1–11. DOI: 10.1177/2158244015607352.

Liu D (2012) The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes* 31(1): 25–35. DOI: 10.1016/j.esp.2011.07.002.

Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. TESOL Quarterly, 45(4), 661–688. DOI: 10.5054/tq.2011.247707

Long MH (1991) Focus on form: A design feature in language teaching methodology. In: de Bot K, Ginsberg RB and Kramsch C (eds) *Foreign language research in cross-cultural perspective*. Amsterdam: John Benjamins, pp. 39–52.

Long, M. H. (1983). Does second language instruction make a difference? A review of research. *TESOL Quarterly*, *17*, 359–382.

Lu X (2018) Natural language processing and intelligent computer-assisted language learning (ICALL). In: Liontas JI (ed.) *The TESOL Encyclopedia of English Language Teaching*. Chichester, UK: Wiley Blackwell. DOI: 10.1002/9781118784235.eelt0422.

MacWhinney B (2017) A shared platform for studying second language acquisition. *Language Learning* 67(S1): 254–275. DOI: 10.1111/lang.12220.

Malone J (2018) Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*: 1–25. DOI: 10.1017/S0272263117000341.

Meurers, D., De Kuthy, K., Nuxoll, F., Rudzewitz, B. and Ziai, R. (2019). Scaling up intervention studies to investigate real-life foreign language learning in school. *Annual Review of Applied Linguistics*, 39, pp.161-188. DOI:10.1017/S0267190519000126.

Meurers, D. and Dickinson, M (2017) Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning*, *67*(S1), pp.66-95. DOI:

Meurers, D., Ziai, R., Amaral, L., Boyd, A., Dimitrov, A., Metcalf, V. and Ott, N., (2010). Enhancing authentic web pages for language learners. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 10-18). Association for Computational Linguistics.

Mitchell C (2018) Teaching phrasal verbs. In: Liontas JI (ed.) The TESOL *Encyclopedia of English Language Teaching*. Chichester, UK: Wiley Blackwell. DOI: 10.1002/9781118784235.eelt0759.

Morgan-Short K, Faretta-Stutenberg M, Brill-Schuetz KA, Carpenter H and Wong PCM (2014) Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition* 17(1): 56–72. DOI: 10.1017/S1366728912000715.

Murakami A (2016) Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning* 66(4):834–871. DOI: 10.1111/lang.12166.

Neagu, M. (2007). English verb particles and their acquisition: A cognitive approach. Revista Española de Lingüística Aplicada, *20*, 121–138.

Norris JM and Ortega L (2000) Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* 50(3): 417–528. DOI: 10.1111/0023-8333.00136.

Pawlak M (2017) Overview of learner individual differences and their mediating effects on the process and outcome of L2 interaction. In: Gurzynski-Weiss L (ed.) *Expanding individual difference research in the interaction approach*. Amsterdam: John Benjamins, pp. 19–40.

Pino-Silva J (2009) Extensive reading through the Internet: Is it worth the while? *International Journal of English Studies* 9(2).

Poldrack, R. A., and Packard, M. G. (2003) Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245–251.

R Development Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 45–73.

Robinson P, Mackey A, Gass SM and Schmidt R (2012) Attention and awareness in second language acquisition. In: Gass, S., Mackey, A. (eds) *The Routledge Handbook of Second Language Acquisition*. New York: Routledge, pp. 247–267.

Roehr K (2008) Linguistic and metalinguistic categories in second language learning. *Cognitive Linguistics*

19(1): 67–106. DOI: 10.1515/COG.2008.005.

Ruiz, S., Tagarelli, K.M. and Rebuschat, P., (2018). Simultaneous acquisition of words and syntax: Effects of exposure condition and declarative memory. *Frontiers in Psychology*, 9, p.1168. DOI: 10.3389/fpsyg.2018.01168.

Schmitt, N. (2010). Researching vocabulary: A vocabulary research manual. Basingstoke:Palgrave.

Schmitt N and Redwood S (2011) Learner knowledge of phrasal verbs: A corpus-informed study. In: Meunier F, De Cock S, Gilquin G and Paquot M (eds) *A taste for corpora: In honour of Sylviane* Granger. Amsterdam: John Benjamins Publishing Company, pp. 173–209.

Sharwood Smith M (1991) Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research* 7(2): 118–32. http://www.jstor.org/stable/43104426.

Sharwood Smith M (1993) Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition* 15(2): 165–179. DOI: 10.1017/S0272263100011943.

Side, R. (1990) Phrasal verbs: Sorting them out. *ELT Journal*, 2(44), 144–152.

Skeppstedt, M., Ahltorp, M., and Henriksson, A. (2013) Vocabulary expansion by semantic extraction of medical terms. In: Rinaldi F and Kim, J (eds) *Proceedings of the Symposium on Languages in Biology and Medicine (LBM)*. Tokyo: Database Center for Life Science.

Sok S, Kang EY and Han Z (2018) Thirty-five years of ISLA on form-focused instruction: A methodological synthesis. *Language Teaching Research* DOI: 10.1177/1362168818776673.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, 63(1), 121–159.

Spada N (2011) Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching* 44(2): 225–236. DOI:10.1017/S0261444810000224.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263–308.

Squire LR (2004) Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory* 82(3): 171–177. DOI: 10.1016/j.nlm.2004.06.005.

Sung MC (2017) Underuse of English verb-particle constructions in an L2 learner corpus: Focus on structural patterns and one-word preference. *Corpus Linguistics and Linguistic Theory* DOI: 10.1515/cllt-2017-0002.

Suzuki Y and DeKeyser RM (2017) Exploratory research on second language practice distribution: An aptitude-by-treatment interaction. *Applied Psycholinguistics* 38(1): 27–56. DOI: 10.1017/S0142716416000084.

Suzuki, Y. , Nakata, T. and DeKeyser, R. (2019) Optimizing second language practice in the classroom: perspectives from cognitive psychology. *The Modern Language Journal*. DOI: 10.1111/modl.12582

Swain M (1995) Three functions of output in second language learning. In: Cook G and Seidlhofer B (eds) *Principles and practice in applied linguistics*. Oxford: Oxford University Press, pp. 125–144.

Tagarelli, K.M., Borges Mota, M., & Rebuschat, P. (2015). Working memory, learning conditions and the acquisition of L2 syntax. In: Z. E. Wen, MB Mota and A. McNeill (eds) Working memory in second language acquisition and processing. Bristol: Multilingual Matters, pp. 224–247.

Tagarelli, K.M., Ruiz, S., Moreno-Vega, J.L. and Rebuschat, P., (2016). Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity. *Studies in Second Language Acquisition*, 38(2), pp. 293-316. DOI: 10.1017/S0272263116000036.

Thibeau, T. J. (1999). *English prepositions in phrasal verbs: A study in second language acquisition*. Arizona: The University of Arizona.

Thim S (2012) *Phrasal verbs: The English verb-particle construction and its history*. New York: Mouton de Gruyter.

Trahan DE and Larrabee GJ (1988) *Continuous Visual Memory Test*. Odessa, FL: Psychological Assessment Resources.

Ullman MT (2015) The declarative/procedural model: A neurobiologically motivated theory of first and second language. In: VanPatten B and Williams J (eds) *Theories in second language acquisition: An introduction*, 2nd edition. New York: Routledge, pp. 135–158.

Ullman MT (2016) The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In: Hickok G and Small SA (eds) *Neurobiology of Language*. Amsterdam: Elsevier, pp. 953–968.

Ullman MT and Lovelett JT (2018) Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research* 34(1): 39–65. DOI: 10.1177/0267658316675195.

Unsworth N, Heitz RP, Schrock JC and Engle RW (2005) An automated version of the operation span task. *Behavior Research Methods* 37(3): 498–505. DOI: 10.3758/BF03192720.

VanPatten B (1990) Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition* 12(3): 287–301. DOI: 10.1017/S0272263100009177.

Vatz K, Tare M, Jackson SR and Doughty CJ (2013) Aptitude-treatment interaction studies in second language acquisition: Findings and methodology. In: Granena G and Long M (eds) *Sensitive periods, language aptitude, and ultimate L2 attainment*. Amsterdam: John Benjamins, pp. 273–292.

Wen, Z. (2019). Working memory as language aptitude: The Phonological/Executive Model. In: Wen Z., Skehan, P., Biedron, A., Li, S., & Sparks, R. (eds.), *Language aptitude: Advancing theory, testing, research and practice.* New York: Routledge, pp.187–214.

Wen, Z. (2015). Working memory in second language acquisition and processing: The phonological/ executive model. In: Z. Wen, M. Borges Mota, and A. McNeill (eds) *Working memory in second language acquisition and processing*. Bristol: Multilingual Matters, pp. 41–62.

Williams JN (2012) Working memory and SLA. In: Robinson P, Mackey A, Gass SM and Schmidt R (eds) *The Routledge handbook of second language acquisition*. New York: Routledge, pp. 427–441.

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

Ziegler, N., Meurers, D., Rebuschat, P., Ruiz, S., Moreno-Vega, J. L., Chinkina, M., Li, W., and Grey, S. (2017). Interdisciplinary research at the intersection of CALL, NLP, and SLA: Methodological implications from an input enhancement project. Language Learning, 67(S1), 210–232. DOI: 10.1111/lang.12227

Appendix 1

Target phrasal verbs in the 40-text reading corpus

Rank	Phrasal verb	Target meaning	Semantic category	Raw frequency	% of total phrasal verbs
1	carry out (36)	complete	figurative	9	5.45
2	go on (1)	proceed	literal	8	4.85
3	take on (15)	accept	figurative	8	4.85
4	set up (11)	present	figurative	7	4.24
5	take over (37)	assume	literal	7	4.24
6	end up (18)	finish	figurative	6	3.64
7	open up	expand	figurative	6	3.64
8	give up (16)	abandon	literal	5	3.03
9	grow up (10)	mature	literal	5	3.03
10	make up (17)	constitute	figurative	5	3.03
11	set out (64)	establish	literal	5	3.03
12	blow up (99)	explode	literal	4	2.42
13	clean up (65)	sanitize	figurative	4	2.42
14	come up (4)	produce	figurative	4	2.42
15	pay off (78)	liquidate	literal	4	2.42
16	sign up	register	literal	4	2.42
17	buy up	acquire	literal	3	1.82
18	figure out (21)	understand	literal	3	1.82
19	lay off	disemploy	figurative	3	1.82
20	lay out (73)	present	figurative	3	1.82
21	point out (9)	indicate	figurative	3	1.82
22	rule out (123)	exclude	figurative	3	1.82
23	scale up	maximize	literal	3	1.82
24	shore up	support	figurative	3	1.82
25	shut down (66)	close	figurative	3	1.82
26	water down	attenuate	figurative	3	1.82
27	build up (84)	increase	figurative	2	1.21
28	gobble up	consume	figurative	2	1.21
29	ramp up	intensify	figurative	2	1.21
30	turn out (12)	prove	figurative	2	1.21

Note. Numbers in parentheses indicate the frequency ranking order in the list of 150 most frequently used phrasal verbs in English (at least 10 tokens per million words) (Liu, 2011).

Appendix 2

Registration window in the ICALL system

Appendix 3

Sample of test items (based on Schmitt and Redwood, 2011)

A. Productive item on the phrasal verb *put up*:

Some skeptical people have p[____] o[____] that self-driving cars can be dangerous. (indicate)

B. Receptive item on the phrasal verb *put up:*

Some skeptical people have _____ _____ that self-driving cars can be dangerous. (indicate)

- ○ ?
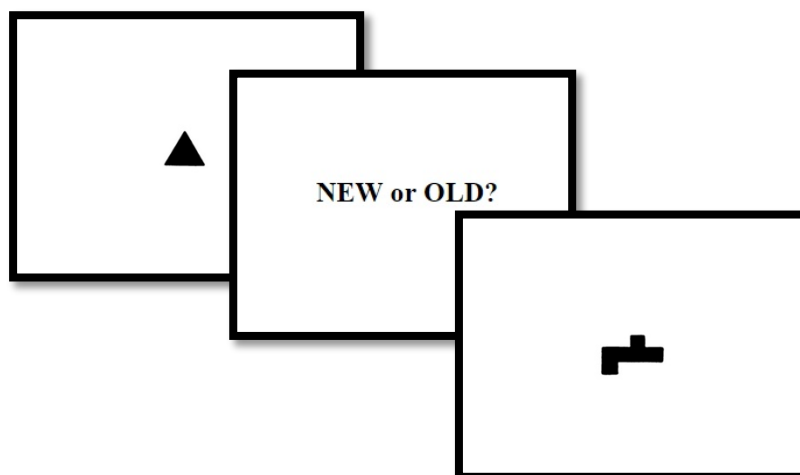- ○ ruled out
- ○ held out
- ○ pointed out
- ○ come out

Appendix 4

Cognitive tests

A. *Working memory*
Automated Operation Span (OSpan; Unsworth et al., 2005). Online version with Klingon characters based on Hicks et al. (2016)

Remember

Solve

(9 * 2) - 9

When you have solved the math problem, click on Continue.

Continue

Remember

Recall

Select the symbols in the order presented.

Clear   Exit
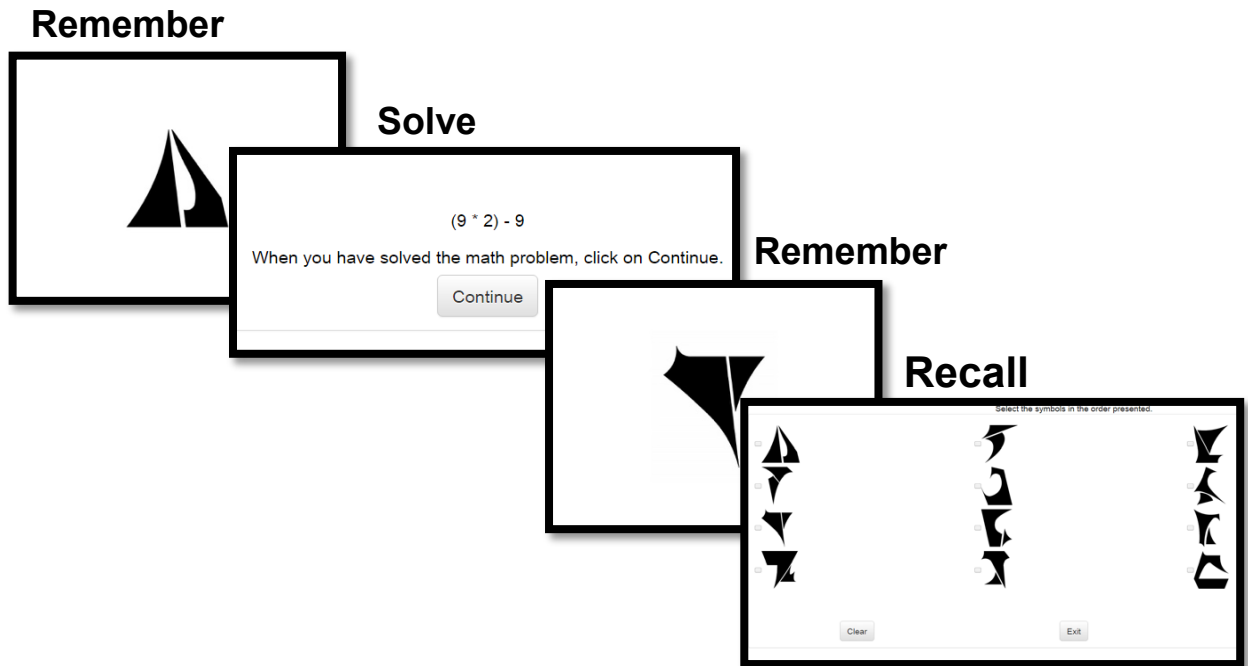
B. *Visual declarative memory:*
Continuous Visual Memory Task (CVMT; Trahan and Larrabee, 1988)

NEW or OLD?

C. *Verbal declarative memory*
Modern Language Aptitude Test, subtest 5, word pair associates (MLAT5; Carroll and Sapon, 1959)

MODERN LANGUAGE
APTITUDE TEST

You are going to take a test to determine how easily and rapidly you learn a foreign language. Do your best, regardless of whether or not you think you are good at learning foreign languages.

Your first task is to MEMORIZE some Kurdish-English vocabulary. You will be given TWO minutes to study the vocabulary on the next page. At the end of the two minutes, you will complete some filling-in-the-blank exercises, which will help you memorize the vocabulary. After filling the blanks, continue studying if there is still time.

Go ahead now and start memorizing the vocabulary. You will be tested on this later.

Please click on the "Start" button to proceed.

> Start

Appendix 5

Example of C-test

Example:

How Lakeland wind farm plan has environmentalists in a spin
On maps, you can already see it: a circle, almost a collar, clearly beginning to take shape on the borders of Britain's largest and most visited national park, the Lake District. It i____ a cir_____of st_____ , consisting o____ wind fa____, working a___ planned: congre____ of enor____ windmills s____ on hill____ where th____ can ca____ the wi___ to pro____ electricity, a____which a____ visible f____ miles.
So____ are alr____ operating, so___ are un____ construction, oth_____ are be_____ evaluated, a f____ have be____ turned down. But what is certain is that they represent only the first of many more to come on the Lake District borders in the Cumbrian hills, and an increasing number of local people fear they will ruin huge swaths of some of the loveliest landscape in England, and have a dire effect on the national park itself.

Remember, you must not consult any materials at all when you are taking the test.
You now have **30 minutes** to complete **5 texts**.

START THE TEST

Appendix 6

Sample of treatment pages

A. Meaning-focused condition:

B. Form-focused condition:

Appendix 7

Sample of reading questionnaire

Reading Questionnaire

What is the main idea of the article?

- ○ In a survey, parents indicated that they were more stressed than ever about college costs.
- ○ According to the article, parents may be saving too much money for their children's college.
- ○ In a survey, over 50 percent of parents indicated that they were losing sleep over college costs.
- ○ According to the article, parents may not be saving money enough for their children's college.

Feedback Questionnaire

Reading this text was:

very easy ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 very difficult

How much did you enjoy reading this text?

very little ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 very much

How familiar was the general topic to you?

very little ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 very much

Feel free to write any other comments

Submit